TI-83/84 PLUS

To do the mechanics of a hypothesis test for a proportion:

- Select **5:1-PropZTest** from the **STAT TESTS** menu.
- Specify the hypothesized proportion.
- Enter the observed value of x.
- Specify the sample size.
- Indicate what kind of test you want: one-tail lower tail, two-tail, or one-tail upper tail.
- Calculate the result.

COMMENTS

Beware: When you enter the value of x, you need the *count*, not the percentage. The count must be a whole number. If the number of successes is given as a percent, you must first multiply np and round the result to obtain x.

Exercises

Section 17.1

1. Better than aspirin? A very large study showed that aspirin reduced the rate of first heart attacks by 44%. A pharmaceutical company thinks they have a drug that will be more effective than aspirin, and plans to do a randomized clinical trial to test the new drug.

a) What is the null hypothesis the company will use?
b) What is their alternative hypothesis?

2. Psychic A friend of yours claims to be psychic. You are skeptical. To test this you take a stack of 100 playing cards and have your friend try to identify the suit (hearts, diamonds, clubs, or spades), without looking, of course!

a) State the null hypothesis for your experiment.
b) State the alternative hypothesis.

Section 17.2

3. Better than aspirin 2? A clinical trial compares the new drug described in Exercise 1 to aspirin. The group using the new drug had somewhat fewer heart attacks than those in the aspirin group.

a) The P-value from the hypothesis test was 0.28. What do you conclude?
b) What would you have concluded if the P-value had been 0.004?

4. Psychic after all? The "psychic" friend from Exercise 2 correctly identified more than 25% of the cards.

a) A hypothesis test gave a P-value of 0.014. What do you conclude?
b) What would you conclude if the P-value had been 0.245?

Section 17.3

5. Hispanic origin According to the 2010 Census, 16% of the people in the United States are of Hispanic or Latino origin. One county supervisor believes her county has a different proportion of Hispanic people than the nation as a whole. She looks at their most recent survey data, which was a random sample of 437 county residents, and found that 44 of those surveyed are of Hispanic origin.

a) State the hypotheses.
b) Name the model and check appropriate conditions for a hypothesis test.
c) Draw and label a sketch, and then calculate the test statistic and P-value.
d) State your conclusion.

6. Empty houses According to the 2010 Census, 11.4% of all housing units in the United States were vacant. A county supervisor wonders if her county is different from this. She randomly selects 850 housing units in her county and finds that 129 of the housing units are vacant.

a) State the hypotheses.
b) Name the model and check appropriate conditions for a hypothesis test.
c) Draw and label a sketch, and then calculate the test statistic and P-value.
d) State your conclusion.

Section 17.4

7. Psychic again (you should have seen this coming) If you were to do a hypothesis test on your experiment with your "psychic" friend from Exercise 2, would your alternative hypothesis be one-sided or two-sided? Explain why.

8. Hispanic origin II The county supervisor in Exercise 5 is going to perform a hypothesis test to see if she has evidence that the proportion of people in her county that are of Hispanic or Latino origin is different from that for the nation. Should her alternative hypothesis be one-sided or two-sided? Explain.

Section 17.5

9. Bad medicine Occasionally, a report comes out that a drug that cures some disease turns out to have a nasty side effect. For example, some antidepressant drugs may cause suicidal thoughts in younger patients. A researcher wants to study such a drug and look for evidence that such side effects exist.

a) If the test yields a low P-value and the researcher rejects the null hypothesis, but there is actually no ill side effect of the drug, what are the consequences of such an error?

b) If the test yields a high P-value and the researcher fails to reject the null hypothesis, but there *is* a bad side effect of the drug, what are the consequences of such an error?

10. Expensive medicine Developing a new drug can be an expensive process, resulting in high costs to patients. A pharmaceutical company has developed a new drug to reduce cholesterol, and it will conduct a clinical trial to compare the effectiveness to the most widely used current treatment. The results will be analyzed using a hypothesis test.

a) If the test yields a low P-value and the researcher rejects the null hypothesis that the new drug is not more effective, but it actually is not better, what are the consequences of such an error?

b) If the test yields a high P-value and the researcher fails to reject the null hypothesis, but the new drug *is* more effective, what are the consequences of such an error?

Chapter Exercises

11. Hypotheses Write the null and alternative hypotheses you would use to test each of the following situations:

a) A governor is concerned about his "negatives"—the percentage of state residents who express disapproval of his job performance. His political committee pays for a series of TV ads, hoping that they can keep the negatives below 30%. They will use follow-up polling to assess the ads' effectiveness.

b) Is a coin fair?

c) Only about 20% of people who try to quit smoking succeed. Sellers of a motivational tape claim that listening to the recorded messages can help people quit.

12. More hypotheses Write the null and alternative hypotheses you would use to test each situation.

a) In the 1950s, only about 40% of high school graduates went on to college. Has the percentage changed?

b) Twenty percent of cars of a certain model have needed costly transmission work after being driven between 50,000 and 100,000 miles. The manufacturer hopes that a redesign of a transmission component has solved this problem.

c) We field-test a new-flavor soft drink, planning to market it only if we are sure that over 60% of the people like the flavor.

13. Negatives After the political ad campaign described in Exercise 11, part a, pollsters check the governor's negatives. They test the hypothesis that the ads produced no change against the alternative that the negatives are now below 30% and find a P-value of 0.22. Which conclusion is appropriate? Explain.

a) There's a 22% chance that the ads worked.

b) There's a 78% chance that the ads worked.

c) There's a 22% chance that their poll is correct.

d) There's a 22% chance that natural sampling variation could produce poll results like these if there's really no change in public opinion.

14. Dice The seller of a loaded die claims that it will favor the outcome 6. We don't believe that claim, and roll the die 200 times to test an appropriate hypothesis. Our P-value turns out to be 0.03. Which conclusion is appropriate? Explain.

a) There's a 3% chance that the die is fair.

b) There's a 97% chance that the die is fair.

c) There's a 3% chance that a loaded die could randomly produce the results we observed, so it's reasonable to conclude that the die is fair.

d) There's a 3% chance that a fair die could randomly produce the results we observed, so it's reasonable to conclude that the die is loaded.

15. Relief A company's old antacid formula provided relief for 70% of the people who used it. The company tests a new formula to see if it is better and gets a P-value of 0.27. Is it reasonable to conclude that the new formula and the old one are equally effective? Explain.

16. Cars A survey investigating whether the proportion of today's high school seniors who own their own cars is higher than it was a decade ago finds a P-value of 0.017. Is it reasonable to conclude that more high schoolers have cars? Explain.

17. He cheats? A friend of yours claims that when he tosses a coin he can control the outcome. You are skeptical and want him to prove it. He tosses the coin, and you call heads; it's tails. You try again and lose again.

a) Do two losses in a row convince you that he really can control the toss? Explain.

b) You try a third time, and again you lose. What's the probability of losing three tosses in a row if the process is fair?

c) Would three losses in a row convince you that your friend controls the outcome? Explain.

d) How many times in a row would you have to lose to be pretty sure that this friend really can control the toss? Justify your answer by calculating a probability and explaining what it means.

18. Candy Someone hands you a box of a dozen chocolate-covered candies, telling you that half are vanilla creams and the other half peanut butter. You pick candies at random and discover the first three you eat are all vanilla.

a) If there really were 6 vanilla and 6 peanut butter candies in the box, what is the probability that you would have picked three vanillas in a row?

b) Do you think there really might have been 6 of each? Explain.

c) Would you continue to believe that half are vanilla if the fourth one you try is also vanilla? Explain.

19. Smartphones Many people have trouble setting up all the features of their smartphones, so a company has developed what it hopes will be easier instructions. The goal is to have at least 96% of customers succeed. The company tests the new system on 200 people, of whom 188 were successful. Is this strong evidence that the new system fails to meet the company's goal? A student's test of this hypothesis is shown. How many mistakes can you find?

$H_0: \hat{p} = 0.96$

$H_A: \hat{p} \neq 0.96$

SRS, $0.96(200) > 10$

$$\frac{188}{200} = 0.94; \quad SD(\hat{p}) = \sqrt{\frac{(0.94)(0.06)}{200}} = 0.017$$

$$z = \frac{0.96 - 0.94}{0.017} = 1.18$$

$$P = P(z > 1.18) = 0.12$$

There is strong evidence the new instructions don't work.

20. Obesity 2008 In 2008, the Centers for Disease Control and Prevention reported that 34% of adults in the United States are obese. A county health service planning a new awareness campaign polls a random sample of 750 adults living there. In this sample, 228 people were found to be obese based on their answers to a health questionnaire. Do these responses provide strong evidence that the 34% figure is not accurate for this region? Correct the mistakes you find in a student's attempt to test an appropriate hypothesis.

$H_0: \hat{p} = 0.34$

$H_A: \hat{p} < 0.34$

SRS, $750 \geq 10$

$$\frac{228}{750} = 0.304; \quad SD(\hat{p}) = \sqrt{\frac{(0.304)(0.696)}{750}} = 0.017$$

$$z = \frac{0.304 - 0.34}{0.017} = -2$$

$$P = P(z > -2) = 0.977$$

There is more than a 97% chance that the stated percentage is correct for this region.

21. Dowsing In a rural area, only about 30% of the wells that are drilled find adequate water at a depth of 100 feet or less. A local man claims to be able to find water by "dowsing"—using a forked stick to indicate where the well should be drilled. You check with 80 of his customers and find that 27 have wells less than 100 feet deep. What do you conclude about his claim?

a) Write appropriate hypotheses.

b) Check the necessary assumptions and conditions.

c) Perform the mechanics of the test. What is the P-value?

d) Explain carefully what the P-value means in context.

e) What's your conclusion?

22. Abnormalities In the 1980s, it was generally believed that congenital abnormalities affected about 5% of the nation's children. Some people believe that the increase in the number of chemicals in the environment has led to an increase in the incidence of abnormalities. A recent study examined 384 children and found that 46 of them showed signs of an abnormality. Is this strong evidence that the risk has increased?

a) Write appropriate hypotheses.

b) Check the necessary assumptions and conditions.

c) Perform the mechanics of the test. What is the P-value?

d) Explain carefully what the P-value means in context.

e) What's your conclusion?

f) Do environmental chemicals cause congenital abnormalities?

23. Absentees The National Center for Education Statistics monitors many aspects of elementary and secondary education nationwide. Their 1996 numbers are often used as a baseline to assess changes. In 1996, 34% of students had not been absent from school even once during the previous month. In the 2000 survey, responses from 8302 students showed that this figure had slipped to 33%. Officials would, of course, be concerned if student attendance were declining. Do these figures give evidence of a change in student attendance?

a) Write appropriate hypotheses.
b) Check the assumptions and conditions.
c) Perform the test and find the P-value.
d) State your conclusion.
e) Do you think this difference is meaningful? Explain.

24. Educated mothers The National Center for Education Statistics monitors many aspects of elementary and secondary education nationwide. Their 1996 numbers are often used as a baseline to assess changes. In 1996, 31% of students reported that their mothers had graduated from college. In 2000, responses from 8368 students found that this figure had grown to 32%. Is this evidence of a change in education level among mothers?

a) Write appropriate hypotheses.
b) Check the assumptions and conditions.
c) Perform the test and find the P-value.
d) State your conclusion.
e) Do you think this difference is meaningful? Explain.

25. Contributions, please, part II In Exercise 23 of Chapter 16, you learned that the Paralyzed Veterans of America is a philanthropic organization that relies on contributions. They send free mailing labels and greeting cards to potential donors on their list and ask for a voluntary contribution. To test a new campaign, the organization recently sent letters to a random sample of 100,000 potential donors and received 4781 donations. They've had a contribution rate of 5% in past campaigns, but a staff member worries that the rate will be lower if they run this campaign as currently designed.

a) What are the hypotheses?
b) Are the assumptions and conditions for inference met?
c) Do you think the rate would drop? Explain.

26. Take the offer, part II In Exercise 24 of Chapter 16, you learned that First USA, a major credit card company, is planning a new offer for their current cardholders. First USA will give double airline miles on purchases for the next 6 months if the cardholder goes online and registers for this offer. To test the effectiveness of this campaign, the company recently sent out offers to a random sample of 50,000 cardholders. Of those, 1184 registered. A staff member suspects that the success rate for the full campaign will be comparable to the standard 2% rate that they are used to seeing in similar campaigns. What do you predict?

a) What are the hypotheses?
b) Are the assumptions and conditions for inference met?
c) Do you think the rate would change if they use this fundraising campaign? Explain.

27. Law school 2007 According to the Law School Admission Council, in the fall of 2007, 66% of law school applicants were accepted to some law school.[4] The training program *LSATisfaction* claims that 163 of the 240 students trained in 2007 were admitted to law school. You can safely consider these trainees to be representative of the population of law school applicants. Has *LSATisfaction* demonstrated a real improvement over the national average?

a) What are the hypotheses?
b) Check the conditions and find the P-value.
c) Would you recommend this program based on what you see here? Explain.

28. Med school 2011 According to the Association of American Medical Colleges, only 46% of medical school applicants were admitted to a medical school in the fall of 2011.[5] Upon hearing this, the trustees of Striving College expressed concern that only 77 of the 180 students in their class of 2011 who applied to medical school were admitted. The college president assured the trustees that this was just the kind of year-to-year fluctuation in fortunes that is to be expected and that, in fact, the school's success rate was consistent with the national average. Who is right?

a) What are the hypotheses?
b) Check the conditions and find the P-value.
c) Are the trustees right to be concerned, or is the president correct? Explain.

29. Pollution A company with a fleet of 150 cars found that the emissions systems of 7 out of the 22 they tested failed to meet pollution control guidelines. Is this strong evidence that more than 20% of the fleet might be out of compliance? Test an appropriate hypothesis and state your conclusion. Be sure the appropriate assumptions and conditions are satisfied before you proceed.

30. Scratch and dent An appliance manufacturer stockpiles washers and dryers in a large warehouse for shipment to retail stores. Sometimes in handling them the appliances get damaged. Even though the damage may be minor, the company must sell those machines at drastically reduced prices. The company goal is to keep the level

[4]As reported by the Cornell office of career services in their *Class of 2007 Postgraduate Report.*

[5]www.aamc.org/data/facts/applicantmatriculant/

of damaged machines below 2%. One day an inspector randomly checks 60 washers and finds that 5 of them have scratches or dents. Is this strong evidence that the warehouse is failing to meet the company goal? Test an appropriate hypothesis and state your conclusion. Be sure the appropriate assumptions and conditions are satisfied before you proceed.

31. **Twins** In 2009, a national vital statistics report indicated that about 3% of all births produced twins. Is the rate of twin births the same among very young mothers? Data from a large city hospital found that only 7 sets of twins were born to 469 teenage girls. Test an appropriate hypothesis and state your conclusion. Be sure the appropriate assumptions and conditions are satisfied before you proceed.

32. **Football 2010** During the 2010 season, the home team won 143 of the 246 regular-season National Football League games. Is this strong evidence of a home field advantage in professional football? Test an appropriate hypothesis and state your conclusion. Be sure the appropriate assumptions and conditions are satisfied before you proceed.

33. **WebZine** A magazine is considering the launch of an online edition. The magazine plans to go ahead only if it's convinced that more than 25% of current readers would subscribe. The magazine contacted a simple random sample of 500 current subscribers, and 137 of those surveyed expressed interest. What should the company do? Test an appropriate hypothesis and state your conclusion. Be sure the appropriate assumptions and conditions are satisfied before you proceed.

34. **Seeds** A garden center wants to store leftover packets of vegetable seeds for sale the following spring, but the center is concerned that the seeds may not germinate at the same rate a year later. The manager finds a packet of last year's green bean seeds and plants them as a test. Although the packet claims a germination rate of 92%, only 171 of 200 test seeds sprout. Is this evidence that the seeds have lost viability during a year in storage? Test an appropriate hypothesis and state your conclusion. Be sure the appropriate assumptions and conditions are satisfied before you proceed.

35. **Women executives** A company is criticized because only 13 of 43 people in executive-level positions are women. The company explains that although this proportion is lower than it might wish, it's not a surprising value given that only 40% of all its employees are women. What do you think? Test an appropriate hypothesis and state your conclusion. Be sure the appropriate assumptions and conditions are satisfied before you proceed.

36. **Jury** Census data for a certain county show that 19% of the adult residents are Hispanic. Suppose 72 people are called for jury duty and only 9 of them are Hispanic. Does this apparent underrepresentation of Hispanics call into question the fairness of the jury selection system? Explain.

37. **Dropouts** Some people are concerned that new tougher standards and high-stakes tests adopted in many states have driven up the high school dropout rate. The National Center for Education Statistics reported that the high school dropout rate for the year 2004 was 10.3%. One school district whose dropout rate has always been very close to the national average reports that 210 of their 1782 high school students dropped out last year. Is this evidence that their dropout rate may be increasing? Explain.

38. **Acid rain** A study of the effects of acid rain on trees in the Hopkins Forest shows that 25 of 100 trees sampled exhibited some sort of damage from acid rain. This rate seemed to be higher than the 15% quoted in a recent *Environmetrics* article on the average proportion of damaged trees in the Northeast. Does the sample suggest that trees in the Hopkins Forest are more susceptible than trees from the rest of the region? Comment, and write up your own conclusions based on an appropriate confidence interval as well as a hypothesis test. Include any assumptions you made about the data.

39. **Lost luggage** An airline's public relations department says that the airline rarely loses passengers' luggage. It further claims that on those occasions when luggage is lost, 90% is recovered and delivered to its owner within 24 hours. A consumer group that surveyed a large number of air travelers found that only 103 of 122 people who lost luggage on that airline were reunited with the missing items by the next day. Does this cast doubt on the airline's claim? Explain.

40. **TV ads** A start-up company is about to market a new computer printer. It decides to gamble by running commercials during the Super Bowl. The company hopes that name recognition will be worth the high cost of the ads. The goal of the company is that over 40% of the public recognize its brand name and associate it with computer equipment. The day after the game, a pollster contacts 420 randomly chosen adults and finds that 181 of them know that this company manufactures printers. Would you recommend that the company continue to advertise during Super Bowls? Explain.

41. **John Wayne** Like a lot of other Americans, John Wayne died of cancer. But is there more to this story? In 1955, Wayne was in Utah shooting the film *The Conqueror*. Across the state line, in Nevada, the United States

military was testing atomic bombs. Radioactive fallout from those tests drifted across the filming location. A total of 46 of the 220 people working on the film eventually died of cancer. Cancer experts estimate that one would expect only about 30 cancer deaths in a group this size.

a) Is the death rate among the movie crew unusually high?
b) Does this prove that exposure to radiation increases the risk of cancer?

42. AP Stats The College Board reported that 58.7% of all students who took the 2010 AP Statistics exam earned scores of 3 or higher. One teacher wondered if the performance of her school was better. She believed that year's students to be typical of those who will take AP Stats at that school and was pleased when 34 of her 54 students achieved scores of 3 or better. Can she claim that her school is better? Explain.

✓ Just Checking ANSWERS

1. You can't conclude that the null hypothesis is true. You can conclude only that the experiment was unable to reject the null hypothesis. They were unable, on the basis of 12 patients, to show that aspirin was effective.
2. The null hypothesis is H_0: $p = 0.75$.
3. With a P-value of 0.0001, this is very strong evidence against the null hypothesis. We can reject H_0 and conclude that the improved version of the drug gives relief to a higher proportion of patients.
4. The parameter of interest is the proportion, p, of all delinquent customers who will pay their bills. H_0: $p = 0.30$ and H_A: $p > 0.30$.
5. The very low P-value leads us to reject the null hypothesis. There is strong evidence that the DVD is more effective in getting people to start paying their debts than just sending a letter had been.
6. All we know is that there is strong evidence to suggest that $p > 0.30$. We don't know how much higher than 30% the new proportion is. We'd like to see a confidence interval to see if the new method is worth the cost.

STATCRUNCH

To do inference for a mean using summaries:

- Click on **Stat**.
- Choose T **Statistics » One sample » with summary**.
- Enter the **Sample mean, Sample std dev**, and **Sample size.**
- Click on **Next**.
- Indicate **HypothesisTest**, then enter the hypothesized **Null mean**, and choose the **Alternative** hypothesis.

  OR

  Indicate **Confidence Interval**, and then enter the **Level** of confidence.
- Click on **Calculate.**

To do inference for a mean using data:

- Click on **Stat**.
- Choose T **Statistics » One sample » with data**.
- Choose the variable **Column**.
- Click on **Next**.
- Indicate **HypothesisTest**, then enter the hypothesized **Null mean**, and choose the **Alternative** hypothesis.

  OR

  Indicate **Confidence Interval**, and then enter the **Level** of confidence.
- Click on **Calculate**.

TI-83/84 PLUS

Finding a confidence interval:

- In the **STAT TESTS** menu, choose **8:TInterval**.
- Specify whether you are using data stored in a list or whether you will enter the mean, standard deviation, and sample size.
- You must also specify the desired level of confidence.

Testing a hypothesis:

- In the **STATTESTS** menu, choose **2:T-Test**. You may specify that you are using data stored in a list, or you may enter the mean, standard deviation, and size of your sample. You must also specify the hypothesized model mean and whether the test is to be two-tail, lower-tail, or upper-tail.

Exercises

Section 18.1

1. Salmon A specialty food company sells whole King Salmon to various customers. The mean weight of these salmon is 35 pounds with a standard deviation of 2 pounds. The company ships them to restaurants in boxes of 4 salmon, to grocery stores in cartons of 16 salmon, and to discount outlet stores in pallets of 100 salmon. To forecast costs, the shipping department needs to estimate the standard deviation of the mean weight of the salmon in each type of shipment

a) Find the standard deviations of the mean weight of the salmon in each type of shipment.

b) The distribution of the salmon weights turns out to be skewed to the high end. Would the distribution of shipping weights be better characterized by a Normal model for the boxes or pallets? Explain.

2. LSAT The LSAT (a test taken for law school admission) has a mean score of 151 with a standard deviation of 9 and a unimodal, symmetric distribution of scores. A test preparation organization teaches small classes of 9 students at a time. A larger organization teaches classes of 25 students at a time. Both organizations publish the mean scores of all their classes.

a) What would you expect the distribution of mean class scores to be for each organization?

b) If either organization has a graduating class with a mean score of 160, they'll take out a full-page ad in the local school paper to advertise. Which organization is more likely to have that success? Explain.

c) Both organizations advertise that if any class has an average score below 145, they'll pay for everyone to retake the LSAT. Which organization is at greater risk to have to pay?

Section 18.2

3. *t*-models, part I Using the t tables, software, or a calculator, estimate

a) the critical value of t for a 90% confidence interval with $df = 17$.

b) the critical value of t for a 98% confidence interval with df $= 88$.

4. *t*-models, part II Using the t tables, software, or a calculator, estimate

a) the critical value of t for a 95% confidence interval with df $= 7$.

b) the critical value of t for a 99% confidence interval with df $= 102$.

5. *t*-models, part III Describe how the shape, center, and spread of t-models change as the number of degrees of freedom increases.

6. *t*-models, part IV Describe how the critical value of t for a 95% confidence interval changes as the number of degrees of freedom increases.

Section 18.3

7. Home sales The housing market has recovered slowly from the economic crisis of 2008. Recently, in one large community, realtors randomly sampled 36 bids from potential buyers to estimate the average loss in home value. The sample showed the average loss was $9,560 with a standard deviation of $1500.

a) What assumptions and conditions must be checked before finding a confidence interval? How would you check them?

b) Find a 95% confidence interval for the mean loss in value per home.

c) Interpret this interval and explain what 95% confidence means in this context.

8. Home sales again In the previous exercise, you found a 95% confidence interval to estimate the average loss in home value.

a) Suppose the standard deviation of the losses had been $3000 instead of $1500. What would the larger standard deviation do to the width of the confidence interval (assuming the same level of confidence)?

b) Your classmate suggests that the margin of error in the interval could be reduced if the confidence level were changed to 90% instead of 95%. Do you agree with this statement? Why or why not?

c) Instead of changing the level of confidence, would it be more statistically appropriate to draw a bigger sample?

Section 18.4

9. *t*-models, again Using the t tables, software, or a calculator, estimate

a) the P-value for $t \geq 2.09$ with 4 degrees of freedom.

b) the P-value for $|t| > 1.78$ with 22 degrees of freedom.

10. *t*-models, last time Using the t tables, software, or a calculator, estimate

a) the P-value for $t \leq 2.19$ with 41 degrees of freedom.

b) the P-value for $|t| > 2.33$ with 12 degrees of freedom.

11. Home prices In 2011, the average home in the region of the country studied in Exercise 7 lost $9010. Was the community studied in Exercise 7 unusual? Use a t-test to decide if the average loss observed was significantly different from the regional average.

12. Home prices II Suppose the standard deviation of home price losses had been $3000, as in Exercise 8? What would your conclusion be then?

Section 18.5

13. Jelly A consumer advocate wants to collect a sample of jelly jars and measure the actual weight of the product in the container. He needs to collect enough data to construct a confidence interval with a margin of error of no more than 2 grams with 99% confidence. The standard deviation of these jars is usually 4 grams. What do you recommend for his sample size?

14. A good book An English professor is attempting to estimate the mean number of novels that the student body reads during their time in college. He is conducting an exit survey with seniors. He hopes to have a margin of error of 3 books with 95% confidence. From reading previous studies, he expects a large standard deviation and is going assume it is 10. How many students should he survey?

Chapter Exercises

15. Cattle Livestock are given a special feed supplement to see if it will promote weight gain. Researchers report that the 77 cows studied gained an average of 56 pounds, and that a 95% confidence interval for the mean weight gain this supplement produces has a margin of error of ± 11 pounds. Some students wrote the following conclusions. Did anyone interpret the interval correctly? Explain any misinterpretations.

a) 95% of the cows studied gained between 45 and 67 pounds.

b) We're 95% sure that a cow fed this supplement will gain between 45 and 67 pounds.

c) We're 95% sure that the average weight gain among the cows in this study was between 45 and 67 pounds.

d) The average weight gain of cows fed this supplement will be between 45 and 67 pounds 95% of the time.

e) If this supplement is tested on another sample of cows, there is a 95% chance that their average weight gain will be between 45 and 67 pounds.

16. Teachers Software analysis of the salaries of a random sample of 288 Nevada teachers produced the confidence interval shown below. Which conclusion is correct? What's wrong with the others?

t-Interval for μ: with 90.00% Confidence,
$43454 < \mu(\text{TchPay}) < 45398$

(continued)

a) If we took many random samples of 288 Nevada teachers, about 9 out of 10 of them would produce this confidence interval.
b) If we took many random samples of Nevada teachers, about 9 out of 10 of them would produce a confidence interval that contained the mean salary of all Nevada teachers.
c) About 9 out of 10 Nevada teachers earn between \$43,454 and \$45,398.
d) About 9 out of 10 of the teachers surveyed earn between \$43,454 and \$45,398.
e) We are 90% confident that the average teacher salary in the United States is between \$43,454 and \$45,398.

17. Meal plan After surveying students at Dartmouth College, a campus organization calculated that a 95% confidence interval for the mean cost of food for one term (of three in the Dartmouth trimester calendar) is (\$1372, \$1562). Now the organization is trying to write its report and is considering the following interpretations. Comment on each.

a) 95% of all students pay between \$1372 and \$1562 for food.
b) 95% of the sampled students paid between \$1372 and \$1562.
c) We're 95% sure that students in this sample averaged between \$1372 and \$1562 for food.
d) 95% of all samples of students will have average food costs between \$1372 and \$1562.
e) We're 95% sure that the average amount all students pay is between \$1372 and \$1562.

18. Snow Based on meteorological data for the past century, a local TV weather forecaster estimates that the region's average winter snowfall is 23", with a margin of error of ± 2 inches. Assuming he used a 95% confidence interval, how should viewers interpret this news? Comment on each of these statements:

a) During 95 of the past 100 winters, the region got between 21" and 25" of snow.
b) There's a 95% chance the region will get between 21" and 25" of snow this winter.
c) There will be between 21" and 25" of snow on the ground for 95% of the winter days.
d) Residents can be 95% sure that the area's average snowfall is between 21" and 25".
e) Residents can be 95% confident that the average snowfall during the past century was between 21" and 25" per winter.

19. Pulse rates A medical researcher measured the pulse rates (beats per minute) of a sample of randomly selected adults and found the following Student's t-based confidence interval:

With 95.00% Confidence,
$70.887604 < \mu(\text{Pulse}) < 74.497011$

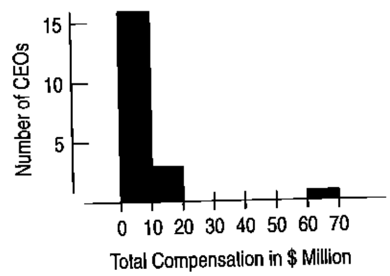
a) Explain carefully what the software output means.
b) What's the margin of error for this interval?
c) If the researcher had calculated a 99% confidence interval, would the margin of error be larger or smaller? Explain.

20. Crawling Data collected by child development scientists produced this confidence interval for the average age (in weeks) at which babies begin to crawl:

t-Intervval for μ (95.00% Confidence): $29.202 < \mu(\text{age}) < 31.844$

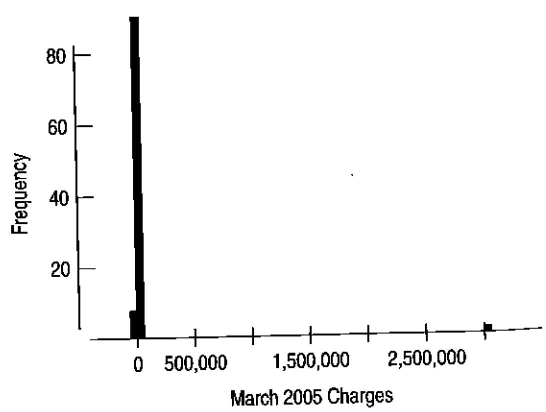
a) Explain carefully what the software output means.
b) What is the margin of error for this interval?
c) If the researcher had calculated a 90% confidence interval, would the margin of error be larger or smaller? Explain.

21. CEO compensation A sample of 20 CEOs from the Forbes 500 shows total annual compensations ranging from a minimum of \$0.1 to \$62.24 million. The average for these 20 CEOs is \$7.946 million. Here's a histogram:



Based on these data, a computer program found that a 95% confidence interval for the mean annual compensation of all Forbes 500 CEOs is (1.69, 14.20) \$ million. Why should you be hesitant to trust this confidence interval?

22. Credit card charges A credit card company takes a random sample of 100 cardholders to see how much they charged on their card last month. Here's a histogram.

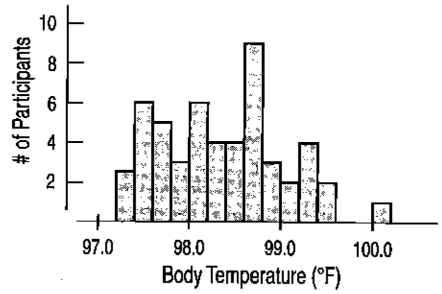


A computer program found that the resulting 95% confidence interval for the mean amount spent in

March 2011 is (−\$28,366.84, \$90,691.49). Explain why the analysts didn't find the confidence interval useful, and explain what went wrong.

23. Normal temperature The researcher described in Exercise 19 also measured the body temperatures of that randomly selected group of adults. Here are summaries of the data he collected. We wish to estimate the average (or "normal") temperature among the adult population.

Summary	Temperature
Count	52
Mean	98.285
Median	98.200
MidRange	98.600
StdDev	0.6824
Range	2.800
IntQRange	1.050



a) Check the conditions for creating a t-interval.
b) Find a 98% confidence interval for mean body temperature.
c) Explain the meaning of that interval.
d) Explain what "98% confidence" means in this context.
e) 98.6°F is commonly assumed to be "normal." Do these data suggest otherwise? Explain.

24. Parking Hoping to lure more shoppers downtown, a city builds a new public parking garage in the central business district. The city plans to pay for the structure through parking fees. During a two-month period (44 weekdays), daily fees collected averaged \$126, with a standard deviation of \$15.

a) What assumptions must you make in order to use these statistics for inference?
b) Write a 90% confidence interval for the mean daily income this parking garage will generate.
c) Interpret this confidence interval in context.
d) Explain what "90% confidence" means in this context.
e) The consultant who advised the city on this project predicted that parking revenues would average \$130 per day. Based on your confidence interval, do you think the consultant was correct? Why?

25. Normal temperatures, part II Consider again the statistics about human body temperature in Exercise 23.

a) Would a 90% confidence interval be wider or narrower than the 98% confidence interval you calculated before? Explain. (Don't compute the new interval.)
b) What are the advantages and disadvantages of the 98% confidence interval?
c) If we conduct further research, this time using a sample of 500 adults, how would you expect the 98% confidence interval to change? Explain.
d) How large a sample might allow you to estimate the mean body temperature to within 0.1 degrees with 98% confidence?

26. Parking II Suppose that, for budget planning purposes, the city in Exercise 24 needs a better estimate of the mean daily income from parking fees.

a) Someone suggests that the city use its data to create a 95% confidence interval instead of the 90% interval first created. How would this interval be better for the city? (You need not actually create the new interval.)
b) How would the 95% interval be worse for the planners?
c) How could they achieve an interval estimate that would better serve their planning needs?
d) How many days' worth of data should they collect to have 95% confidence of estimating the true mean to within \$3?

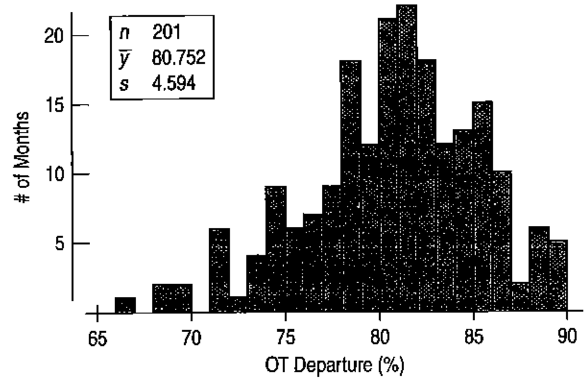
27. Speed of light In 1882, Michelson measured the speed of light (usually denoted c as in Einstein's famous equation $E = mc^2$). His values are in km/sec and have 299,000 subtracted from them. He reported the results of 23 trials with a mean of 756.22 and a standard deviation of 107.12.

a) Find a 95% confidence interval for the true speed of light from these statistics.
b) State in words what this interval means. Keep in mind that the speed of light is a physical constant that, as far as we know, has a value that is true throughout the universe.
c) What assumptions must you make in order to use your method?

28. Better light After his first attempt to determine the speed of light (described in Exercise 27), Michelson conducted an "improved" experiment. In 1897, he reported results of 100 trials with a mean of 852.4 and a standard deviation of 79.0.

a) What is the standard error of the mean for these data?
b) Without computing it, how would you expect a 95% confidence interval for the second experiment to differ from the confidence interval for the first? Note at least three specific reasons why they might differ, and indicate the ways in which these differences would change the interval.
c) According to Stigler (who reports these values), the true speed of light is 299,710.5 km/sec, corresponding to a value of 710.5 for Michelson's 1897 measurements. What does this indicate about Michelson's two experiments? Find a new confidence interval and explain using your confidence interval.

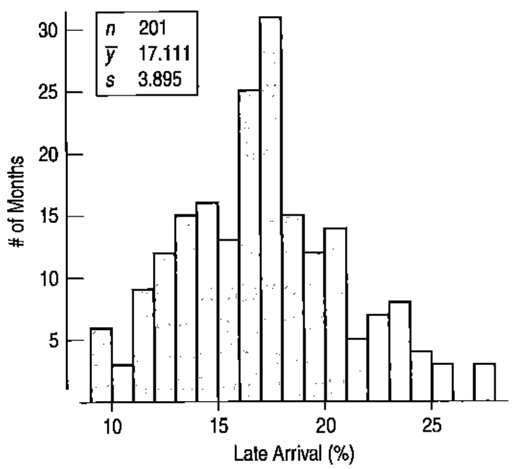
29. Departures 2011 What are the chances your flight will leave on time? The U.S. Bureau of Transportation Statistics of the Department of Transportation publishes information about airline performance. Here are a histogram and summary statistics for the percentage of flights departing on time each month from 1995 thru September 2011. (www.transtats.bts.gov/HomeDrillChart.asp)



There is no evidence of a trend over time.

a) Check the assumptions and conditions for inference.
b) Find a 90% confidence interval for the true percentage of flights that depart on time.
c) Interpret this interval for a traveler planning to fly.

30. Arrivals 2011 Will your flight get you to your destination on time? The U.S. Bureau of Transportation Statistics reported the percentage of flights that were late each month from 1995 through September of 2011. Here's a histogram, along with some summary statistics:



We can consider these data to be a representative sample of all months. There is no evidence of a time trend ($r = 0.07$).

a) Check the assumptions and conditions for inference about the mean.
b) Find a 99% confidence interval for the true percentage of flights that arrive late.
c) Interpret this interval for a traveler planning to fly.

31. Salmon, second look This chapter's For Examples looked at mirex contamination in farmed salmon. We first found a 95% confidence interval for the mean concentration to be 0.0834 to 0.0992 parts per million. Later we rejected the null hypothesis that the mean did not exceed the EPA's recommended safe level of 0.08 ppm based on a P-value of 0.0027. Explain how these two results are consistent. Your explanation should discuss the confidence level, the P-value, and the decision.

32. Hot dogs A nutrition lab tested 40 hot dogs to see if their mean sodium content was less than the 325-mg upper limit set by regulations for "reduced sodium" franks. The lab failed to reject the hypothesis that the hot dogs did not meet this requirement, with a P-value of 0.142. A 90% confidence interval estimated the mean sodium content for this kind of hot dog at 317.2 to 326.8 mg. Explain how these two results are consistent. Your explanation should discuss the confidence level, the P-value, and the decision.

33. Pizza A researcher tests whether the mean cholesterol level among those who eat frozen pizza exceeds the value considered to indicate a health risk. She gets a P-value of 0.07. Explain in this context what the "7%" represents.

34. Golf balls The United States Golf Association (USGA) sets performance standards for golf balls. For example, the initial velocity of the ball may not exceed 250 feet per second when measured by an apparatus approved by the USGA. Suppose a manufacturer introduces a new kind of ball and provides a sample for testing. Based on the mean speed in the test, the USGA comes up with a P-value of 0.34. Explain in this context what the "34%" represents.

35. Fuel economy In Chapter 6, we examined the average fuel economy of several 2010 model vehicles.

Car	mpg
Audi A4	30
BMW 3 series	28
Buick LaCrosse	30
Chevy Cobalt	37
Chevy Suburban 1500	21
Ford Expedition	20
GMC Yukon	21
Honda Civic	34
Honda Accord	31
Hyundai Elantra	35
Lexus IS 350	25
Lincoln Navigator	20
Mazda Tribute	28
Toyota Camry	33
Volkswagen Beetle	28

a) Find and interpret a 95% confidence interval for the gas mileage of 2010 vehicles.
b) Do you think that this confidence interval captures the mean gas mileage for all 2010 vehicles?

36. Computer lab fees The technology committee has stated that the average time spent by students per lab visit has increased, and the increase supports the need for increased lab fees. To substantiate this claim, the committee randomly samples 12 student lab visits and notes the amount of time spent using the computer. The times in minutes are as follows:

Time	
52	74
57	53
54	136
76	73
62	8
52	62

a) Plot the data. Are any of the observations outliers? Explain.
b) The previous mean amount of time spent using the lab computer was 55 minutes. Find a 95% confidence interval for the true mean. What do you conclude about the claim? If there are outliers, find intervals with and without the outliers present.

37. Marriage In 1960, census results indicated that the age at which American men first married had a mean of 23.3 years. It is widely suspected that young people today are waiting longer to get married. We want to find out if the mean age of first marriage has increased during the past 40 years.

a) Write appropriate hypotheses.
b) We plan to test our hypothesis by selecting a random sample of 40 men who married for the first time last year. Do you think the necessary assumptions for inference are satisfied? Explain.
c) Describe the approximate sampling distribution model for the mean age in such samples.
d) The men in our sample married at an average age of 24.2 years, with a standard deviation of 5.3 years. What's the P-value for this result?
e) Explain (in context) what this P-value means.
f) What's your conclusion?

38. Saving gas Congress regulates corporate fuel economy and sets an annual gas mileage for cars. A company with a large fleet of cars hopes to meet the 2011 goal of 30.2 mpg or better for their fleet of cars. To see if the goal is being met, they check the gasoline usage for 50 company trips chosen at random, finding a mean of 32.12 mpg and a standard deviation of 4.83 mpg. Is this strong evidence that they have attained their fuel economy goal?

a) Write appropriate hypotheses.
b) Are the necessary assumptions to make inferences satisfied?
c) Describe the sampling distribution model of mean fuel economy for samples like this.
d) Find the P-value.
e) Explain what the P-value means in this context.
f) State an appropriate conclusion.

39. Ruffles Students investigating the packaging of potato chips purchased 6 bags of Lay's Ruffles marked with a net weight of 28.3 grams. They carefully weighed the contents of each bag, recording the following weights (in grams): 29.3, 28.2, 29.1, 28.7, 28.9, 28.5.

a) Do these data satisfy the assumptions for inference? Explain.
b) Find the mean and standard deviation of the weights.
c) Create a 95% confidence interval for the mean weight of such bags of chips.
d) Explain in context what your interval means.
e) Comment on the company's stated net weight of 28.3 grams.

40. Doritos Some students checked 6 bags of Doritos marked with a net weight of 28.3 grams. They carefully weighed the contents of each bag, recording the following weights (in grams): 29.2, 28.5, 28.7, 28.9, 29.1, 29.5.

a) Do these data satisfy the assumptions for inference? Explain.
b) Find the mean and standard deviation of the weights.
c) Create a 95% confidence interval for the mean weight of such bags of chips.
d) Explain in context what your interval means.
e) Comment on the company's stated net weight of 28.3 grams.

41. Popcorn Yvon Hopps ran an experiment to test optimum power and time settings for microwave popcorn. His goal was to find a combination of power and time that would deliver high-quality popcorn with less than 10% of the kernels left unpopped, on average. After experimenting with several bags, he determined that power 9 at 4 minutes was the best combination. To be sure that the method was successful, he popped 8 more bags of popcorn (selected at random) at this setting. All were of high quality, with the following percentages of uncooked popcorn: 7, 13.2, 10, 6, 7.8, 2.8, 2.2, 5.2. Does this provide evidence that he met his goal of an average of no more than 10% uncooked kernels? Explain.

42. Ski wax Bjork Larsen was trying to decide whether to use a new racing wax for cross-country skis. He decided that the wax would be worth the price if he could average less than 55 seconds on a course he knew well, so he planned to test the wax by racing on the course 8 times. His 8 race times were 56.3, 65.9, 50.5, 52.4, 46.5, 57.8, 52.2, and 43.2 seconds. Should he buy the wax? Explain.

43. Chips Ahoy! In 1998, as an advertising campaign, the Nabisco Company announced a "1000 Chips Challenge," claiming that every 18-ounce bag of their Chips Ahoy! cookies contained at least 1000 chocolate chips. Dedicated Statistics students at the Air Force Academy (no kidding) purchased some randomly selected bags of cookies, and counted the chocolate chips. Some of their data are given below. (Source: *Chance*, 12, no. 1[1999])

1219	1214	1087	1200	1419	1121	1325	1345
1244	1258	1356	1132	1191	1270	1295	1135

a) Check the assumptions and conditions for inference. Comment on any concerns you have.
b) Create a 95% confidence interval for the average number of chips in bags of Chips Ahoy! cookies.
c) What does this evidence say about Nabisco's claim? Use your confidence interval to test an appropriate hypothesis and state your conclusion.

44. Yogurt *Consumer Reports* tested 11 brands of vanilla yogurt and found these numbers of calories per serving:

130	160	150	120	120	110	170	160	110	130	90

a) Check the assumptions and conditions for inference.
b) Create a 95% confidence interval for the average calorie content of vanilla yogurt.
c) A diet guide claims that you will get an average of 120 calories from a serving of vanilla yogurt. What does this evidence indicate? Use your confidence interval to test an appropriate hypothesis and state your conclusion.

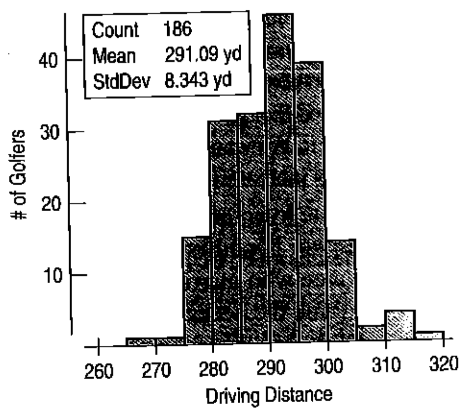
45. Maze Psychology experiments sometimes involve testing the ability of rats to navigate mazes. The mazes are classified according to difficulty, as measured by the mean length of time it takes rats to find the food at the end. One researcher needs a maze that will take rats an average of about one minute to solve. He tests one maze on several rats, collecting the data shown.

Time (sec)	
38.4	57.6
46.2	55.5
62.5	49.5
38.0	40.9
62.8	44.3
33.9	93.8
50.4	47.9
35.0	69.2
52.8	46.2
60.1	56.3
55.1	

a) Plot the data. Do you think the conditions for inference are satisfied? Explain.
b) Test the hypothesis that the mean completion time for this maze is 60 seconds. What is your conclusion?
c) Eliminate the outlier, and test the hypothesis again. What is your conclusion?
d) Do you think this maze meets the "one-minute average" requirement? Explain.

46. Braking A tire manufacturer is considering a newly designed tread pattern for its all-weather tires. Tests have indicated that these tires will provide better gas mileage and longer tread life. The last remaining test is for braking effectiveness. The company hopes the tire will allow a car traveling at 60 mph to come to a complete stop within an average of 125 feet after the brakes are applied. They will adopt the new tread pattern unless there is strong evidence that the tires do not meet this objective. The distances (in feet) for 10 stops on a test track were 129, 128, 130, 132, 135, 123, 102, 125, 128, and 130. Should the company adopt the new tread pattern? Test an appropriate hypothesis and state your conclusion. Explain how you dealt with the outlier and why you made the recommendation you did.

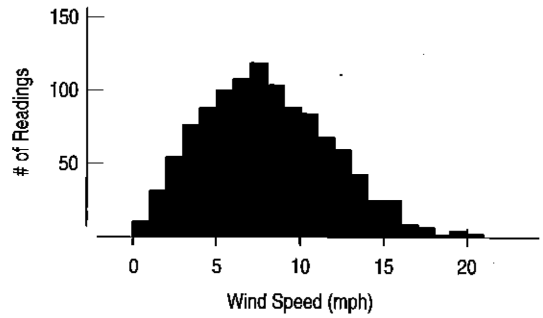
47. Driving distance 2011 How far do professional golfers drive a ball? (For non-golfers, the drive is the shot hit from a tee at the start of a hole and is typically the longest shot.) Here's a histogram of the average driving distances of the 186 leading professional golfers by end of November 2011 along with summary statistics (www.pgatour.com).

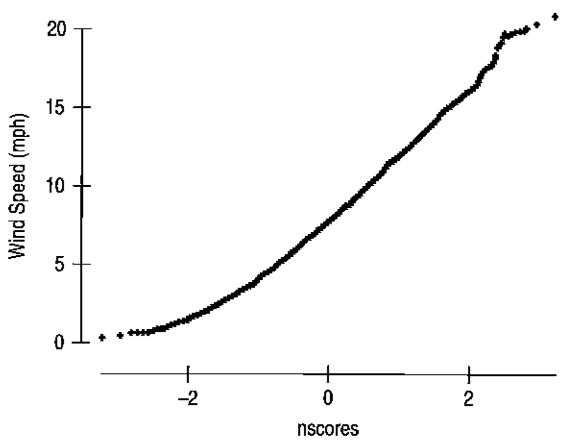


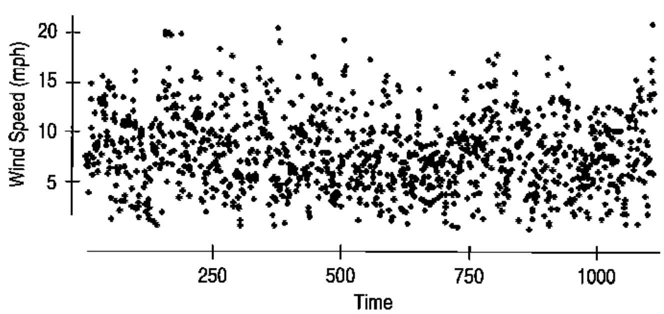
a) Find a 95% confidence interval for the mean drive distance.
b) Interpreting this interval raises some problems. Discuss.
c) The data are the mean driving distance for each golfer. Is that a concern in interpreting the interval? (*Hint:* Review the What Can Go Wrong warnings of Chapter 8. Chapter 8?! Yes, Chapter 8.)

48. Wind power Should you generate electricity with your own personal wind turbine? That depends on whether you have enough wind on your site. To produce enough energy, your site should have an annual average wind speed above 8 miles per hour, according to the Wind Energy Association. One candidate site was monitored for a year, with wind speeds recorded every 6 hours. A total of 1114 readings of wind speed averaged 8.019 mph with a standard deviation of 3.813 mph. You've been asked to make a statistical report to help the landowner decide whether to place a wind turbine at this site.

a) Discuss the assumptions and conditions for using Student's t inference methods with these data. Here are some plots that may help you decide whether the methods can be used:







b) What would you tell the landowner about whether this site is suitable for a small wind turbine? Explain.

Just Checking ANSWERS

1. Questions on the short form are answered by everyone in the population. This is a census, so means or proportions *are* the true population values. The long forms are given just to a sample of the population. When we estimate parameters from a sample, we use a confidence interval to take sample-to-sample variability into account.

2. They don't know the population standard deviation, so they must use the sample standard deviation as an estimate. The additional uncertainty is taken into account by t-models.

3. The margin of error for a confidence interval for a mean depends, in part, on the standard error,

$$SE(\bar{y}) = \frac{s}{\sqrt{n}}.$$

Since n is in the denominator, smaller sample sizes lead to larger *SE*s and correspondingly wider intervals. Long forms returned by one in every six or seven households in a less populous area will be a smaller sample.

4. The critical values for t with fewer degrees of freedom would be slightly larger. The $\sqrt{n}$ part of the standard error changes a lot, making the *SE* much larger. Both would increase the margin of error.

5. The smaller sample is one fourth as large, so the confidence interval would be roughly twice as wide.

6. We expect 95% of such intervals to cover the true value, so we would expect about 5 of the 100 intervals to miss.

Exercises

Section 19.1

1. **Parameters and hypotheses** For each of the following situations, define the parameter (proportion or mean) and write the null and alternative hypotheses in terms of parameter values. Example: We want to know if the proportion of up days in the stock market is 50%. Answer: Let p = the proportion of up days. $H_0: p = 0.5$ vs. $H_A: p \neq 0.5$.
   a) A casino wants to know if their slot machine really delivers the 1 in 100 win rate that it claims.
   b) Last year, customers spent an average of $35.32 per visit to the company's website. Based on a random sample of purchases this year, the company wants to know if the mean this year has changed.
   c) A pharmaceutical company wonders if their new drug has a cure rate different from the 30% reported by the placebo.
   d) A bank wants to know if the percentage of customers using their website has changed from the 40% that used it before their system crashed last week.

2. **Hypotheses and parameters** As in Exercise 1, for each of the following situations, define the parameter and write the null and alternative hypotheses in terms of parameter values.
   a) Seat-belt compliance in Massachusetts was 65% in 2008. The state wants to know if it has changed.
   b) Last year, a survey found that 45% of the employees were willing to pay for on-site day care. The company wants to know if that has changed.
   c) Regular card customers have a default rate of 6.7%. A credit card bank wants to know if that rate is different for their Gold card customers.
   d) Regular card customers have been with the company for an average of 17.3 months. The credit card bank wants to know if their Gold card customers have been with the company on average the same amount of time.

Section 19.2

3. **P-values** Which of the following are true? If false, explain briefly.
   a) A very high P-value is strong evidence that the null hypothesis is false.
   b) A very low P-value proves that the null hypothesis is false.
   c) A high P-value shows that the null hypothesis is true.
   d) A P-value below 0.05 is always considered sufficient evidence to reject a null hypothesis.

4. **More P-values** Which of the following are true? If false, explain briefly.
   a) A very low P-value provides evidence against the null hypothesis.
   b) A high P-value is strong evidence in favor of the null hypothesis.
   c) A P-value above 0.10 shows that the null hypothesis is true.
   d) If the null hypothesis is true, you can't get a P-value below 0.01.

5. **Hypotheses** For each of the following, write out the null and alternative hypotheses, being sure to state whether the alternative is one-sided or two-sided.
   a) A company reports that last year, 40% of their reports in accounting were on time. From a random sample this year, they want to know if that proportion has changed.
   b) A company wants to know if average click-through rates (in minutes) are shorter than the 5.4 minutes that customers spent on their old website before making a purchase.
   c) Last year, 42% of the employees enrolled in at least one wellness class at the company's site. Using a survey, they want to know if a greater percentage is planning to take a wellness class this year.
   d) A political candidate wants to know from recent polls if she's going to garner a majority of votes in next week's election.

6. **More hypotheses** For each of the following, write out the alternative hypothesis, being sure to indicate whether it is one-sided or two-sided.
   a) *Consumer Reports* discovered that 20% of a certain computer model had warranty problems over the first three months. From a random sample, the manufacturer wants to know if a new model has improved that rate.
   b) The last time a philanthropic agency requested donations, 4.75% of people responded. From a recent pilot mailing, they wonder if that rate has increased.
   c) The average age of a customer of a clothing store is 35.2 years. The company wants to know if customers who use their website are younger on average.
   d) A student wants to know if other students on her campus prefer Coke or Pepsi.

Section 19.3

7. **Alpha true and false** Which of the following statements are true? If false, explain briefly.
   a) Using an alpha level of 0.05, a P-value of 0.04 results in rejecting the null hypothesis.
   b) The alpha level depends on the sample size.
   c) With an alpha level of 0.01, a P-value of 0.10 results in rejecting the null hypothesis.
   d) Using an alpha level of 0.05, a P-value of 0.06 means the null hypothesis is true.

8. Alpha false and true Which of the following statements are true? If false, explain briefly.

a) It is better to use an alpha level of 0.05 than an alpha level of 0.01.
b) If we use an alpha level of 0.01, then a P-value of 0.001 is statistically significant.
c) If we use an alpha level of 0.01, then we reject the null hypothesis if the P-value is 0.001.
d) If the P-value is 0.01, we reject the null hypothesis for any alpha level greater than 0.01.

Section 19.4

9. Critical values For each of the following situations, find the critical value(s) for z or t.

a) H_0: $p = 0.5$ vs. H_A: $p \neq 0.5$ at $\alpha = 0.05$.
b) H_0: $p = 0.4$ vs. H_A: $p > 0.4$ at $\alpha = 0.05$.
c) H_0: $\mu = 10$ vs. H_A: $\mu \neq 10$ at $\alpha = 0.05$; $n = 36$.
d) H_0: $p = 0.5$ vs. H_A: $p > 0.5$ at $\alpha = 0.01$; $n = 345$.
e) H_0: $\mu = 20$ vs. H_A: $\mu < 20$ at $\alpha = 0.01$; $n = 1000$.

10. More critical values For each of the following situations, find the critical value for z or t.

a) H_0: $\mu = 105$ vs. H_A: $\mu \neq 105$ at $\alpha = 0.05$; $n = 61$.
b) H_0: $p = 0.05$ vs. H_A: $p > 0.05$ at $\alpha = 0.05$.
c) H_0: $p = 0.6$ vs. H_A: $p \neq 0.6$ at $\alpha = 0.01$.
d) H_0: $p = 0.5$ vs. H_A: $p < 0.5$ at $\alpha = 0.01$; $n = 500$.
e) H_0: $p = 0.2$ vs. H_A: $p < 0.2$ at $\alpha = 0.01$.

Section 19.5

11. Errors For each of the following situations, state whether a Type I, a Type II, or neither error has been made. Explain briefly.

a) A bank wants to know if the enrollment on their website is above 30% based on a small sample of customers. They test H_0: $p = 0.3$ vs. H_A: $p > 0.3$ and reject the null hypothesis. Later they find out that actually 28% of all customers enrolled.
b) A student tests 100 students to determine whether other students on her campus prefer Coke or Pepsi and finds no evidence that preference for Coke is not 0.5. Later, a marketing company tests all students on campus and finds no difference.
c) A human resource analyst wants to know if the applicants this year score, on average, higher on their placement exam than the 52.5 points the candidates averaged last year. She samples 50 recent tests and finds the average to be 54.1 points. She fails to reject the null hypothesis that the mean is 52.5 points. At the end of the year, they find that the candidates this year had a mean of 55.3 points.
d) A pharmaceutical company tests whether a drug lifts the headache relief rate from the 25% achieved by the placebo. They fail to reject the null hypothesis because the P-value is 0.465. Further testing shows that the drug actually relieves headaches in 38% of people.

12. More errors For each of the following situations, state whether a Type I, a Type II, or neither error has been made.

a) A test of H_0: $\mu = 25$ vs. H_A: $\mu > 25$ rejects the null hypothesis. Later it is discovered that $\mu = 24.9$.
b) A test of H_0: $p = 0.8$ vs. H_A: $p < 0.8$ fails to reject the null hypothesis. Later it is discovered that $p = 0.9$.
c) A test of H_0: $p = 0.5$ vs. H_A: $p \neq 0.5$ rejects the null hypothesis. Later is it discovered that $p = 0.65$.
d) A test of H_0: $p = 0.7$ vs. H_A: $p < 0.7$ fails to reject the null hypothesis. Later is it discovered that $p = 0.6$.

Chapter Exercises

13. One sided or two? In each of the following situations, is the alternative hypothesis one-sided or two-sided? What are the hypotheses?

a) A business student conducts a taste test to see whether students prefer Diet Coke or Diet Pepsi.
b) PepsiCo recently reformulated Diet Pepsi in an attempt to appeal to teenagers. They run a taste test to see if the new formula appeals to more teenagers than the standard formula.
c) A budget override in a small town requires a two-thirds majority to pass. A local newspaper conducts a poll to see if there's evidence it will pass.
d) One financial theory states that the stock market will go up or down with equal probability. A student collects data over several years to test the theory.

14. Which alternative? In each of the following situations, is the alternative hypothesis one-sided or two-sided? What are the hypotheses?

a) A college dining service conducts a survey to see if students prefer plastic or metal cutlery.
b) In recent years, 10% of college juniors have applied for study abroad. The dean's office conducts a survey to see if that's changed this year.
c) A pharmaceutical company conducts a clinical trial to see if more patients who take a new drug experience headache relief than the 22% who claimed relief after taking the placebo.
d) At a small computer peripherals company, only 60% of the hard drives produced passed all their performance tests the first time. Management recently invested a lot of resources into the production system and now conducts a test to see if it helped.

15. P-value A medical researcher tested a new treatment for poison ivy against the traditional ointment. He concluded that the new treatment is more effective. Explain what the P-value of 0.047 means in this context.

16. Another P-value Have harsher penalties and ad campaigns increased seat-belt use among drivers and passengers? Observations of commuter traffic failed to find evidence of a significant change compared with three years ago. Explain what the study's P-value of 0.17 means in this context.

17. Alpha A researcher developing scanners to search for hidden weapons at airports has concluded that a new device is significantly better than the current scanner. He made this decision based on a test using $\alpha = 0.05$. Would he have made the same decision at $\alpha = 0.10$? How about $\alpha = 0.01$? Explain.

18. Alpha, again Environmentalists concerned about the impact of high-frequency radio transmissions on birds found that there was no evidence of a higher mortality rate among hatchlings in nests near cell towers. They based this conclusion on a test using $\alpha = 0.05$. Would they have made the same decision at $\alpha = 0.10$? How about $\alpha = 0.01$? Explain.

19. Significant? Public health officials believe that 98% of children have been vaccinated against measles. A random survey of medical records at many schools across the country found that, among more than 13,000 children, only 97.4% had been vaccinated. A statistician would reject the 98% hypothesis with a P-value of $P < 0.0001$.

a) Explain what the P-value means in this context.
b) The result is statistically significant, but is it important? Comment.

20. Significant again? A new reading program may reduce the number of elementary school students who read below grade level. The company that developed this program supplied materials and teacher training for a large-scale test involving nearly 8500 children in several different school districts. Statistical analysis of the results showed that the percentage of students who did not meet the grade-level goal was reduced from 15.9% to 15.1%. The hypothesis that the new reading program produced no improvement was rejected with a P-value of 0.023.

a) Explain what the P-value means in this context.
b) Even though this reading method has been shown to be significantly better, why might you not recommend that your local school adopt it?

21. Groceries In January 2011, Yahoo surveyed 2400 U.S. men. 1224 of the men identified themselves as the primary grocery shopper in their household.

a) Estimate the percentage of all American males who identify themselves as the primary grocery shopper. Use a 98% confidence interval. Check the conditions first.
b) A grocery store owner believed that only 45% of men are the primary grocery shopper for their family, and targets his advertising accordingly. He wishes to conduct a hypothesis test to see if the fraction is in fact higher than 45%. What does your confidence interval indicate? Explain.
c) What is the level of significance of this test? Explain.

22. Is the Euro fair? Soon after the Euro was introduced as currency in Europe, it was widely reported that someone had spun a Euro coin 250 times and gotten heads 140 times. We wish to test a hypothesis about the fairness of spinning the coin.

a) Estimate the true proportion of heads. Use a 95% confidence interval. Don't forget to check the conditions.
b) Does your confidence interval provide evidence that the coin is unfair when spun? Explain.
c) What is the significance level of this test? Explain.

23. Approval 2011 In November 2011, Barack Obama's approval rating stood at 45% in Rasmussen's daily tracking poll of 1500 randomly surveyed U.S. adults.

a) Make a 95% confidence interval for his approval rating by all U.S. adults.
b) Based on the confidence interval, test the null hypothesis that Obama's approval rating was no worse than his November 2009 approval rating of 50%.

24. Hard times In June 2010, a random poll of 800 working men found that 9% had taken on a second job to help pay the bills. (www.careerbuilder.com)

a) Estimate the true percentage of men that are taking on second jobs by constructing a 95% confidence interval.
b) A pundit on a TV news show claimed that only 6% of working men had a second job. Use your confidence interval to test whether his claim is plausible given the poll data.

25. Dogs Canine hip dysplasia is a degenerative disease that causes pain in many dogs. Sometimes advanced warning signs appear in puppies as young as 6 months. A veterinarian checked 42 puppies whose owners brought them to a vaccination clinic, and she found 5 with early hip dysplasia. She considers this group to be a random sample of all puppies.

a) Explain why we cannot use this information to construct a confidence interval for the rate of occurrence of early hip dysplasia among all 6-month-old puppies.
*b) Construct a "plus-four" confidence interval and interpret it in this context.

26. Fans A survey of 81 randomly selected people standing in line to enter a football game found that 73 of them were home team fans.

a) Explain why we cannot use this information to construct a confidence interval for the proportion of all people at the game who are fans of the home team.
*b) Construct a "plus-four" confidence interval and interpret it in this context.

27. Loans Before lending someone money, banks must decide whether they believe the applicant will repay the loan. One strategy used is a point system. Loan officers assess information about the applicant, totaling points they award for the person's income level, credit history, current debt burden, and so on. The higher the point total, the more convinced the bank is that it's safe to make the loan. Any applicant with a lower point total than a certain cutoff score is denied a loan.

We can think of this decision as a hypothesis test. Since the bank makes its profit from the interest collected on repaid loans, their null hypothesis is that the applicant will repay the loan and therefore should get the money. Only if the person's score falls below the minimum cutoff will the bank reject the null and deny the loan. This system is reasonably reliable, but, of course, sometimes there are mistakes.

a) When a person defaults on a loan, which type of error did the bank make?
b) Which kind of error is it when the bank misses an opportunity to make a loan to someone who would have repaid it?
c) Suppose the bank decides to lower the cutoff score from 250 points to 200. Is that analogous to choosing a higher or lower value of α for a hypothesis test? Explain.
d) What impact does this change in the cutoff value have on the chance of each type of error?

28. Spam Spam filters try to sort your e-mails, deciding which are real messages and which are unwanted. One method used is a point system. The filter reads each incoming e-mail and assigns points to the sender, the subject, key words in the message, and so on. The higher the point total, the more likely it is that the message is unwanted. The filter has a cutoff value for the point total; any message rated lower than that cutoff passes through to your inbox, and the rest, suspected to be spam, are diverted to the junk mailbox.

We can think of the filter's decision as a hypothesis test. The null hypothesis is that the e-mail is a real message and should go to your inbox. A higher point total provides evidence that the message may be spam; when there's sufficient evidence, the filter rejects the null, classifying the message as junk. This usually works pretty well, but, of course, sometimes the filter makes a mistake.

a) When the filter allows spam to slip through into your inbox, which kind of error is that?
b) Which kind of error is it when a real message gets classified as junk?
c) Some filters allow the user (that's you) to adjust the cutoff. Suppose your filter has a default cutoff of 50 points, but you reset it to 60. Is that analogous to choosing a higher or lower value of α for a hypothesis test? Explain.
d) What impact does this change in the cutoff value have on the chance of each type of error?

29. Second loan Exercise 27 describes the loan score method a bank uses to decide which applicants it will lend money. Only if the total points awarded for various aspects of an applicant's financial condition fail to add up to a minimum cutoff score set by the bank will the loan be denied.

a) In this context, what is meant by the power of the test?
b) What could the bank do to increase the power?
c) What's the disadvantage of doing that?

30. More spam Consider again the points-based spam filter described in Exercise 28. When the points assigned to various components of an e-mail exceed the cutoff value you've set, the filter rejects its null hypothesis (that the message is real) and diverts that e-mail to a junk mailbox.

a) In this context, what is meant by the power of the test?
b) What could you do to increase the filter's power?
c) What's the disadvantage of doing that?

31. Homeowners 2009 In 2009, the U.S. Census Bureau reported that 67.4% of American families owned their homes. Census data reveal that the ownership rate in one small city is much lower. The city council is debating a plan to offer tax breaks to first-time home buyers in order to encourage people to become homeowners. They decide to adopt the plan on a 2-year trial basis and use the data they collect to make a decision about continuing the tax breaks. Since this plan costs the city tax revenues, they will continue to use it only if there is strong evidence that the rate of home ownership is increasing.

a) In words, what will their hypotheses be?
b) What would a Type I error be?
c) What would a Type II error be?
d) For each type of error, tell who would be harmed.
e) What would the power of the test represent in this context?

32. Alzheimer's Testing for Alzheimer's disease can be a long and expensive process, consisting of lengthy tests and medical diagnosis. A group of researchers (Solomon *et al.*, 1998) devised a 7-minute test to serve as a quick screen for the disease for use in the general population of senior citizens. A patient who tested positive would then go through the more expensive battery of tests and medical diagnosis. The authors reported a false positive rate of 4% and a false negative rate of 8%.

a) Put this in the context of a hypothesis test. What are the null and alternative hypotheses?
b) What would a Type I error mean?
c) What would a Type II error mean?
d) Which is worse here, a Type I or Type II error? Explain.
e) What is the power of this test?

33. Testing cars A clean air standard requires that vehicle exhaust emissions not exceed specified limits for various pollutants. Many states require that cars be tested annually to be sure they meet these standards. Suppose state regulators double-check a random sample of cars that a suspect repair shop has certified as okay. They will revoke the shop's license if they find significant evidence that the shop is certifying vehicles that do not meet standards.

a) In this context, what is a Type I error?
b) In this context, what is a Type II error?
c) Which type of error would the shop's owner consider more serious?
d) Which type of error might environmentalists consider more serious?

34. Quality control Production managers on an assembly line must monitor the output to be sure that the level of defective products remains small. They periodically inspect a random sample of the items produced. If they find a significant increase in the proportion of items that must be rejected, they will halt the assembly process until the problem can be identified and repaired.

a) In this context, what is a Type I error?
b) In this context, what is a Type II error?
c) Which type of error would the factory owner consider more serious?
d) Which type of error might customers consider more serious?

35. Cars, again As in Exercise 33, state regulators are checking up on repair shops to see if they are certifying vehicles that do not meet pollution standards.

a) In this context, what is meant by the power of the test the regulators are conducting?
b) Will the power be greater if they test 20 or 40 cars? Why?
c) Will the power be greater if they use a 5% or a 10% level of significance? Why?
d) Will the power be greater if the repair shop's inspectors are only a little out of compliance or a lot? Why?

36. Production Consider again the task of the quality control inspectors in Exercise 34.

a) In this context, what is meant by the power of the test the inspectors conduct?
b) They are currently testing 5 items each hour. Someone has proposed that they test 10 instead. What are the advantages and disadvantages of such a change?
c) Their test currently uses a 5% level of significance. What are the advantages and disadvantages of changing to an alpha level of 1%?
d) Suppose that, as a day passes, one of the machines on the assembly line produces more and more items that are defective. How will this affect the power of the test?

37. Equal opportunity? A company is sued for job discrimination because only 19% of the newly hired candidates were minorities when 27% of all applicants were minorities. Is this strong evidence that the company's hiring practices are discriminatory?

a) Is this a one-tailed or a two-tailed test? Why?
b) In this context, what would a Type I error be?
c) In this context, what would a Type II error be?
d) In this context, what is meant by the power of the test?
e) If the hypothesis is tested at the 5% level of significance instead of 1%, how will this affect the power of the test?
f) The lawsuit is based on the hiring of 37 employees. Is the power of the test higher than, lower than, or the same as it would be if it were based on 87 hires?

38. Stop signs Highway safety engineers test new road signs, hoping that increased reflectivity will make them more visible to drivers. Volunteers drive through a test course with several of the new- and old-style signs and rate which kind shows up the best.

a) Is this a one-tailed or a two-tailed test? Why?
b) In this context, what would a Type I error be?
c) In this context, what would a Type II error be?
d) In this context, what is meant by the power of the test?
e) If the hypothesis is tested at the 1% level of significance instead of 5%, how will this affect the power of the test?
f) The engineers hoped to base their decision on the reactions of 50 drivers, but time and budget constraints may force them to cut back to 20. How would this affect the power of the test? Explain.

39. Software for learning A Statistics professor has observed that for several years students score an average of 105 points out of 150 on the semester exam. A salesman suggests that he try a statistics software package that gets students more involved with computers, predicting that it will increase students' scores. The software is expensive, and the salesman offers to let the professor use it for a semester to see if the scores on the final exam increase significantly. The professor will have to pay for the software only if he chooses to continue using it.

a) Is this a one-tailed or two-tailed test? Explain.
b) Write the null and alternative hypotheses.
c) In this context, explain what would happen if the professor makes a Type I error.
d) In this context, explain what would happen if the professor makes a Type II error.
e) What is meant by the power of this test?

40. Ads A company is willing to renew its advertising contract with a local radio station only if the station can prove that more than 20% of the residents of the city have heard the ad and recognize the company's product. The radio station conducts a random phone survey of 400 people.

a) What are the hypotheses?
b) The station plans to conduct this test using a 10% level of significance, but the company wants the significance level lowered to 5%. Why?
c) What is meant by the power of this test?
d) For which level of significance will the power of this test be higher? Why?
e) They finally agree to use $\alpha = 0.05$, but the company proposes that the station call 600 people instead of the 400 initially proposed. Will that make the risk of Type II error higher or lower? Explain.

41. Software, part II 203 students signed up for the Stats course in Exercise 39. They used the software suggested by the salesman, and scored an average of 108 points on the final with a standard deviation of 8.7 points.

a) Should the professor spend the money for this software? Support your recommendation with an appropriate test.
b) Does this improvement seem to be practically significant?

42. Testing the ads The company in Exercise 40 contacts 600 people selected at random, and only 133 remember the ad.

a) Should the company renew the contract? Support your recommendation with an appropriate test.
b) Explain what your P-value means in this context.

43. TV safety The manufacturer of a metal stand for home TV sets must be sure that its product will not fail under the weight of the TV. Since some larger sets weigh nearly 300 pounds, the company's safety inspectors have set a standard of ensuring that the stands can support an average of over 500 pounds. Their inspectors regularly subject a random sample of the stands to increasing weight until they fail. They test the hypothesis H_0: $\mu = 500$ against H_A: $\mu > 500$, using the level of significance $\alpha = 0.01$. If the sample of stands fails to pass this safety test, the inspectors will not certify the product for sale to the general public.

a) Is this an upper-tail or lower-tail test? In the context of the problem, why do you think this is important?
b) Explain what will happen if the inspectors commit a Type I error.
c) Explain what will happen if the inspectors commit a Type II error.

44. Catheters During an angiogram, heart problems can be examined via a small tube (a catheter) threaded into the heart from a vein in the patient's leg. It's important that the company that manufactures the catheter maintain a diameter of 2.00 mm. (The standard deviation is quite small.) Each day, quality control personnel make several measurements to test H_0: $\mu = 2.00$ against H_A: $\mu \neq 2.00$ at a significance level of $\alpha = 0.05$. If they discover a problem, they will stop the manufacturing process until it is corrected.

a) Is this a one-sided or two-sided test? In the context of the problem, why do you think this is important?
b) Explain in this context what happens if the quality control people commit a Type I error.
c) Explain in this context what happens if the quality control people commit a Type II error.

45. TV safety, revisited The manufacturer of the metal TV stands in Exercise 43 is thinking of revising its safety test.

a) If the company's lawyers are worried about being sued for selling an unsafe product, should they increase or decrease the value of α? Explain.
b) In this context, what is meant by the power of the test?
c) If the company wants to increase the power of the test, what options does it have? Explain the advantages and disadvantages of each option.

46. Catheters, again The catheter company in Exercise 44 is reviewing its testing procedure.

a) Suppose the significance level is changed to $\alpha = 0.01$. Will the probability of a Type II error increase, decrease, or remain the same?
b) What is meant by the power of the test the company conducts?
c) Suppose the manufacturing process is slipping out of proper adjustment. As the actual mean diameter of the catheters produced gets farther and farther above the desired 2.00 mm, will the power of the quality control test increase, decrease, or remain the same?
d) What could they do to improve the power of the test?

47. Two coins In a drawer are two coins. They look the same, but one coin produces heads 90% of the time when spun while the other one produces heads only 30% of the time. You select one of the coins. You are allowed to spin it *once* and then must decide whether the coin is the 90%- or the 30%-head coin. Your null hypothesis is that your coin produces 90% heads.

a) What is the alternative hypothesis?
b) Given that the outcome of your spin is tails, what would you decide? What if it were heads?
c) How large is α in this case?
d) How large is the power of this test? (*Hint:* How many possibilities are in the alternative hypothesis?)
e) How could you lower the probability of a Type I error and increase the power of the test at the same time?

48. Faulty or not? You are in charge of shipping computers to customers. You learn that a RAM chip was put into some of the machines. There's a simple test you can perform, but it's not perfect. All but 4% of the time, a good chip passes the test, but unfortunately, 35% of the bad chips pass the test, too. You have to decide on the basis of one test whether the chip is good or bad. Make this a hypothesis test.

a) What are the null and alternative hypotheses?
b) Given that a computer fails the test, what would you decide? What if it passes the test?
c) How large is α for this test?
d) What is the power of this test? (*Hint:* How many possibilities are in the alternative hypothesis?)

49. Hoops A basketball player with a poor foul-shot record practices intensively during the off-season. He tells the coach that he has raised his proficiency from 60% to 80%. Dubious, the coach asks him to take 10 shots, and is surprised when the player hits 9 out of 10. Did the player prove that he has improved?

a) Suppose the player really is no better than before—still a 60% shooter. What's the probability he can hit at least 9 of 10 shots anyway? (*Hint:* Use a Binomial model.)
b) If that is what happened, now the coach thinks the player has improved when he has not. Which type of error is that?

(continued)

c) If the player really can hit 80% now, and it takes at least 9 out of 10 successful shots to convince the coach, what's the power of the test?

d) List two ways the coach and player could increase the power to detect any improvement.

50. Pottery An artist experimenting with clay to create pottery with a special texture has been experiencing difficulty with these special pieces. About 40% break in the kiln during firing. Hoping to solve this problem, she buys some more expensive clay from another supplier. She plans to make and fire 10 pieces and will decide to use the new clay if at most one of them breaks.

a) Suppose the new, expensive clay really is no better than her usual clay. What's the probability that this test convinces her to use it anyway? (*Hint:* Use a Binomial model.)

b) If she decides to switch to the new clay and it is no better, what kind of error did she commit?

c) If the new clay really can reduce breakage to only 20%, what's the probability that her test will not detect the improvement?

d) How can she improve the power of her test? Offer at least two suggestions.

Just Checking ANSWERS

1. With a z-score of 0.62, you can't reject the null hypothesis. The experiment shows no evidence that the wheel is not fair.

2. At $\alpha = 0.05$, you can't reject the null hypothesis because 0.30 is contained in the 90% confidence interval—it's plausible that sending the DVDs is no more effective than just sending letters.

3. The confidence interval is from 29% to 45%. The DVD strategy is more expensive and may not be worth it. We can't distinguish the success rate from 30% given the results of this experiment, but 45% would represent a large improvement. The bank should consider another trial, increasing their sample size to get a narrower confidence interval.

4. A Type I error would mean deciding that the DVD success rate is higher than 30% when it really isn't. They would adopt a more expensive method for collecting payments that's no better than the less expensive strategy.

5. A Type II error would mean deciding that there's not enough evidence to say that the DVD strategy works when in fact it does. The bank would fail to discover an effective method for increasing their revenue from delinquent accounts.

6. 60%; the larger the effect size, the greater the power. It's easier to detect an improvement to a 60% success rate than to a 32% rate.