- The reasoning of a significance test is as follows. Suppose that the null hypothesis is true. If we repeated our data production many times, would we often get data as inconsistent with H_0, in the direction specified by H_a, as the data we actually have? If the data are unlikely when H_0 is true, they provide evidence against H_0 and in favor of H_a.
- The **P-value** of a test is the probability, computed supposing H_0 to be true, that the statistic will take a value at least as extreme as the observed result in the direction specified by H_a.
- Small *P*-values indicate strong evidence against H_0. To calculate a *P*-value, we must know the sampling distribution of the test statistic when H_0 is true.
- If the *P*-value is smaller than a specified value α (called the **significance level**), the data are **statistically significant at level α**. In that case, we can **reject** H_0 and say that we have convincing evidence for H_a. If the *P*-value is greater than or equal to α, we **fail to reject** H_0 and say that we do *not* have convincing evidence for H_a.
- A **Type I error** occurs if we reject H_0 when it is in fact true. In other words, the data give convincing evidence for H_a when the null hypothesis is correct. A **Type II error** occurs if we fail to reject H_0 when H_a is true. In other words, the data don't give convincing evidence for H_a, even though the alternative hypothesis is correct.
- In a fixed level α significance test, the probability of a Type I error is the significance level α.

Section 9.1 Exercises

In Exercises 1 to 6, each situation calls for a significance test. State the appropriate null hypothesis H_0 and alternative hypothesis H_a in each case. Be sure to define your parameter each time.

1. **Attitudes** The Survey of Study Habits and Attitudes (SSHA) is a psychological test that measures students' attitudes toward school and study habits. Scores range from 0 to 200. The mean score for U.S. college students is about 115. A teacher suspects that older students have better attitudes toward school. She gives the SSHA to an SRS of 45 of the over 1000 students at her college who are at least 30 years of age.

pg 540

2. **Anemia** Hemoglobin is a protein in red blood cells that carries oxygen from the lungs to body tissues. People with less than 12 grams of hemoglobin per deciliter of blood (g/dl) are anemic. A public health official in Jordan suspects that Jordanian children are at risk of anemia. He measures a random sample of 50 children.

3. **Lefties** Simon reads a newspaper report claiming that 12% of all adults in the United States are left-handed. He wonders if the proportion of lefties at his large community college is really 12%. Simon chooses an SRS of 100 students and records whether each student is right- or left-handed.

4. **Don't argue!** A Gallup Poll report revealed that 72% of teens said they seldom or never argue with their friends.[5] Yvonne wonders whether this result holds true in her large high school. So she surveys a random sample of 150 students at her school.

5. **Cold cabin?** During the winter months, the temperatures at the Colorado cabin owned by the Starnes family can stay well below freezing (32°F or 0°C) for weeks at a time. To prevent the pipes from freezing, Mrs. Starnes sets the thermostat at 50°F. The manufacturer claims that the thermostat allows variation in home temperature of $\sigma = 3°F$. Mrs. Starnes suspects that the manufacturer is overstating how well the thermostat works.

6. **Ski jump** When ski jumpers take off, the distance they fly varies considerably depending on their speed, skill, and wind conditions. Event organizers

must position the landing area to allow for differences in the distances that the athletes fly. For a particular competition, the organizers estimate that the variation in distance flown by the athletes will be $\sigma = 10$ meters. An experienced jumper thinks that the organizers are underestimating the variation.

In Exercises 7 to 10, explain what's wrong with the stated hypotheses. Then give correct hypotheses.

7. **Better parking** A change is made that should improve student satisfaction with the parking situation at a local high school. Right now, 37% of students approve of the parking that's provided. The null hypothesis $H_0: p > 0.37$ is tested against the alternative $H_a: p = 0.37$.

8. **Better parking** A change is made that should improve student satisfaction with the parking situation at your school. Right now, 37% of students approve of the parking that's provided. The null hypothesis $H_0: \hat{p} = 0.37$ is tested against the alternative $H_a: \hat{p} \neq 0.37$.

9. **Birth weights** In planning a study of the birth weights of babies whose mothers did not see a doctor before delivery, a researcher states the hypotheses as

$$H_0: \bar{x} = 1000 \text{ grams}$$

$$H_a: \bar{x} < 1000 \text{ grams}$$

10. **Birth weights** In planning a study of the birth weights of babies whose mothers did not see a doctor before delivery, a researcher states the hypotheses as

$$H_0: \mu < 1000 \text{ grams}$$

$$H_a: \mu = 900 \text{ grams}$$

11. **Attitudes** In the study of older students' attitudes from Exercise 1, the sample mean SSHA score was 125.7 and the sample standard deviation was 29.8. A significance test yields a *P*-value of 0.0101.

pg 544

(a) Explain what it would mean for the null hypothesis to be true in this setting.

(b) Interpret the *P*-value in context.

12. **Anemia** For the study of Jordanian children in Exercise 2, the sample mean hemoglobin level was 11.3 g/dl and the sample standard deviation was 1.6 g/dl. A significance test yields a *P*-value of 0.0016.

(a) Explain what it would mean for the null hypothesis to be true in this setting.

(b) Interpret the *P*-value in context.

13. **Lefties** Refer to Exercise 3. In Simon's SRS, 16 of the students were left-handed. A significance test yields a *P*-value of 0.2184. What conclusion would you make if $\alpha = 0.10$? If $\alpha = 0.05$? Justify your answers.

pg 546

14. **Don't argue!** Refer to Exercise 4. For Yvonne's survey, 96 students in the sample said they rarely or never argue with friends. A significance test yields a *P*-value of 0.0291. What conclusion would you make if $\alpha = 0.05$? If $\alpha = 0.01$? Justify your answers.

15. **Attitudes** Refer to Exercise 11. What conclusion would you make if $\alpha = 0.05$? If $\alpha = 0.01$? Justify your answers.

16. **Anemia** Refer to Exercise 12. What conclusion would you make if $\alpha = 0.05$? If $\alpha = 0.01$? Justify your answers.

17. **Interpreting a *P*-value** When asked to explain the meaning of the *P*-value in Exercise 13, a student says, "This means there is about a 22% chance that the null hypothesis is true." Explain why the student's explanation is wrong.

18. **Interpreting a *P*-value** When asked to explain the meaning of the *P*-value in Exercise 14, a student says, "There is a 0.0291 probability of getting a sample proportion of $\hat{p} = 96/150 = 0.64$ by chance alone." Explain why the student's explanation is wrong.

19. **Drawing conclusions** A student performs a test of $H_0: p = 0.75$ versus $H_a: p > 0.75$ and gets a *P*-value of 0.99. The student writes: "Because the *P*-value is greater than 0.75, we reject H_0. The data prove that H_a is true." Explain what is wrong with this conclusion.

20. **Drawing conclusions** A student performs a test of $H_0: p = 0.5$ versus $H_a: p \neq 0.5$ and gets a *P*-value of 0.63. The student writes: "Because the *P*-value is greater than $\alpha = 0.05$, we accept H_0. The data provide convincing evidence that the null hypothesis is true." Explain what is wrong with this conclusion.

Exercises 21 and 22 refer to the following setting. Slow response times by paramedics, firefighters, and policemen can have serious consequences for accident victims. In the case of life-threatening injuries, victims generally need medical attention within 8 minutes of the accident. Several cities have begun to monitor emergency response times. In one such city, the mean response time to all accidents involving life-threatening injuries last year was $\mu = 6.7$ minutes. Emergency personnel arrived within 8 minutes on 78% of all calls involving life-threatening injuries last year. The city manager shares this information and encourages these first responders to "do better." At the end of the year, the city manager selects an SRS of 400 calls involving life-threatening injuries and examines the response times.

21. **Awful accidents**

pg 548

(a) State hypotheses for a significance test to determine whether the average response time has decreased. Be sure to define the parameter of interest.

(b) Describe a Type I error and a Type II error in this setting, and explain the consequences of each.

(c) Which is more serious in this setting: a Type I error or a Type II error? Justify your answer.

22. Awful accidents

(a) State hypotheses for a significance test to determine whether first responders are arriving within 8 minutes of the call more often. Be sure to define the parameter of interest.

(b) Describe a Type I error and a Type II error in this setting and explain the consequences of each.

(c) Which is more serious in this setting: a Type I error or a Type II error? Justify your answer.

23. **Opening a restaurant** You are thinking about opening a restaurant and are searching for a good location. From research you have done, you know that the mean income of those living near the restaurant must be over $85,000 to support the type of upscale restaurant you wish to open. You decide to take a simple random sample of 50 people living near one potential location. Based on the mean income of this sample, you will decide whether to open a restaurant there.[6]

(a) State appropriate null and alternative hypotheses. Be sure to define your parameter.

(b) Describe a Type I and a Type II error, and explain the consequences of each.

24. **Blood pressure screening** Your company markets a computerized device for detecting high blood pressure. The device measures an individual's blood pressure once per hour at a randomly selected time throughout a 12-hour period. Then it calculates the mean systolic (top number) pressure for the sample of measurements. Based on the sample results, the device determines whether there is convincing evidence that the individual's actual mean systolic pressure is greater than 130. If so, it recommends that the person seek medical attention.

(a) State appropriate null and alternative hypotheses in this setting. Be sure to define your parameter.

(b) Describe a Type I and a Type II error, and explain the consequences of each.

Multiple choice: Select the best answer for Exercises 25 to 28.

25. Experiments on learning in animals sometimes measure how long it takes mice to find their way through a maze. The mean time is 18 seconds for one particular maze. A researcher thinks that a loud noise will cause the mice to complete the maze faster. She measures how long each of 10 mice takes with a noise as stimulus. The appropriate hypotheses for the significance test are

(a) $H_0: \mu = 18; H_a: \mu \neq 18.$

(b) $H_0: \mu = 18; H_a: \mu > 18.$

(c) $H_0: \mu < 18; H_a: \mu = 18.$

(d) $H_0: \mu = 18; H_a: \mu < 18.$

(e) $H_0: \bar{x} = 18; H_a: \bar{x} < 18.$

Exercises 26–28 refer to the following setting. Members of the city council want to know if a majority of city residents supports a 1% increase in the sales tax to fund road repairs. To investigate, they survey a random sample of 300 city residents and use the results to test the following hypotheses:

$$H_0: p = 0.50$$
$$H_a: p > 0.50$$

where p is the proportion of all city residents who support a 1% increase in the sales tax to fund road repairs.

26. A Type I error in the context of this study occurs if the city council

(a) finds convincing evidence that a majority of residents supports the tax increase, when in reality there isn't convincing evidence that a majority supports the increase.

(b) finds convincing evidence that a majority of residents supports the tax increase, when in reality at most 50% of city residents support the increase.

(c) finds convincing evidence that a majority of residents supports the tax increase, when in reality more than 50% of city residents do support the increase.

(d) does not find convincing evidence that a majority of residents supports the tax increase, when in reality more than 50% of city residents do support the increase.

(e) does not find convincing evidence that a majority of residents supports the tax increase, when in reality at most 50% of city residents do support the increase.

27. In the sample, $\hat{p} = 158/300 = 0.527$. The resulting *P*-value is 0.18. What is the correct interpretation of this *P*-value?

(a) Only 18% of the city residents support the tax increase.

(b) There is an 18% chance that the majority of residents supports the tax increase.

(c) Assuming that 50% of residents support the tax increase, there is an 18% probability that the sample proportion would be 0.527 or higher by chance alone.

(d) Assuming that more than 50% of residents support the tax increase, there is an 18% probability that the sample proportion would be 0.527 or higher by chance alone.

(e) Assuming that 50% of residents support the tax increase, there is an 18% chance that the null hypothesis is true by chance alone.

28. Based on the P-value in Exercise 27, which of the following would be the most appropriate conclusion?
(a) Because the P-value is large, we reject H_0. We have convincing evidence that more than 50% of city residents support the tax increase.
(b) Because the P-value is large, we fail to reject H_0. We have convincing evidence that more than 50% of city residents support the tax increase.
(c) Because the P-value is large, we reject H_0. We have convincing evidence that at most 50% of city residents support the tax increase.
(d) Because the P-value is large, we fail to reject H_0. We have convincing evidence that at most 50% of city residents support the tax increase.
(e) Because the P-value is large, we fail to reject H_0. We do not have convincing evidence that more than 50% of city residents support the tax increase.

29. **Women in math** (5.3) Of the 24,611 degrees in mathematics given by U.S. colleges and universities in a recent year, 70% were bachelor's degrees, 24% were master's degrees, and the rest were doctorates. Moreover, women earned 43% of the bachelor's degrees, 41% of the master's degrees, and 29% of the doctorates.[7]
(a) How many of the mathematics degrees given in this year were earned by women? Justify your answer.
(b) Are the events "degree earned by a woman" and "degree was a bachelor's degree" independent? Justify your answer using appropriate probabilities.
(c) If you choose 2 of the 24,611 mathematics degrees at random, what is the probability that at least 1 of the 2 degrees was earned by a woman? Show your work.

30. **Explaining confidence** (8.2) Here is an explanation from a newspaper concerning one of its opinion polls. Explain what is wrong with the following statement.

For a poll of 1,600 adults, the variation due to sampling error is no more than three percentage points either way. The error margin is said to be valid at the 95 percent confidence level. This means that, if the same questions were repeated in 20 polls, the results of at least 19 surveys would be within three percentage points of the results of this survey.

9.2 Tests about a Population Proportion

WHAT YOU WILL LEARN By the end of the section, you should be able to:

- State and check the Random, 10%, and Large Counts conditions for performing a significance test about a population proportion.
- Perform a significance test about a population proportion.
- Interpret the power of a test and describe what factors affect the power of a test.
- Describe the relationship among the probability of a Type I error (significance level), the probability of a Type II error, and the power of a test.

Confidence intervals and significance tests are based on the sampling distributions of statistics. That is, both use probability to say what would happen if we used the inference method many times. Section 9.1 presented the reasoning of significance tests, including the idea of a P-value. In this section, we focus on the details of testing a claim about a population proportion.

Carrying Out a Significance Test

In Section 9.1, we met a virtual basketball player who claimed to make 80% of ... In an SRS of 50 shots