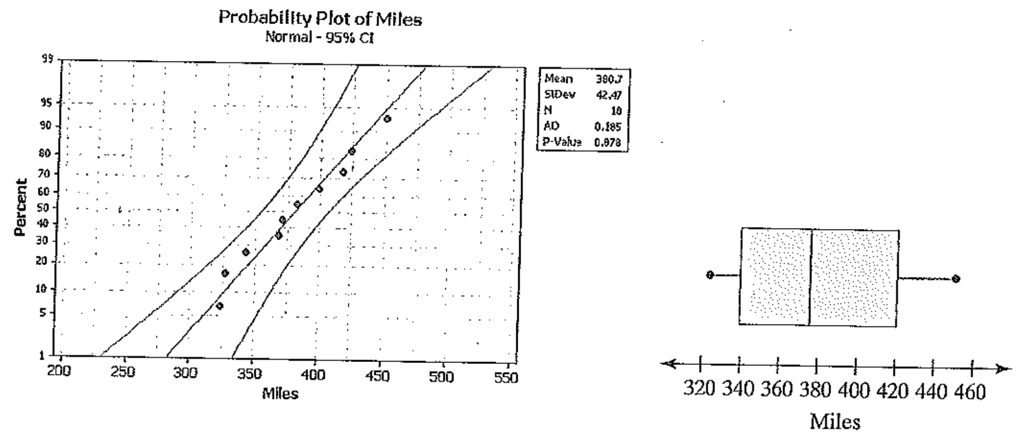
Solution We are asked to construct a 95% confidence interval for the *mean* number of miles driven. We will treat the data as a simple random sample from a large population. Because the sample size is small, we verify that the data come from a population that is normally distributed with no outliers by drawing a normal probability plot and boxplot. See Figure 18. The normal probability plot shows the data could come from a population that is normally distributed, and the boxplot shows there are no outliers. We may proceed with constructing the confidence interval for the population mean.

Figure 18



By-Hand Solution
From the sample data in Table 5, we have $n = 10$, $\bar{x} = 380.67$, and $s = 42.47$. For a 95% confidence interval with $n - 1 = 10 - 1 = 9$ degrees of freedom, we have

$$t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025} = 2.262$$

Lower bound:

$$\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 380.67 - 2.262 \cdot \frac{42.47}{\sqrt{10}} = 350.29$$

Upper bound:

$$\bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 380.67 + 2.262 \cdot \frac{42.47}{\sqrt{10}} = 411.05$$

Technology Solution
Figure 19 shows the results from StatCrunch.

Figure 19

95% confidence interval results:
μ : mean of Variable

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
Miles	380.67	13.429769	9	350.28976	411.05023

So

Lower bound: 350.29 Upper bound: 411.05

Interpretation We are 95% confident that the population mean miles driven on a full tank of gas is between 350.29 and 411.05.

9.3 ASSESS YOUR UNDERSTANDING

SKILL BUILDING

In Problems 1–8, construct the appropriate confidence interval.

1. A simple random sample of size $n = 300$ individuals who are currently employed is asked if they work at home at least once per week. Of the 300 employed individuals surveyed, 35 responded that they did work at home at least once per week. Construct a 99% confidence interval for the population proportion of employed individuals who work at home at least once per week.

2. A simple random sample of size $n = 785$ adults was asked if they follow college football. Of the 785 surveyed, 275 responded that they did follow college football. Construct a 95% confidence interval for the population proportion of adults who follow college football.

3. A simple random sample of size $n = 12$ is drawn from a population that is normally distributed. The sample mean is found to be $\bar{x} = 45$, and the sample standard deviation is found to be $s = 14$. Construct a 90% confidence interval for the population mean.

4. A simple random sample of size $n = 17$ is drawn from a population that is normally distributed. The sample mean is found to be $\bar{x} = 3.25$, and the sample standard deviation is found to be $s = 1.17$. Construct a 95% confidence interval for the population mean.

5. A simple random sample of size $n = 40$ is drawn from a population. The sample mean is found to be $\bar{x} = 120.5$, and the sample standard deviation is found to be $s = 12.9$. Construct a 99% confidence interval for the population mean.

6. A simple random sample of size $n = 210$ is drawn from a population. The sample mean is found to be $\bar{x} = 20.1$, and the sample standard deviation is found to be $s = 3.2$. Construct a 90% confidence interval for the population mean.

APPLYING THE CONCEPTS

7. **Aggravated Assault** In a random sample of 40 felons convicted of aggravated assault, it was determined that the mean length of sentencing was 54 months, with a standard deviation of 8 months. Construct and interpret a 95% confidence interval for the mean length of sentence for an aggravated assault conviction.
Source: Based on data from the U.S. Department of Justice.

8. **Click It** Based on a poll conducted by the Centers for Disease Control, 862 of 1013 randomly selected adults said that they always wear seat belts. Construct and interpret a 95% confidence interval for the proportion of adults who always wear seat belts.

9. **Estate Tax Returns** In a random sample of 100 estate tax returns that was audited by the Internal Revenue Service, it was determined that the mean amount of additional tax owed was \$3421 with a standard deviation of \$2583. Construct and interpret a 90% confidence interval for the mean additional amount of tax owed for estate tax returns.

10. **Muzzle Velocity** Fifty rounds of a new type of ammunition were fired from a test weapon, and the muzzle velocity of the projectile was measured. The sample had a mean muzzle velocity of 863 meters per second and a standard deviation of 2.7 meters per second. Construct and interpret a 99% confidence interval for the mean muzzle velocity.

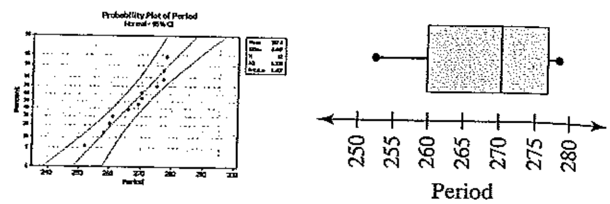
11. **Worried about Retirement?** In a survey of 1008 adult Americans the Gallup organization asked, "When you retire, do you think you will have enough money to live comfortably or not?" Of the 1008 surveyed, 526 stated that they were worried about having enough money to live comfortably in retirement. Construct a 90% confidence interval for the proportion of adult Americans who are worried about having enough money to live comfortably in retirement.

12. **Theme Park Spending** In a random sample of 40 visitors to a certain theme park, it was determined that the mean amount of money spent per person at the park (including ticket price) was \$93.43 per day with a standard deviation of \$15. Construct and interpret a 99% confidence interval for the mean amount spent daily per person at the theme park.

In Problems 13–16, decide if a 95% t-interval about the population mean can be constructed. If it can, do so. If it cannot be constructed, state the reason why. For convenience, a normal probability plot and boxplot are given.

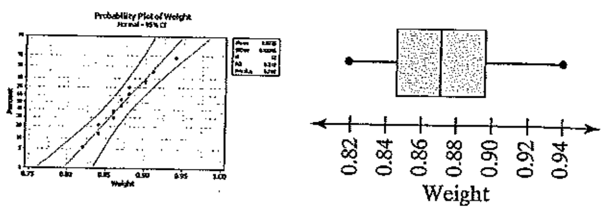
13. **Gestation Period** The following data represent the gestation period of a simple random sample of $n = 12$ live births.

266	270	277	278	258	275
261	260	270	269	252	277



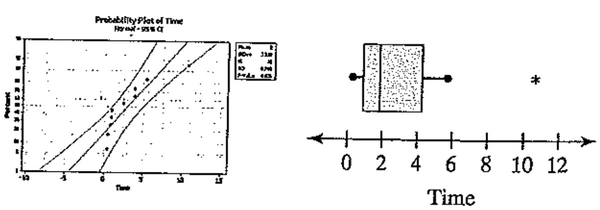
14. **M&Ms** A quality-control engineer wanted to estimate the mean weight (in grams) of a plain M&M candy.

0.87	0.88	0.82	0.90	0.86	0.86
0.84	0.84	0.91	0.94	0.88	0.87



15. **Officer Friendly** A police officer hides behind a billboard to catch speeders. The following data represent the number of minutes he waits before first observing a car that is exceeding the speed limit by more than 10 miles per hour on 10 randomly selected days:

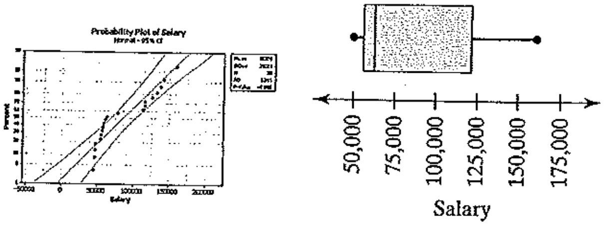
1.0	5.4	0.8	10.7	0.5
0.9	3.9	0.4	2.5	3.9



16. **Law Graduate Salaries** A random sample of recent graduates of law school was obtained in which the graduates were asked to report their starting salary. The data, based on results reported by the National Association for Law Placement, are as follows:

54,400	115,000	132,000	45,000	137,500
60,500	56,250	63,500	125,000	47,250
47,000	160,000	112,500	78,000	58,000
142,000	46,500	57,000	55,000	115,000

Source: NALP, July 2010.



3 State Conclusions to Hypothesis Tests

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.

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.

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

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

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

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

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

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

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

15. **Teenage Mothers** According to the U.S. Census Bureau, 10.5% of registered births in the United States in 2007 were to teenage mothers. A sociologist believes that this percentage has increased since then.

16. **Charitable Contributions** According to the Center on Philanthropy at Indiana University, the mean charitable contribution per household among households with income of $1 million or more in the United States in 2005 was $17,072. A researcher believes that the level of giving has changed since then.

NW 17. **Single-Family Home Price** According to the National Association of Home Builders, the mean price of an existing single-family home in 2009 was $218,600. A real estate broker believes that because of the recent credit crunch, the mean price has decreased since then.

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

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

20. **Overweight** According to the Centers for Disease Control and Prevention, 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 higher in her school district.

21. **Cell Phone Service** According to the *CTIA–The Wireless Association*, the mean monthly cell phone bill was $47.47 in 2010. A researcher suspects that the mean monthly cell phone bill is different today.

22. **SAT Math Scores** In 2010, the standard deviation 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. **Popcorn Consumption** According to popcorn.org, the mean consumption of popcorn annually by Americans is 54 quarts. The marketing division of popcorn.org 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 800 Americans provides enough evidence to conclude that the marketing campaign was effective. Provide a statement that should be put out by the marketing department.

(c) Suppose, in fact, that the mean annual consumption of popcorn after the marketing campaign is 53.4 quarts. Has a Type I or Type II error been made by the marketing department? If they tested the hypothesis at the $\alpha = 0.05$ level of significance, what is the probability of making a Type I error?

36. **Test Preparation** The mean score on the SAT Math Reasoning exam is 516. A test preparation company states that the mean score of students who take its course is higher than 516.

(a) Determine the null and alternative hypotheses.

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

(c) Suppose, in fact, that the mean score of students taking the preparatory course is 522. Has a Type I or Type II error been made? If we tested this hypothesis at the $\alpha = 0.01$ level, what is the probability of committing a Type I error?

(d) If we wanted to decrease the probability of making a Type II error, would we need to increase or decrease the level of significance?

37. **Marijuana Use** According to the Centers for Disease Control and Prevention, in 2008, 6.7% of 12- to 17-year-olds had used marijuana in the previous 6 months. The Drug Abuse and Resistance Education (DARE) program underwent several major changes to keep up with technology and issues facing students in the 21st century. After the changes, a school resource officer (SRO) thinks that the proportion of 12- to 17-year-olds who have used marijuana in the past 6 months has decreased.

(a) Determine the null and alternative hypotheses.

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

(c) Suppose, in fact, that the proportion of 12- to 17-year-olds who have used marijuana in the past 6 months is 6.4%. 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. Consumer Reports The following is an excerpt from a *Consumer Reports* article.

The platinum Gasaver makes some impressive claims. The device, $188 for two, is guaranteed to increase gas mileage by 22% says the manufacturer, National Fuelsaver. Also, the company quotes "the government" as concluding, "Independent testing shows greater fuel savings with Gasaver than the 22 percent claimed by the developer." Readers have told us they want to know more about it.

The Environmental Protection Agency (EPA), after its lab tests of the Platinum Gasaver, concluded, "Users of the device would not be expected to realize either an emission or fuel economy benefit." The Federal Trade Commission says, "No government agency endorses gas-saving products for cars."

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

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

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

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

41. Refer to 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.

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....*

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. 235)
- z_α notation (Section 7.2, pp. 344–345)
- Sampling distribution of the sample proportion (Section 8.2, pp. 379–384)
- Computing normal probabilities (Section 7.2, pp. 338–342)
- Binomial probability distribution (Section 6.2, pp. 309–319)

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

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

Classical Approach

Step 3(continued) 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 VI 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

Step 3 (continued) Because this is a left-tailed test, the P-value is P-value $= P(t_0 < -1.765)$. From Table VI, 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 27.0 minutes and 27.3 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?

CAUTION

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

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

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

10.3 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 for a right-tailed test of a population mean at the $\alpha = 0.1$ level of significance with 22 degrees of freedom.

(b) Determine the critical value for a left-tailed test of a population mean at the $\alpha = 0.01$ level of significance based on a sample size of $n = 40$.

(c) Determine the critical values for a two-tailed test of a population mean at the $\alpha = 0.01$ level of significance based on a sample size of $n = 33$.

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) Determine and interpret the P-value.

(e) If the researcher decides to test this hypothesis at the $\alpha = 0.01$ level of significance, will the researcher reject the null hypothesis? Why?

(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., manufacturers 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? In 1990, the mean height of women 20 years of age or older was 63.7 inches based on data obtained from the Centers for Disease Control and Prevention's *Advance Data Report*, No. 347. Suppose that a random sample of 45 women who are 20 years of age or older today results in a mean height of 63.9 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. In addition to administering this exam, researchers at ACT gauge high school students' readiness for college-level subjects. For example, 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 in high school. This core is 1 year of credit each in 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 as well as 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 as well as 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 development of the hippocampal region in adolescents. The hippocampus is the portion of the brain responsible for long-term memory storage. The researchers randomly selected 12 adolescents with alcohol use disorders. They wanted to determine whether the hippocampal volumes in the alcoholic adolescents were less than the normal volume of 9.02 cubic centimeters (cm^3). An analysis of the sample data revealed that the hippocampal volume is approximately normal with $\bar{x} = 8.10$ and $s = 0.7$. Conduct the appropriate test at the $\alpha = 0.01$ level of significance.

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 (μg/L) of para-nonylphenol in drinking water for 4 weeks. After 4 weeks, the mice were bred. After 100 days, the offspring of the exposed parents were sacrificed and the kidney weights were determined. The mean 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. Its value ranges from 300 to 850. An individual with a FICO score over 700 is 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. TValioliics According to the American Time Use Survey, the typical American spends 154.8 minutes (2.58 hours) per day watching television. Do Internet users spend less time each 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)

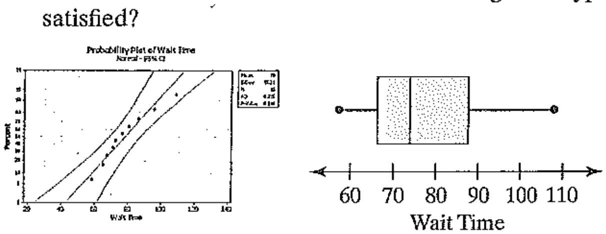
19. Age of Death-Row Inmates In 2002, the mean age of an inmate on death row was 40.7 years, according to data obtained from the U.S. Department of Justice. A sociologist wondered whether the mean age of a death-row inmate has changed since then. She randomly selects 32 death-row inmates and finds that their mean age is 38.9, with a standard deviation of 9.6. Construct a 95% confidence interval about the mean age. What does the interval imply?

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

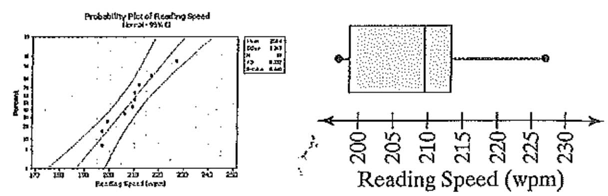


(b) Is the new system effective? Use the $\alpha = 0.1$ level of significance.

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. Are the conditions for testing the hypothesis satisfied?



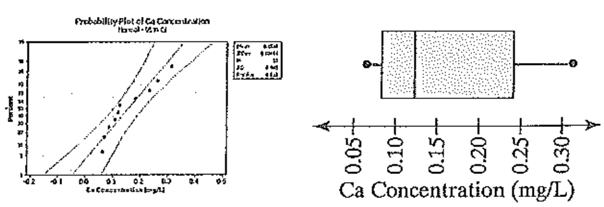
(b) Was the class effective? Use the $\alpha = 0.10$ level of significance.

23. Calcium in Rainwater Calcium is essential to tree growth because it promotes the formation of wood and maintains cell walls. 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 2010 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) Because the sample size is small, we must verify that calcium concentrations are normally distributed and the sample does not have any outliers. The normal probability plot and boxplot are shown. Are the conditions for conducting the hypothesis test satisfied?

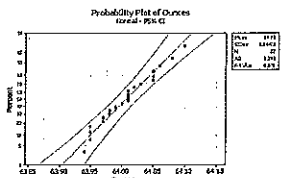


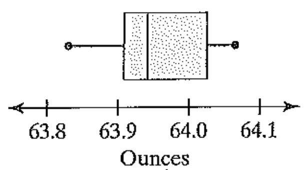
(b) Does the sample evidence suggest that calcium concentrations have changed since 1990? Use the $\alpha = 0.05$ level of significance.

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



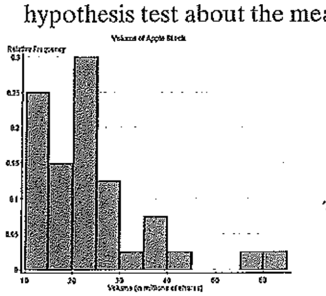


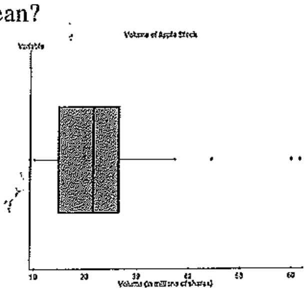
(b) Should the assembly line be shut down so that the machine can be recalibrated? Use a 0.01 level of significance.

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

25. Volume of Apple Stock In 2007, Apple stock had a mean daily volume of 35.1 million shares, according to Yahoo!Finance. A random sample of 40 trading days in 2010 resulted in a sample mean of 23.6 million shares with a standard deviation of 11.7 million shares.

(a) Based on the histogram and boxplot shown (from StatCrunch), why is a large sample necessary to conduct a hypothesis test about the mean?

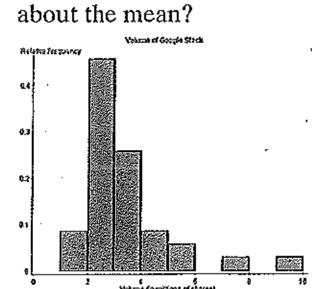


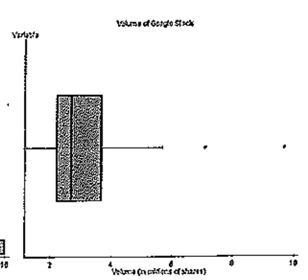


(b) Does the evidence suggest that the volume of Apple stock has changed since 2007? Use the $\alpha = 0.05$ level of significance.

26. Volume of Google Stock Google became a publicly traded company in August 2004. Initially, the stock traded over 10 million shares each day! Since the initial offering, the volume of stock traded daily has decreased substantially. In 2007, the mean daily volume in Google stock was 5.44 million shares, according to Yahoo!Finance. A random sample of 35 trading days in 2010 resulted in a sample mean of 3.28 million shares with a standard deviation of 1.68 million shares.

(a) Based on the histogram and boxplot shown (from StatCrunch), why is a large sample necessary to conduct a hypothesis test about the mean?





(b) Does the evidence suggest that the volume of Google stock has changed since 2007? Use the $\alpha = 0.05$ level of significance.

27. Using Confidence Intervals to Test Hypotheses Test the hypotheses in Problem 23 by constructing a 99% confidence interval.

28. Using Confidence Intervals to Test Hypotheses Test the hypotheses in Problem 24 by constructing a 95% confidence interval.

29. Using Confidence Intervals to Test Hypotheses Test the hypotheses in Problem 25 by constructing a 95% confidence interval.

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

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 P-value ?

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 P-value?

33. Sullivan Statistics Survey and Credit Card Debt According to the credit reporting agency Transunion, the mean credit card debt in the United States among individuals with credit cards for the period between April and June of 2010 was $4951. Treating the results of the StatCrunch survey as a random sample of U.S. residents, determine whether there is evidence to suggest that credit card debt is lower than $4951. Use the $\alpha = 0.05$ level of significance.

34. Sullivan Statistics Survey and Televisions According to the Nielsen's Television Audience Report, the mean number of televisions per household in 2009 was 2.9. Treating the results of the StatCrunch survey as a random sample of U.S. residents, determine whether there is evidence to suggest the number of televisions per household has risen since then. Use the $\alpha = 0.05$ level of significance.

(a) Retrieve the results of the number of televisions per household from the Sullivan Statistics Survey. What is the mean and standard deviation number of televisions per household in the survey?

(b) Draw a boxplot of the data. Describe the shape of the distribution. Are there any outliers? Do you believe it is reasonable to use the normal model to describe the distribution of the sample mean? Why?

(c) Conduct the appropriate test to determine if the results of the survey suggest the number of televisions per household has increased since 2009.

35. 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.

36. 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.

37. 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?

38. Putting It Together: Analyzing a Study An abstract in a journal article is a short synopsis of the study. The following abstract is from an article in the *British Medical Journal* (BMJ). The article uses *relative risk* in the analysis, which is common in health-related studies. **Relative risk** represents how many times more likely an individual is to have some condition when compared to an individual who does not have the condition.

Childhood Cancer in Relation to Distance from High Voltage Power Lines in England and Wales: A Case-Control Study

Objective To determine whether there is an association between distance of home address at birth from high voltage power lines and the incidence of leukemia and other cancers in children in England and Wales.

Design Case-control study.

Setting Cancer registry and National Grid records.

Subjects Records of 29,081 children with cancer, including 9,700 with leukemia. Children were aged 0–14 years and born in England and Wales, 1962–95. Controls were individually matched for sex, approximate date of birth, and birth registration district. No active participation was required.

Main outcome measures Distance from home address at birth to the nearest high voltage overhead power line in existence at the time.

Results Compared with those who lived >600 m from a line at birth, children who lived within 200 m had a relative risk of leukemia of 1.69 (95% confidence interval 1.13 to 2.53); those born between 200 and 600 m had a relative risk of 1.23 (1.02 to 1.49). There was a significant (*P*-value < 0.01) trend in risk in relation to the reciprocal of distance from the line. No excess risk in relation to proximity to lines was found for other childhood cancers.

Conclusions There is an association between childhood leukemia and proximity of home address at birth to high voltage power lines, and the apparent risk extends to a greater distance than would have been expected from previous studies. About 4% of children in England and Wales live within 600 m of high voltage lines at birth. If the association is causal, about 1% of childhood leukemia in England and Wales would be attributable to these lines, though this estimate has considerable statistical uncertainty. There is no accepted biological mechanism to explain the epidemiological results; indeed, the relation may be due to chance or confounding. *Source: BMJ* 2005;330:1290 (4 June), doi:10.1136/bmj.330.7503.1290

(a) Describe the statistical process. Include an interpretation of the confidence interval and *P*-value.

(b) Why is this study an observational study? What is a case-control study? Explain why a research objective such as this does not lend itself to a designed experiment.

EXPLAINING THE CONCEPT

39. What's the Problem? The head of institutional research at a university believed that the mean age of full-time students was 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?

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

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

42. 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 lifespan 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.

Consumer Reports® Eyeglass Lenses

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

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

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

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

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

(a) Suppose it is known that the closest competitor to the manufacturer's lens has a mean haze difference of 0.6. Does the evidence indicate that this manufacturer's lenses are more scratch resistant?

(b) Write a paragraph for the readers of *Consumer Reports* magazine that explains your findings.

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