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.

Refer to the three hypotheses made on the previous page. In Situation A, the null hypothesis is expressed using the notation H_0: $p = 0.80$. This is a statement of *status quo* or no difference. The Latin phrase *status quo* means "the existing state or condition." So, the statement in the null hypothesis means that American opinions have not changed from 2008. We are trying to show that the proportion is different today, so the alternative hypothesis is H_1: $p \neq 0.80$. In Situation B, the null hypothesis is H_0: $\mu = 500$ hours. This is a statement of no difference between the population mean and the lifetime stated on the label. We are trying to show that the mean lifetime is less than 500 hours, so the alternative hypothesis is H_1: $\mu < 500$ hours. In Situation C, the null hypothesis is H_0: $\sigma = 0.08$ percent. 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 alternative hypothesis is H_1: $\sigma < 0.08$ percent. Do you see why?

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: $\mu < 12$ ounces. 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: $\mu \neq 12$ ounces. 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: $\mu = 12$ ounces.

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. Two percent of children taking competing antibiotics 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* price of a used three-year-old Chevy Corvette ZR1 is \$86,012. Grant wonders if the mean price of a used three-year-old Chevy Corvette ZR1 in the Miami metropolitan area is different from \$86,012.

(c) The standard deviation of the contents in a 64-ounce bottle of detergent using an old filling machine is 0.23 ounce. The manufacturer wants to know if a new filling machine has less variability.

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

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

Solution

(a) The hypothesis deals with a population proportion, p. If the new drug is no different from competing drugs, the proportion of individuals taking it who experience a headache will be 0.02; so the null hypothesis is H_0: $p = 0.02$. We want to determine if the proportion of individuals who experience a headache is more than 0.02, so 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 three-year-old Corvette ZR1 in Miami is no different from the *Blue Book* price, then the population mean in Miami will be \$86,012, so the null hypothesis is

(continued)

In Other Words
Look for key phrases when forming the alternative hypothesis. For example, ***more than*** means $>$; ***different from*** means $\neq$ and ***less than*** means $<$. See Table 9 on page 360 for a list of key phrases and the symbols they translate into.

• Now Work Problem 17(a)

H_0: $\mu = \$86{,}012$. Grant wants to know if the mean price is different from \$86,012, so the alternative hypothesis is H_1: $\mu \neq \$86{,}012$. 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 one, 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: $\sigma = 0.23$ ounce. The company wants to know if the new machine has *less* variability than the old machine, so the alternative hypothesis is H_1: $\sigma < 0.23$ ounce. This is a left-tailed test because the alternative hypothesis contains a $<$ symbol. •

2 Explain Type I and Type II Errors

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.

Sample data is used to decide whether or not to reject the statement in the null hypothesis. Because this decision is based on incomplete (sample) information, there is the possibility of making an incorrect decision. In fact, there are four possible outcomes from hypothesis testing.

Four Outcomes from Hypothesis Testing

1. Reject the null hypothesis when the alternative hypothesis is true. This decision would be correct.
2. Do not reject the null hypothesis when the null hypothesis is true. This decision would be correct.
3. 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. 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_o Is True	H_1 Is True
Conclusion	Do Not Reject H_o	Correct Conclusion	Type II Error
	Reject H_o	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 collect and present evidence proving that the defendant is guilty beyond all reasonable doubt.

Because we are seeking evidence for guilt, it becomes the alternative hypothesis. Innocence is assumed, so it is the null hypothesis.

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

In a trial, the jury obtains information (sample data). It then deliberates about the evidence (the data analysis). Finally, it 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, the null hypothesis is never declared true. The two correct decisions are to declare an innocent person not guilty or declare a guilty person to be 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. Two percent of children taking competing antibiotics 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: $p > 0.02$. Explain what it would mean to make a (a) Type I error and (b) Type II error.

Approach A Type I error occurs if we reject the null hypothesis when it is true. A Type II error occurs if we do not reject the null hypothesis when 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. In other words, the sample evidence led the researcher to believe $p = 0.02$ when in fact the true proportion is some value larger than 0.02. •

• Now Work Problems 17(b) and (c)

The Probability of Making a Type I or Type II Error

When we studied confidence intervals, we learned 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 the conclusion of a hypothesis test is correct. However, just as we place a level of confidence in the construction of a confidence interval, we can assign probabilities to making Type I or Type II errors when testing hypotheses. 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 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 $\alpha = 0.01$? Reducing the probability of making a Type I error increases 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 that the probability of convicting an innocent person 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.

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 hypothesis. 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.

3 State Conclusions to Hypothesis Tests

Once the decision whether or not to 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's verdict is "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. By not rejecting the null hypothesis, we are saying that the evidence indicates the null hypothesis *could* be true. That is, there is not enough evidence to reject our assumption that the null hypothesis is true.

EXAMPLE 4 Stating the Conclusion

Problem The Medco pharmaceutical company has just developed a new antibiotic. Two percent of children taking competing antibiotics 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: $p > 0.02$.

Suppose that the sample evidence indicates that

(a) the null hypothesis is rejected. State the conclusion.
(b) the null hypothesis is not rejected. State the conclusion.

In Other Words
The conclusion to a hypothesis test is always as follows: There (***is/is not***) sufficient evidence to conclude that ***insert statement in alternative hypothesis***.

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, 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, 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

Vocabulary and Skill Building

1. A ____________ is a statement regarding a characteristic of one or more populations.

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

3. The _____ ___________ is a statement of no change, no effect, or no difference.

4. The ___________ ___________ is a statement we are trying to find evidence to support.

5. If we reject the null hypothesis when the statement in the null hypothesis is true, we have made a Type _____ error.

6. If we do not reject the null hypothesis when the statement in the alternative hypothesis is true, we have made a Type _____ error.

7. The ______ _ ____________ is the probability of making a Type I error.

8. *True or False*: Sample evidence can prove a null hypothesis is true.

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

9. H_0: $\sigma = 9$
H_1: $\sigma < 9$

10. H_0: $\mu = 7$
H_1: $\mu > 7$

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

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

13. H_0: $\mu = 110$
H_1: $\mu < 110$

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. Complete College For students who first enrolled in two year public institutions in a recent semester, the proportion who earned a bachelor's degree within six years was 0.386. The president of a certain college believes that the proportion of students who enroll in her institution have a lower completion rate.

16. Pizza Historically, the time to order and deliver a pizza at Jimbo's pizza was 48 minutes. Jim, the owner, implements a new system for ordering and delivering pizzas that he believes will reduce the time required to get a pizza to his customers.

NW **17. Single-Family Home Price** Three years ago, the mean price of an existing single-family home was \$243,787. A real estate broker believes that existing home prices in her neighborhood are higher.

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.

19. Valve Pressure The standard deviation in the pressure required to open a certain valve is known to be $\sigma = 1.1$ psi. Due to changes in the manufacturing process, the quality-control manager feels that the pressure variability has increased.

20. Overweight According to the Centers for Disease Control and Prevention, 19.6% 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 different in her school district.

21. Cell Phone Bills According to a report, the standard deviation of monthly cell phone bills was \$49.11 three years ago. A researcher suspects that the standard deviation of monthly cell phone bills is different today.

22. SAT Reading Scores In 2014, the standard deviation of SAT score on the Critical Reading Test for all students taking the exam was 112. A teacher believes that, due to changes to high school curricula, the standard deviation of SAT math scores has decreased.

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

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

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

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

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

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

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

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

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

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

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

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

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

Applying the Concepts

35. Quality Control A can of soda is labeled as containing 16 fluid ounces. The quality control manager wants to verify that the filling machine is neither overfilling nor underfilling the cans.

(a) Determine the null and alternative hypotheses that would be used to determine if the filling machine is calibrated correctly.

(b) The quality control manager obtains a sample of 75 cans and measures the contents. The sample evidence leads the manager to reject the null hypothesis. Write a conclusion for this hypothesis test.

(c) Suppose, in fact, the machine is not out of calibration. Has a Type I or Type II error been made?

(d) Management has informed the quality control department that it does not want to shut down the filling machine unless the evidence is overwhelming that the machine is out of calibration. What level of significance would you recommend the quality control manager to use? Explain.

36. Popcorn Consumption According to a food website, the mean consumption of popcorn annually by Americans is 59 quarts. The marketing division of the food website unleashes an aggressive campaign designed to get Americans to consume even more popcorn.

(a) Determine the null and alternative hypotheses that would be used to test the effectiveness of the marketing campaign.

(b) A sample of 801 Americans provides enough evidence to conclude that marketing campaign was effective. Provide a statement that should be put out by the marketing department.

(c) Suppose, in fact, the mean annual consumption of popcorn after the marketing campaign is 59 quarts. Has a Type I or Type II error been made by the marketing department? If we tested this hypothesis at the $\alpha = 0.01$ level of significance, what is the probability of committing this error? Select the correct choice below and fill in the answer box within your choice.

37. E-Cigs According to the Centers for Disease Control and Prevention, 9.8% of high school students currently use electronic cigarettes. A high school counselor is concerned the use of e-cigs at her school is higher.

(a) Determine the null and alternative hypotheses.

(b) If the sample data indicate that the null hypothesis should not be rejected, state the conclusion of the high school counselor.

(c) Suppose, in fact, that the proportion of students at the counselor's high school who use electronic cigarettes is 0.222. Was a type I or type II error committed?

38. Migraines According to the Centers for Disease Control, 15.2% of American adults experience migraine headaches. Stress is a major contributor to the frequency and intensity of headaches. A massage therapist feels that she has a technique that can reduce the frequency and intensity of migraine headaches.

(a) Determine the null and alternative hypotheses that would be used to test the effectiveness of the massage therapist's techniques.

(b) A sample of 500 American adults who participated in the massage therapist's program results in data that indicate that the null hypothesis should be rejected. Provide a statement that supports the massage therapist's program.

(c) Suppose, in fact, that the percentage of patients in the program who experience migraine headaches is 15.3%. Was a Type I or Type II error committed?

39. Engine Treatment The manufacturer of a certain engine treatment claims that if you add their product to your engine, it will be protected from excessive wear. An infomercial claims that a woman drove 5 hours without oil, thanks to the engine treatment. A magazine tested engines in which they added the treatment 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 that the magazine will test.

(b) Both engines took exactly 15 minutes to seize. What conclusion might the magazine make based on this evidence?

40. Refer to 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.

Retain Your Knowledge

41. Retirement Savings Designed by Bill Bengen, the 4 percent rule says that a retiree may withdraw 4% of savings during the first year of retirement, and then each year after that withdraw the same amount plus an adjustment for inflation. Under this rule, your retirement savings should be expected to last 30 years, which is longer than most retirements.

(a) If your retirement savings is $750,000, how much may you withdraw in your first year of retirement if you want the retirement savings to last 30 years?

(b) According to the American College of Financial Services, the proportion of people 60 to 75 years of age who believe it would be safe to withdraw 6 to 8 percent of their retirement savings annually is 0.16. Suppose you conduct a survey of twenty 60 to 75 year olds and ask them if it is safe to withdraw 6 to 8 percent of retirement savings annually if they wish their retirement savings to last 30 years. Explain why this is a binomial experiment. What are the values of n and p?

(c) In a random sample of twenty 60 to 75 year olds, what is the probability exactly 8 individuals will believe it is safe to withdraw 6 to 8 percent of retirement savings annually if they wish their retirement savings to last 30 years.

(d) In a random sample of twenty 60 to 75 year olds, what is the probability fewer than 8 individuals will believe it is safe to withdraw 6 to 8 percent of retirement savings annually if they wish their retirement savings to last 30 years.

(e) Suppose you obtain a random sample of five hundred 60 to 75 year olds. Explain why the normal model may be used to describe the sampling distribution of $\hat{p}$ the sample proportion of 60 to 75 year olds who believe it is safe to withdraw 6 to 8 percent of their retirement savings annually. Describe this sampling distribution. That is, find the shape, center, and spread of the sampling distribution of the sample proportion.

(f) Use the normal model from part (e) to approximate the probability of obtaining a random sample of at least one hundred 60 to 75 years olds who believe it would be safe to withdraw 6 to 8 percent of their retirement savings annually assuming the true proportion is 0.16. Is this result unusual? Explain.

Explaining the Concepts

42. If 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?

43. What happens to the probability of making a Type II error, β, as the level of significance, α, decreases? Why?

44. 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. . . .*

Source: Letter by Ronald A Fisher in Nature. Copyright © by Nature Publishing Group.

In your own words, explain what this quotation means.

45. In your own words, explain the difference between "beyond all reasonable doubt" and "beyond all doubt." Use these phrases to explain why we never "accept" the statement in the null hypothesis.

10.2 Hypothesis Tests for a Population Proportion

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

- Using probabilities to identify unusual events (Section 5.1, p. 277)
- z_α notation (Section 7.2, pp. 400–401)
- Sampling distribution of the sample proportion (Section 8.2, pp. 436–440)
- Computing normal probabilities (Section 7.2, pp. 394–398)
- Binomial probability distribution (Section 6.2, pp. 356–366)

Objectives

1. Explain the logic of hypothesis testing
2. Test hypotheses about a population proportion
3. Test hypotheses about a population proportion using the binomial probability distribution

1 Explain the Logic of Hypothesis Testing

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. We learned in Section 8.2 that the sampling distribution of $\hat{p}$ is approximately normal, with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$, provided that the following requirements are satisfied.

1. The sample is a simple random sample.
2. $np(1-p) \geq 10$.
3. The sampled values are independent of each other $(n \leq 0.05N)$.

We will present three methods for testing hypotheses. The first method is called 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.

First, we lay out a scenario that will be used to understand both the classical and P-value approaches. Suppose a politician wants to know if a majority (more than 50%) of her constituents are in favor of a certain policy. We are therefore testing the following hypotheses:

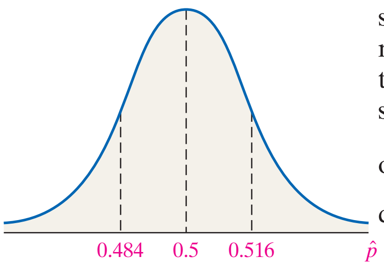
$$H_0: p = 0.5 \quad \text{versus} \quad H_1: p > 0.5$$

The politician hires a polling firm to obtain a random sample of 1000 registered voters in her district and finds that 534 are in favor of the policy, so $\hat{p} = \frac{534}{1000} = 0.534$. Do these results suggest that among *all* registered voters more than 50% favor the policy? Or is it possible that the true proportion of registered voters who favor the policy is 0.5 and we just happened to survey a majority in favor of the policy? In other words, would it be unusual to obtain a sample proportion of 0.534 or higher from a population whose proportion is 0.5? 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** and we reject the null hypothesis.

To determine if a sample proportion of 0.534 is statistically significant, we build a probability model. After all, a second random sample of 1000 registered voters will likely result in a different sample proportion, and we want to describe this variability so we can determine if the results we obtained are unusual. Since $np(1-p) = 1000(0.5)(1-0.5) = 250 \geq 10$ and the sample size $(n = 1000)$ is sufficiently smaller than the population size (provided there are at least $N = 20{,}000$ registered voters in the politician's district), we can use a normal model to describe the variability in $\hat{p}$. The mean of the distribution of $\hat{p}$ is $\mu_{\hat{p}} = 0.5$ (since we assume the statement in the null hypothesis is true) and the standard deviation of the distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5(1-0.5)}{1000}} \approx 0.016$. Figure 2 shows the sampling distribution of the sample proportion for the "politician" example.

Figure 2



Now that we have a model that describes the distribution of the sample proportion, we can use it to look at the logic of the classical and P-value approaches to test if a majority of the politician's constituents are in favor of the policy.

The Logic of the Classical Approach

We may consider the sample evidence to be statistically significant (or sufficient) if the sample proportion is too many standard deviations, say 2, above the assumed population proportion of 0.5.

Recall that $z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}}$ represents the number of standard deviations that $\hat{p}$ is from the population proportion, p. Our simple random sample of 1000 registered voters results in a sample proportion of 0.534, so under the assumption that the null hypothesis is true we have

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{0.534 - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{1000}}} = 2.15$$

The sample proportion is 2.15 standard deviations above the hypothesized population proportion of 0.5, which is more than 2 standard deviations (that is, "too far") above the hypothesized population proportion. So we will reject the null hypothesis. There is statistically significant (sufficient) evidence to conclude that a majority of registered voters are in favor of the policy.

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

Figure 3

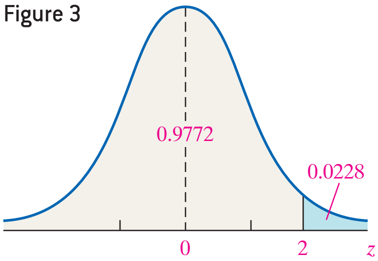


Figure 4

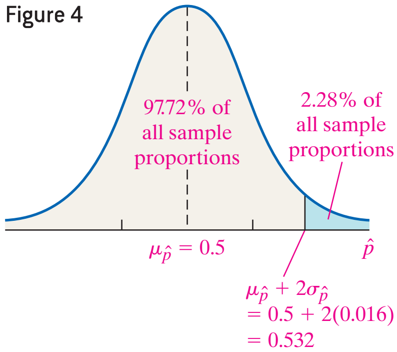


Figure 4 shows that if the null hypothesis is true (that is, if the population proportion is 0.5), then 97.72% of all sample proportions will be 0.532 or less, and only 2.28% of the sample proportions will be more than 0.532 (0.532 is 2 standard deviations above the hypothesized proportion of 0.5). If a sample proportion lies in the blue region, we are inclined to believe it came from a population whose proportion is greater than 0.5, rather than believe that the population proportion equals 0.5 and our sample just happened to result in a proportion of registered voters much higher than 0.5.

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. This is because 2.28% of all sample proportions are more than 0.532, even though the population proportion is 0.5.

This discussion illustrates the following point.

Hypothesis Testing Using the Classical Approach

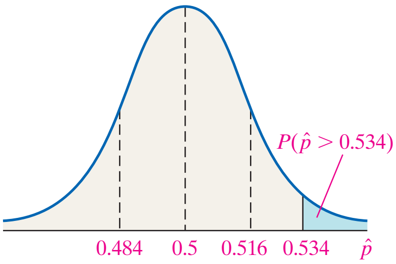
If the sample statistic is too many standard deviations from the population parameter stated in the null hypothesis, we reject the null hypothesis.

The Logic of the *P*-Value Approach

A second criterion we may use for testing hypotheses is to determine how likely it is to obtain a sample proportion of 0.534 or higher from a population whose proportion is 0.5. If a sample proportion of 0.534 or higher is unlikely (or unusual), we have evidence against the statement in the null hypothesis. Otherwise, we do not have sufficient evidence against the statement in the null hypothesis.

We can compute the probability of obtaining a sample proportion of 0.534 or higher from a population whose proportion is 0.5 using the normal model. See Figure 5.

Figure 5



$$P(\hat{p} > 0.534) = P\left(z > \frac{0.534 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{1000}}}\right) = P(z > 2.15) = 0.0158$$

The value 0.0158 is called the *P-value*, which means about 2 samples in 100 will give a sample proportion as high or higher than the one we obtained *if* the population proportion really is 0.5. Because these results are unusual, we take this as evidence against the statement in the null hypothesis.

Definition

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

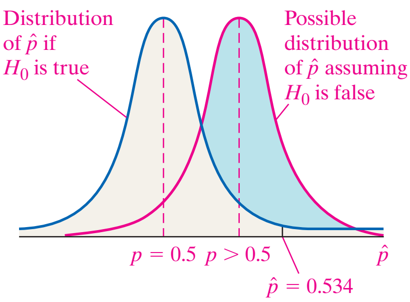
This discussion illustrates the idea behind hypothesis testing using the *P*-value approach.

Hypothesis Testing Using the *P*-Value Approach

If the probability of getting a sample statistic as extreme or more extreme than the one obtained is small under the assumption the statement in the null hypothesis is true, reject the null hypothesis.

Figure 6 illustrates the situation for both the classical and *P*-value approaches. The distribution in blue shows the distribution of the sample proportion assuming the statement in the null hypothesis is true. The sample proportion of 0.534 is too far from the assumed population proportion of 0.5. Therefore, we reject the null hypothesis that $p = 0.5$ and conclude that $p > 0.5$, as indicated by the distribution in red. We do not know what the population proportion of registered voters who are in favor of the policy is, but we have evidence to say it is greater than 0.5 (a majority).

Figure 6



2 Test Hypotheses about a Population Proportion

We now formalize the procedure for testing hypotheses regarding a population proportion.

Testing Hypotheses Regarding a Population Proportion, *p*

Use Steps 1 through 5, provided that

- the sample is obtained by simple random sampling or the data result from a randomized experiment.
- $np_0(1 - p_0) \geq 10$.
- the sampled values are independent of each other.

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.

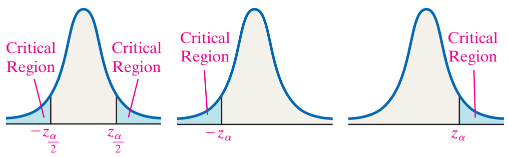
Classical Approach

Step 3 Compute the **test statistic**

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

Use Table V 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



Step 4 Compare the critical value with the test statistic.

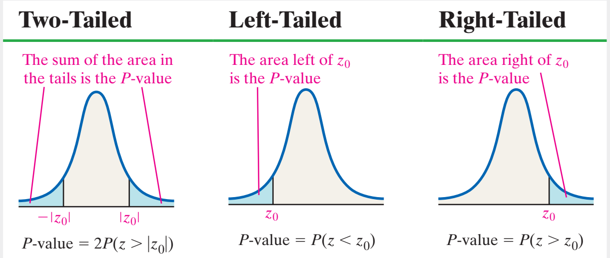
Two-Tailed	Left-Tailed	Right-Tailed
If $z_0 < -z_{\frac{\alpha}{2}}$ or $z_0 > z_{\frac{\alpha}{2}}$, reject the null hypothesis.	If $z_0 < -z_\alpha$, reject the null hypothesis.	If $z_0 > z_\alpha$, reject the null hypothesis.

P-Value Approach

By Hand Step 3 Compute the **test statistic**

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

Use Table V to determine the *P*-value.



Technology Step 3 Use a statistical spreadsheet or calculator with statistical capabilities to obtain the *P*-value. The directions for obtaining the *P*-value using the TI-83/84 Plus graphing calculators, Minitab, Excel, and StatCrunch are in the Technology Step-by-Step on pages 517–518.

Step 4 If *P*-value $< \alpha$, reject the null hypothesis.

Step 5 State the conclusion.

Notice in Step 3 that we are using p_0 (the proportion stated in the null hypothesis) in computing the standard error rather than $\hat{p}$, as we did in constructing confidence intervals about p. This is because H_0 is assumed to be true when performing a hypothesis test, so the assumed mean of the distribution of $\hat{p}$ is $\mu_{\hat{p}} = p_0$ and the assumed standard error is $\sigma_{\hat{p}} = \sqrt{\dfrac{p_0(1 - p_0)}{n}}$.

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

Problem The two major college entrance exams that a majority of colleges accept for admission are the SAT and ACT. ACT looked at historical records and established 22 as the minimum ACT math score for a student to be considered prepared for college mathematics. [**Note:** "Being prepared" means there is a 75% probability of successfully completing College Algebra in college.] An official with the Illinois State Department of Education wonders whether less than half of the students in her state are prepared for College Algebra. She obtains a simple random sample of 500 records of students who have taken the ACT and finds that 219 are prepared for college mathematics (that is, scored at least 22 on the ACT math test). Does this represent significant evidence that less than half of Illinois students who have taken the ACT are prepared for college mathematics upon graduation from a high school? Use the $\alpha = 0.05$ level of significance. *Source:* ACT High School Profile Report.

Approach This problem deals with a hypothesis test of a proportion. We want to determine if the sample evidence shows that less than half of the students are prepared for college mathematics. Symbolically, we represent this as $p < \frac{1}{2}$ or $p < 0.5$.

Verify the three requirements to perform the hypothesis test: 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 (for independence). Then follow Steps 1 through 5.

Solution Assume that $p = 0.5$. The sample is a simple random sample. Also, $np_0(1 - p_0) = 500(0.5)(1 - 0.5) = 125 > 10$. Provided that there are over 10,000 students in the state, the sample size is less than 5% of the population size. Assuming that this is the case, the requirements are satisfied. Now proceed with Steps 1 through 5.

Step 1 The burden of proof lies in showing $p < 0.5$. We assume there is no difference between the proportion of students *ready* for college math and the proportion of students *not ready* for college math. Therefore, the statement in the null hypothesis is that $p = 0.5$. So we have

$$H_0: p = 0.5 \qquad \text{versus} \qquad H_1: p < 0.5$$

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

Classical Approach

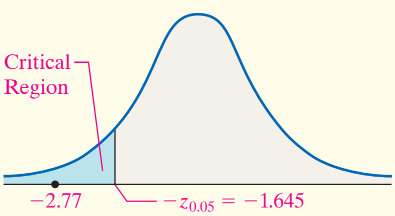
Step 3 The assumed value of the population proportion is $p_0 = 0.5$. The sample proportion is $\hat{p} = \frac{x}{n} = \frac{219}{500} = 0.438$. We want to know if it is unusual to obtain a sample proportion of 0.438 or less from a population whose proportion is assumed to be 0.5.

The test statistic is

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.438 - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{500}}} = -2.77$$

Because this is a left-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 shown in Figure 7.

Figure 7



Step 4 The test statistic, $z_0 = -2.77$, is labeled in Figure 7. Because the test statistic is less than the critical value $(-2.77 < -1.645)$, we reject the null hypothesis.

P-Value Approach

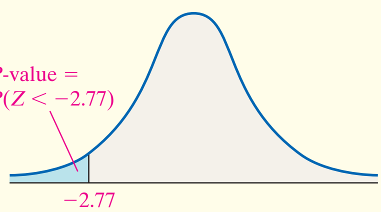
By Hand Step 3 The assumed value of the population proportion is $p_0 = 0.5$. The sample proportion is $\hat{p} = \frac{x}{n} = \frac{219}{500} = 0.438$. We want to know how likely it is to obtain a sample proportion of 0.438 or less from a population whose proportion is assumed to be 0.5.

The test statistic is

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.438 - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{500}}} = -2.77$$

Because this is a left-tailed test, the *P*-value is the area under the standard normal distribution to the left of the test statistic, $z_0 = -2.77$, as shown in Figure 8. So, $P\text{-value} = P(z < z_0) = P(z < -2.77) = 0.0028$.

Figure 8



Technology Step 3 Using Minitab, we find the *P*-value is 0.003. See Figure 9 on the next page.

(continued)

Figure 9

```
Test and CI for One Proportion

Test of p = 0.5 vs p < 0.5

Sample    X    N  Sample p  95% Upper Bound  Z-Value  P-Value
1       219  500  0.438000         0.474496    -2.77    0.003
```

Step 4 The P-value of 0.003 means that *if* the null hypothesis that $p = 0.5$ is true, we expect 219 or fewer successes in 500 trials in 3 out of 1000 repetitions of this study! The observed results are unusual, indeed. Because the P-value is less than the level of significance, $\alpha = 0.05$ $(0.003 < 0.05)$, we reject the null hypothesis.

Step 5 There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that fewer than half of the Illinois students are prepared for college mathematics. In other words, the data suggest less than a majority of the students in the state of Illinois who take the ACT are prepared for college mathematics. •

• **Now Work Problem 17**

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

Problem What do you think is more important—to protect the right of Americans to own guns or to control gun ownership? When asked this question, 46% of all Americans said protecting the right to own guns is more important. The Pew Research Center surveyed 1267 randomly selected Americans with at least a bachelor's degree and found that 559 believed that protecting the right to own guns is more important. Does this result suggest the proportion of Americans with at least a bachelor's degree feel differently than the general American population when it comes to gun control? Use the $\alpha = 0.1$ level of significance.

Approach This problem deals with a hypothesis test of a proportion. Verify the three requirements to perform the hypothesis test. Then follow Steps 1 through 5.

Solution We want to know if the proportion of Americans with at least a bachelor's degree who believe protecting the right of Americans to own guns is more important is *different from* 0.46. To conduct the test, assume the sample comes from a population with $p = 0.46$. The sample is a simple random sample. Also, $np_0(1 - p_0) = 1267(0.46)(1 - 0.46) = 314.7 \geq 10$. Because there are well over 10 million Americans with at least a bachelor's degree, the sample size is less than 5% of the population. We now proceed with Steps 1 through 5.

Step 1 We want to know whether the proportion is *different from* 0.46, which can be written $p \neq 0.46$, so this is a two-tailed test.

$$H_0\text{: } p = 0.46 \quad \text{versus} \quad H_1\text{: } p \neq 0.46$$

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

Classical Approach

Step 3 Assume the sample comes from a population with $p_0 = 0.46$. The sample proportion is $\hat{p} = \dfrac{x}{n} = \dfrac{559}{1267} = 0.441$. Is obtaining a sample proportion of 0.441 from a population whose proportion is 0.46 likely, or is it unusual?

The test statistic is

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} = \frac{0.441 - 0.46}{\sqrt{\dfrac{0.46(1 - 0.46)}{1267}}} = -1.36$$

P-Value Approach

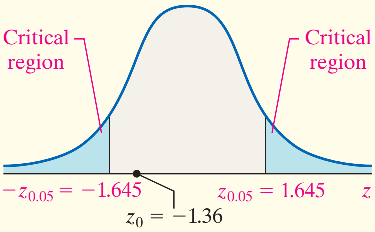
By Hand Step 3 Assume the sample comes from a population with $p_0 = 0.46$. The sample proportion is $\hat{p} = \dfrac{x}{n} = \dfrac{559}{1267} = 0.441$. What is the likelihood of obtaining a sample proportion of 0.441 from a population whose proportion is 0.46?

The test statistic is

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} = \frac{0.441 - 0.46}{\sqrt{\dfrac{0.46(1 - 0.46)}{1267}}} = -1.36$$

Because this is a two-tailed test, we determine the critical values at the $\alpha = 0.10$ level of significance to be $-z_{0.1/2} = -z_{0.05} = -1.645$ and $z_{0.1/2} = z_{0.05} = 1.645$. The critical regions are shown in Figure 10.

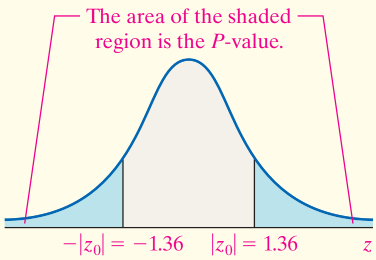
Figure 10



Step 4 The test statistic, $z_0 = -1.36$, is labeled in Figure 10. Because the test statistic does not lie in the critical region, do not reject the null hypothesis.

Because this is a two-tailed test, the P-value is the area under the standard normal distribution to the left of $-|z_0| = -1.36$ and to the right of $|z_0| = 1.36$, as shown in Figure 11.

Figure 11



$$\begin{aligned} P\text{-value} &= P(Z < -|z_0|) + P(Z > |z_0|) \\ &= 2P(Z < -1.36) \quad \text{Use symmetry} \\ &= 2(0.0869) \\ &= 0.1738 \end{aligned}$$

Technology Step 3 Using StatCrunch, we find the P-value is 0.1794. See Figure 12.

Figure 12

Hypothesis test results:
p : proportion of successes for population
$H_0 : p = 0.46$
$H_A : p \neq 0.46$

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	559	1267	0.4411997	0.014001917	-1.3426958	0.1794

Step 4 The P-value of 0.1738 [Technology: 0.1794] means that *if* the null hypothesis that $p = 0.46$ is true, we expect the type of results we observed (or more extreme results) in about 17 or 18 out of 100 samples. The observed results are not unusual. Because the P-value is greater than the level of significance, $\alpha = 0.1$ $(0.1738 > 0.1)$, we do not reject the null hypothesis.

Step 5 There is not sufficient evidence at the $\alpha = 0.1$ level of significance to conclude that Americans with at least a bachelor's degree feel differently than the general American population when it comes to gun control. •

CAUTION!
In Example 2, we do not have enough evidence to reject the statement in the null hypothesis. In other words, it is not unusual to obtain a sample proportion of 0.441 from a population whose proportion is 0.46. However, this does not imply that we are accepting the statement in the null hypothesis (that is, we are not saying that the proportion equals 0.46). We are only saying we do not have enough evidence to conclude that the proportion is different from 0.46. Be sure that you understand the difference between "accepting" and "not rejecting." It is similar to the difference between being declared "innocent" versus "not guilty."

Also, be sure you understand that the *P*-value is the probability of obtaining a sample statistic as extreme or more extreme than the one observed *if* the statement in the null hypothesis is true. The *P*-value does not represent the probability that the null hypothesis is true. The statement in the null hypothesis is either true or false, we just don't know which.

In practice, the level of significance is not reported using the P-value approach. Instead, only the P-value is given, and the reader of the study must interpret its value and judge its significance.

• **Now Work Problem 21**

Test a Hypothesis Using a Confidence Interval

Recall, the level of confidence, $(1 - \alpha) \cdot 100\%$, in a confidence interval represents the percentage of intervals that will contain the unknown parameter if repeated samples are obtained.

Two-Tailed Hypothesis Testing Using Confidence Intervals

When testing H_0: $p = p_0$ versus H_1: $p \neq p_0$, if a $(1 - \alpha) \cdot 100\%$ confidence interval contains p_0, do not reject the null hypothesis. However, if the confidence interval does not contain p_0, conclude that $p \neq p_0$ at the level of significance, α.

EXAMPLE 3 Testing a Hypothesis Using a Confidence Interval

Problem A 2009 study by Princeton Survey Research Associates International found that 34% of teenagers text while driving. Does a recent survey conducted by *Consumer Reports*, which found that 353 of 1200 randomly selected teens had texted while driving, suggest that the proportion of teens who text while driving has changed since 2009? Use a 95% confidence interval to answer the question.

Approach Construct a 95% confidence about the proportion of teens who text while driving based on the *Consumer Reports* survey. If the interval does not include 0.34, reject the null hypothesis H_0: $p = 0.34$ in favor of H_1: $p \neq 0.34$.

Solution The 95% confidence interval for p based on the *Consumer Reports* survey has a lower bound of 0.268 and an upper bound of 0.320. Because 0.34 is not within the bounds of the confidence interval, there is sufficient evidence to conclude that the proportion of teens who text while driving has changed since 2009. •

• Now Work Problem 23

Note: The same *Consumer Reports* article cited in Example 3 states that 75% of teens have friends who text while driving. What does this say about the difficulty in finding truthful responses to questions while conducting a survey?

3 Test Hypotheses about a Population Proportion Using the Binomial Probability Distribution

For the sampling distribution of $\hat{p}$ to be approximately normal, we require that $np(1 - p)$ be at least 10. What if this requirement is not satisfied? In Section 6.2, we used the binomial probability formula to identify unusual events. We stated that an event was unusual if the probability of observing the event was less than 0.05. This criterion is based on the P-value approach to testing hypotheses; the probability that we computed was the P-value. We use this same approach to test hypotheses regarding a population proportion for small samples.

EXAMPLE 4 Hypothesis Test for a Population Proportion: Small Sample Size

Problem According to the U.S. Department of Agriculture (USDA), 48.9% of males aged 20 to 39 years consume the recommended daily requirement of calcium. After an aggressive "Got Milk" advertising campaign, the USDA conducts a survey of 35 randomly selected males aged 20 to 39 and finds that 21 of them consume the recommended daily allowance (RDA) of calcium. At the $\alpha = 0.10$ level of significance, is there evidence to conclude that the percentage of males aged 20 to 39 who consume the RDA of calcium has increased?

Approach We use the following steps:

Step 1 Determine the null and alternative hypotheses.

Step 2 Check whether $np_0(1 - p_0)$ is greater than or equal to 10, where p_0 is the proportion stated in the null hypothesis. If it is, then the sampling distribution of $\hat{p}$ is approximately normal and we can use the steps on page 512. Otherwise, we use Steps 3 and 4, presented next.

Step 3 Compute the P-value. For right-tailed tests, the P-value is the probability of obtaining x or more successes. For left-tailed tests, the P-value is the probability of obtaining x or fewer successes.* The P-value is always computed with the proportion given in the null hypothesis. Remember, assume that the null is true until we have evidence to the contrary.

Step 4 If the P-value is less than the level of significance, α, reject the null hypothesis.

Solution

Step 1 The status quo or no change proportion of 20- to 39-year-old males who consume the recommended daily requirement of calcium is 0.489. We wish to know whether the advertising campaign increased this proportion. Therefore,

$$H_0: p = 0.489 \quad \text{and} \quad H_1: p > 0.489$$

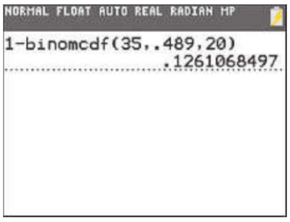
Step 2 From the null hypothesis, we have $p_0 = 0.489$. There were $n = 35$ individuals surveyed, so $np_0(1 - p_0) = 35(0.489)(1 - 0.489) = 8.75$. Because $np_0(1 - p_0) < 10$, the sampling distribution of $\hat{p}$ is not approximately normal.

Step 3 Let the random variable X represent the number of individuals who consume the daily requirement of calcium. We have $x = 21$ successes in $n = 35$ trials, so $\hat{p} = \dfrac{21}{35} = 0.6$. We want to judge whether the larger proportion is due to an increase in the population proportion or to sampling error. We obtained $x = 21$ successes in the survey and this is a right-tailed test, so the P-value is $P(X \geq 21)$.

$$P\text{-value} = P(X \geq 21) = 1 - P(X < 21) = 1 - P(X \leq 20)$$

We will compute this P-value using a TI-84 Plus C graphing calculator, with $n = 35$ and $p = 0.489$. Figure 13 shows the results.

Figure 13



The P-value is 0.1261. Minitab, StatCrunch, and Excel will compute exact P-values using this approach as well.

Step 4 The P-value is greater than the level of significance $(0.1261 > 0.10)$, so we do not reject H_0. There is not sufficient evidence (at the $\alpha = 0.1$ level of significance) to conclude that the proportion of 20- to 39-year-old males who consume the recommended daily allowance of calcium has increased. •

• **Now Work Problem 27**

Technology Step-by-Step Hypothesis Tests Regarding a Population Proportion

TI-83/84 Plus

1. Press STAT, highlight TESTS, and select 5:1-PropZTest.
2. For the value of p_0, enter the value of the population proportion stated in the null hypothesis.
3. Enter the number of successes, x, and the sample size, n.
4. Select the direction of the alternative hypothesis.
5. Highlight **Calculate** or **Draw**, and press ENTER.

Minitab

1. If you have raw data, enter them in C1, using 0 for failure and 1 for success.
2. Select the **Stat** menu, highlight **Basic Statistics**, then highlight **1-Proportion**.
3. If you have raw data, select "One or more samples, each in a column" from the drop-down menu. Place the cursor in the box, highlight the column containing the raw data, and click "Select." If you have summarized data, select "Summarized data" from the drop-down menu. Enter the number of successes in the "Number of events" box and enter the number of trials. Check the "Perform hypothesis test" box and enter the value of the population proportion stated in the null hypothesis.
4. Click Options. Enter the direction of the alternative hypothesis. Assuming $np_0(1 - p_0) \geq 10$, select "Normal approximation" from the drop-down menu. Click OK twice.

Excel

1. Load the XLSTAT Add-In.
2. Select the XLSTAT menu and select **Parametric Tests**. From the drop-down menu, select **Tests for one proportion**.

*We will not address P-values for two-tailed hypothesis tests. For those who are interested, the P-value is two times the probability of obtaining x or more successes if $\hat{p} > p$ and two times the probability of obtaining x or fewer successes if $\hat{p} < p$.

3. In the cell marked Frequency, enter the number of successes. In the cell marked Sample size, enter the number of trials. In the cell marked Test proportion, enter the proportion stated in the null hypothesis. Check the Frequency radio button. Click Options. Choose the appropriate direction for the alternative hypothesis. Be sure Hypothesize difference (D): is set to zero. Enter the level of significance. Click OK.

StatCrunch

1. If you have raw data, enter them into the spreadsheet. Name the column variable.
2. Select **Stat**, highlight **Proportion Stats**, select **One Sample**, and then choose either **With Data** or **With Summary**.
3. If you chose **With Data**, select the column that has the observations, choose which outcome represents a success. If you chose **With Summary**, enter the number of successes and the number of trials. Choose the hypothesis test radio button. Enter the value of the proportion stated in the null hypothesis and choose the direction of the alternative hypothesis from the pull-down menu. Click Compute!.

10.2 Assess Your Understanding

Vocabulary and Skill Building

1. When observed results are unlikely under the assumption that the null hypothesis is true, we say the result is __________ __________ and we reject the null hypothesis.

2. *True or False:* When testing a hypothesis using the Classical Approach, if the sample proportion is too many standard deviations from the proportion stated in the null hypothesis, we reject the null hypothesis.

3. *True or False:* When testing a hypothesis using the P-value Approach, if the P-value is large, we reject the null hypothesis.

4. Determine the critical value for a right-tailed test regarding a population proportion at the $\alpha = 0.01$ level of significance.

5. Determine the critical value for a left-tailed test regarding a population proportion at the $\alpha = 0.1$ level of significance.

6. Determine the critical value for a two-tailed test regarding a population proportion at the $\alpha = 0.05$ level of significance.

In Problems 7–12, test the hypothesis using (a) the classical approach and (b) the P-value approach. Be sure to verify the requirements of the test.

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

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

9. H_0: $p = 0.45$ versus H_1: $p < 0.45$
$n = 150$; $x = 62$; $\alpha = 0.05$

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

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

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

13. You Explain It! Stock Analyst Throwing darts at the stock pages to decide which companies to invest in could be a successful stock-picking strategy. Suppose a researcher decides to test this theory and randomly chooses 100 companies to invest in. After 1 year, 53 of the companies were considered winners; that is, they outperformed other companies in the same investment class. To assess whether the dart-picking strategy resulted in a majority of winners, the researcher tested H_0: $p = 0.5$ versus H_1: $p > 0.5$ and obtained a P-value of 0.2743. Explain what this P-value means and write a conclusion for the researcher.

14. You Explain It! ESP Suppose an acquaintance claims to have the ability to determine the birth month of randomly selected individuals. To test such a claim, you randomly select 80 individuals and ask the acquaintance to state the birth month of the individual. If the individual has the ability to determine birth month, then the proportion of correct birth months should exceed $\frac{1}{12}$, the rate one would expect from simply guessing.

(a) State the null and alternative hypotheses for this experiment.
(b) Suppose the individual was able to guess nine correct birth months. The P-value for such results is 0.1726. Explain what this P-value means and write a conclusion for the test.

Applying the Concepts

15. Cramer Correct Less Than Half the Time? The website pundittracker.com keeps track of predictions made by individuals in finance, politics, sports, and entertainment. Jim Cramer is a famous TV financial personality and author. Pundittracker monitored 678 of his stock predictions (such as a recommendation to buy the stock) and found that 320 were correct predictions. Treat these 678 predictions as a random sample of all of Cramer's predictions.

(a) Determine the sample proportion of predictions Cramer got correct.
(b) Suppose that we want to know whether the evidence suggests Cramer is correct less than half the time. State the null and alternative hypotheses.
(c) Verify the normal model may be used to determine the P-value for this hypothesis test.
(d) Draw a normal model with area representing the P-value shaded for this hypothesis test.
(e) Determine the P-value based on the model from part (d).
(f) Interpret the P-value.
(g) Based on the P-value, what does the sample evidence suggest? That is, what is the conclusion of the hypothesis test? Assume an $\alpha = 0.05$ level of significance.

16. Political Pundits In his book, "The Signal and the Noise," Nate Silver analyzed 733 predictions made by experts regarding political events. Of the 733 predictions, 338 were mostly true.

(a) Determine the sample proportion of political predictions that were mostly true.
(b) Suppose that we want to know whether the evidence suggests the political predictions were mostly true less than half the time. State the null and alternative hypotheses.
(c) Verify the normal model may be used to determine the *P*-value for this hypothesis test.
(d) Draw a normal model with the area representing the *P*-value shaded for this hypothesis test.
(e) Determine the *P*-value based on the model from part (d).
(f) Interpret the *P*-value.
(g) Based on the *P*-value, what does the sample evidence suggest? That is, what is the conclusion of the hypothesis test? Assume an $\alpha = 0.1$ level of significance.

NW **17. Lipitor** The drug Lipitor is meant to reduce 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 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 significance?

18. 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. 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.

19. Fatal Traffic Accidents According to a certain government agency for a large country, the proportion of fatal traffic accidents in the country in which the driver had a positive blood alcohol concentration (BAC) is 0.37. Suppose a random sample of 106 traffic fatalities in a certain region results in 51 that involved a positive BAC. Does the sample evidence suggest that the region has a higher proportion of traffic fatalities involving a positive BAC than the country at the $\alpha = 0.1$ level of significance?

20. Eating Together In a previous poll, 29% of adults with children under the age of 18 reported that their family ate dinner together seven nights a week. Suppose that, in a more recent poll, 318 of 1157 adults with children under the age of 18 reported that their family ate dinner together seven nights a week. Is there sufficient evidence that the proportion of families with children under the age of 18 who eat dinner together seven nights a week has decreased? Use the $\alpha = 0.1$ significance level.

NW **21. Taught Enough Math?** In 1994, 52% of parents with children in high school felt it was a serious problem that high school students were not being taught enough math and science. A recent survey found that 256 of 800 parents with children in high school felt it was a serious problem that high school students were not being taught enough math and science. Do parents feel differently today than they did in 1994? Use the $\alpha = 0.05$ level of significance?
Source: Based on "Reality Check: Are Parents and Students Ready for More Math and Science?" *Public Agenda,* 2006.

22. Living Alone? In 2000, 58% of females aged 15 and older lived alone, according to the U.S. Census Bureau. A sociologist tests whether this percentage is different today by conducting a random sample of 500 females aged 15 and older and finds that 285 are living alone. Is there sufficient evidence at the $\alpha = 0.1$ level of significance to conclude the proportion has changed since 2000?

NW **23. Quality of Education** Several years ago, 39% of parents who had children in grades K–12 were satisfied with the quality of education the students receive. A recent poll asked 1,065 parents who have children in grades K–12 if they were satisfied with the quality of education the students receive. Of the 1,065 surveyed, 459 indicated that they were satisfied. Construct a 90% confidence interval to assess whether this represents evidence that parents' attitudes toward the quality of education have changed.

24. Infidelity According to menstuff.org, 22% of married men have "strayed" at least once during their married lives.
(a) Describe how you might go about administering a survey to assess the accuracy of this statement.
(b) A survey of 500 married men indicated that 122 have "strayed" at least once during their married life. Construct a 95% confidence interval for the population proportion of married men who have strayed. Use this interval to assess the accuracy of the statement made by menstuff.org.

25. Accuracy of the Drive Thru According to QSR Magazine, Chick-fil-A has the best accuracy of drive thru orders with 96.4% of all its drive thru orders filled correctly. The manager of a competing fast food restaurant wants to advertise that her drive thru is more accurate that Chick-fil-A. In a random sample of 350 drive thru orders, how many accurate orders would the manager need out of 350 to be able to claim her drive thru has a statistically significantly better accuracy record than Chick-fil-A at the 0.1 level of significance?

26. Talk to the Animals In an American Animal Hospital Association survey, 37% of respondents stated that they talk to their pets on the 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 telephone. Does the veterinarian have the right to be skeptical? Use a 0.05 level of significance.

NW **27. Small-Sample Hypothesis Test** Professors Honey Kirk and Diane Lerma of Palo Alto College developed a "learning community curriculum that blended the developmental mathematics and the reading curriculum with a structured emphasis on study skills." In a typical developmental mathematics course at Palo Alto College, 50% of the students complete the course with a letter grade of A, B, or C. In the experimental course, of the 16 students enrolled, 11 completed the course with a letter grade of A, B, or C. Do you believe the experimental course was effective at the $\alpha = 0.05$ level of significance?
(a) State the appropriate null and alternative hypotheses.
(b) Verify that the normal model may not be used to estimate the *P*-value.
(c) Explain why this is a binomial experiment.
(d) Determine the *P*-value using the binomial probability distribution. State your conclusion to the hypothesis test.
(e) Suppose the course is taught with 48 students and 33 complete the course with a letter grade of A, B, or C. Verify the normal model may now be used to estimate the *P*-value.
(f) Use the normal model to obtain and interpret the *P*-value. State your conclusion to the hypothesis test.
(g) Explain the role that sample size plays in the ability to reject statements in the null hypothesis.
Source: Kirk, Honey & Lerma, Diane, "Reading Your Way to Success in Mathematics: A Paired Course of Developmental Mathematics and Reading." *MathAMATYC Educator,* Vol. 1. No. 2, 2010.

28. Small-Sample Hypothesis Test In 1997, 4% of mothers smoked more than 21 cigarettes during their pregnancy. An obstetrician believes that the percentage of mothers who smoke 21 cigarettes or more is less than 4% today. She randomly selects

120 pregnant mothers and finds that 3 of them smoked 21 or more cigarettes during pregnancy. Does the sample data support the obstetrician's belief? Use the $\alpha = 0.05$ level of significance.

29. Small Sample Hypothesis Test: Super Bowl Investing From Super Bowl I (1967) through Super Bowl XXXI (1997), the stock market increased if an NFL team won the Super Bowl and decreased if an AFL team won. This condition held 28 out of 31 years.

(a) Suppose the likelihood of predicting the direction of the stock market (increasing or decreasing) in any given year is 0.50. Decide on the appropriate null and alternative hypotheses to test whether the outcome of the Super Bowl can be used to predict the direction of the stock market.

(b) Use the binomial probability distribution to determine the P-value for the hypothesis test from part (a).

(c) Comment on the dangers of using the outcome of the hypothesis test to judge investments. Be sure your comment includes a discussion of circumstances in which associations have a causal relationship.

30. Statistics in the Media A headline read, "More Than Half of Americans Say Federal Taxes Too High." The headline was based on a random sample of 1026 adult Americans in which 534 stated the amount of federal tax they have to pay is too high. Is this an accurate headline?

Source: Gallup Organization, April 14, 2014

31. Gender Income Inequality The Sullivan Statistics Survey II asked, "Do you believe there is an income inequality discrepancy between males and females when each has the same experience and education?" Go to the book's website to obtain the data file SullivanStatsSurveyII using the file format of your choice for the version of the text you are using. The data may be found under the column "GenderIncomeInequality." Treat the sample as a random sample of adult Americans. Do the survey results suggest a supermajority (more than 60%) of adult Americans believe there is income inequality among males and females with the same experience and education? Use an $\alpha = 0.05$ level of significance.

32. Political Philosophy According to Gallup, 21% of adult Americans consider themselves to be liberal. Respondents of the Sullivan Statistics Survey I were asked to disclose their political philosophy: Conservative, Liberal, Moderate. Go to the book's website to obtain the data file SullivanStatsSurveyI using the file format of your choice for the version of the text you are using. The data may be found under the column "Political philosophy." Treat the results of the survey as a random sample of adult Americans. Do the survey results suggest the proportion is higher than that reported by Gallup? Use an $\alpha = 0.05$ level of significance.

33. Are Spreads Accurate? For every NFL game, there is a team that is expected to win by a certain number of points. In betting parlance, this is called the spread. For example, if the Chicago Bears are expected to beat the Green Bay Packers by three points, a sports bettor would say, "Chicago is minus three." So, if the Bears lose to Green Bay, or do not win by more than three points, a bet on Chicago would be a loser. If point spreads are accurate, we would expect about half of all games played to result in the favored team winning (beating the spread) and about half of all games to result in the team favored to not beat the spread. The following data represent the results of 45 randomly selected games where a 0 indicates the favored team did not beat the spread and a 1 indicates the favored team beat the spread. Do the data suggest that sport books establish accurate spreads?

0	0	0	0	0
1	0	0	1	0
0	0	0	1	0
1	1	0	0	1
0	0	1	1	1
0	0	1	1	0
1	0	0	1	0
0	1	1	0	1
0	0	1	1	1

Source: http://www.vegasinsider.com

34. Accept versus Do Not Reject In the United States, historically, 40% of registered voters are Republican. Suppose you obtain a simple random sample of 320 registered voters and find 142 registered Republicans.

(a) Consider the hypotheses H_0: $p = 0.4$ versus H_1: $p > 0.4$. Explain what the researcher would be testing. Perform the test at the $\alpha = 0.05$ level of significance. Write a conclusion for the test.

(b) Consider the hypotheses H_0: $p = 0.41$ versus H_1: $p > 0.41$. Explain what the researcher would be testing. Perform the test at the $\alpha = 0.05$ level of significance. Write a conclusion for the test.

(c) Consider the hypotheses H_0: $p = 0.42$ versus H_1: $p > 0.42$. Explain what the researcher would be testing. Perform the test at the $\alpha = 0.05$ level of significance. Write a conclusion for the test.

(d) Based on the results of parts (a)–(c), write a few sentences that explain the difference between "accepting" the statement in the null hypothesis versus "not rejecting" the statement in the null hypothesis.

35. Interesting Results Suppose you wish to find out the answer to the age-old question, "Do Americans prefer Coke or Pepsi?" You conduct a blind taste test in which individuals are randomly asked to drink one of the colas first, followed by the other cola, and then asked to disclose which drink they prefer. Results of your taste test indicate that 53 of 100 individuals prefer Pepsi.

(a) Conduct a hypothesis test (preferably using technology) H_0: $p = p_0$ versus H_1: $p \neq p_0$ for $p_0 = 0.42, 0.43, 0.44, \ldots, 0.64$ at the $\alpha = 0.05$ level of significance. For which values of p_0 do you not reject the null hypothesis? What do each of the values of p_0 represent?

(b) Construct a 95% confidence interval for the proportion of individuals who prefer Pepsi.

(c) Suppose you changed the level of significance in conducting the hypothesis test to $\alpha = 0.01$? What would happen to the range of values of p_0 for which the null hypothesis is not rejected? Why does this make sense?

36. Simulation Simulate drawing 100 simple random samples of size $n = 40$ from a population whose proportion is 0.3.

(a) Test the null hypothesis H_0: $p = 0.3$ versus H_1: $p \neq 0.3$ for each simulated sample.

(b) If we test the hypothesis at the $\alpha = 0.1$ level of significance, how many of the 100 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) How do we know that a rejection of the null hypothesis results in making a Type I error in this situation?

37. Simulation: Predicting the Future Parapsychology (psi) is a field of study that deals with clairvoyance or precognition. Psi made its way back into the news when a professional, refereed journal published an article by Cornell psychologist Daryl Bem, in which he claimed to demonstrate that psi is a real phenomenon. In the article Bem stated that certain individuals behave today as if they already know what is going to happen in the future. That is, individuals adjust current behavior in anticipation of events that are going to happen in the future. Here, we will present a simplified version of Bem's research.

(a) Suppose an individual claims to have the ability to predict the color (red or black) of a card from a standard 52-card deck. Of course, simply by guessing we would expect the individual to get half the predictions correct, and half incorrect. What is the statement of no change or no effect in this type of experiment? What statement would we be looking to demonstrate? Based on this, what would be the null and alternative hypotheses?

(b) Suppose you ask the individual to guess the correct color of a card 40 times, and the alleged savant (wise person) guesses the correct color 24 times. Would you consider this to be convincing evidence that that individual can guess the color of the card at better than a 50/50 rate? To answer this question, we want to determine the likelihood of getting 24 or more colors correct even if the individual is simply guessing. To do this, we assume the individual is guessing so that the probability of a successful guess is 0.5. Explain how 40 coins flipped independently with heads representing a successful guess can be used to model the card-guessing experiment.

(c) Now, use a random number generator, or applet such as the Coin-Flip applet in StatCrunch to flip 40 fair coins, 1000 different times. What proportion of time did you observe 24 or more heads due to chance alone? What does this tell you? Do you believe the individual has the ability to guess card color based on the results of the simulation, or could the results simply have occurred due to chance?

(d) Explain why guessing card color (or flipping coins) 40 times and recording the number of correct guesses (or heads) is a binomial experiment.

(e) Use the binomial probability function to find the probability of at least 24 correct guesses in 40 trials assuming the probability of success is 0.5.

(f) Look at the graph of the outcomes of the simulation from part (c). Explain why the normal model might be used to estimate the probability of obtaining at least 24 correct guesses in 40 trials assuming the probability of success is 0.5. Use the model to estimate the *P*-value.

(g) Based on the probabilities found in parts (c), (e), and (f), what might you conclude about the alleged savants ability to predict card color?

38. Putting It Together: Lupus Based on historical birthing records, the proportion of males born worldwide is 0.51. In other words, the commonly held belief that boys are just as likely as girls is false. Systematic lupus erythematosus (SLE), or lupus for short, is a disease in which one's immune system attacks healthy cells and tissue by mistake. It is well known that lupus tends to exist more in females than in males. Researchers wondered, however, if families with a child who had lupus had a lower ratio of males to females than the general population. If this were true, it would suggest that something happens during conception that causes males to be conceived at a lower rate when the SLE gene is present. To determine if this hypothesis is true, the researchers obtained records of families with a child who had SLE. A total of 23 males and 79 females were found to have SLE. The 23 males with SLE had a total of 23 male siblings and 22 female siblings. The 79 females with SLE had a total of 69 male siblings and 80 female siblings.
Source: L.N. Moorthy, M.G.E. Peterson, K.B. Onel, and T.J.A. Lehman. "Do Children with Lupus Have Fewer Male Siblings" *Lupus* 2008 17:128–131, 2008.

(a) Explain why this is an observational study.

(b) Is the study retrospective or prospective? Why?

(c) There are a total of $23 + 69 = 92$ male siblings in the study. How many female siblings are in the study?

(d) Draw a relative frequency bar graph of gender of the siblings.

(e) Find a point estimate for the proportion of male siblings in families where one of the children has SLE.

(f) Does the sample evidence suggest that the proportion of male siblings in families where one of the children has SLE is less than 0.51, the accepted proportion of males born in the general population? Use the $\alpha = 0.05$ level of significance.

(g) Construct a 95% confidence interval for the proportion of male siblings in a family where one of the children has SLE.

39. Putting It Together: Naughty or Nice? Yale University graduate student J. Kiley Hamlin conducted an experiment in which 16 ten-month-old babies were asked to watch a climber character attempt to ascend a hill. On two occasions, the baby witnesses the character fail to make the climb. On the third attempt, the baby witnesses either a helper toy push the character up the hill or a hinderer toy prevent the character from making the ascent. The helper and hinderer toys were shown to each baby in a random fashion for a fixed amount of time. The baby was then placed in front of each toy and allowed to choose which toy he or she wished to play with. In 14 of the 16 cases, the baby chose the helper toy.
Source: J. Kiley Hamlin et al., "Social Evaluation by Preverbal Infants." *Nature*, Nov. 2007.

(a) Why is it important to randomly expose the baby to the helper or hinderer toy first?

(b) What would be the appropriate null and alternative hypotheses if the researcher is attempting to show that babies prefer helpers over hinderers?

(c) Use the binomial probability formula to determine the *P*-value for this test.

(d) In testing 12 six-month-old babies, all 12 preferred the helper toy. The *P*-value was reported as 0.0002. Interpret this result.

Explaining the Concepts

40. Explain what a *P*-value is. What is the criterion for rejecting the null hypothesis using the *P*-value approach?

41. Suppose a researcher is testing the hypothesis $H_0: p = 0.3$ versus $H_1: p > 0.3$ and she finds the *P*-value to be 0.19. Explain what this means. Would she reject the null hypothesis? Why?

42. Suppose we are testing the hypothesis $H_0: p = 0.65$ versus $H_1: p \neq 0.65$ and we find the *P*-value to be 0.02. Explain what this means. Would you reject the null hypothesis? Why?

43. 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.

44. The headline reporting the results of a poll conducted by the Gallup organization stated "Majority of Americans at Personal Best in the Morning." The results indicated that a survey of 1100 Americans resulted in 55% stating they were at their personal best in the morning. The poll's results were reported with a margin of error of 3%. Explain why the Gallup organization's headline is accurate.

45. Explain what "statistical significance" means.

10.3 Hypothesis Tests for a Population Mean

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

- Sampling distribution of $\bar{x}$ (Section 8.1, pp. 423–431)
- The t-distribution (Section 9.2, pp. 463–467)
- Using probabilities to identify unusual results (Section 5.1, p. 277)
- Confidence intervals for a mean (Section 9.2, pp. 467–469)

Objectives

1. Test hypotheses about a mean
2. Understand the difference between statistical significance and practical significance

1 Test Hypotheses about a Mean

In Section 8.1, we learned that the distribution of $\bar{x}$ is approximately normal with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ provided the population from which the sample was drawn is normally distributed or the sample size is sufficiently large (because of the Central Limit Theorem). So $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ follows a standard normal distribution. However, it is unreasonable to expect to know σ without knowing μ. This problem was resolved by William Gosset, who determined that $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ follows Student's t-distribution with $n - 1$ degrees of freedom. We use this distribution to perform hypothesis tests on a mean.

Testing hypotheses about a mean follows the same logic as testing a hypothesis about a population proportion. The only difference is that we use Student's t-distribution, rather than the normal distribution.

Testing Hypotheses Regarding a Population Mean

To test hypotheses regarding the population mean, use the following steps, provided that

- the sample is obtained using simple random sampling or from a randomized experiment.
- the sample has no outliers and the population from which the sample is drawn is normally distributed, or the sample size, n, is large $(n \geq 30)$.
- the sampled values are independent of each other.

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.

Classical Approach

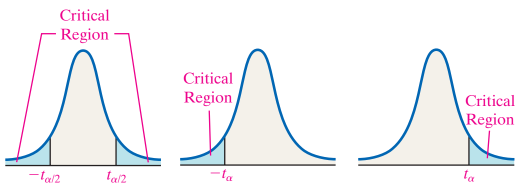
Step 3 Compute the **test statistic**

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

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

Use Table VII to determine the critical value.

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



Step 4 Compare the critical value to 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

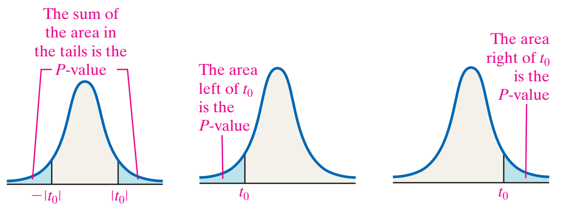
By Hand Step 3 Compute the **test statistic**

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

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

Use Table VII to approximate the P-value.

Two-Tailed	Left-Tailed	Right-Tailed



P-value $= 2P(t > \lvert t_0 \rvert)$	P-value $= P(t < t_0)$	P-value $= P(t > t_0)$

Technology Step 3 Use a statistical spreadsheet or calculator with statistical capabilities to obtain the P-value. The directions for obtaining the P-value using the TI-83/84 Plus graphing calculators, Minitab, Excel, and StatCrunch are in the Technology Step-by-Step on page 528.

Step 4 If the P-value $< \alpha$, reject the null hypothesis.

Step 5 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.

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 (Chapter 15).

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 VII) 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, use statistical software or a graphing calculator with advanced statistical features.

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

Problem The mean height of American males is 69.5 inches. The heights of the 43 male U.S. presidents* (Washington through Obama) have a mean 70.78 inches and a standard deviation of 2.77 inches. Treating the 43 presidents as a simple random sample, determine if there is evidence to suggest that U.S. presidents are taller than the average American male. Use the $\alpha = 0.05$ level of significance.

Approach Assume that all U.S. presidents come from a population whose height is 69.5 inches (that is, there is no difference between heights of U.S. presidents and the general American male population). Then determine the likelihood of obtaining a sample mean of 70.78 inches or higher from a population whose mean is 69.5 inches.

(continued)

*Grover Cleveland was elected to two non-consecutive terms, so there have technically been 44 presidents of the United States.

If the result is unlikely, reject the assumption stated in the null hypothesis in favor of the more likely notion that the mean height of U.S. presidents is greater than 69.5 inches. However, if obtaining a sample mean of 70.78 inches from a population whose mean is assumed to be 69.5 inches is not unusual, do not reject the null hypothesis (and attribute the difference to sampling error). Assume the population of potential U.S. presidents is large (for independence). Because the sample size is large, the distribution of $\bar{x}$ is approximately normal. Follow Steps 1 through 5.

Solution

Step 1 We want to know if U.S. presidents are taller than the typical American male who is 69.5 inches. We assume there is no difference between the height of a typical American male and U.S. presidents, so

$$H_0: \mu = 69.5 \text{ inches} \quad \text{versus} \quad H_1: \mu > 69.5 \text{ inches}$$

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

Classical Approach

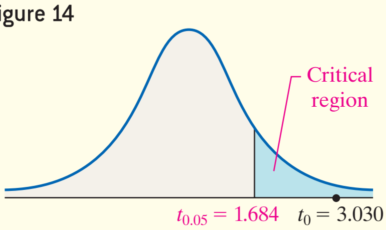
Step 3 The sample mean is $\bar{x} = 70.78$ inches and the sample standard deviation is $s = 2.77$ inches.

The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{70.78 - 69.5}{\frac{2.77}{\sqrt{43}}} = 3.030$$

Because this is a right-tailed test, determine the critical value at the $\alpha = 0.05$ level of significance with $43 - 1 = 42$ degrees of freedom to be $t_{0.05} = 1.684$ (using 40 degrees of freedom since this is closest to 42). The critical region is shown in Figure 14.

Figure 14



Step 4 The test statistic, $t_0 = 3.030$, is labeled in Figure 14. Because the test statistic lies in the critical region, reject the null hypothesis.

P-Value Approach

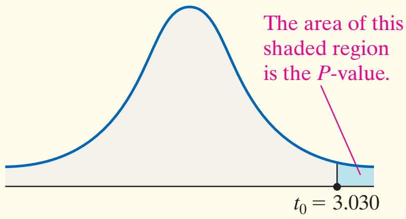
By Hand Step 3 The sample mean is $\bar{x} = 70.78$ inches and the sample standard deviation is $s = 2.77$ inches.

The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{70.78 - 69.5}{\frac{2.77}{\sqrt{43}}} = 3.030$$

Because this is a right-tailed test, the *P*-value is the area under the *t*-distribution with 42 degrees of freedom to the right of $t_0 = 3.030$ as shown in Figure 15.

Figure 15



Using Table VII, find the row that corresponds to 40 degrees of freedom (we use 40 degrees of freedom because it is closest to the actual degrees of freedom, $43 - 1 = 42$). The value 3.030 lies between 2.971 and 3.307. The value of 2.971 has an area of 0.0025 to the right under the *t*-distribution with 40 degrees of freedom. The area under the *t*-distribution with 40 degrees of freedom to the right of 3.307 is 0.001.

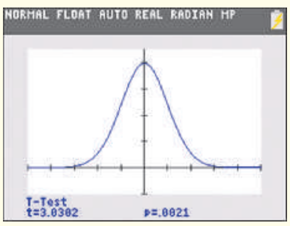
Because 3.030 is between 2.971 and 3.307, the *P*-value is between 0.001 and 0.0025. So

$$0.001 < P\text{-value} < 0.0025$$

There is an alternate form of Table VII (see Table XVII in Appendix A) that is useful for finding more accurate *P*-values. The table is set up similarly to Table V (the standard normal table). To use this alternate version, find the column that corresponds to the degrees of freedom, and the row that corresponds to the test statistic (rounded to the nearest tenth). The intersection of the row and column represents the area under the *t*-distribution to the right of the test statistic. Using 40 degrees of freedom (because 42 df is not in the table) with a test statistic of 3.0, the *P*-value is 0.002.

Technology Step 3 Using a TI-84 Plus C graphing calculator, the *P*-value is 0.0021. See Figure 16.

Figure 16



Note
If you are using Table XVII to find P-values for a left-tailed test, use the symmetry of the t-distribution. That is,

$$P(t < -t_0) = P(t > t_0)$$

Step 4 The P-value of 0.0021 [by hand: $0.001 < P\text{-value} < 0.0025$] means that, *if* the null hypothesis that $\mu = 69.5$ inches is true, we expect a sample mean of 70.78 inches or higher in about 2 out of 1000 samples. The results we obtained do not seem to be consistent with the assumption that the mean height of this population is 69.5 inches. Put another way, because the P-value is less than the level of significance, $\alpha = 0.05$ $(0.0021 < 0.05)$, we reject the null hypothesis.

• Now Work Problem 13

Step 5 There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that U.S. presidents are taller than the typical American male.

EXAMPLE 2 Testing a Hypothesis about a Population Mean: Small Sample

Problem The "fun size" of a Snickers bar is supposed to weigh 20 grams. Because the penalty for selling candy bars under their advertised weight is severe, the manufacturer calibrates the machine so the mean weight is 20.1 grams. The quality-control engineer at M&M–Mars, the Snickers manufacturer, is concerned about the calibration. He obtains a random sample of 11 candy bars, weighs them, and obtains the data shown in Table 1. Should the machine be shut down and calibrated? Because shutting down the plant is very expensive, he decides to conduct the test at the $\alpha = 0.01$ level of significance.

Table 1

19.68	20.66	19.56
19.98	20.65	19.61
20.55	20.36	21.02
21.50	19.74	

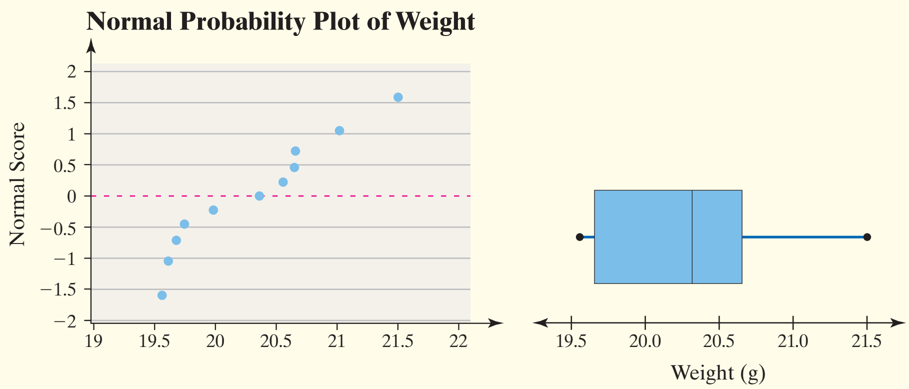
Source: Michael Carlisle, student at Joliet Junior College

Approach Assume that the machine is calibrated correctly. So there is no difference between the actual mean weight and the calibrated weight of the candy. We want to know whether the machine is incorrectly calibrated, which would result in a mean weight that is too high or too low. Therefore, this is a two-tailed test.

Before performing the hypothesis test, verify that the data come from a population that is normally distributed with no outliers by constructing a normal probability plot and boxplot. Then proceed to follow Steps 1 through 5.

Solution Figure 17 displays the normal probability plot and boxplot. The correlation between the weights and expected z-scores is 0.967 [Tech: 0.970]. Because $0.967 > 0.923$ (Table VI), the normal probability plot indicates that the data could come from a population that is approximately normal. The boxplot has no outliers. We can proceed with the hypothesis test.

Figure 17



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

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

This is a two-tailed test.

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

(continued)

Classical Approach

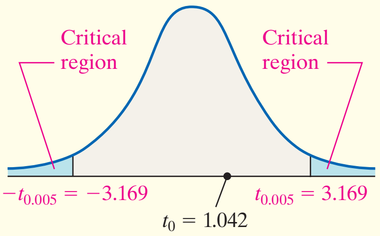
Step 3 From the data in Table 1, the sample mean is $\bar{x} = 20.301$ grams and the sample standard deviation is $s = 0.64$ gram.

The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{20.301 - 20.1}{\frac{0.64}{\sqrt{11}}} = 1.042$$

Because this is a two-tailed test, determine the critical values at the $\alpha = 0.01$ level of significance with $11 - 1 = 10$ 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 shown in Figure 18.

Figure 18



Step 4 Because the test statistic, $t_0 = 1.042$, does not lie in the critical region, do not reject the null hypothesis.

P-Value Approach

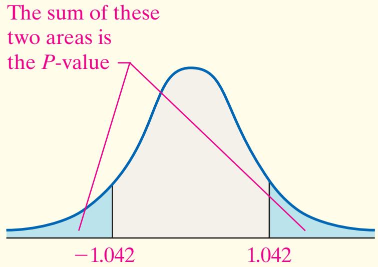
By Hand Step 3 From the data in Table 1, the sample mean is $\bar{x} = 20.301$ grams and the sample standard deviation is $s = 0.64$ gram.

The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{20.301 - 20.1}{\frac{0.64}{\sqrt{11}}} = 1.042$$

Because this is a two-tailed test, the *P*-value is the area under the *t*-distribution with $n - 1 = 11 - 1 = 10$ degrees of freedom to the left of $-t_0 = -1.042$ and to the right of $t_0 = 1.042$, as shown in Figure 19. That is, $P\text{-value} = P(t < -1.042) + P(t > 1.042) = 2P(t > 1.042)$, with 10 degrees of freedom.

Figure 19



Using Table VII, we find the row that corresponds to 10 degrees of freedom. The value 1.042 lies between 0.879 and 1.093. The value of 0.879 has an area of 0.20 to the right under the *t*-distribution. The area under the *t*-distribution to the right of 1.093 is 0.15.

Because 1.042 is between 0.879 and 1.093, the *P*-value is between 2(0.15) and 2(0.20). So

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

Using Table XVII we find $P\text{-value} = 2P(t > 1.0) = 2(0.170) = 0.340$ with 10 degrees of freedom.

Technology Step 3 Using Minitab, the exact *P*-value is 0.323.

Step 4 The *P*-value of 0.323 [by-hand: $0.30 < P\text{-value} < 0.40$] means that, *if* the null hypothesis that $\mu = 20.1$ grams is true, we expect about 32 out of 100 samples to result in a sample mean as extreme or more extreme than the one obtained. The result we obtained is not unusual, so we do not reject the null hypothesis.

Step 5 There is not sufficient evidence to conclude that the Snickers have a mean weight different from 20.1 grams at the $\alpha = 0.01$ level of significance. The machine should not be shut down. •

• **Now Work Problem 21**

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.

2 Understand the Difference between Statistical Significance and Practical Significance

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

EXAMPLE 3 Statistical versus Practical Significance

Problem According to the American Community Survey, the mean travel time to work in Collin County, Texas, in 2013 was 27.5 minutes. The Department of Transportation reprogrammed all the traffic lights in Collin County in an attempt to reduce travel time. To determine if there is evidence that travel time has decreased as a result of the reprogramming, the Department of Transportation obtains a random sample of 2500 commuters, records their travel time to work, and finds a sample mean of 27.2 minutes with a standard deviation of 8.5 minutes. Does this result suggest that travel time has decreased at the $\alpha = 0.05$ level of significance?

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

Solution

Step 1 The Department of Transportation wants to know if the mean travel time to work has decreased from 27.5 minutes. From this, we have

$$H_0: \mu = 27.5 \text{ minutes} \quad \text{versus} \quad H_1: \mu < 27.5 \text{ minutes}$$

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

Step 3 The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{27.2 - 27.5}{\frac{8.5}{\sqrt{2500}}} = -1.765$$

Classical Approach

Because this is a left-tailed test, the critical value with $\alpha = 0.05$ and $2500 - 1 = 2499$ degrees of freedom is $-t_{0.05} \approx -1.645$ (use the last row of Table VII when the degrees of freedom is greater than 1000).

Step 4 Because the test statistic is less than the critical value (the test statistic falls in the critical region), we reject the null hypothesis.

***P*-Value Approach**

Because this is a left-tailed test, $P\text{-value} = P(t_0 < -1.765)$. From Table VII, we find the approximate P-value is $0.025 < P\text{-value} < 0.05$ [Technology: P-value $= 0.0389$].

Step 4 Because the P-value is less than the level of significance, $\alpha = 0.05$, we reject the null hypothesis.

Step 5 There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude the mean travel time to work has decreased.

While the difference between 27.2 minutes and 27.5 minutes is statistically significant, it has no practical meaning. After all, is 0.3 minute (18 seconds) really going to make anyone feel better about his or her commute to work?

The reason that the results from Example 3 were statistically significant had to do with the large sample size. The moral of the story is this:

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

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.

Technology Step-by-Step Hypothesis Tests Regarding μ

TI-83/84 Plus

1. If necessary, enter raw data in L1.
2. Press STAT, highlight TESTS, and select 2:T-Test.
3. If the data are raw, highlight DATA; make sure that List 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.
4. Select the direction of the alternative hypothesis.
5. Highlight **Calculate** or **Draw** and press ENTER. The TI-83/84 gives the P-value.

Minitab

1. Enter raw data in column C1.
2. Select the **Stat** menu, highlight **Basic Statistics**, then highlight **1-Sample t . . . .**
3. If you have raw data, select "One or more samples, each in a column" from the drop-down menu. Place the cursor in the box, highlight the column containing the raw data, and click "Select." If you have summarized data, select "Summarized data" from the drop-down menu. Enter the sample size, sample mean, and sample standard deviation. Check the "Perform hypothesis test" box and enter the value of the population mean stated in the null hypothesis.
4. Click Options. Enter the direction of the alternative hypothesis. Click OK twice.

Excel

1. Enter raw data into Column A.
2. Load the XLSTAT Add-in, if necessary.
3. Select the XLSTAT menu and highlight Parametric tests. Select One-sample t-test and z-test.
4. Place the cursor in the Data: cell and then highlight the data in the spreadsheet. Check Student's t-test.
5. Click the Options tab. Choose the appropriate direction of the alternative hypothesis. Enter the mean stated in the null hypothesis in the Theoretical mean: cell. Enter the level of significance required for a confidence interval. For example, enter 10 for a 90% confidence interval. Click OK.

StatCrunch

1. If you have raw data, enter them into the spreadsheet. Name the column variable.
2. Select **Stat**, highlight **T Stats**, select **One Sample**, and then choose either **With Data** or **With Summary**.
3. If you chose **With Data**, select the column that has the observations. If you chose **With Summary**, enter the mean, standard deviation, and sample size. Choose the hypothesis test radio button. Enter the value of the mean stated in the null hypothesis and choose the direction of the alternative hypothesis from the pull-down menu. Click Compute!.

10.3 Assess Your Understanding

Skill Building

1. (a) Determine the critical value for a right-tailed test of a population mean at the $\alpha = 0.01$ level of significance with 15 degrees of freedom.
(b) Determine the critical value for a left-tailed test of a population mean at the $\alpha = 0.05$ level of significance based on a sample size of $n = 20$.
(c) Determine the critical values for a two-tailed test of a population mean at the $\alpha = 0.05$ level of significance based on a sample size of $n = 13$.

2. (a) Determine the critical value(s) for a right-tailed test of a population mean at the $\alpha = 0.01$ level of significance with 15 degrees of freedom.
(b) Determine the critical value(s) for a left-tailed test of a population mean at the $\alpha = 0.10$ level of significance based on a sample size of $n = 10$.
(c) Determine the critical value(s) for a two-tailed test of a population mean at the $\alpha = 0.01$ level of significance based on a sample size of $n = 14$.

3. 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?

4. 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?

5. 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?
(e) Construct a 99% confidence interval to test the hypothesis.

6. 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?

7. 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?

8. 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?

9. 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? Why?
(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) Approximate 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?

10. 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? Why?
(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) Approximate 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?
(f) Construct a 99% confidence interval to test the hypothesis.

Applying the Concepts

11. You Explain It! ATM Withdrawals According to the Crown ATM Network, the mean ATM withdrawal is \$67. PayEase, Inc., manufactures an ATM that allows one to pay bills (electric, water, parking tickets, and so on), as well as withdraw money. A review of 40 withdrawals shows the mean withdrawal is \$73 from a PayEase ATM machine. Do people withdraw more money from a PayEase ATM machine?

(a) Determine the appropriate null and alternative hypotheses to answer the question.
(b) Suppose the P-value for this test is 0.02. Explain what this value represents.
(c) Write a conclusion for this hypothesis test assuming an $\alpha = 0.05$ level of significance.

12. You Explain It! Are Women Getting Taller? Several years ago, the mean height of women 20 years of age or older was 63.7 inches. Suppose that a random sample of 45 women who are 20 years of age or older today results in a mean height of 64.2 inches.

(a) State the appropriate null and alternative hypotheses to assess whether women are taller today.
(b) Suppose the P-value for this test is 0.35. Explain what this value represents.
(c) Write a conclusion for this hypothesis test assuming an $\alpha = 0.10$ level of significance.

NW **13. Ready for College?** The ACT is a college entrance exam. ACT has determined that a score of 22 on the mathematics portion of the ACT suggests that a student is ready for college-level mathematics. To achieve this goal, ACT recommends that students take a core curriculum of math courses: Algebra I, Algebra II, and Geometry. Suppose a random sample of 200 students who completed this core set of courses results in a mean ACT math score of 22.6 with a standard deviation of 3.9. Do these results suggest that students who complete the core curriculum are ready for college-level mathematics? That is, are they scoring above 22 on the math portion of the ACT?

(a) State the appropriate null and alternative hypotheses.
(b) Verify that the requirements to perform the test using the t-distribution are satisfied.
(c) Use the classical or P-value approach at the $\alpha = 0.05$ level of significance to test the hypotheses in part (a).
(d) Write a conclusion based on your results to part (c).

14. SAT Verbal Scores Do students who learned English and another language simultaneously score worse on the SAT Critical Reading exam than the general population of test takers? The mean score among all test takers on the SAT Critical Reading exam is 501. A random sample of 100 test takers who learned English and another language simultaneously had a mean SAT Critical Reading score of 485 with a standard deviation of 116. Do these results suggest that students who learn English as well as another language simultaneously score worse on the SAT Critical Reading exam?

(a) State the appropriate null and alternative hypotheses.
(b) Verify that the requirements to perform the test using the t-distribution are satisfied.
(c) Use the classical or P-value approach at the $\alpha = 0.1$ level of significance to test the hypotheses in part (a).
(d) Write a conclusion based on your results to part (c).

15. Effects of Alcohol on the Brain 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 hippocampal region, the portion of the brain responsible for long-term memory storage, in adolescents. The researchers randomly selected 12 adolescents with alcohol use disorders to determine whether the hippocampal volumes in the alcoholic adolescents were less than the normal volume of 9.02 cubic centimeters (cm^3). An analysis of the sample data revealed that the hippocampal volume is approximately normal with $\bar{x} = 8.10\ \text{cm}^3$ and $s = 0.7\ \text{cm}^3$. Conduct the appropriate test at the $\alpha = 0.01$ level of significance.

16. 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 micrograms per liter ($\mu\text{g/L}$) 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 kidney weight of the 12 offspring was found to be 396.9 milligrams (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.

17. Credit Scores A Fair Isaac Corporation (FICO) score is used by credit agencies (such as mortgage companies and banks) to assess the creditworthiness of individuals. Values range from 300 to 850, with a FICO score over 700 considered to be a quality credit risk. According to Fair Isaac Corporation, the mean FICO score is 703.5. A credit analyst wondered whether high-income individuals (incomes in excess of \$100,000 per year) had higher credit scores. He obtained a random sample of 40 high-income individuals and found the sample mean credit score to be 714.2 with a standard deviation of 83.2. Conduct the appropriate test to determine if high-income individuals have higher FICO scores at the $\alpha = 0.05$ level of significance.

18. TVaholics According to the American Time Use Survey, the typical American spends 154.8 minutes (2.58 hours) per day watching television. A survey of 50 Internet users results in a mean time watching television per day of 128.7 minutes, with a standard deviation of 46.5 minutes. Conduct the appropriate test to determine if Internet users spend less time watching television at the $\alpha = 0.05$ level of significance.
Source: Norman H. Nie and D. Sunshine Hillygus. "Where Does Internet Time Come From? A Reconnaissance." *IT & Society,* 1(2).

19. Age of Death Row Inmates One year, the mean age of inmates on death row was 40.6 years. 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 39.2, with a standard deviation of 8.1. Construct a 95% confidence interval about the mean age. What does the interval imply?

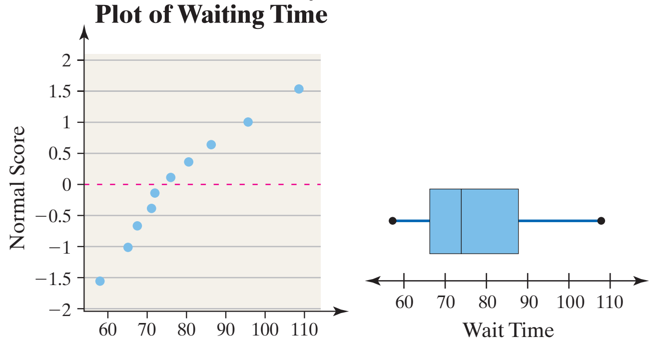
20. Energy Consumption In 2001, the mean household expenditure for energy was \$1493, according to data from the U.S. Energy Information Administration. An economist wanted to know whether this amount has changed significantly from its 2001 level. In a random sample of 35 households, he found the mean expenditure (in 2001 dollars) for energy during the most recent year to be \$1618, with a standard deviation \$321. Construct a 95% confidence interval about the mean energy expenditure. What does the interval imply?

NW **21. 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. The correlation between waiting time and expected z-scores is 0.971. Are the conditions for testing the hypothesis satisfied?

Normal Probability Plot of Waiting Time



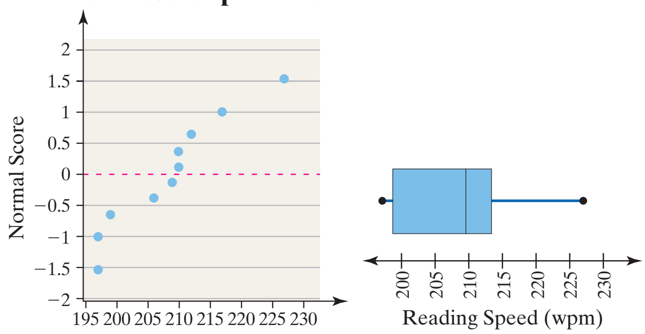
(b) Is the new system effective? Use the $\alpha = 0.1$ level of significance.

22. Reading Rates Michael Sullivan, son of the author, decided to enroll in a reading course that allegedly increases reading speed and comprehension. Prior to enrolling in the class, Michael read 198 words per minute (wpm). The following data represent the words per minute read for 10 different passages read after the course.

206	217	197	199	210
210	197	212	227	209

(a) Because the sample size is small, we must verify that reading speed is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. The correlation between reading rate and expected z-scores is 0.964. Are the conditions for testing the hypothesis satisfied?

Normal Probability Plot of Words per Minute



(b) Was the class effective? Use the $\alpha = 0.10$ level of significance.

23. Calcium in Rainwater Calcium is essential to tree growth. In 1990, the concentration of calcium in precipitation in Chautauqua, New York, was 0.11 milligram per liter (mg/L). A random sample of 10 precipitation dates in 2014 results in the following data:

0.065	0.087	0.070	0.262	0.126
0.183	0.120	0.234	0.313	0.108

Source: National Atmospheric Deposition Program

A normal probability plot suggests the data could come from a population that is normally distributed. A boxplot does not show any outliers. Does the sample evidence suggest that calcium concentrations have changed since 1990? Use the $\alpha = 0.05$ level of significance.

24. Filling Bottles A certain brand of apple juice is supposed to have 64 ounces of juice. Because the penalty 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, 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 normal probability plot suggests the data could come from a population that is normally distributed. A boxplot does not show any outliers.

(a) Should the assembly line be shut down so that the machine can be recalibrated? Use a 0.01 level of significance.

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

25. Starbucks Stock The volume of a stock is the number of shares traded for a given day. In 2011, Starbucks stock had a mean daily volume of 7.52 million shares according to Yahoo!Finance. A random sample of 40 trading days in 2014 was obtained and the volume of shares traded on those days was recorded. Go to the book's website to obtain the data file 10_3_25 using the file format of your choice for the version of the text you are using.

(a) Draw a histogram of the data. Describe the shape of the distribution.

(b) Draw a boxplot of the data. Are there any outliers?

(c) Based on the shape of the histogram and boxplot, explain why a large sample size is necessary to perform inference on the mean using the normal model.

(d) Does the evidence suggest that the volume of Starbucks stock has changed since 2014? Use an $\alpha = 0.05$ level of significance.

26. Study Time Go to the book's website to obtain the data file 10_3_26 using the file format of your choice for the version of the text you are using. The data represent the amount of time students in Sullivan's online statistics course spent studying for Section 4.1—Scatter Diagrams and Correlation.

(a) Draw a histogram of the data. Describe the shape of the distribution.

(b) Draw a boxplot of the data. Are there any outliers?

(c) Based on the shape of the histogram and boxplot, explain why a large sample size is necessary to perform inference on the mean using the normal model.

(d) According to MyStatLab, the mean time students would spend on this assignment nationwide is 95 minutes. Treat the data as a random sample of all Sullivan online statistics students. Do the sample data suggest that Sullivan's students are any different from the country as far as time spent on Section 4.1 goes? Use an $\alpha = 0.05$ level of significance.

Test the hypothesis in the problem given by constructing a 95% confidence interval.

27. Problem 23 **28.** Problem 24

29. Problem 25 **30.** Problem 26

31. 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 $\mu = 515$. 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 519 with a standard deviation of 111.

(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 519 significantly higher than 515?

(c) Do you think that a mean SAT math score of 519 versus 515 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 the same sample statistics. Is a sample mean of 519 significantly more than 515? What do you conclude about the impact of large samples on the hypothesis test?

32. 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 8 weeks to be 0.9 pound with standard deviation weight loss of 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 pound 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 the same sample statistics. Is a sample mean weight loss of 0.9 pound significantly more than 0 pound? What do you conclude about the impact of large samples on the hypothesis test?

33. Accept versus Do Not Reject The mean IQ score of humans is 100. Suppose the director of Institutional Research at Joliet Junior College (JJC) obtains a simple random sample of 40 JJC students and finds the mean IQ is 103.4 with a standard deviation of 13.2.

(a) Consider the hypotheses H_0: $\mu = 100$ versus H_1: $\mu > 100$. Explain what the director of Institutional Research is testing. Perform the test at the $\alpha = 0.05$ level of significance. Write a conclusion for the test.

(b) Consider the hypotheses H_0: $\mu = 101$ versus H_1: $\mu > 101$. Explain what the director of Institutional Research is testing. Perform the test at the $\alpha = 0.05$ level of significance. Write a conclusion for the test.

(c) Consider the hypotheses H_0: $\mu = 102$ versus H_1: $\mu > 102$. Explain what the director of Institutional Research is testing. Perform the test at the $\alpha = 0.05$ level of significance. Write a conclusion for the test.

(d) Based on the results of parts (a)–(c), write a few sentences that explain the difference between "accepting" the statement in the null hypothesis versus "not rejecting" the statement in the null hypothesis.

34. Simulation Simulate drawing 100 simple random samples of size $n = 15$ from a population that is normally distributed with mean 100 and standard deviation 15.

(a) Test the null hypothesis H_0: $\mu = 100$ versus H_1: $\mu \neq 100$ for each of the 100 simple random samples.

(b) If we test this hypothesis at the $\alpha = 0.05$ level of significance, how many of the 100 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.

35. Simulation The *exponential probability distribution* can be used to model waiting time in line or the lifetime of electronic components. Its density function is skewed right. Suppose the wait time in a line can be modeled by the exponential distribution with $\mu = \sigma = 5$ minutes.

(a) Simulate obtaining 100 simple random samples of size $n = 10$ from the population described. That is, simulate obtaining a simple random sample of 10 individuals waiting in a line where the wait time is expected to be 5 minutes.

(b) Test the null hypothesis H_0: $\mu = 5$ versus the alternative H_1: $\mu \neq 5$ for each of the 100 simulated simple random samples.

(c) If we test this hypothesis at the $\alpha = 0.05$ level of significance, how many of the 100 samples would you expect to result in a Type I error?

(d) Count the number of samples that lead to a rejection of the null hypothesis. Is it close to the expected value determined in part (c)? What might account for any discrepancies?

Retain Your Knowledge

36. Reading at Bedtime It is well-documented that watching TV, working on a computer, or any other activity involving artificial light can be harmful to sleep patterns. Researchers wanted to determine if the artificial light from e-Readers also disrupted sleep. In the study, 12 young adults were given either an iPad or printed book for four hours before bedtime. Then, they switched reading devices. Whether the individual received the iPad or book first was determined randomly. Bedtime was 10 P.M. and the time to fall asleep was measured each evening. It was found that participants took an average of 10 minutes longer to fall asleep after reading on an iPad. The *P*-value for the test was 0.009.

Source: Anne-Marie Chang, et. al. "Evening Use of Light-Emitting eReaders Negatively Affects Sleep, Circarian Timing, and Next-Morning Alertness" *PNAS* 2015 112(4) 1232-1277. doi:10.1073/pnas.1418490112

(a) What is the research objective?

(b) What is the response variable? It is quantitative or qualitative?

(c) What is the treatment?

(d) Is this a designed experiment or observational study? What type?

(e) The null hypothesis for this test would be that there is no difference in time to fall asleep with an e-Reader and printed book. The alternative is that there is a difference. Interpret the *P*-value.

Explaining the Concepts

37. What's the Problem? The head of institutional research at a university believes that the mean age of full-time students is declining. In 1995, the mean age of a full-time student was known to be 27.4 years. After looking at the enrollment records of all 4934 full-time students in the current semester, he found that the mean age was 27.1 years, with a standard deviation of 7.3 years. He conducted a hypothesis of H_0: $\mu = 27.4$ years versus H_1: $\mu < 27.4$ years and obtained a *P*-value of 0.0019. He concluded that the mean age of full-time students did decline. Is there anything wrong with his research?

38. The procedures for testing a hypothesis regarding a population mean are robust. What does this mean?

39. Explain the difference between *statistical significance* and *practical significance*.

40. Wanna Live Longer? Become a Chief Justice The life expectancy of a male during the course of the past 100 years is approximately 27,725 days. Go to Wikipedia.com and download the data that represent the life span of chief justices of Canada for those who have died. Conduct a test to determine whether the evidence suggests that chief justices of Canada live longer than the general population of males. Suggest a reason why the conclusion drawn may be flawed.

10.4 Hypothesis Tests for a Population Standard Deviation

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

- Confidence intervals for a population standard deviation (Section 9.3, pp. 479–481)

Objective ❶ Test hypotheses about a population standard deviation

In this section, we discuss methods for testing hypotheses regarding a population variance or standard deviation.

Why might we be interested in testing hypotheses regarding σ^2 or σ? Many production processes require not only accuracy on average (the mean) but also consistency. Consider a filling machine (such as a coffee machine) that over- and underfills cups, but, on average, fills correctly. Customers are not happy about underfilled cups, and they are dissatisfied