1

Solutions

In-Class Activities

Activity 1-1: Student Data

a. Answers will vary.

b. Answers will vary.

c. No, every word did not contain the same number of letters.

d. The variable is the *number of letters in each word*.

e.
- *How many hours you slept in the past 24 hours:* quantitative
- *Whether you slept for at least 7 hours in the past 24 hours:* binary categorical
- *How many states you have visited:* quantitative
- *Handedness:* binary categorical, unless you classify "ambidextrous," in which case it is not binary.
- *Day of the week on which you were born:* categorical
- *Gender:* binary categorical
- *Average study time per week:* quantitative
- *Score on the first exam in this course:* quantitative

f. No, neither *average height of students in the class* nor *percentage of students in the class who have used a cell phone today* can legitimately be considered variables when the observational units are the students in your class. Both of these are numbers that provide summary information about the class as a whole. They do not vary from student to student.

g. If you record the average student height or percentage of student cell phone usage *by class* taught at your school, these would become legitimate variables. Now these numbers would (potentially) take on different values from class to class. The observational units are no longer the students in your class, but rather all classes taught at your school.

Activity 1-2: Variables of State

- **a.** Binary categorical variable
- **b.** Not a variable
- **c.** Quantitative variable
- **d.** Quantitative variable
- **e.** Binary categorical variable
- **f.** Quantitative variable
- **g.** Not a variable

Activity 1-3: Cell Phone Fraud

- **a.** The observational units are the cell phone calls.
- **b.** The binary categorical variables are *direction, location,* and *whether the call took place on a weekday or weekend.* The non-binary categorical variable is *day of week.*
- **c.** The quantitative variables are *duration of the call* and *time of day.*

Activity 1-4: Studies from *Blink*

a. Observational units: 100 CEOs

Variable: height	Type: quantitative

b. Observational units: 50 marriage counselors

Variable: whether the counselor makes the correct prediction about whether a couple will still be married in five years	Type: (binary) categorical

c. Observational units: 200 African-American college students

Variable 1: whether their version of the exam asks them to indicate race	Type: (binary) categorical
Variable 2: score on SAT-like exam	Type: quantitative

d. Observational units: 10 car dealerships

Variable 1: gender of customer	Type: (binary) categorical
Variable 2: race of customer	Type: (binary) categorical
Variable 3: price negotiated for the car	Type: quantitative

Activity 1-5: Student Data

- **a.** Many answers are possible, but some examples include these: Do male and female students differ with regard to the number of states they have visited? Are sleeping times associated with the day of the week on which a student was born?
- **b.** Answers will vary.

Activity 1-6: A Nurse Accused

a. The observational units are the eight-hour shifts.

b. One variable is *whether Gilbert worked on the shift.* This variable is categorical and binary. The other variable is *whether a patient died on the shift.* This variable is also categorical and binary.

Solutions

● ● ● In-Class Activities

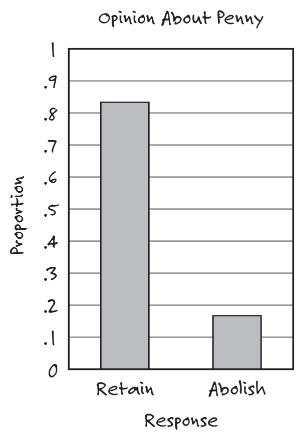
Activity 2-1: Penny Thoughts

a. This is a binary categorical variable.

b. Answers will vary by class. One example is 15/18 or .833 voted to retain the penny.

c. Answers will vary, but for the class in part b, 3/18 or .167 voted to abolish the penny.

d. Answers will vary, but for the class in part b, the bar graph is shown here:



e. For the class in part c, more than 80% favors retaining the penny, but the vote is not unanimous. A significant proportion of the class (.167) favors abolishing the penny.

Activity 2-2: Hand Washing

a. The proportion of men who washed their hands is 2393/3206 or .746. The proportion of women who washed their hands is 2802/3130 or .895.

b. Yes, these proportions are consistent with the bar graphs. The heights of the "washed hands" bars are about .75 and .90.

c. It does appear that women are a little more likely to wash their hands after using a public restroom than men are. About 90% of the women in this sample did, whereas only 75% of the men did, indicating that women are about 1.2 times more likely to wash their hands after using a public restroom.

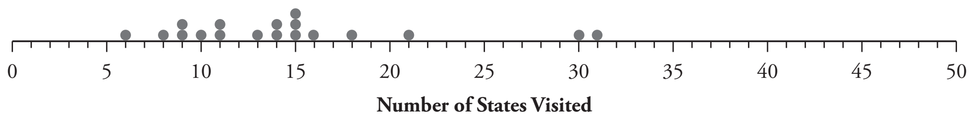
d. Atlanta: .73   Chicago: .88   New York: .79   San Francisco: .88

e. There does not appear to be much difference between these cities in terms of hand washing. Chicago and San Francisco appear to have identical proportions of people who washed their hands after using a public restroom, whereas New York seems to have about 10% fewer hand washers. Atlanta seems to have the smallest percentage of hand washers among these four cities, with just under 75% of the sample from Atlanta washing their hands.

Activity 2-3: Student Travels

a. This is a quantitative variable.

b. Answers will vary. Here is one example:



c. Answers will vary.

d. Answers will vary.

e. The number of states visited by students in this class varies from a minimum of 6 to about 21. There are two high outliers of 30 and 31 states. The typical number of states visited by a student in this class seems to be about 14.

Activity 2-4: Buckle Up!

a. The states are the observational units for these data.

b. The *primary or secondary law* variable is categorical and binary. The *percentage usage* variable is quantitative.

c. Answers will vary. The typical usage percentage for a primary-type seatbelt law state is about 86%; for states with a secondary-type law, the typical usage percentage appears to be about 77%.

d. No, a state with a primary law does not always have a higher usage percentage than a state with a secondary law; for example, Tennessee (p, 74.4%) and Virginia (s, 8.4%).

e. Yes, states with a primary law *tend to* have higher usage percentages than states with a secondary law. You see this in the dotplot because most of the dots for the primary law states are clustered at the high percentage values (from 80–95%), whereas most of the secondary law states have percentages that fall between 65%–85%.

f. Yes, the data seem to support the contention that tougher laws lead to more seatbelt usage, but you cannot draw a definite cause-and-effect conclusion. There might be hidden or "confounding" variables that explain the association between primary seatbelt laws and increased seatbelt usage.

Activity 2-5: February Temperatures

a. San Luis Obispo tended to have the highest temperatures that month (its temperatures are all clustered from 50–90°F, with a significant chunk above 75°F), and Lincoln tended to have the lowest temperatures, with many temperatures below 45°F, whereas Sedona's temperatures all ranged from about 48–68°F.

b. Sedona had the most day-to-day consistency in its high temperatures that month (the temperatures stayed between 48–68°F all month), whereas Lincoln had the least consistency because its temperatures ranged from about 10°F all the way to about 75°F.

Activity 2-6: Sporting Examples

a. The observational units are the students enrolled in one or the other of these sections.

b. The variables are *section* (categorical, binary), *grade* (categorical, not binary), and *total points earned* (quantitative).

c. Students in the regular section tended to score more points than those in the sports section. Scores in the regular section appear to be centered around 340 (85% of the possible points), whereas those in the sports section are centered around 310–320 points (a bit less than 80% of the possible points). Scores in the sports section are more spread out than those in the regular section. Students in the sports section had the six lowest scores, all less than 260 points, but that section also had the highest overall score, greater than 390 points.

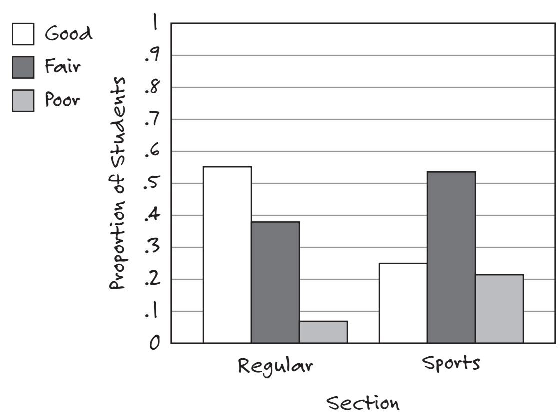
d. Students in the regular section tended to score more points than those in the sports section. Most students in the regular section scored between 300–380 points, with a center of approximately 340 points. In contrast, many students in the sports section scored less than 300 points, and the center was approximately 310–320 points.

e. No, some sports students scored more points than some regular students. The statistical tendency means that a typical student in the regular section scored more points than a typical student in the sports section.

f. The proportions are found by dividing the counts by 29 for the regular section and by 28 for the sports section. These proportions are

Regular section:	.552 good	.379 fair	.069 poor
Sports section:	.250 good	.536 fair	.214 poor

g. The bar graphs follow:



h. The bar graphs reveal similar results to the dotplots: Students in the regular section tended to score higher than those in the sports section. More than half of the regular students were in the good category, compared to only one-fourth of the students in the sports section. At the other extreme, only 6.9% of the regular students did poor work, compared to 21.4% of sports students.

i. You cannot draw a cause-and-effect conclusion between the type of section and student performance. You will study these issues again in the next topic, but one key is that students self-selected which section to take. Perhaps those who chose to take the sports section had lower academic aptitude than those who selected the regular section, or perhaps students were sleepier in the sports section because it met earlier in the day.

Solutions

In-Class Activities

Activity 3-1: Elvis Presley and Alf Landon

a. Population: all adult Americans (record company interest)

Sample: those listening to the radio who called in

b. No, 56% is probably not an accurate reflection of the opinions of all adult Americans on this issue. People who chose to call in (who took the time and were willing to spend the money) probably felt differently and more strongly about the issue than other adult Americans. The timing (on the anniversary of Elvis' death) could also have influenced the opinions of those who called. You also have no indication of how widely distributed across the country the radio stations were (perhaps there could be bias if the stations tended to be mostly from the south).

c. Population: all Americans eligible to vote in 1936

Sample: the 2.4 million who returned the questionnaires

d. The *Literary Digest's* prediction was in error because its sampling method was biased. By sampling people who owned vehicles and telephones in 1936, *Literary Digest* was sampling from a subset of the population who during that period tended to be wealthy. Historically, the wealthy have tended to support the Republican candidate (conservative), whereas those without money have tended to vote Democrat (for social change). Thus, the pollsters contacted primarily

Republican voters, but on election day, there was a heavy Democratic turn out. Furthermore, those who chose to respond were probably more dissatisfied with the incumbent (Roosevelt) than those who chose not to respond.

e.
- *The 56% of callers who believed that Elvis was alive:* statistic
- *The 57% of voters who indicated they would vote for Alf Landon:* statistic
- *The 63% of votes who actually voted for Franklin Roosevelt:* parameter

f.
- *The proportion of students in your class who use instant-messaging or text-messaging on a daily basis:* statistic
- *The proportion of students at your school who use instant-messaging or text-messaging on a daily basis:* parameter
- *The average number of hours students at your school spent watching television last week:* parameter
- *The average number of hours students in your class slept last night:* statistic

g.
- *The proportion of voters who voted for President Bush in the 2004 election:* parameter
- *The proportion of voters surveyed by CNN who voted for John Kerry in the 2004 election:* statistic
- *The proportion of voters among your school's faculty members who voted for Ralph Nader in the 2004 election:* parameter (assuming the population is all of your school's voting faculty members)
- The *average number of points scored in a Super Bowl game:* parameter (assuming the population is all Super Bowl games)

h. A categorical variable leads to a parameter or statistic that is a proportion; a quantitative variable leads to a parameter or statistic that is an average.

Activity 3-2: Self-Injuries

a. Observational units: students

Variable: whether they had injured themselves Type: binary categorical

b. Population: all American college students

Sample: the 2875 students from Cornell and Princeton who responded to the survey

c. The sample size is 2875.

d. The number 17% is a statistic; it is a proportion derived from the sample of students.

e. This percentage is unlikely to be representative of all college students in the world because the sample was taken from two U.S. colleges, and the college experience in the United States is very different from the rest of the world. It is not even clear that it would be representative of all U.S. colleges: Both schools in the survey were Ivy League schools, so their students would hardly be "typical" U.S. college students and might also have distinct types of stress and social reactions to stress.

Activity 3-3: Candy and Longevity

a. No, this is unlikely to be a representative sample of the health habits of all adult Americans. Everyone in the sample was male and college-educated (to some extent) at an Ivy League school at least 30 years before the study—not the picture of the "average" American. In addition to the gender differences, the health knowledge and access to medical care by the men in this sample could differ from the rest of the population.

b. The proportion who consumed candy is 4529/7841 or .5776. This number is a statistic.

c. The proportion of nonconsumers who had died is 247/3312 or .075. The proportion of consumers who had died is 267/4529 or .059.

d. Observational units: 7841 men who entered Harvard between 1916 and 1950

Explanatory variable: whether they consumed candy — Type: binary categorical

Response variable: whether they had died by the end of 1993 — Type: binary categorical

e. Perhaps men who like candy also like to exercise regularly, and perhaps those who do not tend to eat candy do not like to exercise regularly. In this case, it might be the exercise that increases lifespan, rather than the candy. (Other possibilities include differences in diet, differences in family size, and happiness levels.)

f. The proportion of nonconsumers who had never smoked is 1201/3312 or .363. The proportion of consumers who had never smoked is 1852/4529 or .409.

g. A greater percentage of the candy consumers had never smoked. The higher death rate among those who did not tend to consume candy might have been due to smoking rather than not eating candy.

Activity 3-4: Sporting Examples

a. Observational units: statistics students

Explanatory variable: section (exclusively sports examples or not) — Type: binary categorical

Response variable: performance (points earned) — Type: quantitative

b. You know this is an observational study because the students self-selected into the two sections. The researcher (professor) merely passively observed the students' selections and subsequent performances.

c. No, it is not legitimate to conclude that the sports examples caused the lower academic performance. One obvious confounding variable would be the time of the class. The section with exclusively sports examples was offered at a different hour of the morning than the other section. Perhaps there was an honors class being given at the same time as the earlier class, so a larger percentage of honors students (regardless of their interests) had to sign up for the late class. Or maybe the students in the sports-examples section meeting earlier in the day were not as awake during classtime and that was responsible for their lower academic performance.

Activity 3-5: Childhood Obesity and Sleep

a. The explanatory variable is the amount of sleep that a child gets per night. This is a quantitative variable, although it would be categorical if the sleep data were reported only in intervals (more sleep vs. less sleep). The response variable is whether the child is obese, which is a binary categorical variable.

b. This is an observational study because the researchers passively recorded information about the children's sleeping habits. They did not impose a certain amount of sleep on children. Therefore, it is not appropriate to draw a cause-and-effect conclusion that less sleep causes a higher rate of obesity. Children who get less sleep might differ in some other way that could account for the increased rate of obesity. For example, amount of exercise could be a confounding variable. Perhaps children who exercise less have more trouble sleeping, in which case exercise would be confounded with sleep. You have no way of knowing whether the higher rate of obesity is due to less sleep or less exercise, or both, or due to some other variable that is also related to both sleep and obesity.

c. The population from which these children were selected is apparently all children aged 5–10 in primary schools in the city of Trois-Rivières. These Quebec children might not be representative of all children in this age group worldwide, so you should be cautious about generalizing that a relationship between sleep and obesity exists for children around the world.

Solutions

● ● ● In-Class Activities

Activity 4-1: Sampling Words

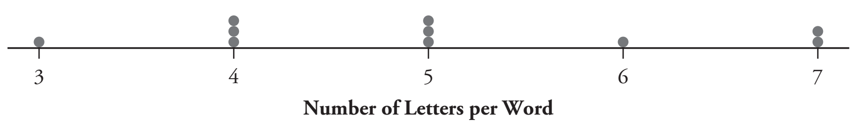
a. Answers will vary. The answers given here are one example.

b. Here are completed tables:

Word	score	forth	whether	have	might
Number of Letters	5	5	7	4	5

Word	did	here	full	resolve	perish
Number of Letters	3	4	4	7	6

c. Dotplot:



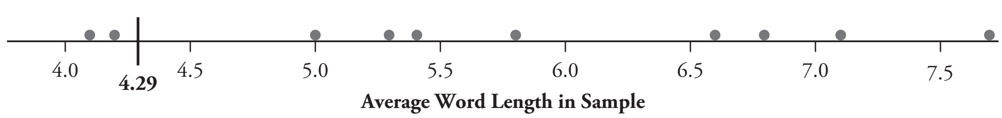
Observational units: words

Variable: number of letters per word

Type: quantitative

d. The average is five letters per word. This number is a statistic.

e. An example set of responses from one class follows.



f. Observational units: samples of 10 words

Variable: average number of letters per word (quantitative)

g. In this example, 8/10 or .8 of the students produced a sample average greater than 4.29 letters per word.

h. Yes, this sampling method appears to be biased. It appears to overestimate the population mean. This is evident from the dotplot because it is centered at about 5.7 (rather than 4.29), and it indicates that a large proportion of the class selected samples with averages greater than 4.29.

i. Your eyes are most likely drawn to the longer words, and you tend to overlook the short, common words such as *a, and, is,* and *or*. Thus, when you try to choose representative samples, you do not select enough short words in your sample.

j. If you use this method, you would also be likely to select too many long words in your sample because the long words take up more space on the page and therefore have a greater chance of being selected when you blindly point to a location.

k. No, increasing the sample size will not make up for the biased sampling *method*. You would still tend to overrepresent the long words.

l. You need to employ a truly random method to select the words. You could write each word on the same size slip of paper, put each slip in a hat, mix them thoroughly, and then draw ten slips from the hat.

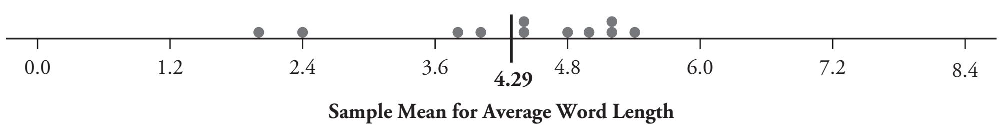
Activity 4-2: Sampling Words

a. Many answers are possible. The following was obtained from the Random Digits Table starting at the beginning of line 60:

	1	2	3	4	5
Random Digits	031	025	052	076	059
Word	now	That	can	A	a
Word Length	3	4	3	1	1

b. The average word length is 2.4 letters per word.

c. Answers will vary. The following is an example from one class.



d. The distribution is much closer to being centered at 4.29 and has a smaller horizontal spread than the previous one did (though the latter is not always the case).

e. The sample averages are roughly split evenly on both sides of 4.29.

f. Yes, random sampling appears to have produced unbiased estimates of the average word length in the population.

Activity 4-3: Sampling Words

a. Answers will vary. The following are from one particular running of the applet.

	1	2	3	4	5
Word	The	These	here	for	Should
Number of Letters	3	5	4	3	6

Average number of letters: 4.2

b. You will probably not obtain the same sample of words or the same average length the second time.

c. The average of the 500 sample averages is 4.31 letters per word.

d. Yes, this appears to be "around" 4.29.

e. Answers will vary according to student expectation.

f. The center of this distribution should also be near 4.29 but the horizontal spread is much smaller.

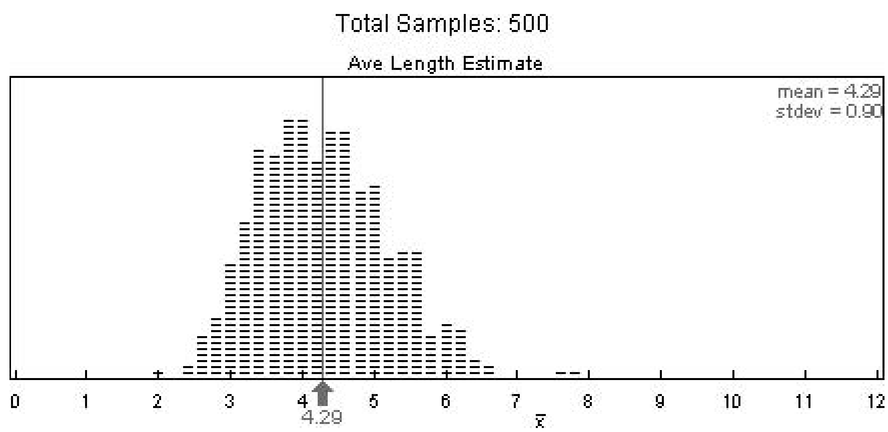
g. The distribution of the samples of size 20 has less variability (more consistency) in the values of the sample average word length.

h. The result of a single sample is more likely to be close to 4.29 with a sample of size 20 than with a sample of size 5.

i. No, increasing the sample size when using a biased sampling method will not reduce the bias. The results from different samples will tend to be closer together but will still be centered in the wrong location (not around the parameter value of interest). If you want to reduce the bias, you must change the sampling *method.*

Activity 4-4: Sampling Words

a. One example set of results follows.



b. Both distributions are roughly bell-shaped, centered at about 4.29 words, with a horizontal spread from about 2 to 7 words.

c. Yes, these distributions seem to have similar variability.

d. No, not much changed when you sampled from the larger population.

Activity 4-5: Back to Sleep

a. The population of interest is all infants younger than eight months in the United States in those years. The sampling frame is the list of households with such infants, generated from birth records, infant photography companies, and infant formula companies. The sample consists of the infants in the 1002 households whose mothers (or other caregivers) participated in the interview.

b. The sample size is 1002. (Actually, a total of 1015 infants were in the sample because some households had twins.)

c. The researchers did not technically obtain a simple random sample of infants. One reason is that the sampling frame did not include the entire population. Another reason is that more than half of the numbers called did not lead to an interview. Infants who were not included in the sampling frame or whose mothers declined to participate might differ systematically in some ways from those who were included. Nevertheless, the researchers did use randomness to select their sample, and they probably obtained as representative a sample as reasonably possible.

d. Perhaps mothers in those groups were in a lower economic class and therefore less likely to have phones in the first place, or perhaps they had to work so their children were in daycare.

e. These comparisons address the issue of bias, not precision. The sampling method was slightly biased with regard to the mother's race and age and the infant's birth weight.

f. These percentages are statistics because they are based on the sample.

g. The large sample size produces high precision. This means that the sample statistics are likely to be close to their population counterparts. For example, the population proportion of infants who sleep on their backs should be close to the sample proportion who sleep on their backs.

h. The sample size for subgroups is smaller than for the whole group, so the sample results would be less precise.

Solutions

In-Class Activities

Activity 5-1: Testing Strength Shoes

a. No, this is anecdotal evidence based on just two of your friends. The friend who wears the strength shoe may be much more athletic in general than your other friend, or may be taller, etc., which would explain why he or she is able to jump farther.

b. Explanatory: whether the individual wears strength shoes — Type: binary categorical

Response: length of jump — Type: quantitative

c. No, you cannot legitimately conclude that strength shoes cause longer jumps because the subjects self-selected which type of shoe they would wear, so your results are from an observational study. There could be something else different about people who choose to wear the strength shoes. For example, males might be more likely to choose the strength shoe than females.

d. Randomly assign six of the subjects to each group.

e. You could flip a coin for each subject. If it lands heads up, that subject will wear ordinary shoes; otherwise, the subject will wear strength shoes. Continue to flip the coin until you have six subjects in the ordinary (or strength) shoe group. If you have not filled both groups, the remaining subjects should all be placed in the unfilled group. Technically, this is not random assignment, even though this approach is used quite frequently. A more correct approach would be to number each subject and put the numbers in a bag, and then the first six numbers drawn out of the hat are assigned to the "strength shoe" group and the others are assigned to the "ordinary shoe" group.

Activity 5-2: Testing Strength Shoes

a. Answers will vary. Below is one example.

Strength Shoe Group			Ordinary Shoe Group		
Name	Gender	Height	Name	Gender	Height
Brad	Male	70	Audrey	Female	67
Mary	Female	66	Michael	Male	71
Peter	Male	69	Russ	Male	68
Kyle	Male	71	Patrick	Male	70
Barbie	Female	63	Anna	Female	61
Matt	Male	73	Shawn	Male	67

b. Strength shoe group: .667 Ordinary shoe group: .667

Difference (*strength* − *ordinary*): 0

c. Strength shoe group: 68.67 inches Ordinary shoe group: 67.33 inches

Difference (*strength* − *ordinary*): 1.33

d. No, the two groups are not identical with regard to both of these variables, but they are similar.

e. Answers will vary from class to class. The dotplot of differences in proportions should be roughly symmetric and centered around zero. The larger the class, the more symmetric the plot should be. The horizontal axis should be labeled "difference in sample proportions" with a scale from −1 to 1, and the vertical axis should display the count/tally of each difference.

The dotplot of the differences in average heights should also be roughly symmetric and centered at zero. The horizontal axis should be labeled "difference in sample heights" with a scale from approximately −6 to 6, and the vertical axis should display the count/tally of each difference. For both plots, the observational units are the random assignments.

f. Both plots should appear to be roughly centered around zero. This indicates that random assignment is effective because it is "balancing out" the proportion of men/women and the heights in both groups. In the long run, both groups are roughly the same with regard to these variables because, on average, the difference between them is zero. In particular, you have no prior suspicion that one group will have certain characteristics that differ from the other group.

Activity 5-3: Testing Strength Shoes

a. Answers will vary. The answers here are from one particular running of the applet.

Group A: .5 Group B: .8333

Difference: −.3333

b. No, you probably will not get the exact same assignment of subjects to the groups nor the same difference in proportions.

c. Answers will vary as this is a prediction.

d. The distribution should be centered at zero. It may or may not be what the students predicted.

e. Random assignment does not *always* balance out the gender variable exactly, but it does *tend to* balance out gender between the two groups. You can see this because the dotplot is centered at zero, and zero is the difference that occurred most frequently. Occasionally, there were differences as large or small as $\pm .8333$.

f. This distribution is roughly symmetric, centered near zero, ranging from roughly -5.5 to 5.5. This indicates that the randomization also tends to balance out the heights between the two groups.

g. Answers will vary; the question asks for student expectation.

h. Both dotplots (for gene and *x*-variable) are roughly symmetric and centered at zero. This indicates that random assignment tends to balance out variables even when they are unseen or unrecorded.

i. By randomly assigning the subjects to the two groups, you have (hopefully) balanced out all other potentially confounding variables, making the only difference between the two groups the type of shoe. Therefore, if you then find that the strength shoe group jumps substantially farther, on average, than the ordinary shoe group, you would be able to conclude that the increase was due to the strength shoe because there should not be any other explanation for a difference between the two groups.

5

Activity 5-4: Botox for Back Pain

a. No, you should not conclude that Botox is an effective treatment for back pain. There was no control group of patients who did not receive the treatment with which to compare. Perhaps the patients' pain decreased simply because time had passed.

b. This is a better design because it allows you to compare the response of a treatment group with that of a control group and because you are randomly assigning the subjects to the two groups so you can control for confounding variables. The only difference between the patients in the two groups should be the Botox treatment.

c. Perhaps the patients who are given the Botox treatment will think they feel better simply because they are given *any* treatment. Back pain is very subjective and hard to measure, so any sort of treatment may cause some of the subjects to believe their pain has decreased.

d. The researchers could inject the control group with saline or some other harmless substance so that patients in both groups believe they are receiving a treatment, equalizing any psychological effects between the two groups.

Activity 5-5: Memorizing Letters

a. This is an experiment because the teacher actively imposed the treatment (grouping of letters) on each subject/student.

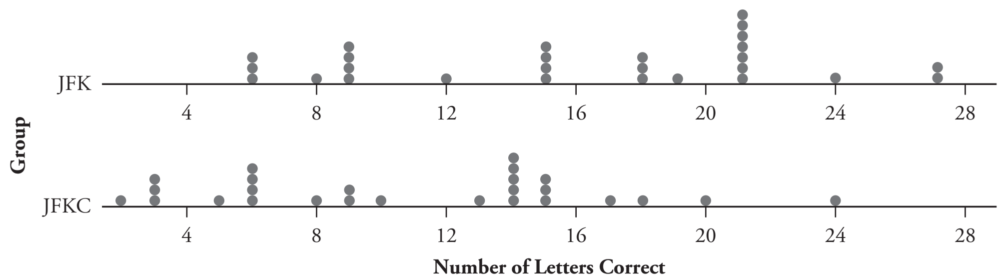
b. Explanatory: which sequence of letters you were given Type: binary categorical

Response: number of letters correctly memorized Type: quantitative

c. The instructor randomly decided which grouping of letters each student would receive. This was important because it prevented self-selection and controlled for confounding variables. You should not expect any differences between the groups prior to the treatment.

d. The students were blind to the fact that there were two different groupings of letters given out initially. They were unaware that you were trying to compare the effect of these familiar and unfamiliar groupings, so they could not unintentionally influence the results.

e. Answers will vary. Here is one representative set of answers.



f. Yes, these data appear to support the conjecture that those who receive the letters in convenient three-letter chunks tend to memorize more letters. The center of this plot is about six letters higher than for the JFCK plot.

g. Yes, because this was a well-designed, randomized controlled experiment you could legitimately conclude that the grouping of letters into familiar chunks *caused* the higher scores. Because you randomly assigned the students to each type of grouping, there should have been roughly an equal number of good memorizers in both the JFK and JFKC groups, so the randomization controlled for this potentially confounding variable.

Activity 5-6: Nicotine Lozenge

a. This is an experiment because the researchers imposed the nicotine lozenge (or placebo) on subjects.

b. The experimental units are the smokers interested in quitting who volunteered to participate in the study.

c. The explanatory variable is whether the smoker was given a nicotine lozenge or a placebo. This variable is categorical and binary. The response variable is whether the smoker successfully quit smoking by the end of the study. This variable is also categorical and binary.

d. This information validates that the random assignment achieved its goal of balancing out all of these variables, which could potentially be related to a smoker's ability to quit, between the nicotine lozenge and placebo groups. Thus, if the nicotine lozenge group has a higher proportion who quit smoking, then the researchers can attribute that to the lozenge, not to any of these background variables, because they were similar between the groups.

e. Yes. Because this was a randomized experiment, it is legitimate to conclude that the nicotine lozenge caused the increase in the proportion of subjects who successfully quit smoking.

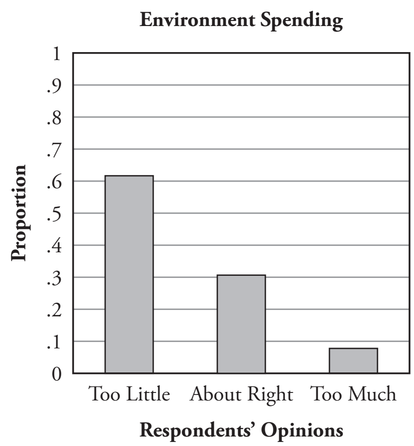
Solutions

In-Class Activities

Activity 6-1: Government Spending

a. Too little: 398/646 = .616 About right: 198/646 = .307
Too much: 50/646 = .077

b. Here is the bar graph for the marginal distribution of the variable *opinion about federal spending on the environment:*



c. Assuming the sample is representative, you can say that Americans tend to believe that the government is spending too little on the environment. More than 60% of the sample felt this way, whereas less than 8% felt that the government is spending too much. About 30% of this sample felt that government spending on the environment is right where it should be.

d. The proportion of liberal respondents who say the federal government spends too little on the environment is 127/155 or .819.

e. The proportion of liberal respondents who say the federal government spends the right amount on the environment is 27/155 or .174.

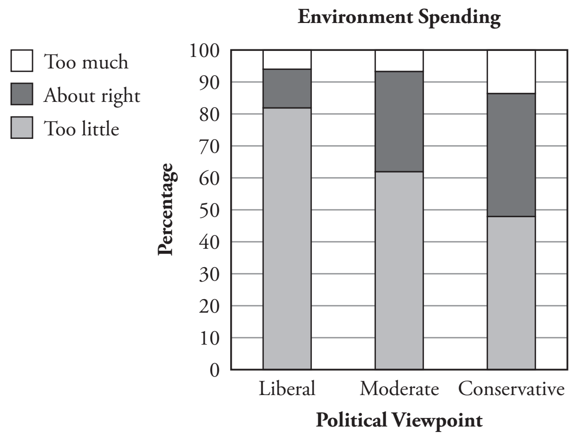
f. The proportion of liberal respondents who say the federal government spends too much on the environment is 1/155 or .006.

g. Yes, these three proportions add up to 1.00 if you don't use the rounded proportions.

h. Here is the conditional distribution for the "liberal" column of the table:

	Liberal	Moderate	Conservative
Too Little	.819	.619	.479
About Right	.174	.314	.385
Too Much	.006	.067	1.36
Total	1.000	1.000	1.000

i. The completed segmented bar graph follows:



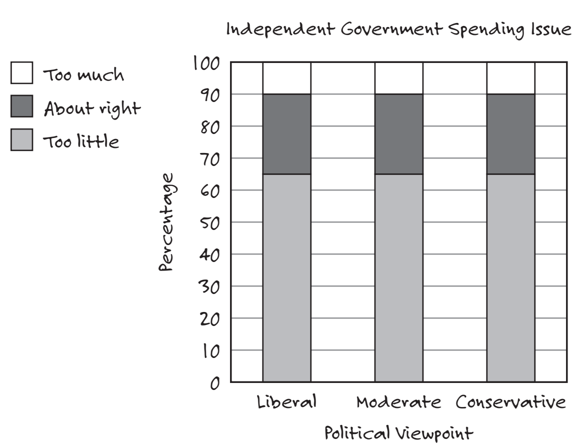
j. Yes, the distribution of spending opinion seems to differ among the three political groups. The more liberal the group, the more likely they are to believe that the government is not spending enough on the environment. Whereas 82% of the liberals in the sample feel the government is spending too little money, 62% of the moderates and only 48% of the conservatives believe this; on the other hand, 14% of conservatives believe the government is spending too much on the environment compared to less than 1% (.6%) of the liberals and 7% of the moderates.

l. Approximately 35% of the liberals in the sample believe the government spends too much on the space program.

m. There is a difference—but not a big difference—in how the three political groups feel about government spending on the space program. Roughly 40% of each group stated the government is spending too much, roughly 10% stated the government is spending too little, and roughly 50% stated the government is spending about the right amount.

n. Government spending is closest to being independent of political viewpoint with the issue of the space program. You can tell because the breakdown of the three segments in the segmented bar graph is nearly identical across the three political parties, whereas they are quite distinct across the three political parties in the environmental spending graph.

o. Answers will vary. Here is an example graph for an issue where *opinion about government spending* is perfectly independent of *political viewpoint:*



The goal is for the three segments to have the same breakdown across the three groups (though not necessarily being an equal size within each political party).

p. The proportion who identify themselves as liberal is 127/398 or .319.

q. No, the proportion in part p is less than half the value found in part d (.819).

Activity 6-2: AZT and HIV

a. This is an experiment because the researchers randomly assigned the subjects to the AZT and control (placebo) groups.

b. Explanatory variable: AZT or placebo

Response variable: whether the baby was born HIV-infected

c. Here is the 2 × 2 table:

	Placebo	AZT	Total
HIV-infected	40	13	53
Not HIV-infected	143	167	310
Total	183	180	363

d. For the placebo group, 40/183 or .219. For the AZT group, 13/180 or .072. Yes, these calculations are consistent with the segmented bar graph.

e. The difference is .219 – .072 or .147 (or .146 if using more than three decimal places for the proportions). This difference does not appear to be terribly large.

f. The ratio is .219/.072 or 3.04 (or 3.03 if using more than three decimal places for the proportions). The risk of a baby being born HIV-infected is more than three times greater for those whose mothers were in the placebo group.

g. Yes, you can legitimately conclude that AZT is the cause of the three-fold reduction in HIV-infection rate compared to the placebo group because this was a well-designed, randomized experiment. These results can be cautiously generalized to HIV-positive pregnant women (you don't have much information about how the women in the sample were selected).

Activity 6-3: Lifetime Achievements

a. Gender: binary categorical

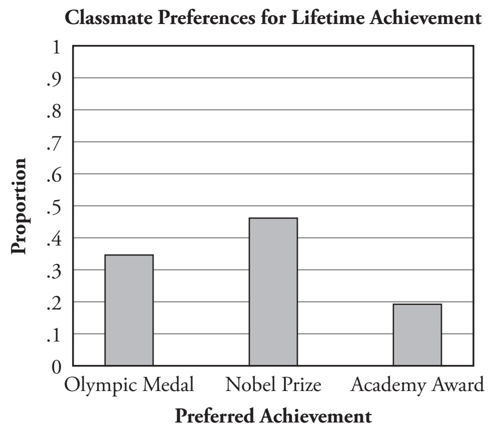
Preferred lifetime achievement: categorical (nonbinary)

b. Explanatory variable: gender

Response variable: preferred lifetime achievement

c. Answers will vary by class. Here is one representative set of answers from a class of 78 students:

Olympic Medal	Nobel Prize	Academy Award	Total
.346	.462	.192	1



More than 45% of this class would prefer to win a Nobel prize, whereas less than 20% aspire to win an Academy Award. Just over one-third of the class would like their greatest lifetime achievement to be winning an Olympic medal.

d. Using the example data from part c gives these counts:

	Male	Female
Olympic Medal	11	16
Nobel Prize	12	24
Academy Award	1	14

e. With this information you would only be able to fill in the totals; you need to know the information on both gender and preferred lifetime achievement simultaneously in order to fill in each cell of the table.

Activity 6-4: Hypothetical Hospital Recovery Rates

a. A: 800/1000 = .8 B: 900/1000 = .9

Hospital B saved the higher percentage of its patients.

b. A: (590 + 210) /(600 + 400) = .8 B: (870 + 30) / (900 + 100) = .9

c. A: 590/600 = .980 B: 879/900 = .967

Hospital A saved the greater percentage of its patients who had been in fair condition.

d. A: 210/400 = .525 B: 30/100 = .300

Hospital A saved the greater percentage of its patients who had been in poor condition.

e. Hospital A tends to treat a higher proportion of patients who are in poor condition to begin with and whose chances of survival are not very high, regardless of where they seek treatment. Hospital A does a better job with both types of patients, but their overall survival percentage is lower because they treat such a higher rate of patients in poor condition than Hospital B does.

f. If you were ill, you should go to Hospital A, which does a better job of treating both types of patients. No matter how sick you are, you stand a better chance of surviving at Hospital A.

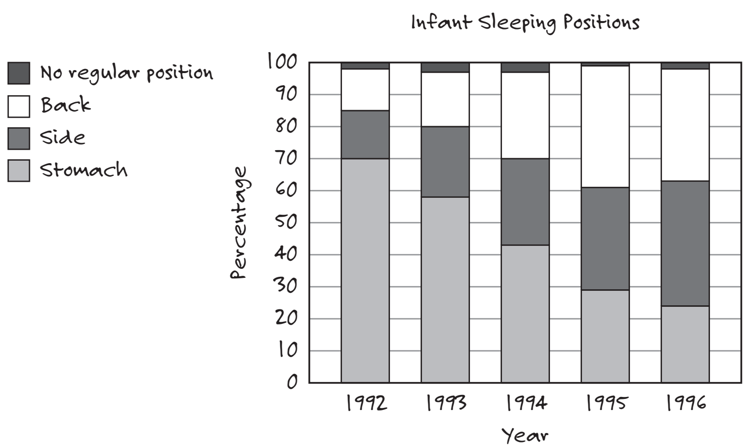
Activity 6-5: Back to Sleep

a. These are statistics because they pertain to the sample of parents interviewed in the study. They are not calculated based on the population of all American parents with infants less than eight months old.

b. These are conditional distributions. Each year has its own percentages sleeping in each of the positions, which sum to 100%.

c. The explanatory variable is *year;* the response variable is *infant's sleep position.*

d. The segmented bar graph is shown here:



e. The segmented bar graph reveals that the distribution of infant sleep positions changed considerably over this five-year period. The percentage of parents who placed infants on their stomachs declined dramatically, from more than two-thirds (70%) to less than one-fourth (24%) in these five years. The proportions of infants placed on their backs and sides both increased during this period. Because this is an observational study and not an experiment, you cannot say that the recommendation or promotional campaign caused these changes, but (assuming the sample is representative) it is nonetheless heartening to find that parents were generally changing their habits and placing infants in safer positions to sleep.

Solutions

In-Class Activities

Activity 7-1: Matching Game

Consider the following seven variables:

a. These are all quantitative variables (except perhaps "jersey number;" even though it is numerical, it does not make sense to determine the "average jersey number").

b. **1.** Variable: point values of letters in the board game Scrabble

Explanation: The values are regularly spaced (point values are whole numbers) with most letters being worth few points.

2. Variable: prices of properties on the Monopoly game board

Explanation: The values are very regularly spaced with two properties having exactly the same prices.

3. Variable: annual snowfall amounts for a sample of cities around the United States

Explanation: Most cities have little or no snow, but some have quite a bit.

4. Variable: jersey numbers of Cal Poly football players in 2006

Explanation: There are virtually no repeats and the distribution covers nearly all the values from 1–99.

5. Variable: blood pressure measurements for a sample of healthy adults

Explanation: The values are fairly symmetric, with both a high and low outlier.

6. Variable: weights of rowers on the 2004 U.S. men's Olympic team

Explanation: There is one low outlier (coxswain) and two weight classes.

7. Variable: quiz percentages for a class of statistics students (quizzes were quite straightforward for most students)

Explanation: Most of the values are high (near 80–90) with a few very low outliers.

c. Dotplots 1 and 3 have a similar shape (they are skewed to the right), whereas dotplots 6 and 7 are both skewed to the left with outliers.

Activity 7-2: Rowers' Weights

a. The weights of the rowers on the men's 2004 Olympic team are skewed to the left, ranging from about 120 to 230 lbs. There are two clusters of weights, one tight cluster around 160 lbs and one that ranges from about 180 to 230 lbs. There is one low outlier who weighs only 120 lbs.

b. Cipollone is the apparent outlier at 120 lbs. His job (the coxswain) is to call out the cadence to the rowers—he does not row himself. So it is important that he not add excess weight to the boat.

c. There is a cluster of rowers whose weights seem to be no more than 160 lbs. These rowers are all involved in "lightweight" events, which require them to weigh below a certain amount on race day. The upper cluster are not involved in the lightweight events and have no upper limit on their weights.

Activity 7-3: British Monarchs' Reigns

a. The current monarch is Queen Elizabeth II. She is not represented because her reign had not ended (as of December 2006).

b. The longest reign was that of Queen Victoria. Her reign was 63 years.

c. The shortest reign was less than 12 months (0). Edward V spent the least time on the throne. He ascended the throne at the age of 12 in 1483 upon the death of his father Edward IV, but was deposed two months later by his uncle Richard, Duke of Gloucester.

d. 6|3 represents 63 years. Victoria reigned for 63 years.

e. *Four* monarchs reigned for 13 years. You can tell this from the stemplot because there are four 3s on the 10s-stem.

f. This distribution is skewed to the right.

g. Half of the monarchs reigned for more than 19.5 years and half of them reigned for less than 19.5 years.

h. One-quarter of the monarchs reigned for fewer than 9.5 years. One-quarter of the monarchs reigned for more than 34 years.

Activity 7-4: Population Growth

a. The completed stemplot is shown here.

Western States		Eastern States
9998888550	0	0333445556889999
43322	1	00014567
982100	2	136
0	3	
0	4	
	5	
6	6	

leaf unit = one percentage point

b. Western states tend to have higher percentage growths than eastern states. The distributions of percentage growth for both the eastern and western states are both skewed to the right. The percentages in the 26 eastern states display less variability than the western states, ranging from near 0% to about 26%, with a center of about 10%. The percentages in the 24 western states range from a minimum near 0% to a maximum 30%, with a high outlier at 40% and a center of about 13%. There is also one western state with an extremely high population growth rate (Nevada)—more than 66%.

Activity 7-5: Diabetes Diagnoses

a. Observational units: 548 subjects with diabetes

Variable: age when diagnosed

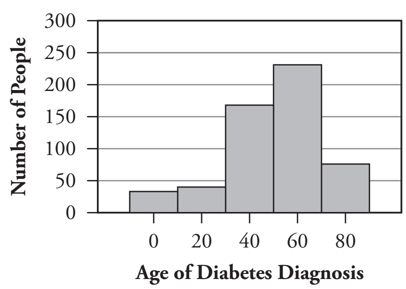
b. No, it is not practical to construct a dotplot or stemplot to display this distribution because there are so many data values.

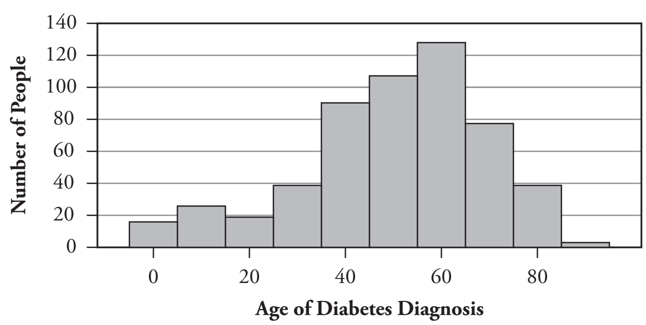
c. The proportion of people who were diagnosed with diabetes before the age of 18 is (4 + 23 + 14 + 7)/548 = 48/548 = .088.

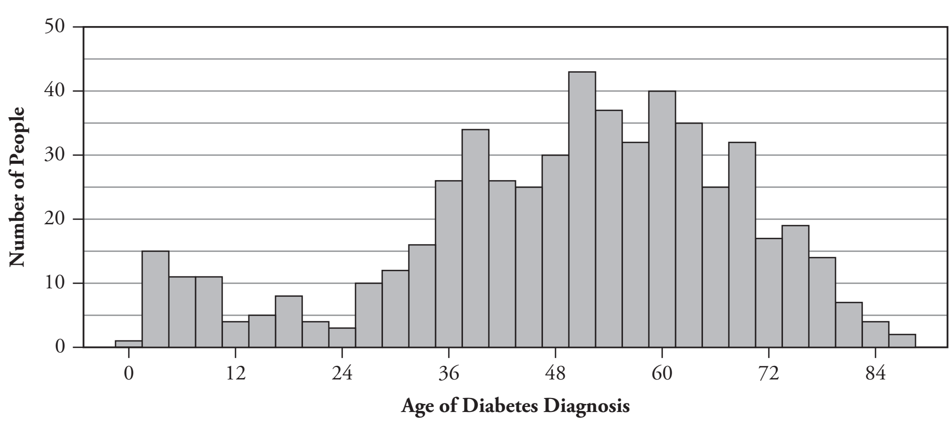
d. The proportion of people who were diagnosed with diabetes after the age of 62 is (44 + 40 + 33 + 16 + 5 + 1)/548 + 139/548 = .254.

e. These ages range from about 1 year to 88 years and are centered at about 50 years. Most of the ages are clustered between 30 and 80 years, but there is another small cluster indicating juvenile onset diabetes—between 1 to 15 years of age.

f. Of these three choices, using 10 intervals seems to provide the most informative histogram for these data. This provides enough information for you to see both clusters without adding too much clutter to the graph.







g. The first graph *appears* to have a uniform shape; the second is slightly skewed to the right. But neither is a legitimate histogram of the Olympic rowers' weights because the variable (*weight*) is not on the horizontal axis—it is on the vertical axis! For example, it would be very misleading to say the "center" of the weights was around Dan Berry's weight—he actually has one of the greatest weight values. The rowers' names are of less interest than the pattern in the weight values to the behavior of the weights.

Activity 7-6: Go Take a Hike!

a. The stemplot is shown here.

```
0 | 68
1 | 00000555555588
2 | 0000000001225555556668      leaf unit = .1 mile
3 | 000002224458
4 | 00000566
5 | 00055668
6 | 0000
7 | 004
8 |
9 | 5
```

b. The distribution of hike distances is sharply skewed to the right, indicating there are many hikes on the short side and only a few longer hikes. A typical hike is between 2 and 3 miles. Most hikes are between 1 and 6 miles, but two hikes are less than a mile and a few are more than 6 miles. The longest hike is 9.5 miles, which is a bit unusual and could be considered an outlier because this hike is more than two miles longer than the next longest hike (7.4 miles). Many hikes have a reported distance that is a multiple of a whole number or a half number of miles.

Solutions

In-Class Activities

Activity 8-1: Sleeping Times

a. Observational units: students in a statistics course

Explanatory variable: section of the course — Type: categorical

Response variable: sleeping time in hours — Type: quantitative

b. The centers are not all similar: The center for Section 1 is noticeably less than the center for Section 2, which is less than the center for Section 3.

c. The center for Section 1 is about 6 hours. The peak of the distribution is about 6 hours. Approximately half of the times are less than this, and approximately half are more.

d. The mean is 106.25/17 or 6.25 hours.

e. The median is 6.25 hours. You calculate (17+1)/2 = 9. The ninth observation, counting from either end, is 6.25 hours.

f. Section 2 mean: 7.000 — Section 3 mean: 7.523

Explanation: The mean for Section 2 should be less because there are many observations at 8 hours in Section 3, but not in Section 2, and the classes behave similarly at 6 hours and less.

g. Section 2 median: 7.0 — Section 3 median: 7.5

h. No, based on this example, the mean and median of a dataset do not always equal each other.

i. The mode of the section variable is Section 3. More students were enrolled in this section than in any other section. Perhaps that is because it was offered much later in the morning!

Activity 8-2: Game Show Prizes

a. The median prize amount is (750 + 1000)/2 or $875.

b. Answers will vary (student prediction).

c. Mean: $131,478 Median: $875

d. Answers will vary regarding the predictions. The mean and median are not close to each other. This makes sense because the data is strongly skewed to the right; there are several extremely large values that "pull" the mean to the right and make it much greater than the median.

e. Six prize amounts are greater than the mean. This proportion is 6/26 or .231.

f. Thirteen prize amounts are greater than the median. This proportion is 13/26 or .5.

g. If the producers want to give the impression that contestants win huge amounts, they should advertise the mean amount won because it is so much greater than the median.

Activity 8-3: Matching Game

a. Symmetric: prices of properties on the Monopoly game board

Skewed to the left: quiz percentages for a class of statistics students

Skewed to the right: annual snowfall amounts for a sampling of U.S. cities

b. Here is the completed table:

	Monopoly Prices	Snowfall Amounts	Quiz Percentages
Mean	208.6	21.65	88.9
Median	210	19.9	95

c. When the distribution is skewed to the left, the mean is less than the median. When the distribution is skewed to the right, the mean is greater than the median. In symmetric distributions, the mean and median are similar.

Activity 8-4: Rowers' Weights

a. Answers will vary (estimates).

b. The distribution is skewed to the left with an extreme low outlier (the coxswain), so it makes sense that the mean will be less than the median.

	Whole Team	Without Coxswain	With Max Weight at 329	With Max Weight at 2229
Mean	188.93	191.58	195.42	268.5
Median	195	195	195	195

c. Predictions. Answers will vary.

d. See table following part b.

e. Predictions. Answers will vary.

f. See table following part b. (Cipollone's weight was not added back in.)

g. Predictions. Answers will vary.

h. See table following part b.

i. The median is resistant. This conclusion makes sense because to find its value you use only the middle of the data, not the ends where outliers would occur. The mean is *not* resistant, which also makes sense: To find the mean you use *all* the data. Outliers will pull the mean toward them—potentially affecting the value of the mean drastically.

j. The mean only has to stay between the minimum and maximum values in the dataset. (And so, if one of those values changes without bound, so does the mean.)

Activity 8-5: Buckle Up!

a. The states with primary seatbelt laws tend to have higher compliance percentages.

b. The primary mean is 86.4%; the primary median is 86%.

The secondary mean is 77.14%; the secondary median is 77.4%.

The compliance percentage in states with primary seatbelt laws is about 9 points higher, on average, than in states with secondary seatbelt laws. This is a fairly striking difference.

c. No, you cannot draw a causal conclusion between stricter laws and the higher percentage of people who wear seatbelts because this is an observational study and not a well-designed, comparative experiment. There could be many confounding variables that explain the strong association here.

Activity 8-6: Wrongful Conclusions

a. The conclusion drawn is not valid because the mean does not divide the dataset in half. It could be the case that most of the house prices are less than the mean, but one house price is extremely high and is the only one greater than the mean.

b. The conclusion drawn is not valid because the median indicates the middle data value in an ordered set. Even though the median business trip cost $600, the most expensive trip may have cost anything, say $5000, which would definitely bring the total cost over $3000.

c. The conclusion drawn is not valid because the mean does not indicate what percentage of the data values fall above or below it. It is possible that the CEO, presidents, and vice-presidents make up 10% of the company and earn enough between them to bring the mean above the salaries of the other 90% of the employees.

d. The conclusion drawn is not valid because the mode indicates the most frequent choice; this choice does not have to be the preference of more than half of the customers when there are more than two choices. It could be the case that 100 customers chose chocolate, 90 chose vanilla, and 80 chose strawberry. The mode is chocolate, but 170/270 = .63, so 63% prefer a flavor other than chocolate.

Activity 8-7: Readability of Cancer Pamphlets

a. To calculate the mean, you need to know *all* of the actual data values, and you do not know the values for those patients with a reading level below grade 3 or above grade 12.

b. There are 63 patients, so the median is the ordered value in position (63 + 1)/2, or the 32nd value. If you start counting from the low end, you find 6 patients read below grade 3, 10 patients at grade 3 or below, 14 patients at grade 4 or below, 17 patients at grade 5 or below, 20 patients at grade 6 or below, 22 patients at grade 7 or below, 28 patients at grade 8 or below, and 33 patients at grade 9 or below. The 32nd value is therefore at grade level 9, which is the median patient reading level.

c. There are 30 pamphlets, so the median readability level is the average of the 15th and 16th pamphlets. Counting in a similar way, the 15th and 16th readability values are both located at grade level 9, so a pamphlet's median readability level is grade 9.

d. These medians are identical.

e. No. The centers of the distributions (as measured by the medians) are well-matched, but you need to look at both distributions in their entirety and consider all values. The problem is that many patients read at a level below that of the simplest pamphlet's readability level. Seventeen patients read at a level below grade 6, which is the lowest readability level of a pamphlet.

f. 17/63 = .27, so 27% of the patients have a reading level below that of the simplest pamphlet.

Solutions

In-Class Activities

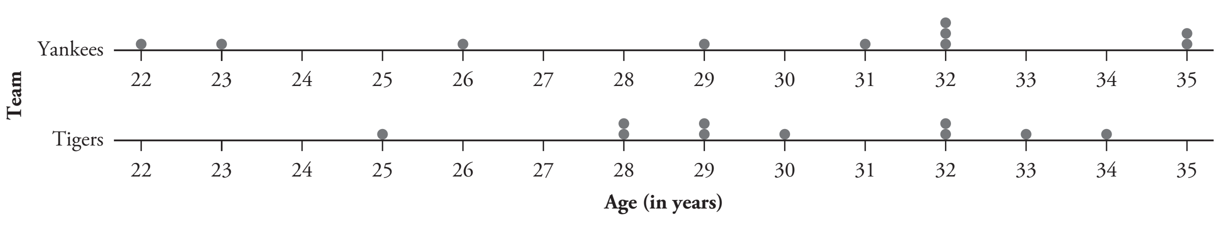
Activity 9-1: Baseball Lineups

a. Observational units: baseball players

Explanatory variable: team — Type: binary categorical

Response variable: age — Type: quantitative

b. Here are the comparative dotplots:



The average age of both teams seems to be about the same (about 30 years), but the spreads are quite different. The 2006 Tigers are much closer together in age than the 2006 Yankees.

c. Yankees Mean: 29.7 years Median: 31.5 years

Tigers Mean: 30 years Median: 29.5 years

The centers of these distributions are relatively similar.

d. No; although the centers are the same, the spreads are very different, with the Yankees having the youngest and oldest players in the two distributions and not much consistency in the ages of their players.

e. The Yankees' lineup appears to have greater variability in its ages.

f. Oldest: 35 years Youngest: 22 years Difference: 13 years

g. Oldest: 34 years Youngest: 25 years Difference: 9 years

h. Lower quartile: 28 years Upper quartile: 32 years IQR: 4 years

i. The Yankees have the greater age range and greater IQR. These values are consistent with the answer to part e.

j. The average age of the starting lineups on both teams is about 30 years, but the Tigers' ages are fairly tightly clustered from 28–34 years, with the exception of one player (Granderson) who is only 25 years old. In comparison, the Yankees' ages range from a low of 22 years to a high of 35 years and also have a larger interquartile range. It is difficult to judge the shape with these small sample sizes, but the distribution of the ages for the Tigers appears more symmetric, whereas the distribution of ages for the Yankees is more skewed to the left.

Activity 9-2: Baseball Lineups

a. Here is the completed table:

Player	Age	Deviation from Mean	Absolute Deviation	Squared Deviation
I. Rodriguez	34	$34 - 30 = 4$	4	16
Casey	32	2	2	4
Perez	33	3	3	9
Inge	29	−1	1	1
Guillen	30	0	0	0
Monroe	29	−1	1	1
Granderson	25	−5	5	25
Gomez	28	−2	2	4
Young	32	2	2	4
Robertson	28	−2	2	4
Total	300	0	22	68

b. This sum in the “deviation from mean” column is zero, which makes sense because the positive deviations from the mean “cancel out” the negative deviations from the mean.

c. See the table in part a. The sum of the absolute deviations is 22 years.

d. The mean of the absolute deviation is 22/10 or 2.2 years.

e. See the table in part a. The sum of the squared deviations is 68 years^2.

f. $68/9 = 7.56$ years^2

g. 2.749 years

h. The standard deviation for the Tiger lineup’s ages is 2.749 years. The standard deviation for the Yankees lineup’s ages is 4.62 years.

As expected, the Yankees’ standard deviation is larger because the ages tend to be located farther from the average age.

i. Answers will vary by student expectation. This change will definitely affect the range and the standard deviation, but it should have little or no effect on the IQR because you are changing only an extreme value (endpoint).

j. See table in part k.

k. Here is the completed table:

	Range	Interquartile Range	Standard Deviation
Original Data	9	4	2.749
With Large Outlier (43)	18	4	4.86
With Huge Outlier (134)	109	4	33.1

l. See table in part k. These results demonstrate that the IQR is resistant, but the range or standard deviation is not. You know this because the value of the IQR does not change when the value of the outlier changes, whereas the range and standard deviation are affected dramatically.

Activity 9-3: Value of Statistics

a. Answers will vary by student prediction. Many students will pick class F, focusing incorrectly on the irregularity in the heights of the bars.

b. Answers will vary by student prediction. Many students will incorrectly predict that class J has the most variability because more of the possible data values appear in the histogram. Students again may incorrectly see class H as having more variability because they are looking at the differences in the heights of the bars.

c. Here is the completed table:

	Class F	Class G	Class H	Class I	Class J
Range	6	8	8	8	8
Interquartile Range	2.75	3	0	8	4.5
Standard Deviation	1.769	2.041	1.18	4	2.657

d. According to these measures of spread, class G has more variability than class F, because class G has more data values farther from the mean.

e. According to these measures of spread, class I has the most variability and class H has the least variability. Class I has more of its data values at the extremes (and far from the center), whereas most of class H's observations are close to the mean. Class J is in between the two.

f. Class F has more bumpiness in its histogram, but has less variability than class G.

g. Class J has the greatest number of distinct values but does not have the most variability among classes H, I, and J.

h. No, based on the two previous questions, variability does not measure either bumpiness or variety; variability measures spread from the center (mean). A distribution can be very "bumpy" without having a great deal of variability, and vice versa. It is more important to consider the overall tendency for data values to be far from the center.

i. Many answers are possible, but all ten values need to be the same so the standard deviation is zero.

j. Only one answer is possible: {1, 1, 1, 1, 1, 9, 9, 9, 9, 9}. This dataset maximizes the distances of observations from the mean and has a standard deviation of 4.22. Any other combination will have a smaller standard deviation. (*Note:* If you did not balance the 1s and 9s, the mean would shift away from 5 and would put the more frequent values closer to the mean.)

Activity 9-4: Placement Exam Scores

a. Yes, this distribution appears to be roughly symmetric and mound-shaped.

b. $\bar{x} + s = 14.08$

$\bar{x} - s = 6.362$

c. The scores in this interval are 7, 8, 9, 10, 11, 12, 13, and 14. There are 16 + 15 + 17 + 32 + 17 + 21 + 12 or 146 of them. The proportion is 146/213 or .685.

d. The scores in this interval are 3, 4, . . . 15, 16, 17. There are 202 scores in this interval and this proportion is 202/213 or .948.

e. This would include scores from 0 to 21, that is, all the scores. Thus 100% of the scores fall within three standard deviations of the mean.

Activity 9-5: SATs and ACTs

a. If applicant Bobby scored 1740 on the SAT, he scored 1740 − 1500 or 240 points greater then the SAT mean.

b. If applicant Kathy scored 30 on the ACT, she scored 30 − 21 or 9 points greater than the ACT mean.

c. No; the scales on these two tests are different, so you cannot conclude that Bobby outperformed Kathy simply because he scored more points above the mean than Kathy did. The 240 and the 9 points cannot be directly compared.

d. Bobby scored 240/240 or 1 standard deviation above the mean.

e. Kathy scored 9/6 or 1.5 standard deviations above the mean.

f. Kathy has the higher z-score.

g. Kathy performed better on her admissions test relative to her peers because her z-score is greater.

h. Peter: $z = (1380 - 1500)/240 = -0.5$ Kelly: $z = (15 - 21)/6 = -1.0$

i. Peter has the higher z-score (less negative as the observation is not as far below the mean).

j. When the observation is less than the mean, the z-score will turn out to be negative.

Activity 9-6: Marriage Ages

a. With 24 people in each group, the median ages are the average of the 12th and 13th ordered values. For husbands, the median age is (30 + 31)/2 or 30.5 years. For wives, the median is (28 + 30)/2 or 29 years. For husbands, the mean age is 35.7 years and for wives, the mean age is 33.8 years. Husbands tend to be a little less than two years older than their wives.

b. The lower quartile is the median of the bottom 12 ordered values, so the average of the 6th and 7th values. For husbands, the lower quartile is (25 + 25)/2 or 25 years and the upper quartile is (51 + 38)/2 or 44.5 years. The IQR is, therefore, 44.5 − 25 or 19.5 years. You can see this by examining the sorted ages for husbands:

19 23 23 25 25 25 ↓ 25 26 26 29 29 30 ↓ 31 31 31 34 35 38 ↓ 51 54 54 60 62 71

For wives, the lower quartile is (24 + 24)/2 or 24 years and the upper quartile is (39 + 44)/2 or 41.5 years, so the IQR is 41.5 − 24 or 17.5 years. The standard deviations are 14.6 and 13.6 years for husbands and wives, respectively. These calculations indicate that the middle 50% of husbands' ages cover a slightly greater distance than the wives' ages by 2 years and that the husbands' ages typically lie slightly farther from the mean, by approximately 1 year on average.

c. The age distributions are quite similar for husbands and wives. Both are skewed to the right, centered around the low 30s or so, with considerable variability from the upper teens through low 70s. The husbands are a bit older on average, and their ages are a bit more spread out than the wives' ages.

d. The ordered difference in couple's ages are

−7, −5, −5, −2, −1, −1, 0, 0, 1, 1, 1, 1, 1, 2, 2, 3, 3, 3, 3, 5, 7, 8, 10, 15

The median is the average of the 12th and 13th ordered values: (1 + 1)/2 or 1 year.

The mean is the sum of these differences divided by 24, which turns out to be 45/24 or 1.9 years.

Notice that the mean of the age differences is equal to the difference in mean ages between husbands and wives: 1.9 = 35.7 − 33.8. But this property does not quite hold for the median.

e. The quartiles are −0.5 and 3, so the IQR is 3.5 years. The standard deviation of these age differences is 4.8 years. The IQR of the differences and the standard deviations of the differences calculated here are less than the individual IQRs (19.5 and 17.5) and the individual standard deviations (14.56 and 13.56) calculated in part b.

f. To be within one standard deviation of the mean is to be within 1.9 ± 4.8 years, which means between –2.9 and 6.7 years. Seventeen of the age differences fall within this interval, which is a proportion of 17/24 or .708, or 70.8%. This percentage is quite close to 68%, which is what the empirical rule predicts. Because the distribution of the age differences does look fairly symmetric and mound-shaped, this outcome is not surprising.

g. The mean and median indicate that, on average, people marry someone within a couple years of their own age. More importantly, the measures of spread are fairly small for the differences, much smaller than for individual ages. This result suggests that there is not much variability in the differences, which suggests that people do tend to marry people of similar ages.

h. The differences have less variability because even though people get married from their teens to seventies (and beyond), they tend to marry people within a few years of their own age.

Solutions

In-Class Activities

Activity 10-1: Natural Selection

a. Observational units: adult male sparrows

Explanatory variable: total length — Type: quantitative

Response variable: whether the sparrow survived — Type: binary categorical

b. This is an observational study because Bumpus simply recorded this information about the sparrows; he did not impose any treatment on them.

c. Median: 159 mm

Lower quartile: 158 mm — Upper quartile: 160 mm

Minimum: 153mm — Maximum: 166 mm

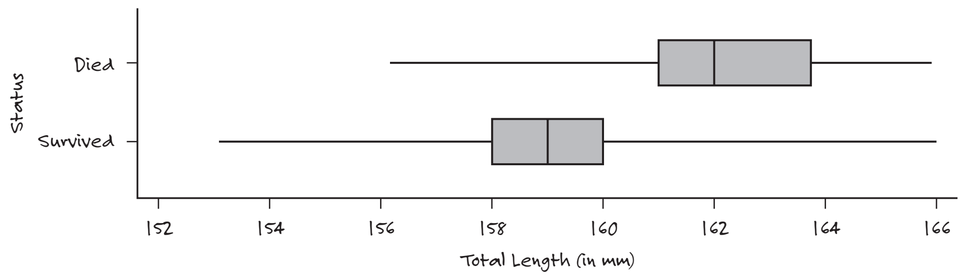
d. Median: 159 mm

Lower quartile: 158 mm — Upper quartile: 160 mm

Minimum: 153 mm — Maximum: 166 mm

e. The sparrows that died tended to be longer than the sparrows that survived. Seventy-five percent of those that died were at least 161 mm long, but 75% of those that survived were shorter than 160 mm. The typical length for the sparrows that died was 162 mm, and the typical length was only 159 mm for those that survived.

f. The following boxplots display the distribution of lengths for the sparrows that survived and died:

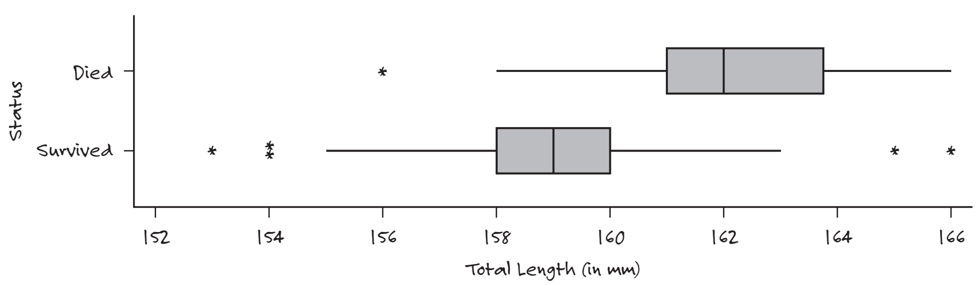


g. Here is the completed table:

Q_L	Q_U	$IQR = Q_U - Q_L$	$1.5 \times IQR$	$Q_L - 1.5 \times IQR$	$Q_U + 1.5 \times IQR$
158	160	2	3	155	163

Outliers would be outside the interval [155, 163], so there are three low outliers (153, 154, and 154) and two high outliers (165 and 166).

h. The following modified boxplots display the lengths for the surviving sparrows and for the sparrows that died:



i. There does appear to be a substantial difference in the lengths between the sparrows that survived and those that did not survive the storm. Those that survived the storm tended to be about 3 inches shorter at each quartile than those that died, although there was more variability and outliers (on both ends) in the group that survived.

j. No, you cannot draw a cause-and-effect conclusion between sparrow length and survival because this is an observational study, not an experiment.

k. Yes, it is clear from this study that shorter sparrows were more likely to survive the storm. This makes no claim about *why* they were more likely to survive the storm—only that the shorter sparrows tended to survive more frequently than did the longer sparrows.

Activity 10-2: Roller Coasters

a. Fifty percent of the steel coasters have a top speed of 60 mph or greater. The boxplot shows that the median for the steel coasters is 60 mph.

b. Twenty-five percent of the wooden coasters have a top speed of 60 mph or greater. The boxplot shows that the upper quartile of the wooden coasters is 60 mph.

c. The lower quartile for both types of coasters is 50 mph, so 75% of both types of coasters have top speeds of 50 mph or greater.

d. You cannot tell which type of coaster has a higher proportion of coasters with a top speed greater than 45 mph because 45 mph is not a quartile for either type of coaster.

e. The steel coasters have more variability. This is obvious from the boxplots because the boxplot for the steel coasters extends from about 25 mph to about 100 mph, whereas the boxplot for the wooden coasters extends from only about 40 mph to about 65 mph. Similarly, the box itself, representing the middle 50% of the distribution, is much longer for the steel coasters than the wooden coasters.

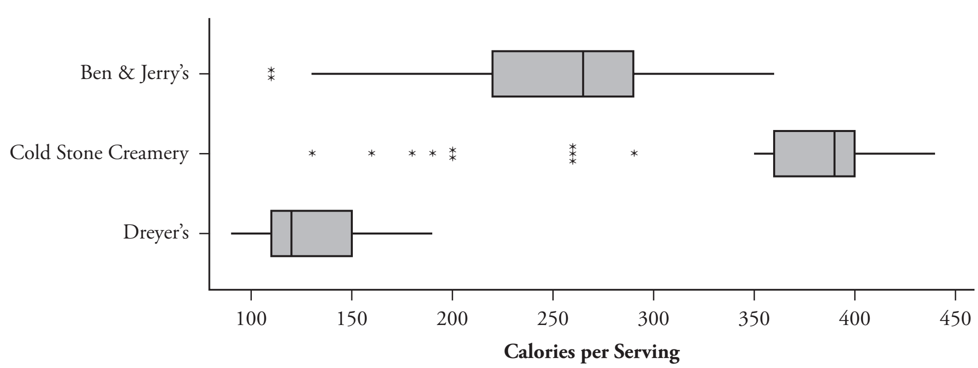
f. The boxplots do not show how many coasters there are of either type; they indicate only where the quartiles (percentages) of the distribution lie.

Activity 10-3: Ice Cream Calories

a. Here is the completed table:

	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
Ben & Jerry's	110	220	265	290	360
Cold Stone	130	360	390	400	440
Dreyer's	90	110	120	150	190

b. The following boxplots display the distribution of calorie amounts for the three brands:



Dreyer's ice cream appears to have significantly fewer calories than the other two brands, as well as less variability among its calorie amounts. All of the Dreyer's ice cream flavors have fewer than 200 calories, whereas only 25% of the Ben & Jerry's flavors have 220 or fewer calories. There is a great deal of variability in the number of calories for the Ben & Jerry flavors as they range from 110 to 360 calories, whereas (excluding outliers) the Cold Stone Creamery flavors range from 360 to 440 calories.

c. The serving sizes may not be the same for all three brands. This would make it difficult to compare the calories as given.

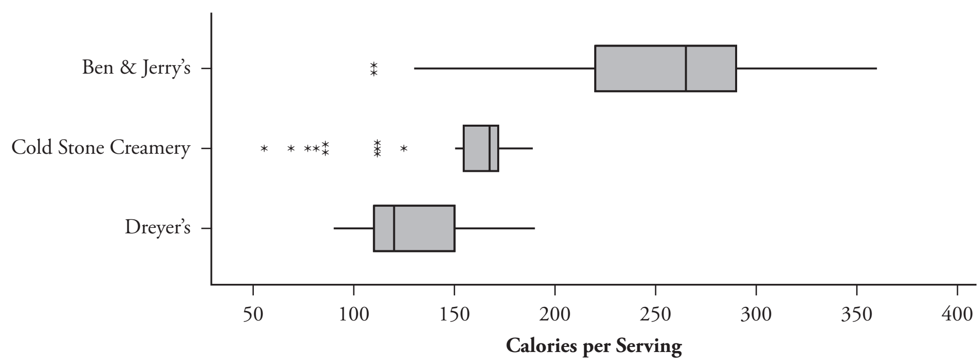
d. You could convert the Cold Stone Creamery serving from 170 grams to the comparable measure of volume in 1/2 cups.

e. Divide each of the Cold Stone Creamery listings by 170 grams, then multiply by 73 grams per 1/2 cup.

f. You calculate *new value* = (*old value*)/170 * 73.

g. Here is the five-number summary for Cold Stone's calorie amounts using the "per half cup" scale: min = 55.82, Q_L = 154.59, median = 167.47, Q_U = 171.76, max = 188.94.

h. The following boxplots display the three distributions of calorie amounts:



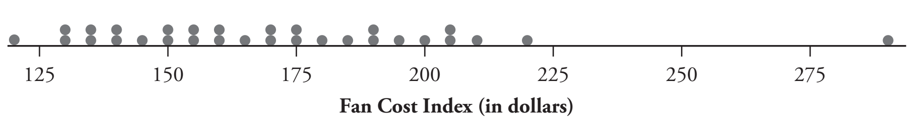
Once you adjust the ice cream servings so that they all have the same serving size, Cold Stone Creamery still has several flavors that are low outliers (meaning that these flavors have an unusually small amount of calories per serving). Excluding the outliers, the Dreyer's flavors are generally the lowest in calorie content, followed by the Cold Stone Creamery flavors, which have a very narrow spread (from only about 155–189 calories per 1/2 cup); at least 75% of the Ben & Jerry flavors have more calories than either of the other two brands.

Activity 10-4: Fan Cost Index

a. Highest FCI team: Boston Value: $287.84

Lowest FCI team: Kansas City Value: $120.34

b. The following dotplot displays the distribution of the FCI values:



The fan cost index ranges from a low of about $120 to a high of about $220, with an outlier of Boston at $287.84. The average FCI is $171, and the median is slightly lower at $166. The standard deviation is $35.05, whereas the IQR is $47.76.

c. Answers will vary.

d. Answers will vary.

e. Answers will vary.

f. Highest team: NY Mets Value: $4.75

Lowest team: Baltimore, Milwaukee, and Kansas Value: $2.00

g. The number of ounces in a "small" soda or beer is not the same in every park.

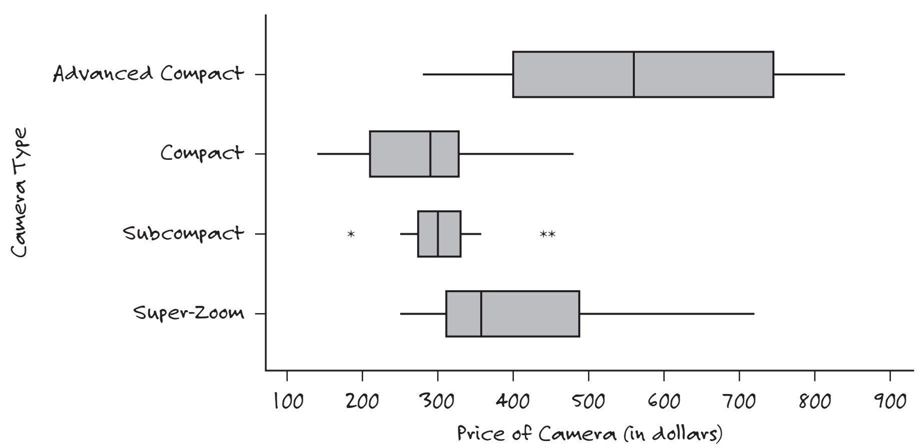
h. Highest team: LA Dodgers Value: $0.354/oz

Lowest team: Pittsburgh Value: $0.1125/oz

Activity 10-5: Digital Cameras

a. The table below reports the five-number summaries, as reported by the software package Minitab. The boxplots of camera prices follow the table.

	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
Advanced Compact	$280	$400	$560	$745	$840
Compact	$140	$210	$290	$327.5	$480
Subcompact	$185	$273.75	$300	$330	$450
Super-Zoom	$250	$311.25	$357.5	$487.5	$720



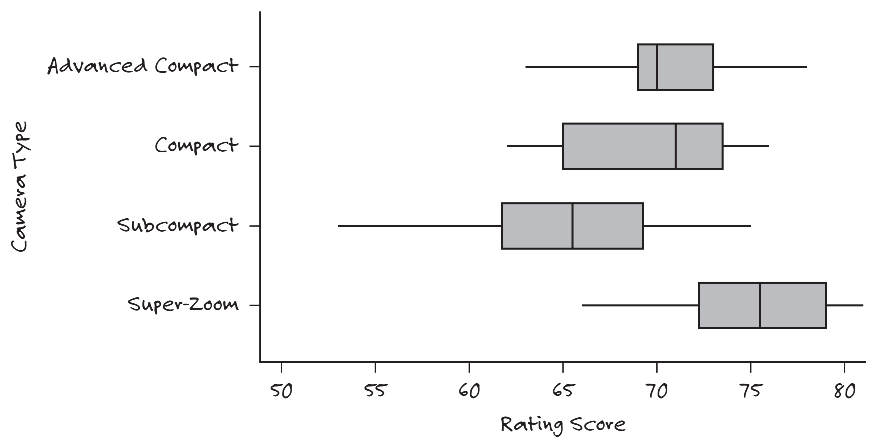
There is considerable overlap in prices among these four groups, but some camera types (e.g. advanced compact) do tend to cost more than other types.

b. Advanced compact cameras tend to cost the most, followed by super-zoom cameras. Compact and subcompact camera prices are similar, but the subcompact cameras cost a bit more on average than the compact cameras.

c. Advanced compact cameras have the most spread in terms of prices, again followed by super-zoom cameras. But compact cameras have more spread in prices than do subcompact cameras, which have the least variability in prices.

d. The boxplots of camera ratings are shown in the following the table. The five-number summaries, as reported by the software package Minitab, are as shown here:

	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
Advanced Compact	63	69	70	73	78
Compact	62	65	71	73.5	76
Subcompact	53	61.75	65.5	69.25	75
Super-Zoom	66	72.25	75.5	79	81



The super-zoom cameras tend to have the highest ratings, and the subcompact ones tend to rate the lowest. The subcompact cameras also have the most variability in ratings.

e. Even though the advanced compact cameras have the highest median price by far, their median rating is surpassed by both super-zoom and compact cameras. In fact, compact cameras have the second-highest median rating despite having the second-lowest median price.

Solutions

In-Class Activities

Activity 11-1: Random Babies

Answers will vary. Here is one representative set of answers:

a. One mother received her own baby.

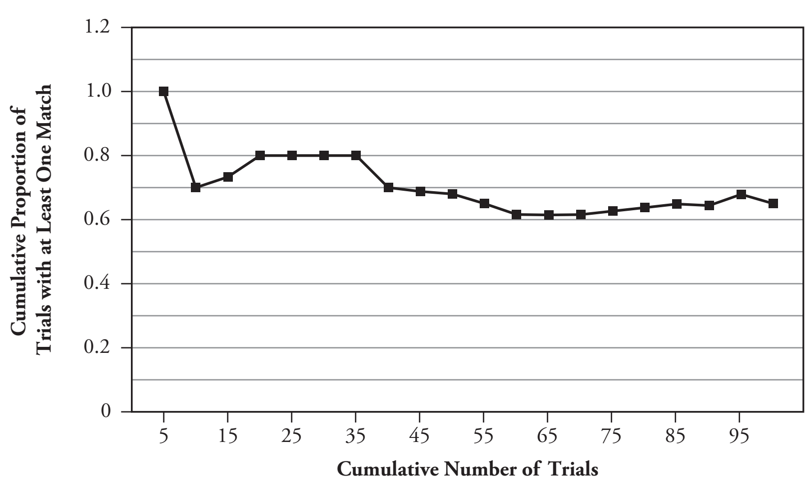
b. Here is a record of the random "dealing":

Number of Repetitions	1	2	3	4	5
Number of Matches	1	1	2	1	2

c. Here is a completed table for the example repetitions:

Student #	Column 1: Number of Trials with at Least One Match	Column 2: Cumulative Number of Trials	Column 3: Cumulative Number of Trials with at Least One Match	Column 4: Cumulative Proportion of Trials with at Least One Match
1	5	5	5	1
2	2	10	7	.7
3	4	15	11	.7333
4	5	20	16	.8
5	4	25	20	.8
6	4	30	24	.8
7	4	35	28	.8
8	0	40	28	.7
9	3	45	31	.6889
10	3	50	34	.6800
11	2	55	36	.6545
12	1	60	37	.6167
13	3	65	40	.6154
14	3	70	43	.6143
15	4	75	47	.6267
16	4	80	51	.6375
17	4	85	55	.6471
18	3	90	58	.6444
19	5	95	63	.6632
20	2	100	65	.6500

d. The following graph displays the cumulative proportion of trials with at least one match vs. the cumulative number of trials.



11

e. The proportion of trials that result in at least one mother getting the correct baby fluctuates more at the beginning of this process.

f. Yes, the relative frequency appears to be settling down and approaching one particular value. Answers will vary as to what that value is, but it should be in the ballpark of .627.

g. Here is a completed table with counts and proportions recorded:

Number of Matches	0	1	2	3	4	Total
Count	34	33	28	0	5	100
Proportion	.34	.33	.28	0	.05	1.00

h. The proportion is .33 + .28 + .05 or .66.

i. Pr(*no matches*) ≈ .34

j. Pr(*at least one match*) ≈ .66

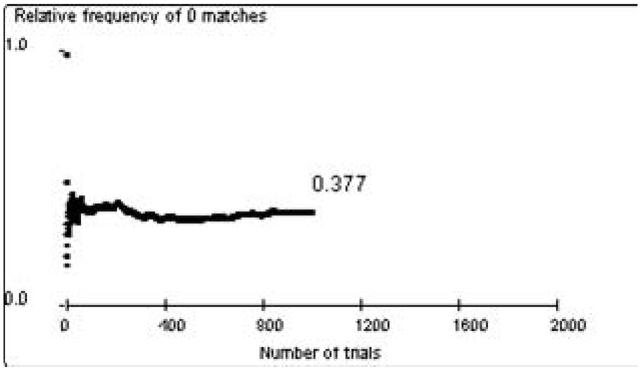
k. Answers will vary. The following are results of one simulation:

Number of Matches	0	1	2	3	4	Total
Count	359	335	267	0	39	1000
Proportion	.359	.335	.267	0	.039	1.00

l. Yes, these simulation results are reasonably consistent with the class results.

m. Pr(*at least one match*) = 1 − .359 = .641.

n. Yes, this graph appears to be fluctuating less as more trials are performed, approaching a limiting value:



o. Getting exactly three matches is impossible because if three mothers and babies matched, then the fourth mother would have to be matched correctly with her baby. There would be no “incorrect” baby left for her to be paired with.

p. It is not impossible to get four matches, but it seems very rare/unlikely because it appears to have happened only about 39 out of 1000 times.

q. No, results of zero, one, or two matches do not seem to be unlikely. Each would appear to occur more than 25% of the time; a zero match and one match even occur more than one-third of the time each.

Activity 11-2: Random Babies

a. There are 24 different arrangements for returning the four babies to the four mothers.

b. Here are the number of mothers who get the correct baby for each arrangement:

1234	4	1243	2	1324	2	1342	1	1423	1	1432	2
2134	2	2143	0	2314	1	2341	0	2413	0	2431	1
3124	1	3142	0	3214	2	3241	1	3412	0	3421	0
4123	0	4132	1	4213	1	4231	2	4312	0	4321	0

c. 4: 1 3: 0 2: 6 1: 8 0: 9

d. 4:1/24 = .04167 3: 0 2:6/24 = .250 1: 8/24 = .333 0: 9/24 = .375

e. The empirical probabilities from our class are reasonably close to these theoretical probabilities. The applet simulation probabilities are even closer.

f. The mean number of matches per repetition is

$$\frac{0.34 + 1.33 + 2.28 + 3.0 + 4.5}{100} = \frac{109}{100} = 1.09$$

(Results will differ.)

g. The expected value for the number of matches from the probability distribution is

$$4\left(\frac{1}{24}\right) + 3\left(\frac{0}{24}\right) + 2\left(\frac{6}{24}\right) + 1\left(\frac{8}{24}\right) + 0\left(\frac{9}{24}\right) = \frac{24}{24} = 1$$

This is very close to the value found in part f.

h. Pr(*number of matches* = 1) = .333. No, you do not expect this value to occur most of the time; you expect it to occur about 1/3 of the time, which is significantly less than half of the time.

Activity 11-3: Family Births

a. Probability does not have a memory. With each birth, there is a 50% chance of a girl, but that probability does not change after the first child is born. If the first child is a girl, the probability that the second child is a girl is still 1/2—not zero. This probability applies "in the long run"—about half of all children born will be girls—but not for every family.

b. There are four equally likely outcomes (not three) for a family with two children. The outcomes are: Boy first and then Boy; Boy first and then Girl; Girl first and then Boy; and Girl first and then Girl. So, the probability that the couple has one boy and one girl (GB or BG) is 2/4 or 50%, and the probability that the couple has two children of the same gender (GG or BB) is also 2/4 or 50%.

c. Answers will vary. The following were obtained using row 31 of the table.

	Family 1		Family 2		Family 3		Family 4	
	Child 1	Child 2	Child 1	Child 2	Child 1	Child 2	Child 1	Child 2
Random Digit	3	4	2	8	2	5	0	7
Gender	B	G	G	G	G	B	G	B
Number of Girls	1		2		1		1	

d. Here are the proportions of each type:

	Two Girls	Two Boys	One of Each
Tally (Count)	4	5	11
Proportion	.20	.25	.55

e. No, it does not appear that the probability of each of these outcomes is 1/3. It appears that the probability of having one child of each gender is about twice that of having two girls or two boys.

f. You can obtain better empirical estimates of these probabilities if you simulate more families.

g. Here is a completed table of simulation results:

	Two Girls	Two Boys	One of Each
Tally (Count)	96	205	99
Proportion	96/400 = .24	205/400 = .5125	99/400 = .2475

h. No, it does not appear that the probability of each of these outcomes is 1/3. It appears that the probability of having one girl and one boy each is about .5, whereas the probability of having two girls or of two boys is about .25.

i. Two girls: $\Pr(GG) = 1/4 = .25$ Two boys: $\Pr(BB) = 1/4 = .25$

One of each: $\Pr(BG \text{ or } GB) = 1/2 = .5$

Yes, these probabilities are reasonably close to the empirical estimates from the class simulations.

Activity 11-4: Jury Selection

a. The probability that he/she will be age 65 or older is .20.

b. Answers will vary by student guess.

c. At least one-third: (135 + 60 + 20)/1000; the approximate probability is .25.

d. Answers will vary by student guess.

e. The approximate probability is 2/1000 or .002.

f. A sample of size 12 is more likely to contain at least one-third senior citizens. This result makes sense because only 20% of the population is senior citizens, so it is easier to select 4 of 12 jurors who are senior citizens than it is to select 25 of 75. There should be more sampling variability (easier to get an unlucky result) with the smaller sample size.

g. Based on the simulations, you see that a sample size of 75 is more likely to contain between 15% and 25% (11.25 and 21) senior citizens than a sample size of 25. This result makes sense because with the larger sample size, the sample proportion should be close to the population proportion more often.

h. The previous questions show that a larger sample size is more likely to produce a sample proportion that is close to the population proportion, and smaller sample sizes are more likely to produce samples with greater variability from sample to sample.

Activity 11-5: Treatment Groups

a. You could assign each subject a number from 1–6. Then, as the die is rolled, you could assign the subjects corresponding to the first three (distinct) numbers rolled to the new treatment group. The process would be similar using a random digit table, where you could simply skip over the digits 7–9 and 0.

b. Answers will vary. (You should find *roughly* 2 randomizations with 0 women in the new group, 18 randomizations with 1 woman, 18 randomizations with 2 women, and 2 randomizations with 3 women.)

c. Answers will vary. Determine the total number of repetitions with 1 woman or 2 women in the new group, and divide that number by 40. (Your empirical estimate should be fairly close to .9.)

d. Performing many more than 40 repetitions of the random assignment should produce a more accurate estimate.

e. Answers will vary. Calculate this empirical estimate of the expected value by multiplying the number of women by the count of repetitions that produced that number, and then divide that value by 40. (Your empirical estimate should be fairly close to 1.5.)

f. Random assignment ensures that all possible ways to assign these subjects to the treatment groups are equally likely to occur.

g. The exact probabilities are 1/20 or .05 for 0 women in the new group, 9/20 or .45 for 1 woman in the new group, 9/20 or .45 for 2 women in the new group, and 1/20 or .05 for 3 women in the new group.

h. The probability of a 1/2 or 2/1 gender breakdown is 18/20 or .9.

i. The probability that the two genders are completely separated into two groups (all men in one group and all women in the other) is 2/20 or .1. This result is small enough to be perhaps mildly surprising, but it is not small enough to be very surprising. Although randomization *should* balance out the gender breakdown between the two groups *in the long run,* with such a small study group, you could, simply by chance, still end up with all males in one group and all females in the other group.

j. The expected value is $0(1/20) + 1(9/20) + 2(9/20) + 3(1/20)$ or 30/20 or 1.5 women.

k. This expected value makes sense because with three women randomly assigned among two groups, you would expect half of them to be assigned to each group in the long run, and half of 3 is 1.5.

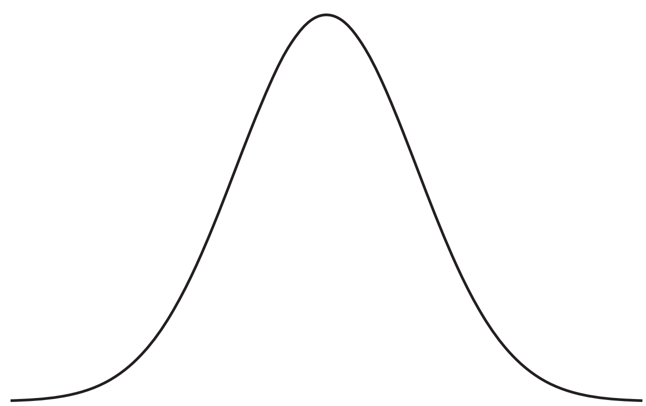
Solutions

In-Class Activities

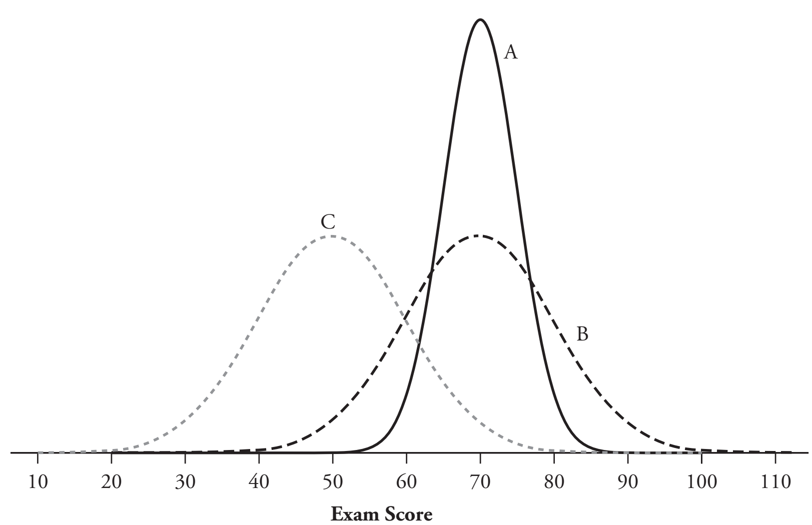
Activity 12-1: Body Temperatures and Jury Selection

a. Both distributions are roughly symmetric and mound-shaped.

b. The following sketch approximates the general shape in the two histograms:



c. The dashed curve (C) has a mean of 50 and a standard deviation of 10. The dashed curve (B) has a mean of 70 and a standard deviation of 10. The solid curve (A) has a mean of 70 and a standard deviation of 5. Here is the labeled drawing:



d. Yes, these predictions are quite close to each other and to what the empirical rule would predict.

e. The z-score for the value 97.5 in the distribution of body temperatures is $z = (97.5 - 98.249)/.733$ or -1.02.

f. The z-score for the value 11.5 in the distribution of number of senior citizens in the jury pool is $z = (11.5 - 14.921)/3.336$ or -1.025.

g. The two z-scores are almost identical.

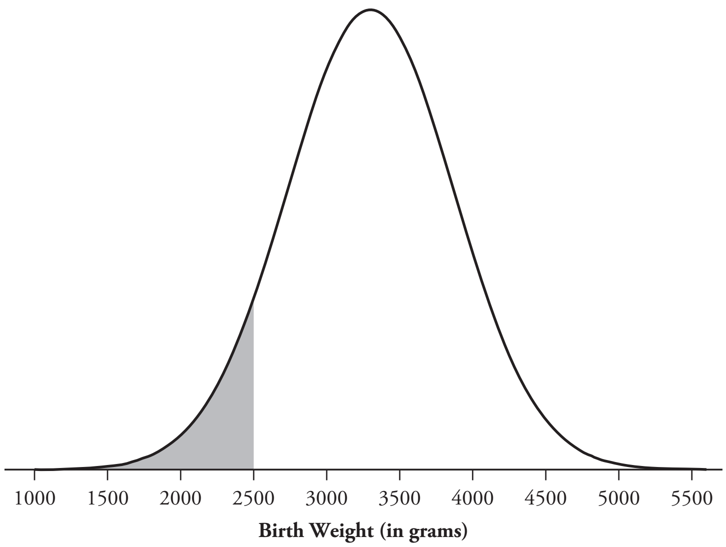
h. Both z-scores indicate that the observations are just over one standard deviation below their respective means.

i. The value is .1515.

j. Yes, this value is reasonably close to .146 and .153.

Activity 12-2: Birth Weights

a. The following graph shows the shaded region whose area corresponds to the probability that a baby will have a low birth weight:



b. Answers will vary by student guess.

c. The z-score for a birth weight of 2500 grams is $z = (2500 - 3300)/570$ or -1.40.

d. $\Pr(Z < -1.40) = .0808$

e. The applet gives the probability of .0802.

f. Using the applet, $\Pr(\textit{birth weight} > 4536) = \Pr(Z > 2.17) = .0151$.

g. One way to use Table II to answer part f is to look up the z-score and subtract the given area from 1.000. A second way is to look up the opposite z-score (-2.17) because the curve is symmetric.

h. Using the applet, $\Pr(3000 < \textit{birth weight} < 4000) = \Pr(-.53 < Z < 1.23) = .5910$. Using Table II, $.8907 - .2981 = .5926$.

i. For babies with low birth weight, you calculate (331772/4112052) or .081, which is very close to the predicted .08 from your calculations in part d. For babies between 3000 and 4000 kg, you calculate (2697819/4112052) or .656, which is not as close to (but not unreasonably far from) the prediction in part h.

j. The lightest 2.5% corresponds to a z-score of -1.96 (using Table II or technology). You calculate $z = -1.96 = (x - 3300)/570$; weight $= x = 2182.8$ kg.

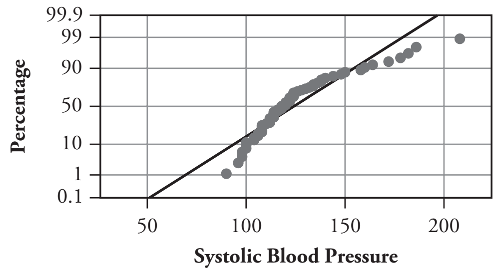
k. The heaviest 10% (or bottom 90%) corresponds to a z-score of 1.28 (using Table II or technology). You calculate $z = 1.28 = (x - 3300)/570$; weight $= x = 4029.6$ kg.

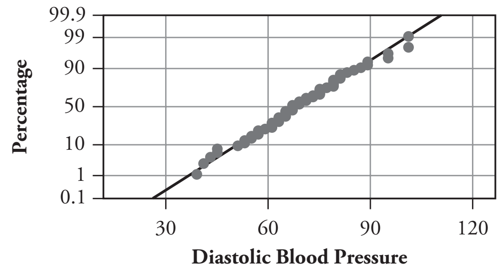
Activity 12-3: Blood Pressures and Pulse Rate Measurements

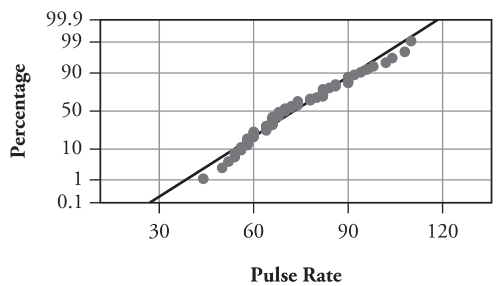
a. The diastolic blood pressure is the most symmetric and mound-shaped of the three dotplots, so it is most likely to have come from a normal population.

b. The systolic blood pressure is least likely to have come from a normal distribution because its distribution is not symmetric or mound-shaped.

c. Here are normal probability plots for systolic blood pressure, diastolic blood pressure, and pulse rate:



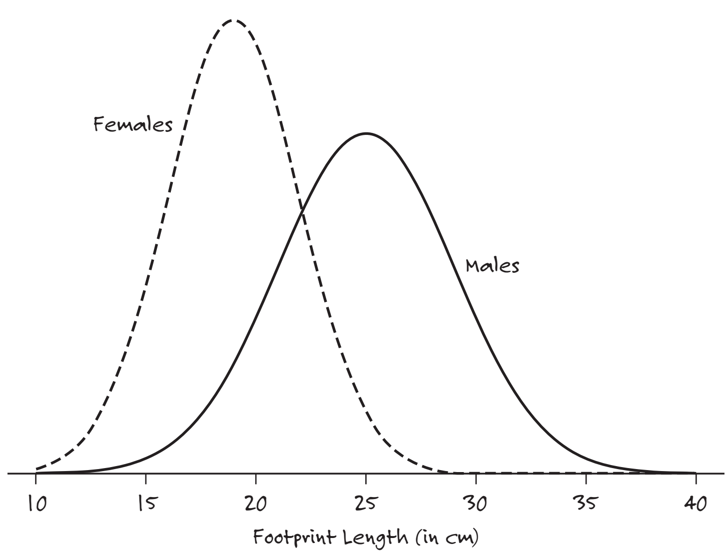




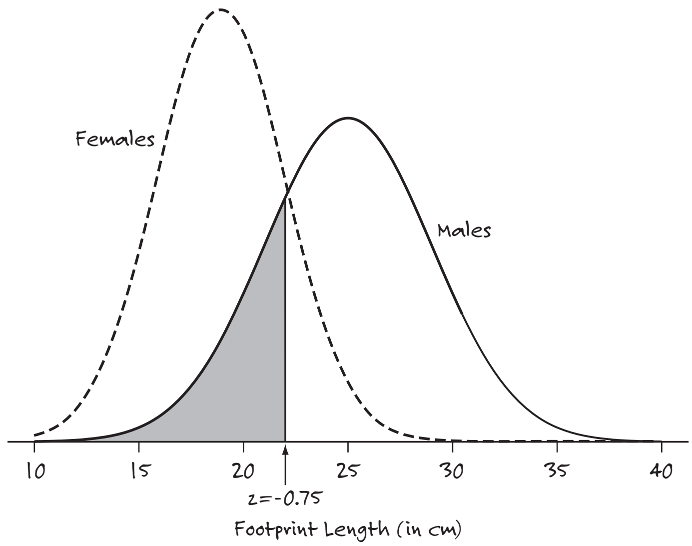
These probability plots confirm that the systolic pressures were unlikely to have come from a normal distribution, and the diastolic pressures could quite possibly have come from a normal distribution.

Activity 12-4: Criminal Footprints

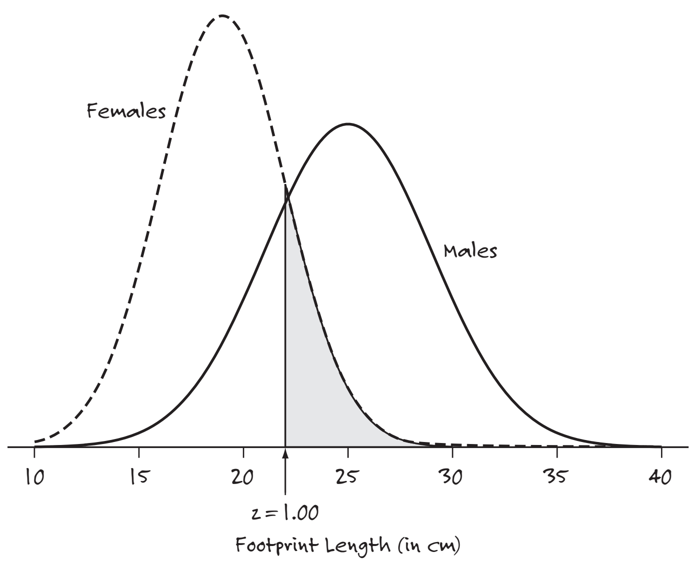
a.



b. Because you are dealing with a male footprint, the z-score is $(22 - 25)/4$ or -0.75. Using Table II (or technology), the probability of the footprint being less than 22 centimeters is .2266. So, roughly 22.66% of men have a footprint smaller than 22 centimeters and would be misclassified as female.

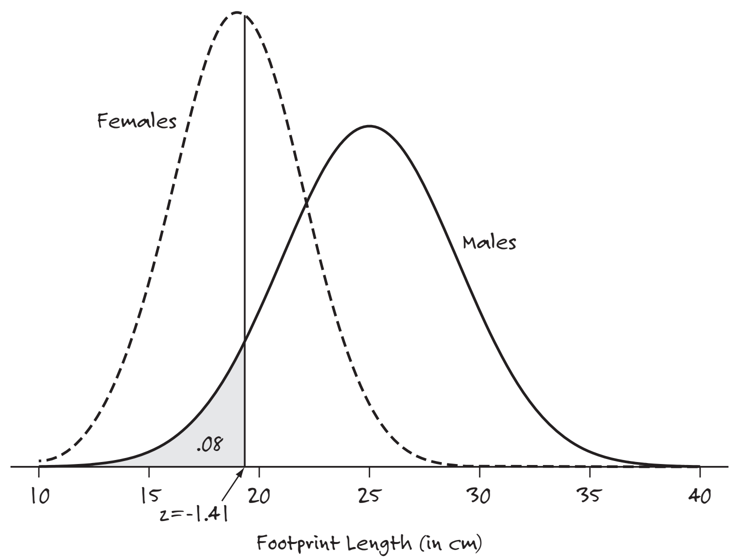


c. Now you are dealing with a female footprint, so the z-score is $(22 - 19)/3$ or 1.00. Using Table II (or technology), the probability of the footprint being longer than 22 centimeters is $1 - .8413$ or .1587. This indicates that about 16% of females would be mistakenly identified as male.

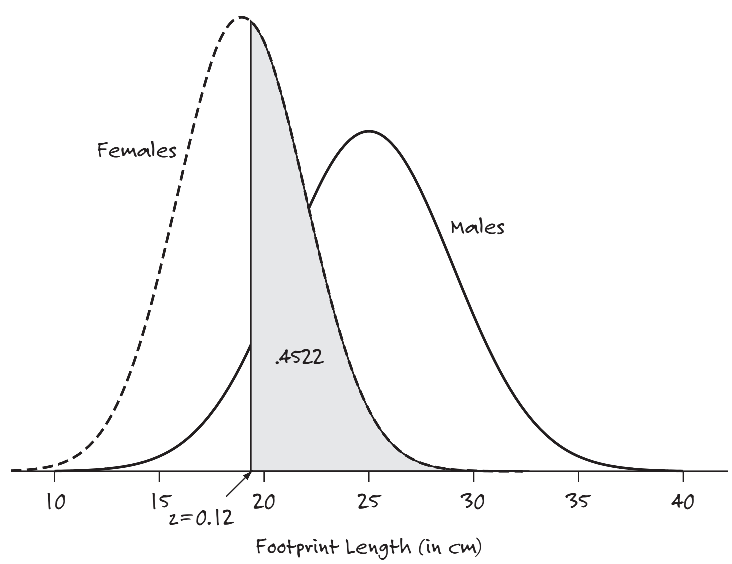


d. Using Table II or technology, the z-score to produce a probability of .08 is -1.41. To find the corresponding male foot length, you need to solve $-1.41 = (x - 25)/4$, which gives you $x = 25 - 1.41(4) = 25 - 5.64 = 19.36$ centimeters. You can also think of this as subtracting 1.41 standard deviations of 4 from the mean of 25. Notice that this new cutoff value (19.36 centimeters) is much smaller than before

(22 centimeters) in order to reduce the probability of classifying a male footprint as having come from a woman.



e. Using this new cutoff value of 19.36 centimeters, the *z*-score for a female footprint is (19.36 – 19)/3 or 0.12. The probability of a female footprint being longer than 19.36 centimeters is 1 – .5478 or .4522 (using Table II or technology).



f. The probability of misclassifying a female footprint as a male footprint is much greater than before (.4522 as opposed to .1587). In order to reduce the probability of one type of error (misclassifying a man's footprint as having come from a woman) from .2266 to .08, the probability of making the other kind of error increases substantially. This exercise reveals that there is a trade-off between the probabilities of making the two kinds of errors that can occur: you can reduce one error probability but only by increasing the other error probability.

13

Solutions

In-Class Activities

Activity 13-1: Candy Colors

Answers will vary. Here is one representative set of answers.

a. The count and proportion of each color in this sample are recorded in the table:

	Orange	Yellow	Brown
Count	13	7	5
Proportion (Count/25)	.52	.28	.2

b. This is a statistic. The symbol used to denote the proportion is $\hat{p}$.

c. This is a parameter. The symbol used to denote the proportion is π.

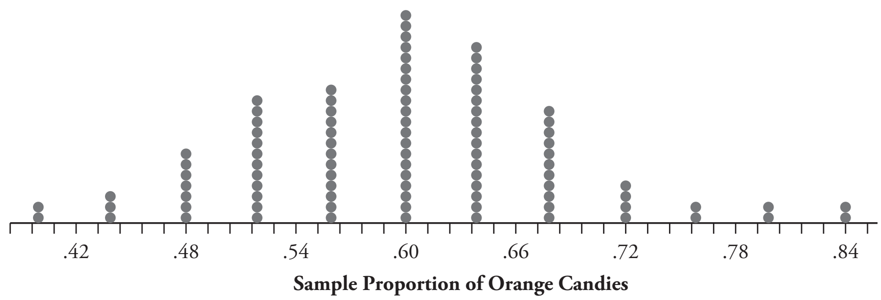
d. No, you do not *know* the proportion of orange candies manufactured by Hershey.

e. Yes, you know the proportion of orange candies among the 25 candies that you individually selected.

f. It is very unlikely that every student in the class obtained the same proportion of orange candies in his or her sample.

Answers will vary. Here is one representative set of answers:

g. Here is a dotplot of the sample proportions of orange candies:



h. The observational units in this graph are the samples of 25 candies. The variable being measured from unit to unit is the *proportion of the sample that is colored orange.*

i. This dotplot of the sample proportions is symmetric and mound-shaped (roughly normal), with a center of .5, and with most of the sample proportions falling between .4 and .6 (min .36, max .68). The standard deviation is .08.

j. Based on the sample results from *this* class, a reasonable guess for π would be .5.

k. Most estimates would be reasonably close to π, but a very few estimates would be way off. You can see this from the dotplot. Most of the class results are the same (near .5), but a few of the class results are quite extreme (far from .5).

l. If each student had taken samples of size 10 instead of size 25, you would expect more variability (greater horizontal spread) in the dotplot.

m. If each student had taken samples of size 75 instead of size 25, you would expect less variability (less horizontal spread) in the dotplot.

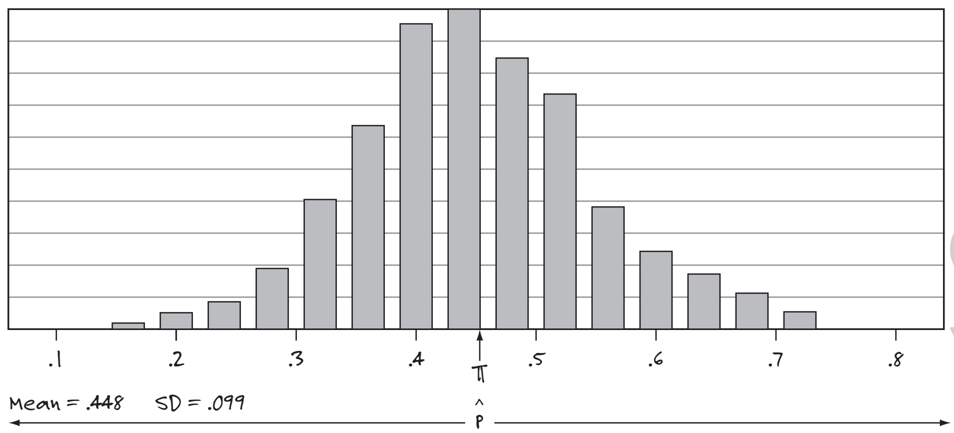
Activity 13-2: Candy Colors

Answers will vary. The following results are from one particular running of the applet.

a. $\hat{p} = .44$

b. $\hat{p} = .52, .48, .52, .56, .40$. No, you did not get the same sample proportion each time.

c. Here is a sketch of the results displayed in the dotplot:



d. Yes, the distribution appears roughly normal, centered at about .45, with a standard deviation of about .1.

e. A normal curve seems to model the simulated sample proportions very well.

f. Mean of $\hat{p}$ values: .449 Standard deviation of $\hat{p}$ values: .100

g. Roughly speaking, more sample proportions are close to .45 than are far away from it.

h. Here is the completed table:

	Number of 500 Sample Proportions	Percentage of 500 Sample Proportions
Within ± .10 of .45	354	71.5%
Within ± .20 of .45	473	95.6%
Within ± .30 of .45	491	99.2%

i. About 95% would capture the actual population proportion.

j. No, you would not have any definite way of knowing whether your sample proportion was within .20 of the population proportion. However, you could be reasonably confident that your sample proportion was within .20 of the population proportion because about 95% of the sample proportions would be within .2 of π.

k. Mean of $\hat{p}$ values: .446 Standard deviation of $\hat{p}$ values: .057

l. The shape is still roughly normal and the center is still about .45. The spread, however, has decreased significantly (from .1 to about .057).

m. The applet reports that 460/500 = 92% of the sample proportions are within .1 of .45.

n. This percentage is much greater (92% versus 71.5%) than it was when the sample size was $n = 25$.

o. The sample proportion is more likely to be closer to the population proportion with a *larger* sample size.

p. You calculate $.057 \times 2 = .114$, and $.45 \pm .114 = [.336, .564]$.

q. The applet reports that 336/500 or 95% of the sample proportions are within .114 of .45.

r. About 95% of the students' intervals would contain the actual population proportion of .45.

s. Theoretical mean of $\hat{p}$ values: .45

Theoretical standard deviation of $\hat{p}$ values: $\sqrt{\frac{(.45)(.55)}{25}} = .0995 \approx .1$

t. Theoretical mean of $\hat{p}$ values: .45

Theoretical standard deviation of $\hat{p}$ values: $\sqrt{\frac{(.45)(.55)}{75}} = .057$

u. No, the normal model does not summarize this distribution well. This is not a contradiction to the Central Limit Theorem because $n\pi = 25(.1) = 2.5 \ngeq 10$.

Activity 13-3: Kissing Couples

a. This is a parameter; $\pi = .5$.

b. You calculate $n\pi = 124(.5) = 62 > 10$ and $n(1 - \pi) = 124(.5) = 62 > 10$, so the CLT does apply.

Shape: approximately normal

Center: $\pi = .5$

Spread: $\sqrt{\frac{(.5)(.5)}{124}} = .0449$

c. Yes, the histogram does appear to be consistent with what the CLT predicts. It is bell-shaped, centered at about .5, and extends from about $.5 - 3(.0449)$ or .3653 to about $.5 + 3(.0449)$ or .6347.

d. $\hat{p} = 80/124 = .645$

e. Yes, it would be very surprising to observe such a sample proportion (.645) if 1/2 of all kissing couples lean their heads to the right; this sample proportion never occurred in 1000 simulations.

f. The z-score for the observed sample proportion is $z = (.645 - .5)/.0449 = 3.23$.

g. Yes, this is a very surprising z-score; $\Pr(Z > 2.33) = .0099$. If 1/2 of all kissing couples lean their heads to the right, you would see a sample result as or more extreme than .645 in less than 1% of random samples.

Activity 13-4: Kissing Couples

a. Recall that the observed sample proportion of kissing couples who lean their heads to the right is $\hat{p} = 80/124 = .645$. This value is not at all uncommon in the first histogram.

b. The CLT says that the sample proportion in this case would vary approximately normally with mean equal to .667 and standard deviation equal to

$$\sqrt{\frac{(.667)(.333)}{124}} \approx .042$$

The z-score for the observed sample proportion of .645 is therefore

$$z = \frac{.645 - .667}{.042} \approx -0.52$$

so the observed sample proportion .645 lies only about half of a standard deviation from the population proportion when $\pi = .667$.

c. The observed sample proportion is barely one-half of a standard deviation away from what you would expect if the population proportion were equal to 2/3, not a surprising result at all. Therefore, the sample data provide no reason to doubt that the population proportion of kissing couples who lean their heads to the right equals 2/3.

d. The value .645 is pretty far along the lower tail of the second histogram. This indicates that the observed sample proportion would rarely occur if the population proportion were equal to 3/4. Further evidence of this result is provided by the rather large negative z-score:

$$z = \frac{.645 - .750}{\sqrt{\frac{(.750)(.250)}{124}}} = \frac{.645 - .750}{.39} \approx -2.69$$

Therefore, the sample data provide fairly strong evidence that the population proportion of kissing couples who lean their heads to the right is not 3/4 (because it would be rather surprising to find a sample proportion so far from this population proportion by chance alone).

e. A reasonable estimate of the population proportion π is the sample proportion .645. An estimate of the standard deviation of $\hat{p}$ would then be

$$\sqrt{\frac{(.645)(.355)}{124}} \approx .043$$

Doubling this standard deviation gives .086. The interval is, therefore, .645 ± .086, which runs from .559 to .731. Notice that 1/2 and 3/4 are not in this interval, but 2/3 is. The interval is consistent with the earlier analysis of the plausibility of the values 1/2, 2/3, and 3/4 for the population proportion of kissing couples who lean their heads to the right.

Solutions

In-Class Activities

Activity 14-1: Coin Ages

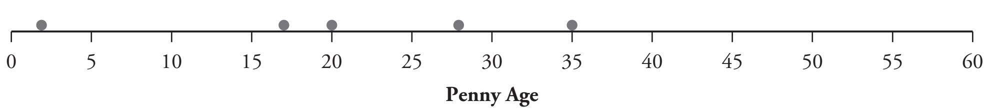
a. Observational units: pennies Variable: age

Quantitative or categorical? quantitative

b. These values are parameters, represented by symbols μ (mean) and σ (standard deviation).

c. No, the distribution of ages does not follow a normal distribution; it is strongly skewed to the right.

d. Answers will vary. One example (using row 48 of the Random Digits Table, three digits at a time) is coins numbered 788, 929, 977, 718, 049, which have ages 20, 28, 35, 17, and 2, respectively. Note that we would delete any three-digit numbers that correspond to a coin that has already been selected (sampling without replacement). You are also free to read the table vertically, backward, selecting a new row, etc. In order for someone to evaluate your procedure though, include enough details so your method is clear.



e. The sample mean of your five penny ages is 20.4 years.

f. The following table gives the sample mean age for each sample:

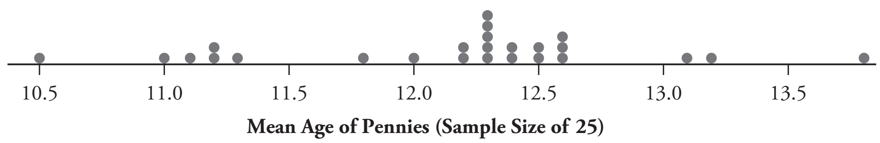
	1	2	3	4	5
Sample Mean ($\bar{x}$)	20.4	18.8	1.2	18	10.6

g. No, you do not get the same value for the sample mean all five times. This reveals the phenomenon of sampling variability. This variable is *quantitative,* whereas the variable in the Candy Colors activity (Activities 13-1 and 13-2) was *binary categorical.*

h. Mean of $\bar{x}$ values: 13.8 years SD of $\bar{x}$ values: 7.99 years

i. The mean of $\bar{x}$ values is reasonably close to the population mean ($\mu = 12.264$ years). The standard deviation is less than the population standard deviation of 9.613 years.

j. The following dotplot displays the sample means:



The observational units are the samples of 25 coins.

k. Yes, this distribution should be roughly centered at 12.264 years.

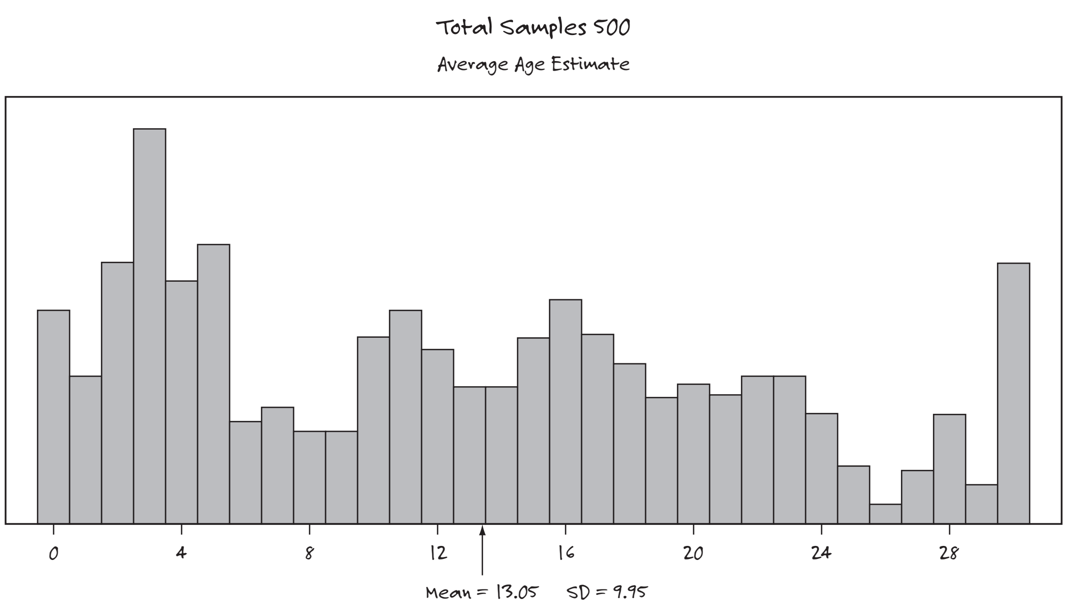
l. These values appear to be less spread out than either the population or the samples of five pennies.

m. Yes, this distribution does appear to be closer to a normal shape than the distribution of ages in the original population, which was so strongly skewed to the right.

Activity 14-2: Coin Ages

Answers will vary. The following are examples from one running of the applet.

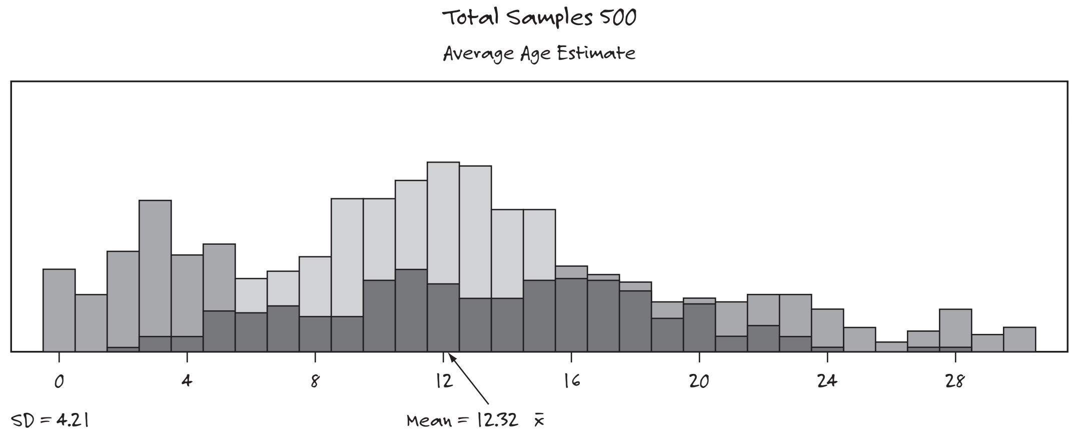
a. A rough sketch of the results displayed in the dotplot follow:



Yes, this dotplot does resemble the distribution of ages in the population; it is definitely skewed to the right. *Note:* Any observations to the right of the applet window are piled in the last bin.

b. The mean is 13.05 years and the standard deviation is 9.95 years. The shape is somewhat skewed to the right.

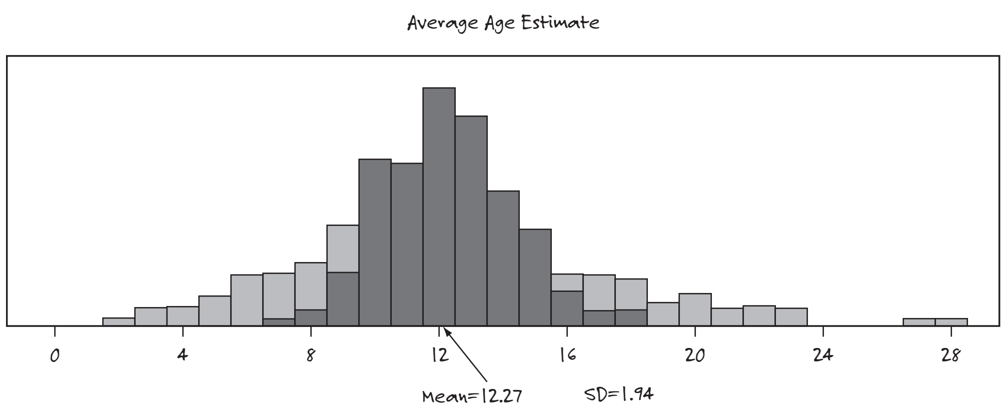
c. A sketch of the results displayed in the dotplot follow (500 sample means with $n = 5$):



Now the shape of the distribution is much more symmetric and less skewed. The center is 12.32 years and the standard deviation is 4.2 years.

d. The mean of the sampling distribution is very close to the population mean, but the standard deviation of the sampling distribution is much smaller.

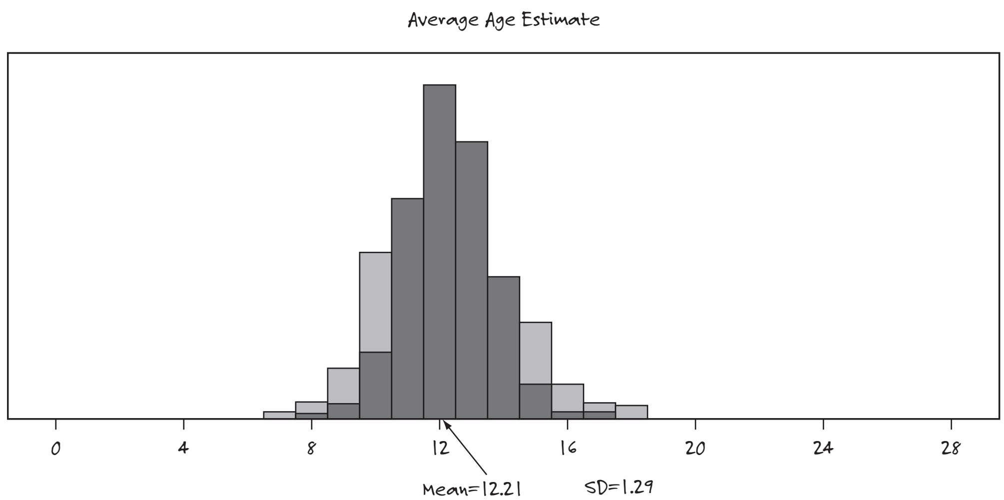
e. A sketch of the results displayed in the dotplot follow (500 sample means with $n = 25$):



The shape is even more symmetric, and the spread is much smaller. (The center has remained the same.)

f. The mean of the sampling distribution is very close to the population mean, but the standard deviation of the sampling distribution is much smaller.

g. A sketch of the results displayed in the dotplot follow (500 sample means with $n = 50$):



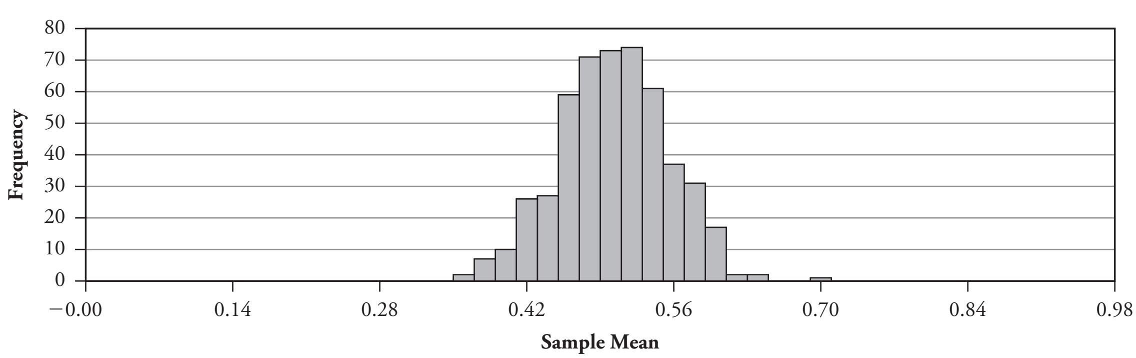
The shape is even more symmetric, and the spread is even smaller, but the center has remained at about 12.2 years.

h. Once again the mean of the sampling distribution is very close to the population mean, but the standard deviation of the sampling distribution is much smaller.

	Population Mean (μ = 12.264 years)	Population SD (σ = 9.613 years)	Population Shape: Skewed to the Right
Sample Size	Mean of Sample Means	SD of Sample Means	Shape of Sample Means
1	13.05	9.95	Skewed to right
5	12.32	4.21	Symmetric
25	12.27	1.94	Normal
50	12.21	1.29	Normal

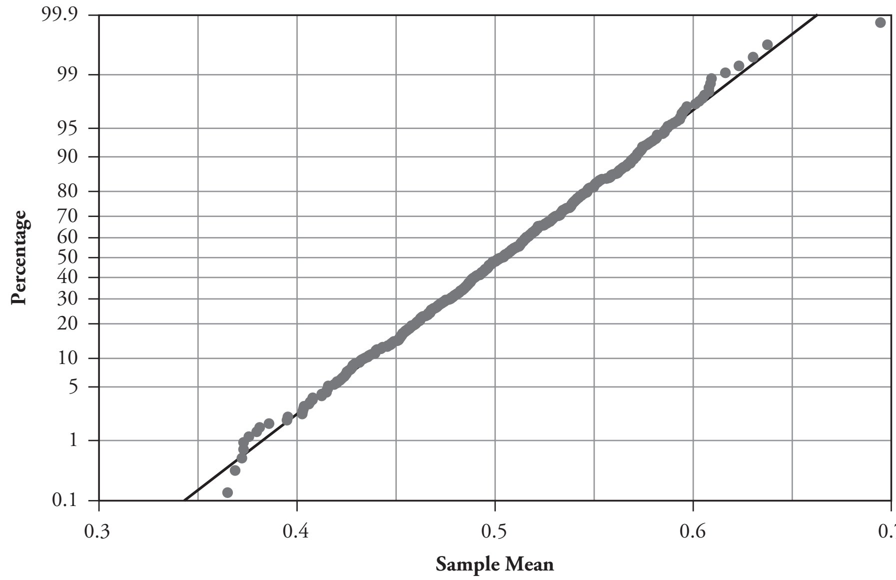
i. No, this population follows a uniform distribution (from $0.01 to $0.99).

j. The following histogram displays the results of 500 sample means with $n = 30$:



The histogram of sample means is fairly bell-shaped (approximately normal). The mean is $0.50 and the standard deviation is $0.0517.

k. Yes, this probability plot indicates the distribution is definitely well modeled by a normal curve:



l. You calculate $\sigma/\sqrt{n} = 9.613/\sqrt{50} = 1.359$ years. This value is reasonably close to the standard deviation of the 500 simulated sample means (1.29).

m. You calculate $\sigma/\sqrt{n} = 28.866/\sqrt{30} = 4.08$¢. This is reasonably close to the simulated standard deviation of 5¢.

Activity 14-3: Christmas Shopping

a. The value \$857 is a statistic, represented with the symbol $\bar{x}$. This value is a statistic because it comes from a sample, not from the entire population of American adults.

b. **i.** Population of interest: all American adults

ii. Sample selected: 922 American adults who expected to buy Christmas gifts in 1999

iii. Parameter of interest: average amount all American adults expected to spent on Christmas gifts in 1999

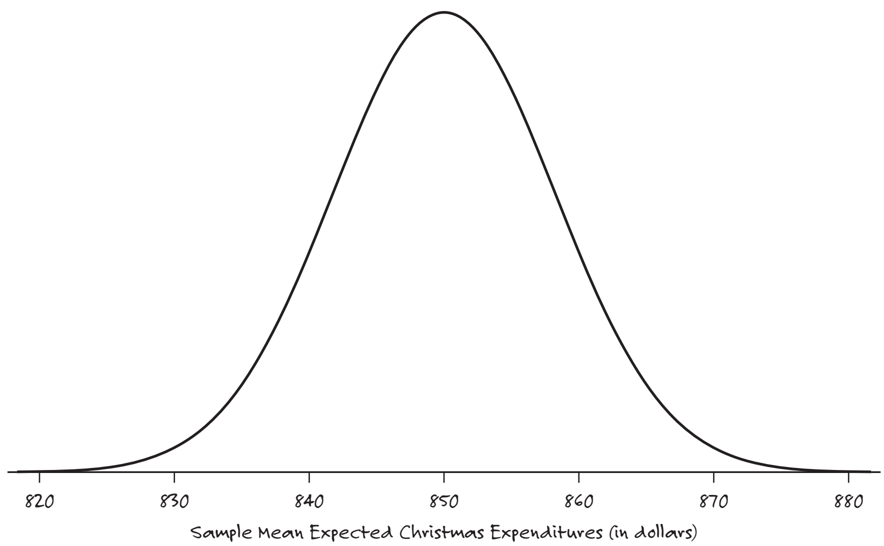
iv. Statistic calculated: \$857, or the average amount this sample of 922 adults expected to spend on Christmas gifts

c. No, μ does not necessarily equal \$857. It is possible that $\bar{x} = \$857$, even if $\mu =$ \$850, or \$800, or \$1000. But such a sample mean is not very likely for $\mu = \$800$, and it is extremely unlikely for $\mu = \$1000$.

d. The CLT says that that sampling distribution of $\bar{x}$ would be approximately normal, with mean \$850 and standard deviation $\$250/\sqrt{922} = \$8.23$. Because the sample size is large, your answer does not depend on the shape of the distribution of expected expenditures in the population.

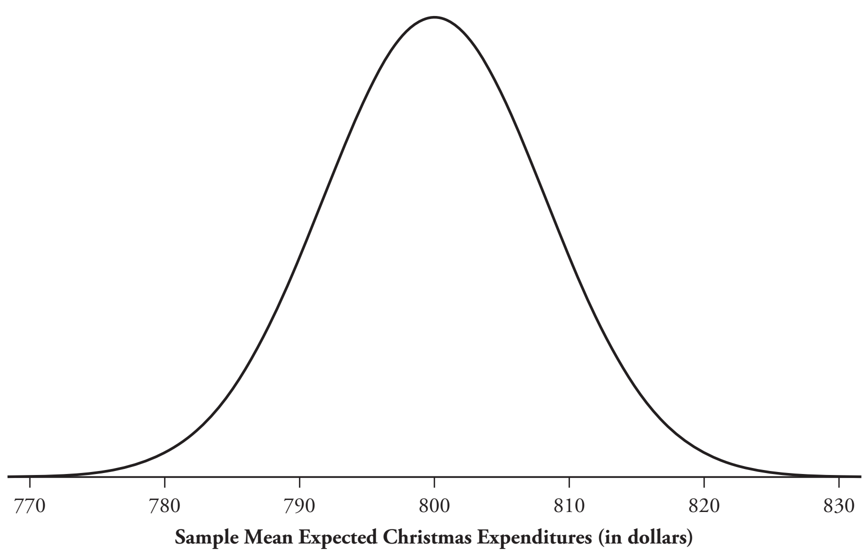
14

e. Here is a sketch for this distribution:



f. No, a sample mean of \$857 would not be at all surprising as it is located close to the center of the distribution and there is a great deal of area to the right of this value in the plot shown in part e.

g. The CLT says that that sampling distribution of $\bar{x}$ would be approximately normal with mean \$800 and standard deviation $\$250/\sqrt{922} = \$8.23$. Because the sample size is large, your answer does not depend on the shape of the distribution of expected expenditures in the population. Here is a sketch for this distribution:



This time a sample mean of \$857 would be very surprising because this value does not appear on the normal curve modeling the behavior of the sample means.

h. The standard deviation should be $\$250/\sqrt{922} = \$8.23$.

i. You calculate 2 × \$8.23 = \$16.46, so \$857 − \$16.46 = \$840.54 and \$857 + \$16.46 = \$873.46. The interval of values is [\$840.54, \$873.46]. *Note:* These could be considered the plausible values of μ based on observing a sample mean of $\bar{x} = 857$.

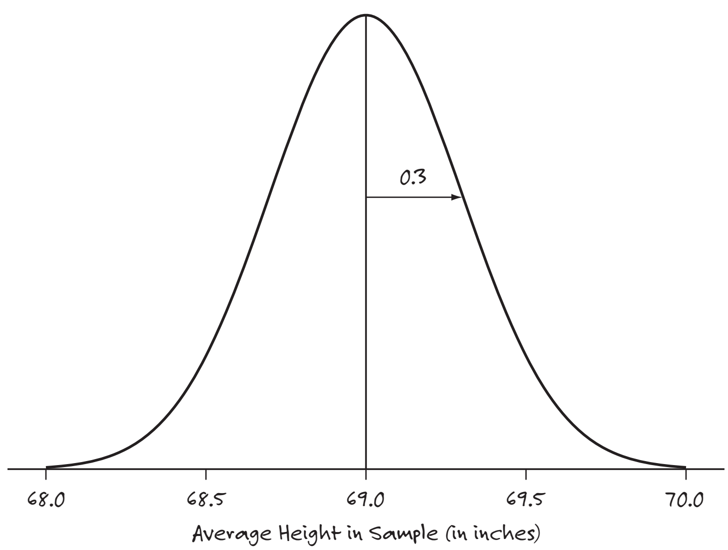
Activity 14-4: Looking Up to CEOs

a. The mean height among all adult American males, 69 inches, is a parameter because it describes the entire population. It is denoted by the symbol μ.

b. The sample size is 100 and is denoted by the symbol n.

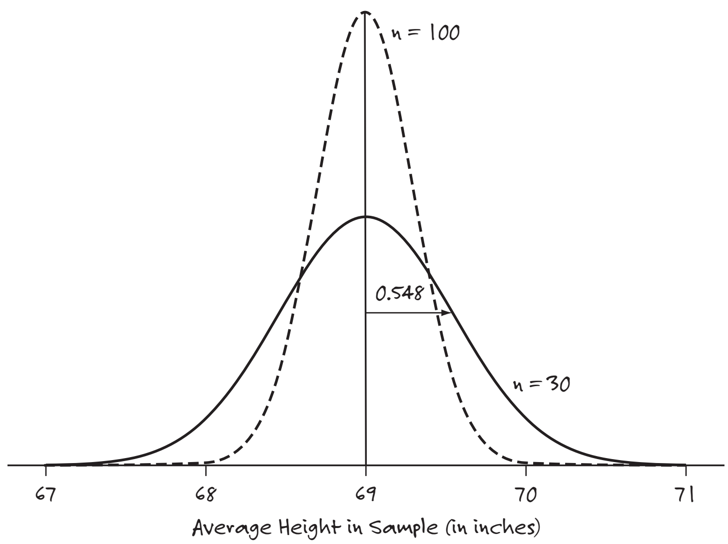
c. You need to know the population standard deviation of the heights of adult American males, denoted by σ. Because the sample size is fairly large (100 is larger than 30), you do not need to know whether the population distribution of heights is normal because the Central Limit Theorem tells you the shape of the sampling distribution of the sample mean will be approximately normal with this sample size, regardless of the shape of the original population distribution.

d. The CLT establishes that the sampling distribution of the sample mean height is approximately normal, with mean 69 inches and standard deviation $3/\sqrt{100} =$ 0.3 inches.



e. Doubling the standard deviation of the sample mean gives you 0.6 inches, so the sample mean height among the CEOs would have to be at least 69.6 inches to persuade the psychologist that, on average, CEOs are indeed taller than the average adult male.

f. If the sample size were 30, the normal approximation should still be valid, and the standard deviation of the sampling distribution would increase to $3/\sqrt{30} =$ 0.548 inches. Doubling this standard deviation gives you 1.096 inches, so the sample mean height would have to be at least 70.096 inches to be persuasive. The smaller sample size produces more sampling variability, hence a larger standard deviation. As a result, the cut-off value needed for a persuasive sample mean height is larger.


n = 100
0.548
n = 30
67
68
69
70
71
Average Height in Sample (in inches)

Solutions

In-Class Activities

Activity 15-1: Smoking Rates

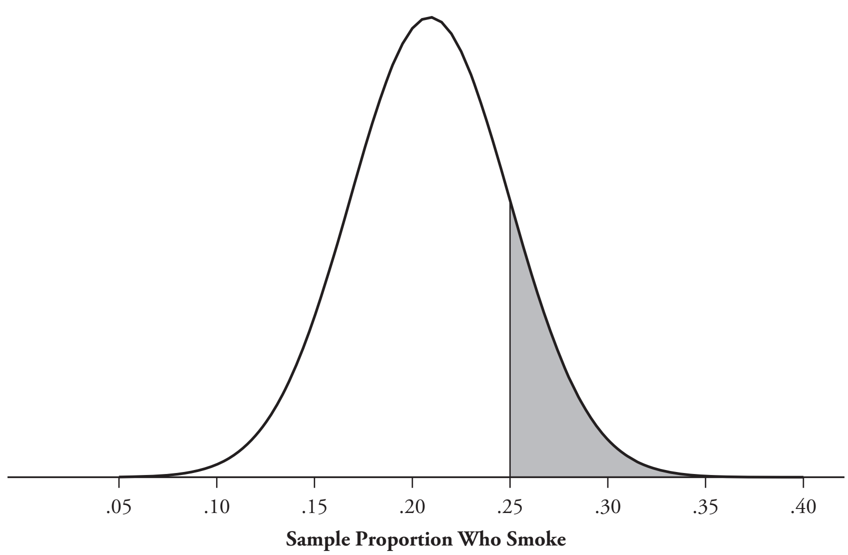
a. The symbol π represents the proportion .209.

b. No; in general, the sample result will not equal .209 exactly because of sampling variability.

c. The CLT predicts the sampling distribution of $\hat{p}$ will be approximately normal, centered at .209 with a standard deviation equal to

$$\sqrt{\frac{(.209)(.791)}{100}} = .04066.$$

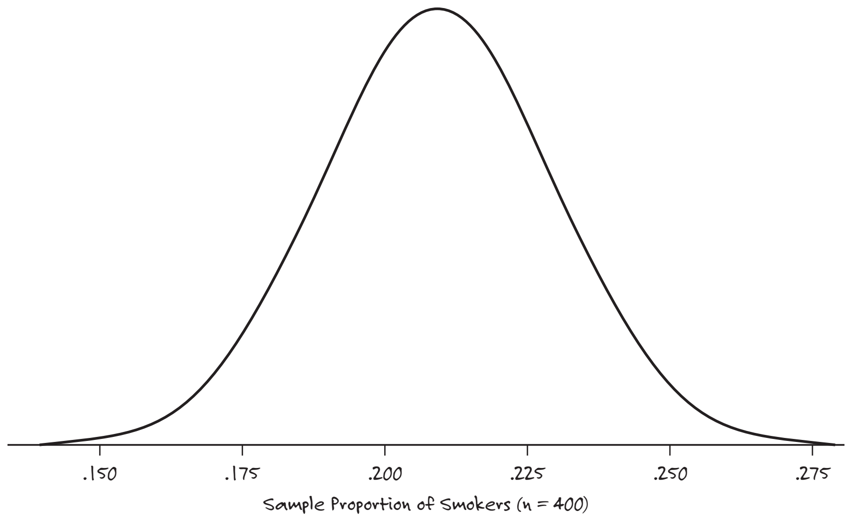
d. The following graph shows the shaded area that corresponds to a sample proportion exceeding .25:



e. Using the CLT result, $z = (.25 - .209)/.04066 = 1.01$.

f. $\Pr(Z > 1.01) = .1562$ (Table II) or .1566 (applet)

g. When the sample size increases to 400, the standard deviation of the sampling distribution will decrease to .0203. This means there will be fewer sample proportions as far from the center of .209, and it will be less likely that you will have a sample proportion greater than .25. The following sketch displays these results:



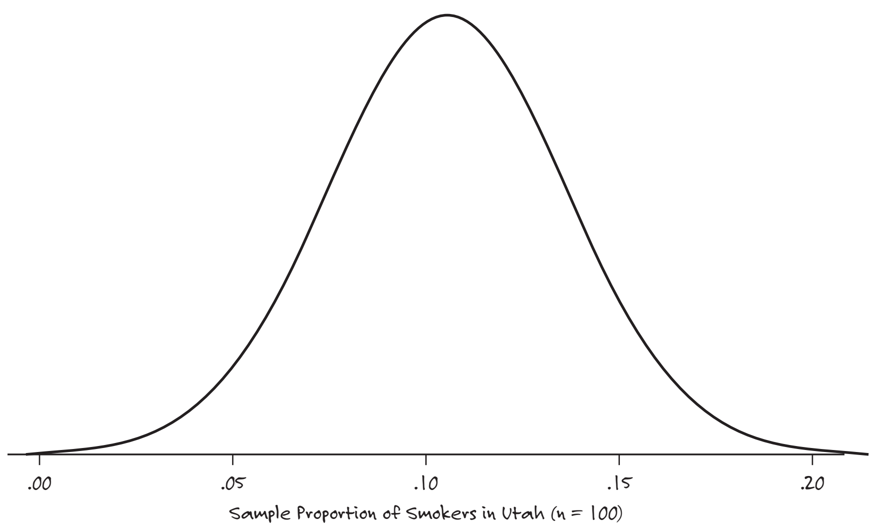
h. You calculate $z = (.25 - .209)/.0203 = 2.02$; $\Pr(Z > 2.02) = .0217$. Yes, this probability has decreased as predicted.

i. No, the population size of the United States did not enter into the calculations.

j. The previous calculations would not change in any way.

Activity 15-2: Smoking Rates

a. You have N(.105, .0307). The following sketch displays these results:



b. You calculate $z = (.25 - .105)/.0307 = 4.72$; $\Pr(Z > 4.72) \approx 0.000$.

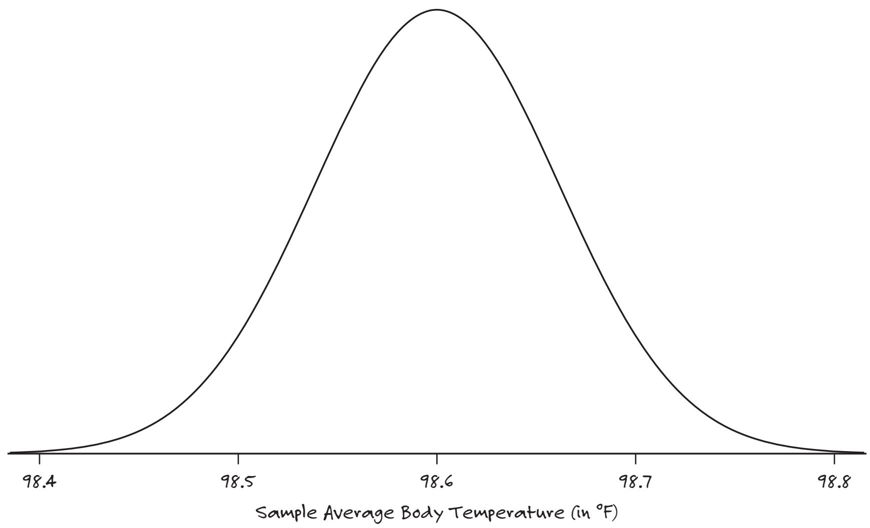
c. Yes, you would have strong reason to doubt that the state was Utah, because the probability of finding a random sample of 100 people from Utah containing 25 smokers is essentially zero—this never happens by chance alone. So if you find a random sample of 25/100 smokers, you have very strong evidence the sample is from some other state where the proportion of smokers is greater than 10.5%.

Activity 15-3: Body Temperatures

a. These numbers are parameters: $\mu = 98.6°F$, $\sigma = 0.7°F$.

b. Yes, the sample size is greater than 30 and you have a simple random sample, so the CLT applies.

c. The CLT says the sampling distribution of the sample means will be approximately normal with a mean of 98.6°F and a standard deviation of $.7/\sqrt{130} = .061$.



d. Pr(98.5 < *sample mean* < 98.7) = $Pr(-1.64 < Z < 1.64) = .9495 - .0505 =$ .8990 (Table II) or .8989 (applet)

e. Pr(98.2 < *sample mean* < 98.4) = $Pr(-1.64 < Z < 1.64) = .9495 - .0505 =$ .8990 (Table II) or .8989 (applet)

f. These answers are the same; you have simply shifted the center of the plot, but the area within ±0.1 degrees of the center has not changed.

g. $Pr(-0.1/.061 < Z < 0.1/.061) = Pr(-1.64 < Z < 1.64) = .9495 - .0505 =$ .8990 (Table II) or .8989 (applet)

So, there is about a 90% chance that a random sample of 130 adults will result in a sample mean body temperature that is within ±.1 degrees of the actual population mean μ if you assume the population standard deviation is $\sigma = 0.7°F$.

Activity 15-4: Solitaire

a. The CLT says the sampling distribution will be approximately N(.1111, .0994). You calculate $z = (.1 - .1111)/.0994 = -.11$. So $Pr(Z < -.11) = .4562$ (Table II) or .4562 (applet).

b. $Pr(\hat{p} < .10) = .308 + .385 = .693$

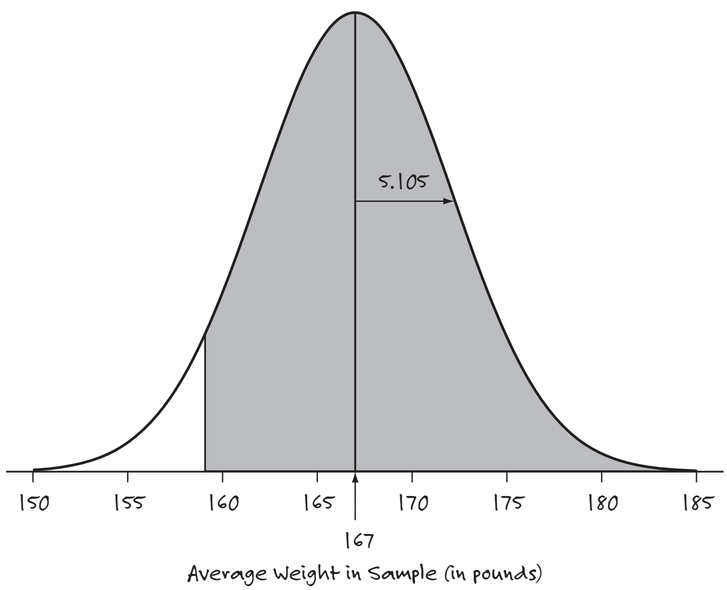
c. No, the probabilities in parts a and b are not close.

d. The technical conditions for the CLT are not satisfied. You calculate $n\pi = 10 \times (1/9) = 1.111 \not\geq 10$ and $n(1 - \pi) = 10 \times (8/9) = 8.888 \not\geq 10$.

Activity 15-5: Capsized Tour Boat

First, weight is a quantitative variable, so the relevant statistic is the sample mean weight of the 47 passengers. Because the question is phrased in terms of the *total* weight in a sample of 47 adults, you must rephrase it in terms of the sample *mean* weight. If total weight exceeds 7500 pounds, then the sample mean weight must exceed 7500/47 or 159.574 pounds. So, you want to find the probability that $\bar{x} > 159.574$ (with $n = 47$ and $\sigma = 35$).

The CLT applies because the sample size ($n = 47$) is fairly large, greater than 30. The sampling distribution of $\bar{x}$ is, therefore, approximately normal with mean 167 pounds and standard deviation equal to $\sigma/\sqrt{n} = 35/\sqrt{47} = 5.105$ pounds. A sketch of this sampling distribution is shown here:



Now you can use the `Normal Probability Calculator` applet or the Standard Normal Probabilities Table to find the probability of interest. The *z*-score corresponding to a sample mean weight of 159.574 pounds is $(159.574 - 167)/5.105 = -1.45$. The probability of the weight being less than 159.574 pounds is found from the table to be .0736, so the probability of exceeding this weight is $1 - .0736 = .9264$. It's not surprising the boat capsized with 47 passengers!

15

16

Solutions

In-Class Activities

Activity 16-1: Generation M

a. The observational units are American youth ages 8–18.

b. The variable is *whether each youth has a television in his or her bedroom* (binary categorical).

c. The value .68 is a statistic, which can be denoted by the symbol $\hat{p}$.

d. The relevant parameter of interest is the population proportion of all American youth ages 8–18 who have a television in their bedrooms, which can be denoted by the symbol π.

e. The Kaiser survey does not allow the researchers to determine the exact value of the parameter because they did not survey all American youth in the population.

f. The parameter value is more likely to be close to the survey's sample proportion than to be far from it, because the surveyed sample was randomly selected. You expect some sampling variability, but with a large sample size like this, you expect the sample proportion to be reasonably close to the true parameter.

g. The value $\hat{p}$ (.68) seems like a reasonable replacement to use as an estimate for π.

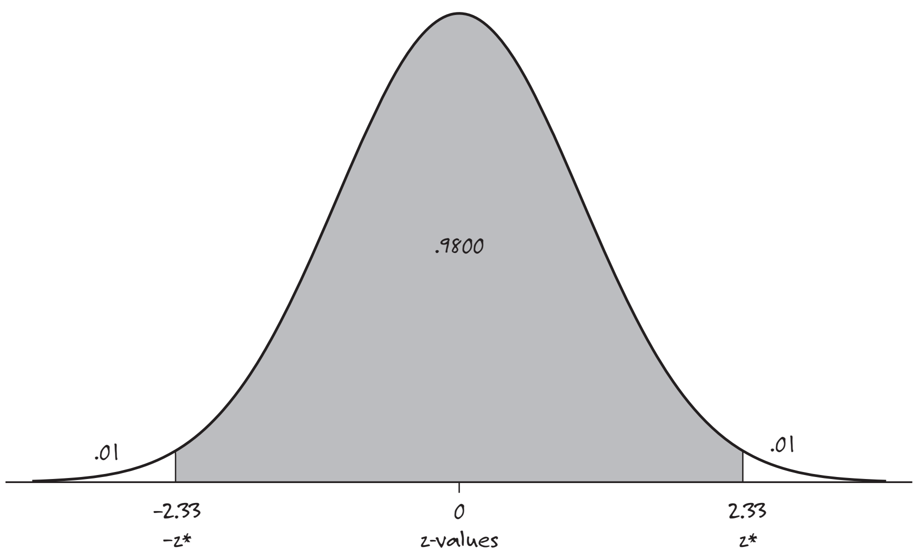
h. The standard error of $\hat{p}$ is $\sqrt{\frac{(.68)(.32)}{2032}} = .01035$.

i. You calculate $.68 \pm 2(.01035) = .68 \pm .0207 = (.6593, .7007)$. Consider the value of the population parameter, π, to be somewhere in this interval.

j. You do not know for sure whether the actual value of π is contained in this interval (as in part e).

Activity 16-2: Critical Values

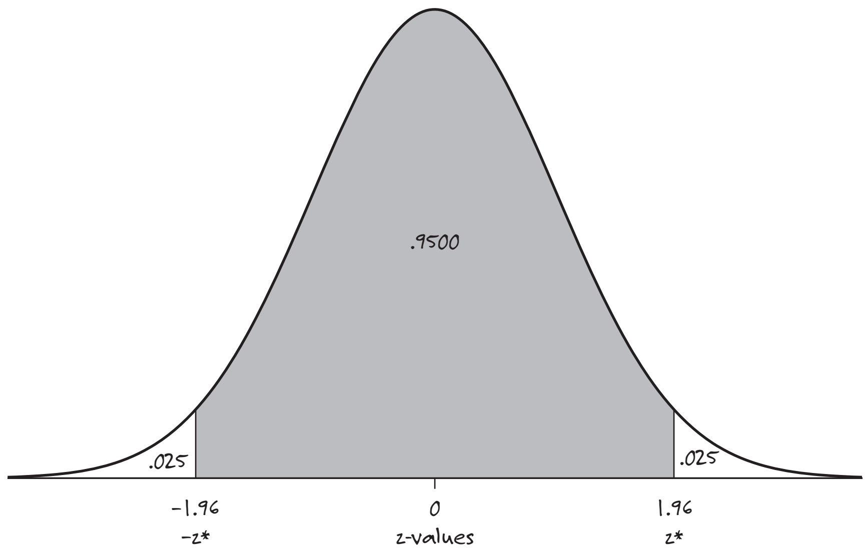
a. Here is the sketch:



b. The total area to the left of z^* is .9900.

c. $z^* = 2.33$

d. Here is the sketch:



The total area to the left of z^* is .975.

$z^* = 1.96$

Activity 16-3: Generation M

a. For a 95% confidence interval, you calculate $\sqrt{.68(.32)/2032} = .01035$. So $.68 \pm 1.96(.01035) = .68 \pm .020286 = (.6597, .7003)$.

b. You are 95% confident the population proportion of all American 8–18-year-olds who have a television in their bedrooms is between .66 and 70.

c. No, you cannot be *certain* that this interval contains the actual value of π.

d. Width = .7003 − .6587 = .0406

e. Half-width = .0406/2 = .0203

f. This half-width is also $z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ or 1.96 × (.01035), which is the margin-of-error.

g. Midpoint = (.7003 + .6597)/2 = .68

h. Yes, this value looks familiar. This value is $\hat{p}$. This value makes sense because the interval is created extending an equal distance above and below the observed sample proportion $\hat{p}$.

i. Technical conditions: (i) The sample is a simple random sample from the population of interest (probably more complicated than a “simple random sample,” but still random), and (ii) the sample size is large relative to π: $(2032)(.68) = 1382 > 10$ and $2032(.32) = 650 > 10$.

j. Answers will vary based on student intuition, but students should expect the interval to be wider; if they specify more values, they will be more confident that the actual value is in the specified range.

k. You calculate $.68 \pm 2.576(.01035) = .68 \pm .0267 = (.653, .707)$.

l. The midpoints of both intervals are the same (.68), but the 99% confidence interval is wider; it has a greater margin-of-error (.267) vs. (.203).

m. .50: not plausible .75: not plausible Two-thirds: plausible

Explanation: Both .50 and .75 do not seem to be plausible values for π as they are not contained in either the 95% or 99% confidence intervals. However, .67 is contained in both confidence intervals, so it seems to be a plausible value for π.

n. Boys: $.72 \pm (1.96)\sqrt{.72(.28)/996} = .72 \pm (1.96)(.0142) = (.692, .748)$

Girls: $.64 \pm (1.96)\sqrt{.64(.36)/1036} = .64 \pm (1.96)(.0149) = (.611, .669)$

o. These intervals do seem to indicate that there is a difference in the population proportion of boys and girls who have a television in their bedrooms. You are 95% confident the population proportion of boys with a television in their bedrooms is at least .69 (and no more than .75), whereas you are 95% confident that the proportion of girls with a television in their bedrooms is between .61 and .67. There is no overlap in these intervals; the values of π that are plausible values for boys are not plausible values for girls.

p. The margins-of-error for these intervals are .028 (boys) and .029 (girls). These are greater than the margin-of-error based on the entire sample, which makes sense because the entire sample is roughly twice as large as the single gender samples, so you would expect it to have less sampling variability and therefore a smaller margin-of-error.

q. You set the margin-of-error formula equal to .01 and then solve for the sample size n, as follows:

$$.01 = 1.96\sqrt{\frac{(.68)(.32)}{n}} \qquad n = (.68)(.32)\left(\frac{1.96}{.01}\right)^2 = 8359.322;\ n = 8{,}360$$

r. Answers will vary by student expectation, but the required sample size will increase.

s. To determine the sample size, you calculate

$$.01 = 2.576\sqrt{\frac{(.68)(.32)}{n}} \qquad n = (.68)(.32)\left(\frac{2.576}{.01}\right)^2 = 14439.448;\ n = 14{,}440$$

Activity 16-4: Candy Colors

Answers will vary. Here is one representative set of answers.

a. Endpoints: .315, .539.

Yes, this interval succeeds in capturing $\pi = .45$.

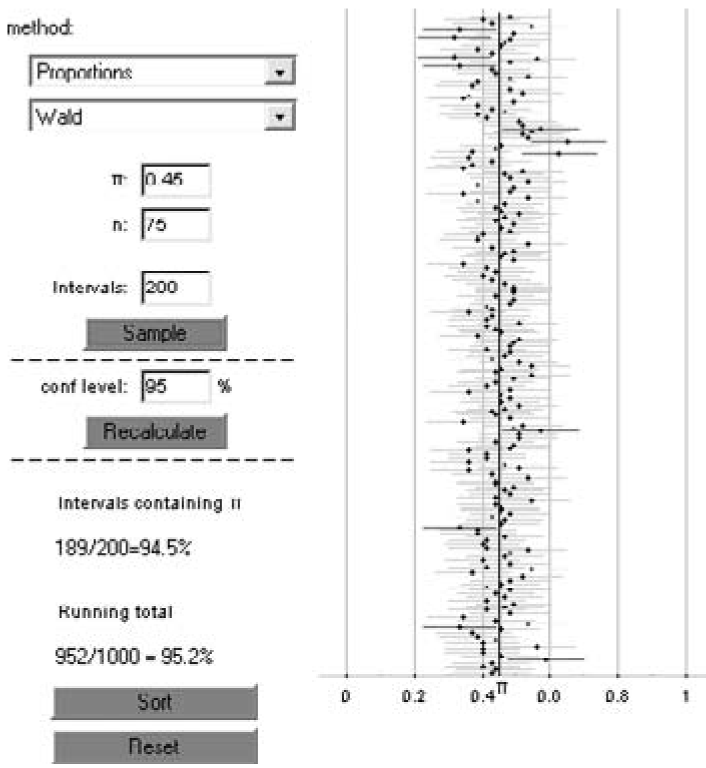
b. Yes, the interval changed each time. No, the intervals did not all succeed in capturing the true value of π (4 out of 5 did).

c. In this simulation, the proportion of intervals that succeed in capturing π is 190/200 = 95% (you expect something close to 95% but it may not match exactly with 200 samples).

d. The intervals that fail to capture π have midpoints that are fairly far away (more than 2 standard deviations) from .45 (either on the low end of the scale or the high end).

e. No; if you had taken a single sample in a real situation, you would have no way of knowing whether the true value of π was contained in your interval because you would not know what π was; you are not guaranteed that the constructed interval will capture the value of π.

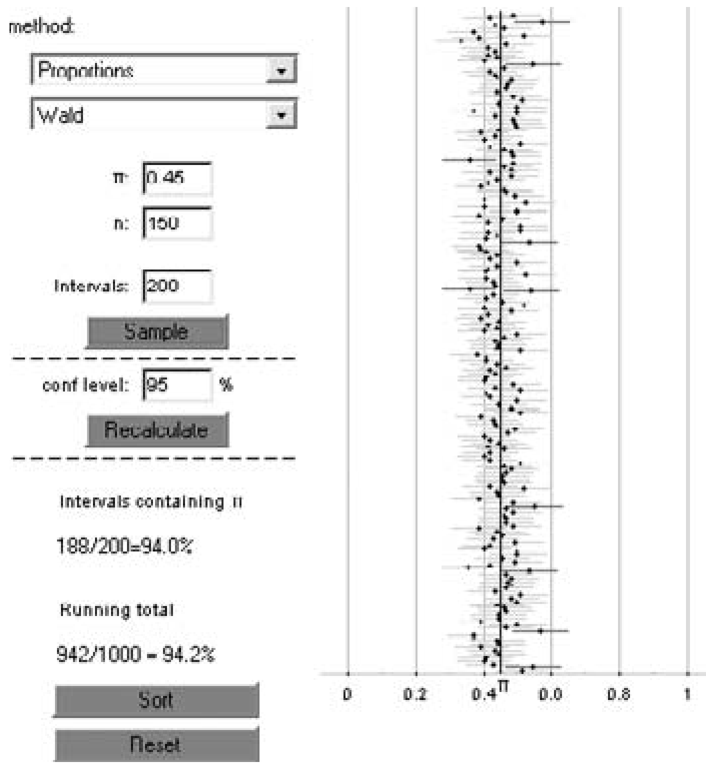
f. In the following simulation, the proportion of intervals that succeed in capturing π is 952/1000 = 95.2%:



g. Yes, this percentage is close to 95%. Yes, it should be close to 95% because these are 95% confidence intervals; the procedure should be "successful" 95% of the time. This simulation reveals that the phrase "95% confidence" indicates that your method of creating confidence intervals is successful in capturing the true population parameter (π) 95% of the time and that it fails 5% of the time in the long run (over many, many intervals).

h. Answers will vary by student prediction, but the intervals will become less wide (more narrow) as the sample size increases.

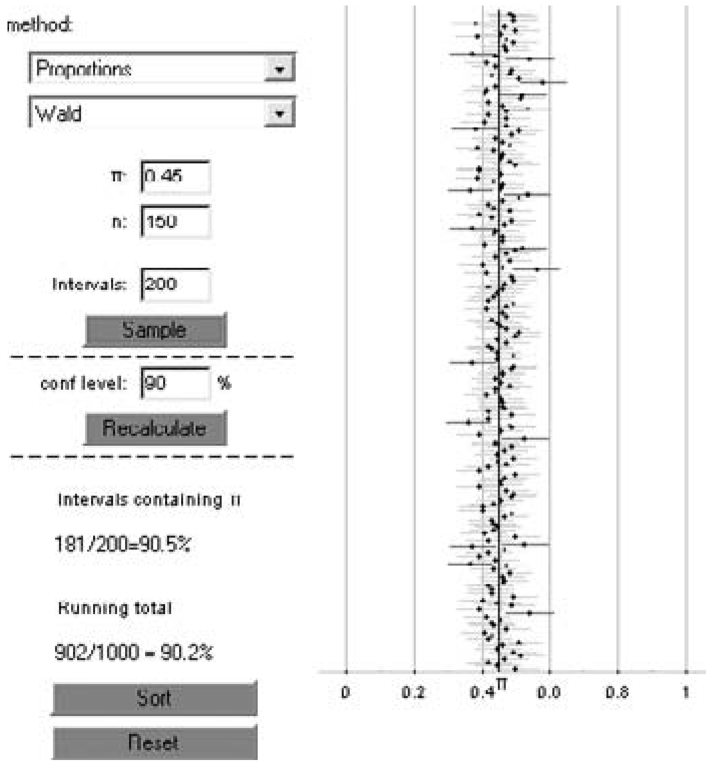
i. In the following simulation, the proportion of intervals that succeed in capturing π is $942/1000 = 94.2\%$:



This is reasonably close to the percentage from part f. The noticeable difference about these intervals is that they are not as wide as those that were generated with samples of 75 candies. [Example interval: (.341, .499).]

j. Answers will vary by student prediction, but students should predict that the length of the intervals will *shorten* and the success rate will *decrease* when the confidence level is changed to 90%.

k. In the following simulation, the proportion of intervals that succeed in capturing π is $902/1000 = 90.2\%$.



This is not particularly close to the percentages from parts f and i, but it shouldn't be because you changed the confidence level. The noticeable differences about these intervals are that they are not as wide as the previous intervals [example interval: (.334, .466)], nor are they as successful in capturing π.

Activity 16-5: Elvis Presley and Alf Landon

a. For a 99.9% confidence interval, you calculate $.57 \pm (3.291)\sqrt{.57(.43)/2400000} = .57 \pm (3.22291)(.00032) = (.569, .571)$.

b. This interval is so narrow because the sample size is so very, very large.

c. The confidence interval did such a poor job predicting the election results because the necessary technical conditions were not satisfied. In particular, the sample was not randomly selected; it was chosen in a very biased fashion and therefore vastly overestimated the support for the Republican candidate.

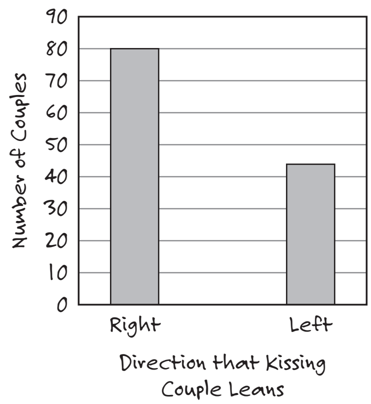
Activity 16-6: Kissing Couples

a. The observational units are the kissing couples. The variable is *which direction the couples lean their heads while kissing.*

b. The sample consists of the 124 kissing couples observed by the researchers in various public places. The statistic is the sample proportion of couples who lean to the right when kissing:

$\hat{p} = 80/124 = .645$

A bar graph is shown here:



c. A 95% confidence interval for the population proportion of all couples who lean to the right is

$$.645 \pm 1.96\sqrt{\frac{(.645)(1 - .645)}{124}}$$

which is .645 ± .084, or (.561, .729). You are 95% confident that the population proportion of all kissing couples who lean to the right is somewhere between .561 and .729. This "95% confidence" means that if you were to take many random samples and generate a 95% confidence interval (CI) from each, then in the long run, 95% of the resulting intervals would succeed in capturing the actual value of the population proportion, in this case the proportion of all kissing couples who lean their heads to the right.

d. A 90% CI is

$$.645 \pm 1.645\sqrt{\frac{(.645)(1 - .645)}{124}}$$

which is .645 ± .071, or (.574, .716).

A 99% CI is

$$.645 \pm 2.576\sqrt{\frac{(.645)(1 - .645)}{124}}$$

which is .645 ± .111, or (.534, .756).

The higher confidence level produces a wider confidence interval. All of these intervals have the same midpoint: the sample proportion .645.

e. Because none of these intervals includes the value .5, it does not appear to be plausible that 50% of all kissing couples lean to the right. In fact, all of the intervals lie entirely above .5, so the data suggest that more than half of all kissing couples lean to the right. The value 2/3 is quite plausible for this population proportion because .667 falls within all three confidence intervals. The value 3/4 is not very plausible because only the 99% CI includes the value .75; the 90% CI and 95% CI do not include .75 as a plausible value.

f. The sample size condition is clearly met, as $n\hat{p} = 80$ is greater than 10, and $n(1 - \hat{p}) = 44$ is also greater than 10. But the other condition is that the sample be randomly drawn from the population of all kissing couples. In this study, the couples selected for the sample were those who happened to be observed in public places while the researchers were watching. Technically, this is not a random sample, and so you should be cautious about generalizing the results of the confidence intervals to a larger population.

Solutions

In-Class Activities

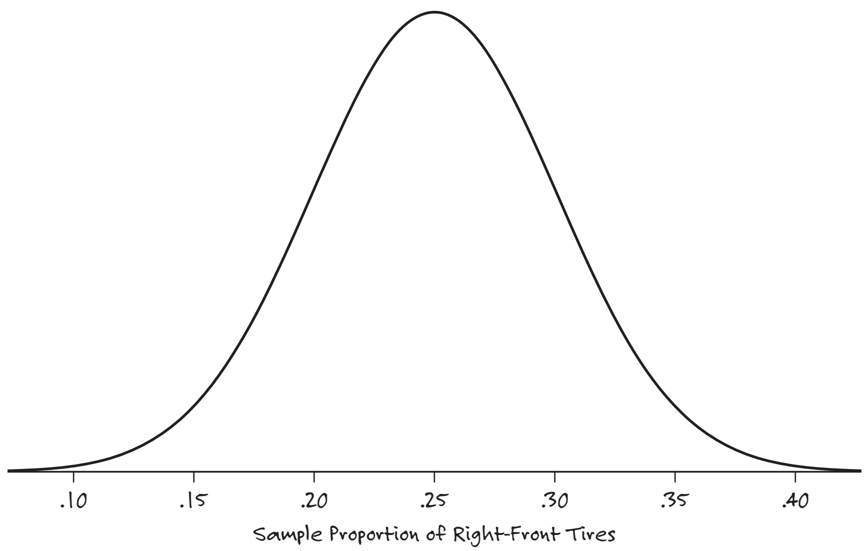
Activity 17-1: Flat Tires

a. 1/4 or .25

b. This value is a parameter because it represents the overall process, not simply results that you observed, and is represented by π.

c. $\pi > .25$

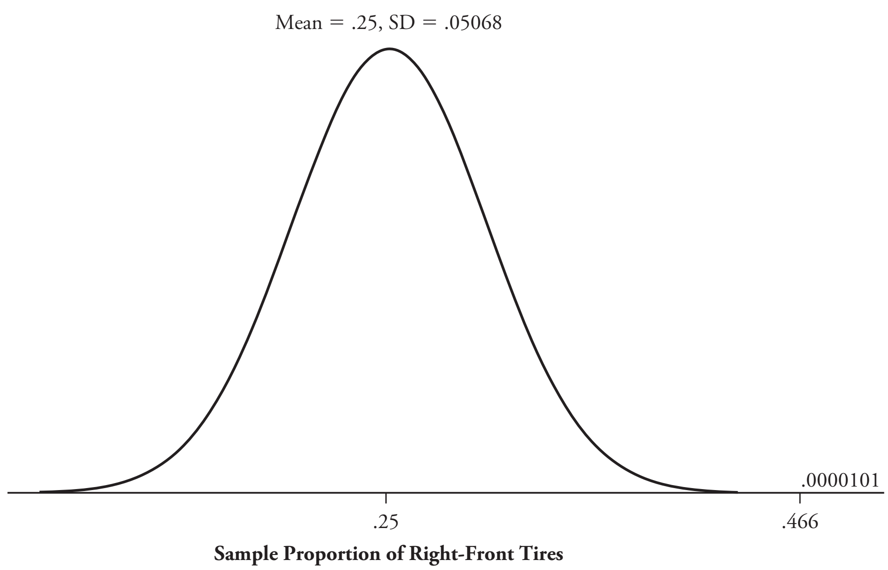
d. The sampling distribution of the sample proportion will be approximately normal, with mean equal to .25, and standard deviation equal to $\sqrt{.25(1 - .25)/73} =$ .0507. Here is a sketch of the sampling distribution:



e. The conditions necessary for the CLT to be valid are that $n\pi \geq 10$ and $n(1 - \pi) \geq 10$. These conditions are met ($73(.25) > 10$ and $73(.75) > 10$). However, you also need to believe that the sample is representative of the larger population process. This is less clear, but there may not be any reason to believe that this professor's class would behave substantially differently on this issue than college students in general, or you may want to think more carefully about how you define the population in this activity (e.g., students of similar age and major).

f. To determine the sample proportion, you calculate $\hat{p} = \frac{34}{73} = .466$. Yes, this sample proportion is greater than 1/4.

g. The following graph displays the shaded area:



For the z-score, you calculate

$$z = \frac{.466 - .25}{.0507} = 4.26$$

p-value $= \Pr(Z > 4.26) < 0.0002$

h. This sample result would be surprising if there were nothing special about the right-front tire. With a *p*-value of approximately zero, you would consider this sample result so surprising that you would conclude the right-front tire would be chosen by more than one-fourth of the population.

i. Let π represent the proportion of the population who will select the right-front tire when asked this "flat tire" question.

j. $H_0: \pi = .25$

k. $H_a: \pi > .25$

l. The test statistic is $z = \dfrac{.466 - .25}{\sqrt{.25(1 - .25)/73}} = \dfrac{.466 - .5}{.0507} = 4.26.$

m. The *p*-value $< .0002$.

n. Yes, this probability suggests that it is very unlikely for 34 or more of 73 randomly selected students to choose the right-front tire if one-fourth of the population would choose the right-front tire. Such a sample result would occur in less than 1% of random samples if $\pi = .25$.

o. Calculate $73(.25) = 18.25 > 10$ and $73(.75) = 54.75 > 10$, so this condition is met. You are not certain this is a simple random sample from the population of interest, but it is likely to be a representative sample of introductory statistics students.

p. You have found very strong statistical evidence that *introductory statistics students* tend to choose the right-front tire more than one-fourth of the time.

Activity 17-2: Flat Tires

a. Define parameter of interest: Let π represent the proportion of all introductory statistics students who will select the right-front tire when asked this "flat tire" question.

$H_0: \pi = .25$

$H_a: \pi > .25$

Check technical conditions: $74(.25) = 18.5 > 10$ and $74(.75) = 55.5 > 10$, so this condition is met. You are not certain this is a simple random sample from the population of interest, but it is likely to be a representative sample of introductory statistics students.

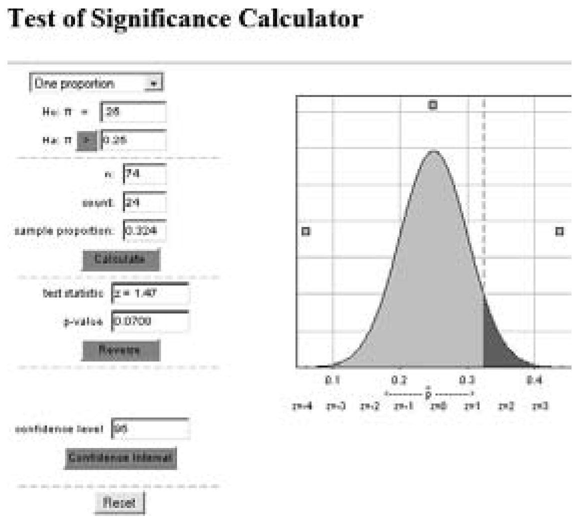
Test statistic: $z = \dfrac{.324 - .25}{\sqrt{\dfrac{(.25)(.75)}{73}}} = 1.47$

p-value $= \Pr(Z > 1.47) = .0708$

Test decision: At $\alpha = .05$ level, do not reject H_0 (*p*-value $= .0708 > .05$).

Conclusion in context: You do not have sufficient statistical evidence at the 5% level (though you do at the 10% level) to conclude that the proportion of statistics students who will choose the right-front tire is greater than one-fourth. You will continue to believe that the right-front tire is no more likely to be chosen than any other tire.

b. The following graph confirms calculation of the test statistic and p-value:



Activity 17-3: Racquet Spinning

a. Population parameter of interest: Let π represent the proportion of times a spun tennis racquet will land "up."

b. H_0: $\pi = .5$ (A spun tennis racquet will land "up" half the time.)

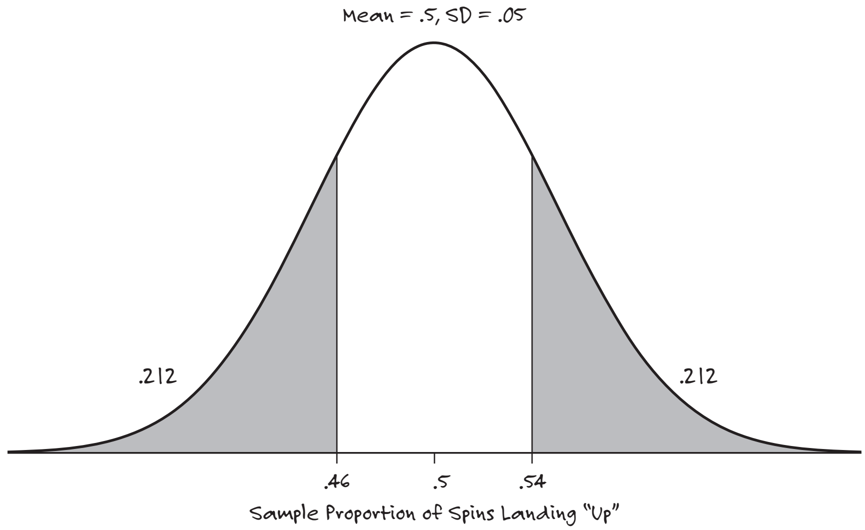
H_a: $\pi \neq .5$ (A spun tennis racquet will not land "up" half the time.)

c. Technical conditions: As long as there was nothing unusual about the spinning process, you will consider these data a "random" sample, and $100(.5) = 100(1 - .5) = 50 > 10$, so the technical conditions for the validity of this test procedure are satisfied.

d. The test statistic is $z = \dfrac{.46 - .5}{\sqrt{\dfrac{(.5)(.5)}{100}}} = -0.80$.

e. p-value $= 2 \times \Pr(Z < -0.80) = 2 \times .2119 = .4238$

Here is a sketch of the standard normal curve with the *p*-value shaded:



f. This *p*-value is very large (.4238 > .05). Therefore, do not reject the null hypothesis at the .05 significance level.

g. Conclusion in context: You do not have statistical evidence that would allow you to conclude that a spun tennis racquet will fail to land "up" half the time.

Activity 17-4: Flat Tires

a. You need more information in order to decide whether this constitutes strong evidence that the right-front tire would be chosen more than one-quarter of the time in the long run. You need to know the number of people surveyed (the sample size).

b. See values in table following part c.

c. Here is the completed table:

Sample Size	# "right-front"	$\hat{p}$	*z* statistic	*p*-value	$\alpha = .10$?	$\alpha = .05$?	$\alpha = .01$?	$\alpha = .001$?
50	15	.30	0.82	.207	No	No	No	No
100	30	.30	1.15	.124	No	No	No	No
150	45	.30	1.41	.079	Yes	No	No	No
250	75	.30	1.83	.034	Yes	Yes	No	No
500	150	.30	2.58	.005	Yes	Yes	Yes	No
1000	300	.30	3.65	.000	Yes	Yes	Yes	Yes

d. When the sample size is small, a sample result of .30 is not statistically significant at any level. But, as the sample size increases, this result becomes more significant—meaning that it becomes more unlikely that you would obtain a sample result of .30 (or more extreme) if the population parameter is actually .25 as you use larger and larger samples.

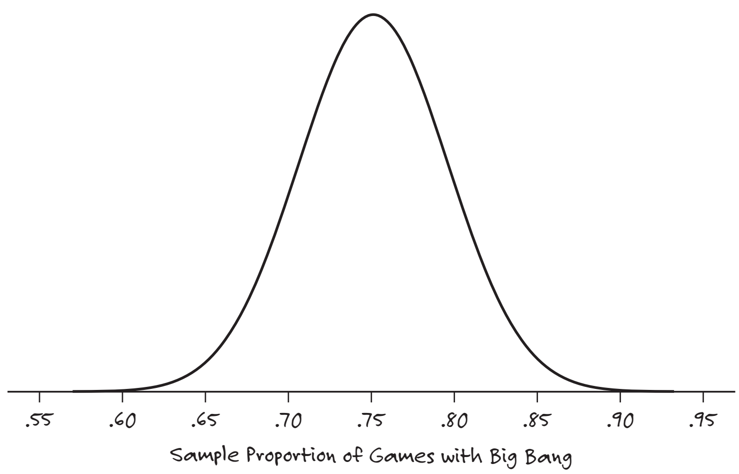
Activity 17-5: Baseball "Big Bang"

a. The null hypothesis is that the proportion of all major-league baseball games that contain a big bang is three-fourths. In symbols, the null hypothesis is H_0: $\pi = .75$.

b. The alternative hypothesis is that less than three-fourths of all major-league baseball games contain a big bang. In symbols, the alternative hypothesis is H_a: $\pi < .75$.

c. A week of games was randomly selected. Although this does not constitute a simple random sample of all games, you do hope it is representative of the scores in such games. The CLT applies here because $95(.75) = 71.25$ is greater than 10, and $95(.25) = 23.75$ is also greater than 10. According to the CLT, the sample proportion would vary approximately normally, with mean .75 and standard deviation equal to

$$\sqrt{\frac{(.75)(.25)}{95}} \approx .0444$$



d. The sample proportion of games in which a big bang occurred is

$$\hat{p} = \frac{47}{95} \approx .495$$

e. Yes, this sample proportion is less than .75, as Marilyn conjectured.

f. The test statistic is

$$z = \frac{.495 - .75}{\sqrt{\frac{(.75)(.25)}{95}}} \approx \frac{.495 - .75}{.0444} \approx -5.74.$$

This test statistic says that the observed sample result is almost six standard deviations below what the grandfather conjectured. This z-score is way off the chart in Table II, indicating that the p-value is virtually zero ($< .0002$).

g. Yes, this very small p-value indicates that the sample data provide extremely strong evidence against the grandfather's claim. There is extremely strong evidence that less than 75% of all major-league baseball games contain a big bang. The null (grandfather's) hypothesis would be rejected at the $\alpha = .01$ level.

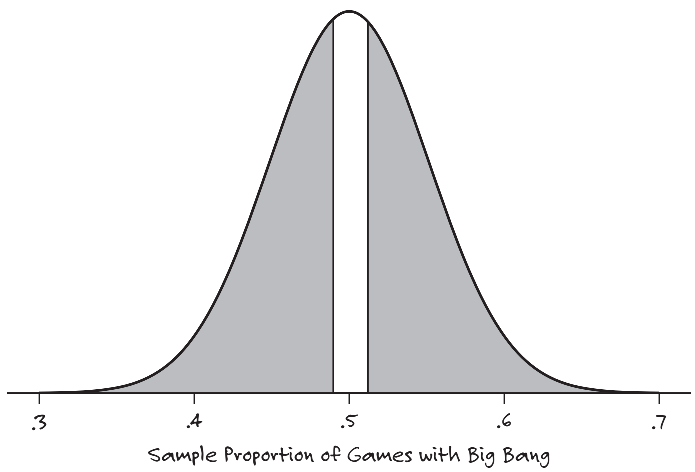
h. The hypotheses for testing Marilyn's claim are H_0: $\pi = .5$ vs. H_a: $\pi \neq .5$.

17

i. The test statistic is

$$z = \frac{.495 - .5}{\sqrt{\frac{(.5)(.5)}{95}}} \approx \frac{.495 - .5}{.0513} \approx -0.10$$

The p-value is $2(.4602) = .9204$.



j. This p-value is not small at all, suggesting that the sample data are quite consistent with Marilyn's hypothesis that half of all games contain a big bang. The sample data provide no reason to doubt Marilyn's hypothesized value for π.

k. A 95% confidence interval for π (the population proportion of games that contain a big bang) is given by

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

with $z^* = 1.96$, which is

$$.495 \pm 1.96\sqrt{\frac{(.495)(.505)}{95}}$$

which is $.495 \pm .101$, which is the interval from .394 to .596. Therefore, you are 95% confident that between 39.4% and 59.6% of all major-league baseball games contain a big bang. The grandfather's claim (75%) is not within this interval or even close to it, which explains why it was so soundly rejected. Marilyn's conjecture (50%) is well within this interval of plausible values, which is consistent with it not being rejected.

18

Solutions

● ● ● In-Class Activities

Activity 18-1: Generation M

a. This is a statistic because it is a number that represents a sample.

b. For a 99% confidence interval, you calculate .68 ± (2.576)(.010348) = .68 ± (.02666) = (.6533, .7066).

c. The values .70 and .6667 are in this interval; .65 and .707 are not.

d. A significance test should reject the hypothesis that $\pi = .65$, but .7 is contained in the interval, so this could be a plausible value for π. A significance test would not reject the hypothesis that $\pi = .7$.

e. Using Minitab's Test and CI for One Proportion:

```
Test of p = 0.65 vs p not = 0.65
Sample     X     N  Sample p        99% CI          Z-Value  P-Value
1       1382  2032   0.680118  (0.653465, 0.706771)     2.85    0.004
Using the normal approximation.
```

f. See table following part g.

g. Here is the completed table:

Hypothesized Value	Contained in 99% Confidence Interval?	Test Statistic	*p*-value	Significant at .01 Level?
.65	No	2.85	.004	Yes
.6667	Yes	1.28	.199	No
.7	Yes	−1.96	.05	No
.707	No	−2.66	.008	Yes (barely)

h. If a hypothesized value is contained in the 99% confidence interval, then this value is *not* significant at the .01 level, and vice versa.

Activity 18-2: Pet Ownership

a. Because this number (.316) describes a sample, it is a statistic, represented by $\hat{p}$.

b. Let π represent the proportion of all American households who own a pet cat.

The null hypothesis is that one-third of all American households own a pet cat. In symbols, the null hypothesis is H_0: $\pi = .333$.

The alternative hypothesis is that the proportion of American households who own a pet cat differs from one-third. In symbols, the alternative hypothesis is H_a: $\pi \neq .333$.

The test statistic is $z = \dfrac{.316 - .333}{\sqrt{\dfrac{(.333)(.667)}{80{,}000}}} = -10.20$.

Using Table II, p-value $= 2 \times \Pr(Z < -10.20) < .0002$.

Reject H_0 with this very small p-value.

You have overwhelming statistical evidence that the proportion of all American households who own a cat differs from one-third.

c. For a 99.9% CI, you calculate $.316 \pm (3.291)\sqrt{.316(1 - .316)/80000} = .316 \pm (3.291)(.001644) = (.310591, .321409)$. You are 99.9% confident the proportion of all American households who own a pet cat is between .311 and .321.

d. Yes, this confidence interval is consistent with the test results because $1/3 \approx .333$ is not contained in the interval.

e. Yes, the sample data provide *very strong* evidence that the population proportion (π) is not one-third. The p-value is what helps you decide this; the p-value is so small (essentially zero) that it easily convinces you that π is not one-third.

f. No, the sample data do not provide strong evidence that the population proportion of households who own a pet cat is very different from one-third. The evidence suggests that this proportion is between .311 and .321, which are awfully close to .33. The confidence interval helps you decide how much π differs from one-third.

Activity 18-3: Racquet Spinning

a. Sample proportion: .565 Test statistic: 1.84 *p*-value: .066 Significant at .05? no

b. Sample proportion: .575 Test statistic: 2.12 *p*-value: .034 Significant at .05? yes

c. Sample proportion: .65 Test statistic: 4.24 *p*-value: .000 Significant at .05? yes

d. The sample results are most similar in parts a and b, where you had almost the same number of "ups."

e. The decisions are the same in parts b and c (where the sample results are quite dissimilar).

Activity 18-4: Female Senators

a. For a 95% CI, you calculate $.16 \pm 1.96(.0367) = (.088, .231)$.

b. No, this confidence interval is not a reasonable estimate of the actual proportion of all humans who are female.

c. The confidence interval procedure fails in this case because the alien did not select a simple random sample of all humans. The U.S. Senate is not representative of the population of all humans with respect to gender, so the sampling method is extremely biased and you cannot legitimately use the confidence interval procedure.

d. You do not need to estimate the proportion of women in the 2007 U.S. Senate. You know this proportion is .16.

Activity 18-5: Hypothetical Baseball Improvements

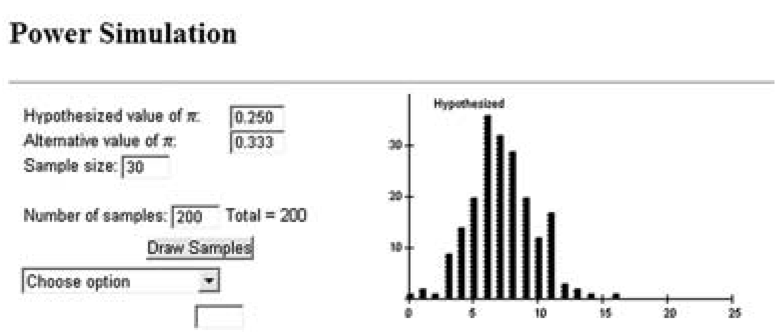
a. The null hypothesis is that this player is still a .250 hitter. In symbols, H_0: $\pi = .250$.

The alternative hypothesis is that this player has improved and is now better than a .250 hitter. In symbols, H_a: $\pi > .250$.

b. A Type I error would be deciding that the player has improved his batting performance when, in fact, he is still batting no better than .250.

c. A Type II error would be failing to realize that the player has improved.

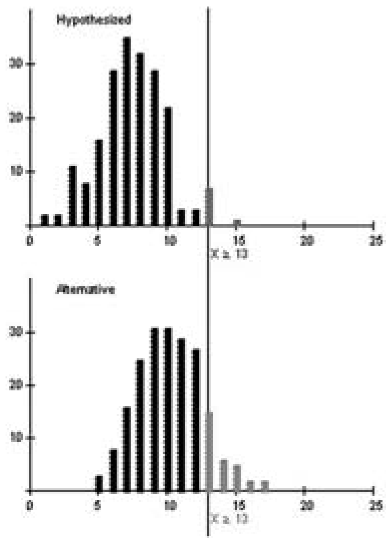
d. Answers will vary. The following is a representative set:



This distribution is roughly normal, centered at about 7.5 hits, and extends from about 1 hit to about 17 hits.

e. A player would need to get at least 13 hits.

f. There is a great deal of overlap between the two distributions.

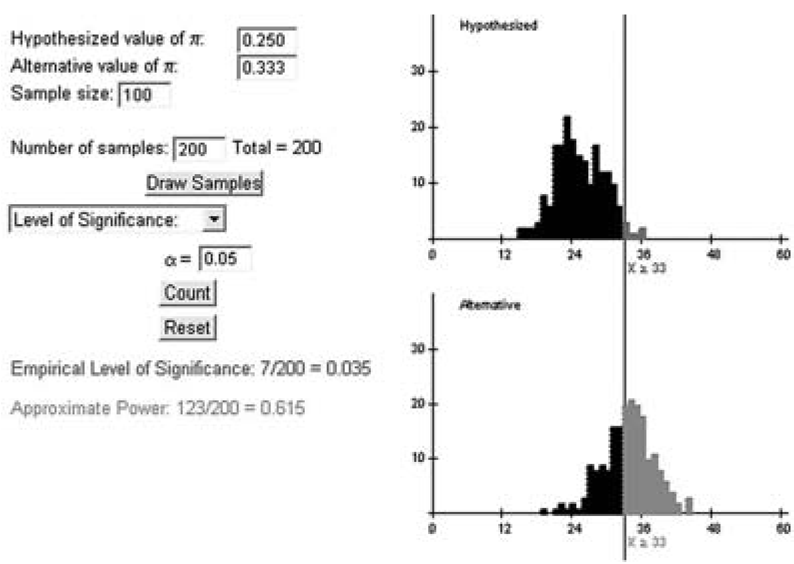


g. 28/200 = 14%

h. No, it does not appear very likely that a .333 hitter will be able to establish that he is better than a .250 hitter in 30 at-bats. Based on this simulation, he had only about a 14% chance of establishing his improvement (performing well enough to convince the manager that his success rate was now greater than .250).

i. Power ≈ .14

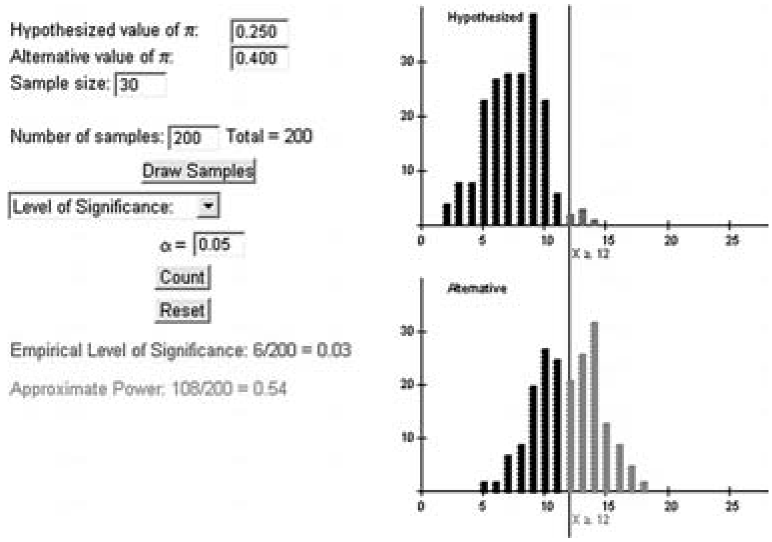
j. The following is based on one representative running of the applet:



Based on this simulation, a player would need at least 33 hits (out of 100) in order for the probability of a .250 hitter to do that well by chance alone to be less than .05. The approximate power of this test is 123/200 or .615.

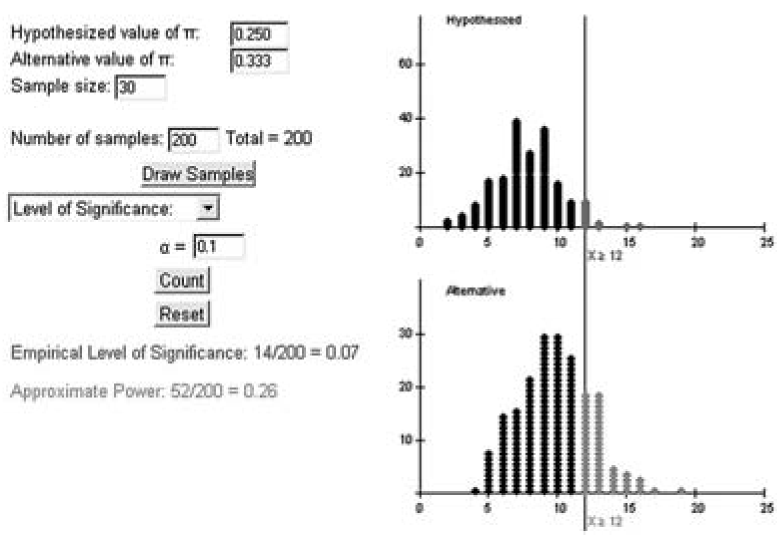
k. Answers will vary by student expectation, but the test will be more powerful if the player improves to a .400 hitter. It should be easier to detect the improvement to .400 because it is farther away from .250 than .333 is.

The simulation confirms this result; the approximate power is 108/200 or .54 (using 30 at-bats).



l. Answers will vary by student expectation, but the test will be more powerful if you use a higher significance level.

The simulation confirms this result; the approximate power is 52/200 or .26 (using 30 at-bats, alternative value of $\pi = .333$ and the significance level .10).



m. i. The magnitude of the difference between the hypothesized value π_0 and the particular alternative value of π

ii. The significance level

Activity 18-6: *West Wing* Debate

a. The population of interest is all adult Americans who are familiar with these fictional candidates. The parameter (call it π) is the proportion of this population who would have supported Santos if they had been asked.

b. The 90% CI for π is $.54 \pm .024$, which is (.516, .564).

The 95% CI for π is .54 ± .028, which is (.512, .568).

The 99% CI for π is .54 ± .037, which is (.503, .577).

c. The midpoints are all the same, namely .54, the sample proportion of Santos supporters. The 99% CI is wider than the 95% CI, and the 90% CI is the narrowest.

d. Yes. All three intervals contain only values greater than .5, so they do suggest, even with 99% confidence, that more than half of the population would have favored Santos.

e. H_0: $\pi = .5$ (half of the population favored Santos)

H_a: $\pi > .5$ (more than half of the population favored Santos)

f. Because all three intervals fail to include the value .5, you know that the *p*-value for a two-sided alternative would be less than .10, .05, and .01. Because you have a one-sided alternative in this case, you know that the *p*-value will be less than .01 divided by 2, or .005 (because the observed sample proportion is in the conjectured direction).

g. A Type I error occurs when the null hypothesis is really true but is rejected. In this case, a Type I error would mean that you conclude that Santos was favored by more than half of the population when in truth he was not favored by more than half. In other words, committing a Type I error means concluding that Santos was ahead (favored by more than half) when he wasn't really. A Type II error occurs when the null hypothesis is not really true but is not rejected (you continue to believe a false null hypothesis). In this case, a Type II error means that you conclude Santos was not favored by more than half of the population when in truth he was favored by more than half of the population. In other words, committing a Type II error means concluding that Santos was not ahead when he really was.

h. The test would be more powerful if Santos really were favored by 55% rather than 52%. The higher population proportion would make it more likely to reject the null hypothesis that only half of the population favored Santos because the distribution of sample proportions would center around .55 rather than .52 (further from .5).

i. The larger sample (10,000) would produce stronger evidence that more than half of the population favored Santos. With less variability in the sampling distribution, the *p*-value would be much smaller.

Solutions

In-Class Activities

Activity 19-1: Christmas Shopping

a. The *amount expected to be spent on Christmas presents in 1999* is a quantitative variable.

b. The value $857 is a statistic because it is a number that describes a sample. This statistic is represented by $\bar{x}$.

c. The parameter is the average (mean) amount expected to be spent by all American adults on Christmas presents in 1999. This parameter is represented by μ.

d. You do not *know* the value of the μ, but it is more likely to be close to $857 than to be far from it.

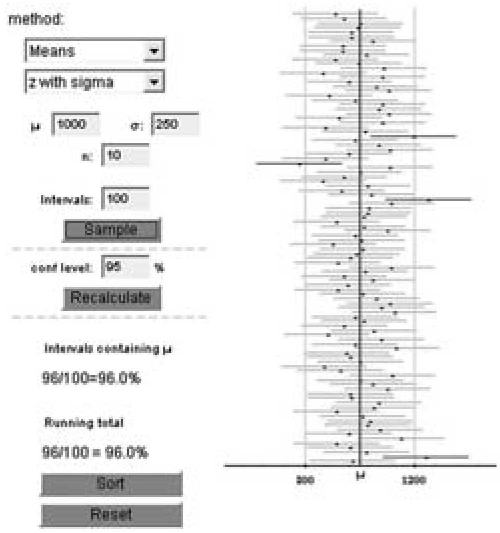
e. The standard deviation of the sample mean $\bar{x}$ is $\sigma/\sqrt{n} = 250/\sqrt{922} = \$8.23$.

f. This interval estimate works out to be $\$857 \pm 2(\$8.23) = \$857 \pm \$16.47 = (\$840.53, \$873.47)$.

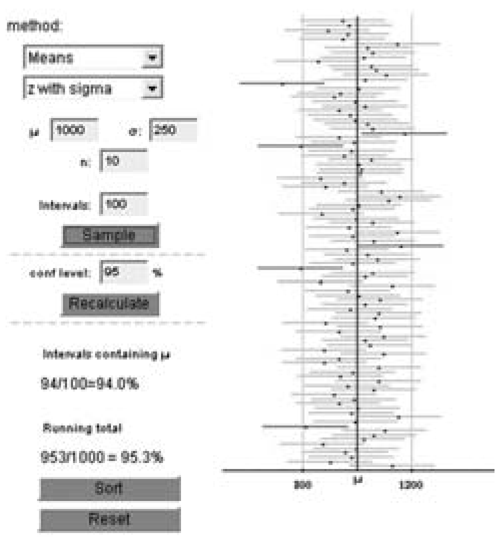
g. The sample standard deviation, s, is a reasonable substitute for σ.

h. Answers will vary. The following is from one representative running of the applet:

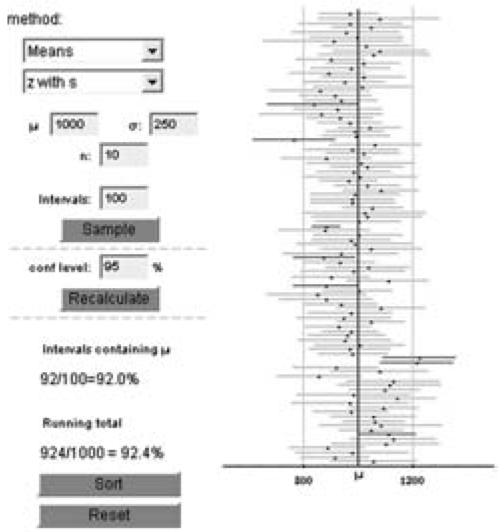
You find 96% of the intervals succeed in capturing the value of μ. Your total should be roughly 95%.



i. The running total percentage of intervals that succeed in capturing the population mean is 95.3%, which is very close to the expected 95%.

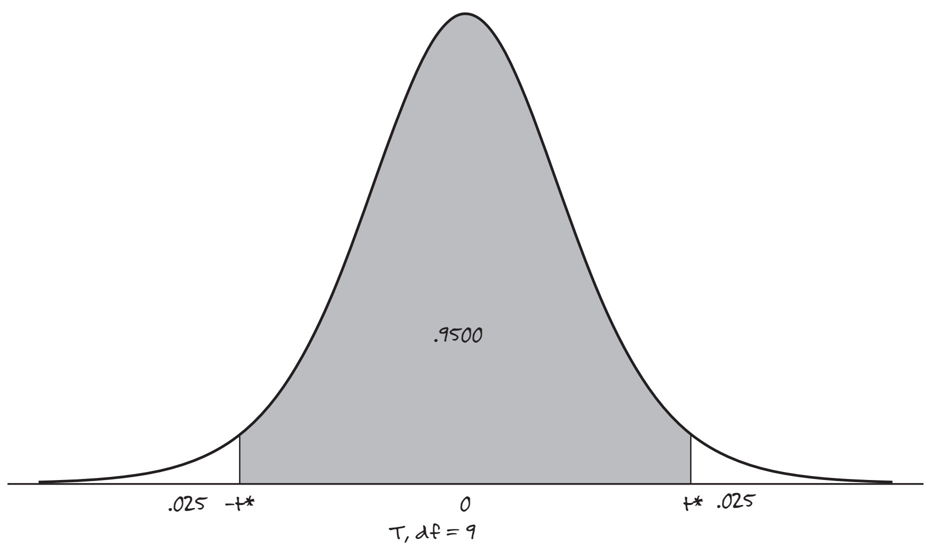


j. The running total percentage of intervals that succeed in capturing the population mean is 92.4%, which is noticeably less than 95%.



Activity 19-2: Exploring the *t*-Distribution

a. Here is a sketch of the $t(9)$ distribution:



b. See shading in plot above.

c. The area to the right of $t^* = 1 - .975 = .025$.

d. Using the *t*-table (Table III), $t^* = 2.262$.

e. This t^* critical value is greater than the z^* critical value for a 95% confidence interval because of the greater uncertainty introduced by estimating with s rather than σ. This will make the margins-of-error of your confidence intervals wider in order to achieve the stated confidence level in the long run.

f. For a 95% confidence interval, $t^* = 2.045$. This value is less than the previous t^* value, which makes sense because you have increased the sample size, decreasing the uncertainty in estimating σ by s.

g. For a 90% confidence interval $t^* = 1.699$; for a 99% confidence interval $t^* = 2.759$. The t^* for a 99% confidence interval is greater, which is appropriate because this interval claims more confidence (more certainty). In order to be more confident, the interval will need to be wider.

h. With 100 degrees of freedom, $t^* = 1.984$.

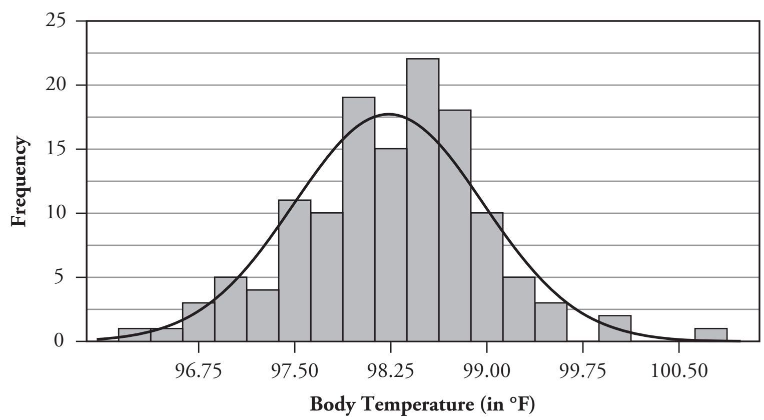
Activity 19-3: Body Temperatures

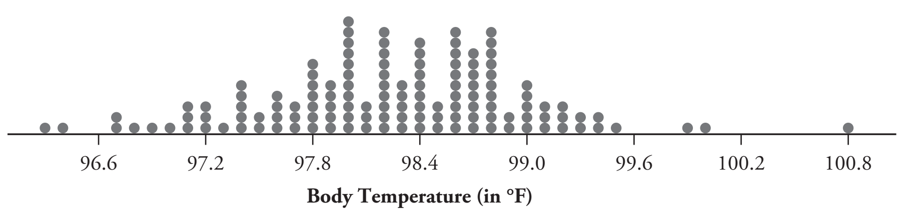
a. For a 95% CI with 129 degrees of freedom, you calculate $98.249 \pm (1.984)(.733)/\sqrt{130} = (98.1215, 98.3765)$.

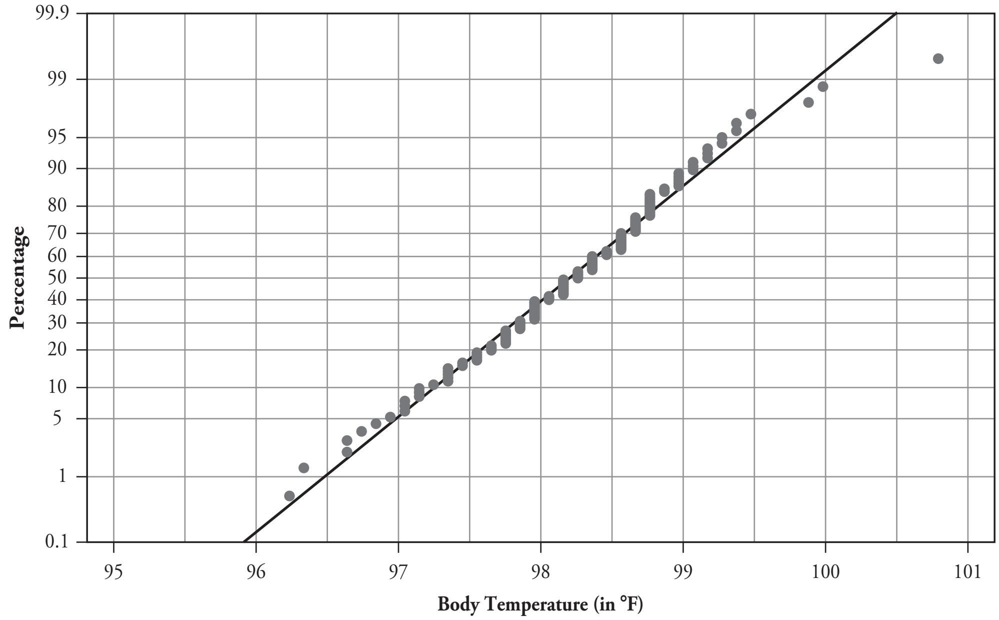
b. You are 95% confident the average body temperature of a sample of 130 healthy adults is between 98.12°F and 98.38°F.

When you say 95% confident, you mean that if you repeated this procedure of creating confidence intervals in this same manner (using random samples of 130 healthy adults), in the long run 95% of the intervals would contain the population mean body temperature of all healthy adults and 5% of the intervals would not contain this parameter.

c. Here are graphs of the sample data:







Yes, the body temperatures appear to be roughly normally distributed.

d. No, the normality of the population of body temperatures is not required for this *t*-procedure to be valid with these data because the sample size is large ($n = 130$).

e. You do not know whether this was a simple random sample of healthy adults. If you assume that it was, then the other technical condition required for the validity of this *t*-interval is satisfied.

f. Using Minitab's One-Sample T procedure,

```
Variable       N      Mean    StDev   SE Mean         95% CI
body temp    130   98.2492   0.7332    0.0643   (98.1220, 98.3765)
```

90% CI: (98.1427, 98.3558) 99% CI: (98.0811, 98.4174)

g. The midpoints of all three intervals are the same: 98.249 (the sample mean). The 90% confidence interval is the most narrow (width = .2131), followed by the 95% confidence interval (width = .2545), whereas the 99% confidence interval is the widest (width = .3363).

h. It does not appear that 98.6 is a plausible value for the mean body temperature for the population of all healthy adults because this value is not contained in any of the confidence intervals.

i. If the sample size had been only 13, but the sample mean and standard deviation had been the same, the 95% CI would be much wider (though it would have the same center) because (i) the t^* value used to create the interval would be much greater and (ii) the standard error would be greater (the square root of 13 is much smaller than the square root of 130).

j. For a 95% confidence interval with 12 degrees of freedom, you calculate $98.249 \pm (2.179)\,(.733)/\sqrt{13} = (97.807, 98.692)$°F.

As predicted, the midpoint is still 98.249, but the width is much greater (.88597). In fact, with this interval, 98.6 would be a plausible value for the mean body temperature for the population of all healthy adults.

19

Activity 19-4: Sleeping Times

a. Here is the completed table:

Sample Number	Sample Size	Sample Mean	Sample SD
3	30	6.6	0.825
1	10	6.6	0.825
2	10	6.6	1.597
4	30	6.6	1.597

b. They all have a sample mean of 6.6 hours.

c. The most important difference between samples 1 and 2 are the spreads (sample SDs). Sample 1 has a much smaller SD than does sample 2.

d. The most important difference between samples 1 and 3 is the sample size. Sample 3 uses a sample of size 30, whereas sample 1 uses a sample of size 10.

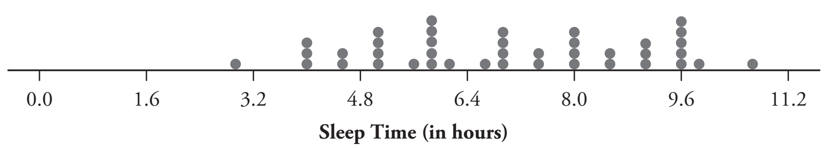
e. Sample 1 produced a more precise estimate of μ. This result makes sense because sample 1 has a smaller SD than sample 2, and so sample 1 will have a smaller margin-of-error.

f. Sample 3 produced a more precise estimate of μ. This result makes sense because sample 3 has a larger sample size, and so sample 3 will have a smaller margin-of-error.

Activity 19-5: Sleeping Times

Results will vary by class, but here are the results from one college class.

a. The following graph displays the results:



These 40 sleep times are roughly normally distributed, centered around 7 hours, with a minimum of 3 hours and a maximum of 10.5 hours. There are no noticeable outliers.

b. The sample size is $n = 40$; the sample mean is $\bar{x} = 6.981$ hours; and the sample standard deviation is $s = 1.981$ hours.

c. The sample size is large ($40 > 30$), but the sample was not randomly selected, because it consisted of the students in this one class. It might not be representative of students at the entire school with regard to sleep hours, as students in a statistics class may tend to be mostly from one type of major who may tend to study and sleep more or less than the typical student.

d. For a 90% CI, with 39 degrees of freedom, you calculate $6.981 \pm (1.685)(.31322) = (6.45322, 7.50878)$ hours.

You are 90% confident the average amount of sleep per night obtained by all students at the school is between 6.45 and 7.51 hours.

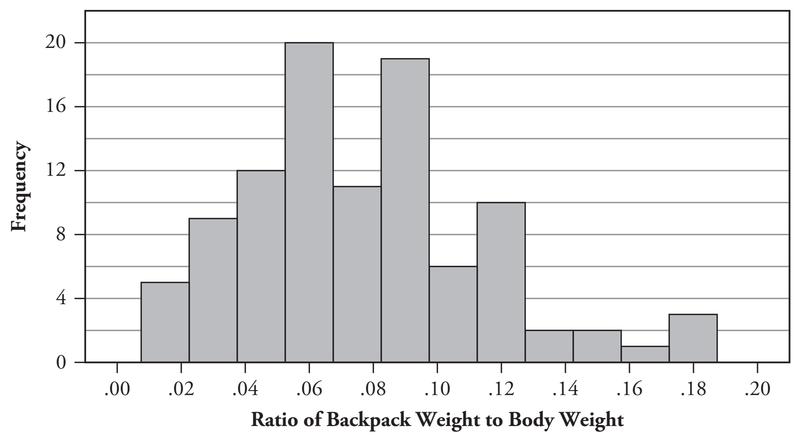
e. Seven of the 40 sleep times fall within this interval. This is 17.5%.

f. No, this percentage is not close to 90%, but there is no reason that it should be. Your interval is designed to estimate the *average* sleep time. It is not telling you anything about the individual sleep times. They may or may not fall within this interval.

Activity 19-6: Backpack Weights

a. The observational units are the students. The variable is the *ratio of backpack weight to body weight,* which is quantitative. The sample is the 100 Cal Poly students whose weights were recorded by the student researchers. The population is all Cal Poly students at the time the study was conducted.

b. The following histogram reveals that the distribution of these weight ratios is a bit skewed to the right. The center is around .07 or .08 (mean $\bar{x} = .077$, median = .071). The five-number summary is (.016, .050, .071, .096, .181), so students in the sample carried as little as 1.6% of their weight in their backpacks and as much as 18.1% of their weight in their backpacks. The standard deviation of these ratios is $s = .037$.



c. Calculating a 99% CI for the population mean by hand using the formula

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

with $100 - 1 = 99$ rounding down to 80 degrees of freedom gives $.0771 \pm 2.639 \times .0366/\sqrt{100}$, which is $.0771 \pm .0097$, which corresponds to the interval from .0674 through .0868. (This t^* value should be based on 99 degrees of freedom, but we used 80 degrees of freedom here, the closest value less than 99 that appears in Table III.) Using technology gives a slightly more accurate 99% CI for μ of .0675 through .0867.

d. You are 99% confident that the mean weight ratio of backpack-to-body weights among all Cal Poly students at the time of this study is between .0674 and .0868. In other words, you are 99% confident that the average Cal Poly student carries between 6.74% and 8.68% of his/her body weight in his/her backpack. By "99% confidence," you mean that 99% of all intervals constructed with this method would succeed in capturing the actual value of the population mean weight ratio.

e. The first condition is that the sample be randomly selected from the population. This is not literally true in this case because the student researchers did not obtain a list of all students at the university and select randomly from that list, but they did try to obtain a representative sample. The second condition is either that the population of weight ratios is normal or that the sample size is large. In this case, the sample size is large ($n = 100$, which is greater than 30), so this condition is satisfied even though the distribution of ratios in the sample is somewhat skewed (and so presumably is the population).

f. You do not expect 99% of the sample, nor 99% of the population, to have a weight ratio between .0673 and .0869. You are 99% confident that the population *mean* weight ratio is between these two endpoints. In fact, only 18 of the 100 students in the sample have a weight ratio in this interval.

Solutions

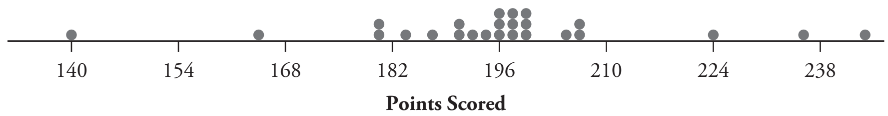
In-Class Activities

Activity 20-1: Basketball Scoring

a. If the rule changes had no effect on scoring, μ would have the value 183.2. This is the null hypothesis.

b. If the rule changes had the desired effect on scoring, then $\mu > 183.2$, which would be the alternative hypothesis.

c. The following graph displays the NBA game data:



Scoring definitely seems to have increased over the previous season's mean of 183.2 points per game. In only 4 of these 25 games were fewer than 183 points scored, and the center of this dotplot is now about 196 points per game.

d. Sample mean $\bar{x}$: 195.88 points Sample standard deviation *s:* 20.27 points

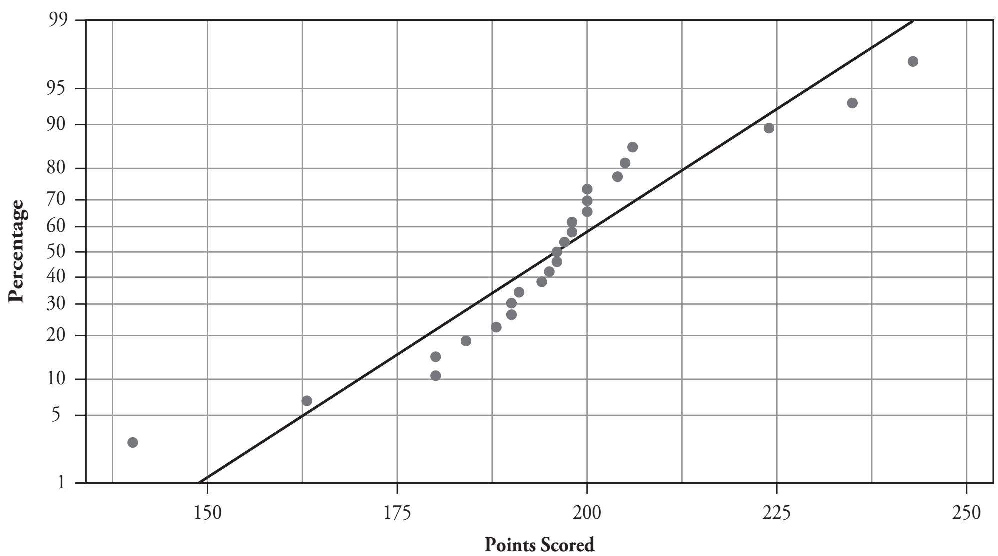
e. Yes, the sample mean is in the direction specified in the alternative hypothesis (greater than 183.20 points).

f. Yes, it is *possible* to have gotten such a large sample mean even if the new rules had no effect on scoring.

g. H_0: $\mu = 183.2$

H_a: $\mu > 183.2$

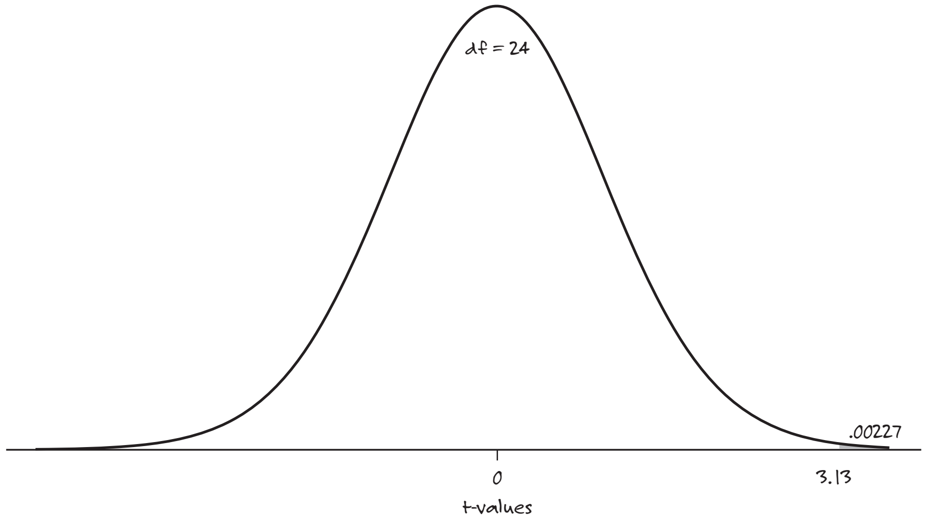
h. Technical conditions: The sample size is not large ($n = 25 < 30$), so the population must follow a normal distribution. A probability plot of the data provides evidence that these sample data are *not* arising from a normal population. Still, the data are reasonably symmetric and the sample size is moderately large, so this condition could be considered met with caution.



However, the data are not a simple random sample gathered from the population as the data are all the NBA games played from December 10–12, 1999. These games occurred relatively early in the season and are probably not representative of scoring over the course of the entire season, especially with respect to a new rule change. So the technical conditions for the validity of this *t*-test have not been met.

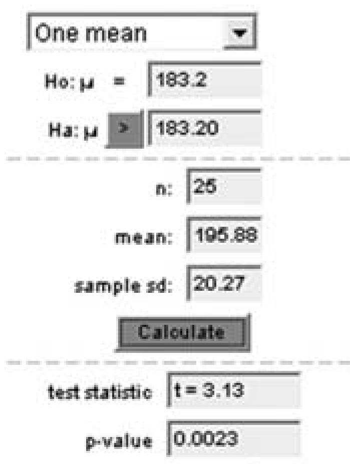
i. The test statistic is $t = \dfrac{195.88 - 183.2}{20.27/\sqrt{25}} = 3.13$.

j. Here is a sketch of the *t*-distribution:



k. You find $2.797 < 3.13 < 3.467$, so $.005 > p\text{-value} > .001$.

l. Using the applet, the *p*-value is .0023. The picture shows the information that should be entered into the applet:



m. If the average number of points per game for all NBA games in this season were still 183.2 points, there is only a .0023 chance that you would find a random sample of 25 games with a mean of at least 195.88 points. Because finding a sample as extreme as this one is so unlikely by chance alone, you conclude that the mean number of points per game this season has increased; it is no longer 183.2 points.

n. Yes, you would reject the null hypothesis at the .10, .05, .01, and .005 levels because the *p*-value is less than each of these significance levels.

o. If these data had been a random sample from a normal population, you would have very strong statistical evidence that the mean points per game in the 1999–2000 season were greater than in the previous season. However, you would not be able to conclude that the rule change caused the average point increase because this is not a randomized experiment.

Activity 20-2: Sleeping Times

Answers will vary by class. The following is one representative set of answers.

a. Let μ represent the mean sleep time of all students at your school.

The null hypothesis is that the mean sleep time of the population is 7 hours. In symbols, the null hypothesis is H_0: $\mu = 7.0$ hours.

The alternative hypothesis is that the mean sleep time of the population is not 7 hours. In symbols, the alternative hypothesis is H_a: $\mu \neq 7.0$ hours.

The test statistic is $t = \dfrac{6.981 - 7}{1.981/\sqrt{40}} = -0.06$.

Using Table III with 39 degrees of freedom, p-value $> 2 \times .20 = .40$.

Using Minitab, p-value $= 2 \times .476231 = .952462$.

Because the p-value is not small, do not reject H_0.

You do not have any statistical evidence to suggest that the mean sleep time of all students at your school differs from 7.0 hours.

b. Technical conditions: The sample size is large ($40 > 30$), but the sample was not randomly selected because it consisted of only the students in your class. It might not be representative of students at your school with regard to sleep hours, as students in a statistics class may tend to have similar majors and may tend to study and sleep more or less than the typical student. Here is the completed table:

Sample Number	Sample Size	Sample Mean	Sample SD	Test Statistic	*p*-value
1	10	6.6	0.825	−1.53	.1596
2	10	6.6	1.597	−0.79	.4487
3	30	6.6	0.825	−2.66	.0127
4	30	6.6	1.597	−1.37	.1806

c. Sample 1 would produce a smaller *p*-value because it has the smaller standard deviation.

d. Sample 3 would produce a smaller *p*-value because it has the larger sample size.

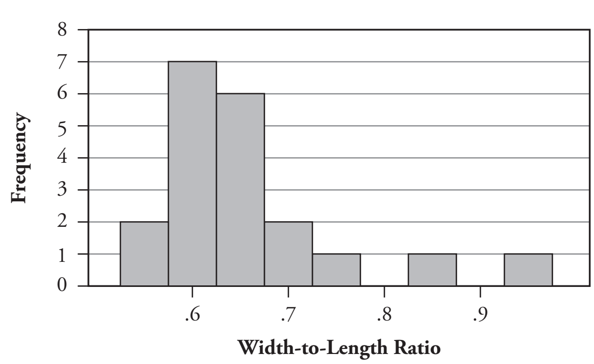
e. See the table following part b.

f. Only sample 3 gives enough evidence to reject the null hypothesis at the .05 level.

g. Answers will vary by student conjecture.

Activity 20-3: Golden Ratio

a. The following histogram displays data for width-to-length ratios for a sample of 20 beaded rectangles:



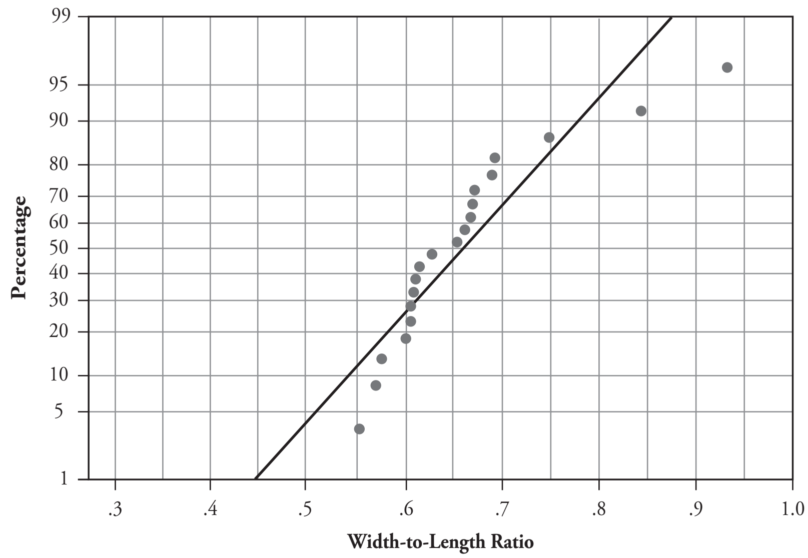
These twenty width-to-length ratios are skewed right, with a minimum of .553 and a maximum of .933. The median is .6410, the mean is .6605, and the standard deviation is .0925.

b. Let μ represent the average width-to-length ratio for all beaded rectangles made by the Shoshoni Indians.

The null hypothesis is that this average ratio is the golden ratio (.618). In symbols, the null hypothesis is H_0: $\mu = .618$.

The alternative hypothesis is that this average ratio is not the golden ratio. In symbols, the alternative hypothesis is H_a: $\mu \neq .618$.

Technical conditions: You don't know whether the sample was randomly selected, but it is small ($n = 20$), and the sample (and therefore the population) does not appear to be normally distributed. So the technical conditions for this procedure to be valid are not satisfied.



The test statistic is $t = \dfrac{.6695 - .618}{.0925/\sqrt{20}} = 2.05$.

The p-value is $2 \times \Pr(T > 2.05)$.

Using Table III with 19 degrees of freedom, $1.729 < 2.05 < 2.093$, so $2 \times .025 < p\text{-value} < 2 \times .05$ or $.05 < p\text{-value} < .10$. You would reject H_0 at the .10 significance level.

Using the applet, the p-value $= .0539 < .10$. You would reject H_0 at the .10 significance level.

You can conclude that the average width-to-length ratio for all beaded rectangles made by the Shoshoni Indians is not the golden ratio.

20

Activity 20-4: Children's Television Viewing

The observational units are third- and fourth-grade students. The sample consists of the 198 students at two schools in San Jose. The population could be considered all American third- and fourth-graders, but it might be more reasonable to restrict the population to be all third- and fourth-graders in the San Jose area at the time the study was conducted.

The variable measured here is the *amount of television the student watches in a typical week,* which is quantitative. The parameter is the mean number of hours of television watched per week among the population of all third- and fourth-graders. This population mean is denoted by μ. The question asked about watching an average of two hours of television per day, so convert that to be 14 hours per week.

The null hypothesis is that third-and fourth-graders in the population watch an average of 14 hours of television per week (H_0: $\mu = 14$). The alternative hypothesis is that these children watch more than 14 hours of television per week on average (H_a: $\mu > 14$).

Check technical conditions:

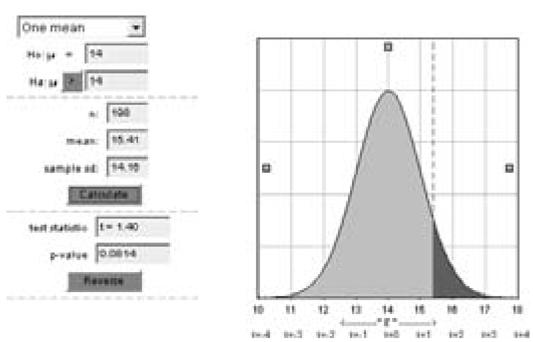
- The sample of children was not chosen randomly; they all came from two schools in San Jose. You might still consider these children to be representative of third- and fourth-graders in San Jose, but you might not be willing to generalize to a broader population.

- The sample size is large enough (198 is far greater than 30) that the second condition holds regardless of whether the data on television watching follow a normal distribution. You do not have access to the child-by-child data in this case, so you cannot examine graphical displays; however, the large sample size assures you that this condition may be considered satisfied.

Test statistic: The sample size is $n = 198$; the sample mean is $\bar{x} = 15.41$ hours; and the sample standard deviation is $s = 14.16$ hours. The test statistic is

$$t = \frac{15.41 - 14}{14.16/\sqrt{198}} \approx 1.401$$

indicating the observed sample mean lies 1.401 standard errors above the conjectured value for the population mean. Using Table III and the 100 degrees of freedom line (rounded down from the actual number of degrees of freedom of $198 - 1 = 197$) reveals the *p*-value (probability to the right of $t = 1.401$) to be between .05 and .10. Technology calculates the *p*-value more exactly to be .081.



Test decision: This *p*-value is not less than the .05 significance level. The sample data, therefore, do not provide sufficient evidence to conclude that the population mean is greater than 14 hours of television watching per week.

Conclusion in context: This conclusion stems from realizing that obtaining a sample mean of 15.41 hours or greater would not be terribly uncommon when the population mean is really 14 hours per week. If you had used a greater significance level (such as .10), which requires less compelling evidence in order to reject a hypothesis, then you would have concluded that the population mean exceeds 14 hours per week.

21

Solutions

In-Class Activities

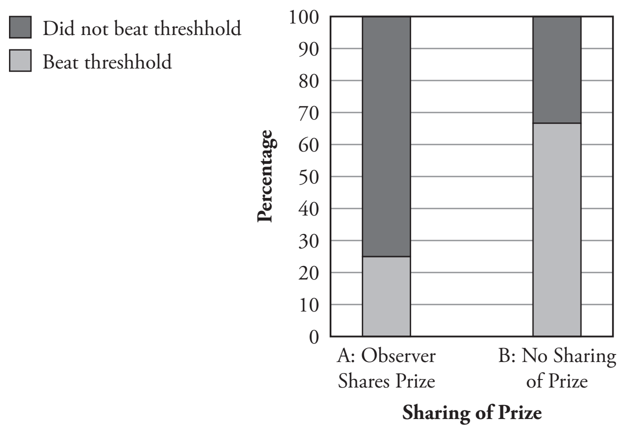
Activity 21-1: Friendly Observers

a. Explanatory: *whether the observer shares the prize* Type: binary categorical

Response: *whether the participants beat the threshold time* Type: binary categorical

b. This is an experiment because the researchers randomly assigned the subjects to the control and treatment groups.

c. The sample proportions of success for each group are $\hat{p}_A = .25$; $\hat{p}_B = .6667$. The following segmented bar graph displays the results:



d. Yes, these sample proportions differ in the direction conjectured by the researchers. The proportion in group A is substantially smaller than in group B.

e. Yes; even if there were no effect due to the observer's incentive, it is possible to have obtained such a big difference between the two groups simply from the random assignment process.

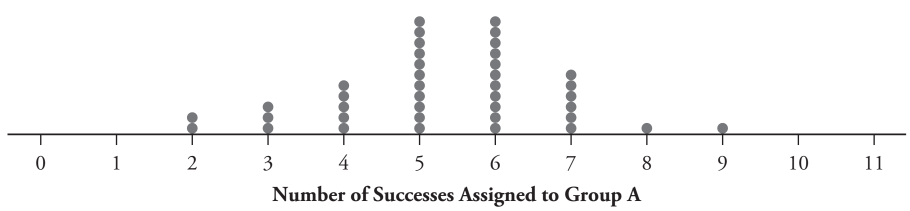
For parts f–l, answers will vary by student. The following are one representative set of answers.

f. Five of the 12 cards are successes. No, this result is not at least as extreme as the result in the actual study.

g. Here is a completed table:

Repetition #	1	2	3	4	5
Number of "Successes" Assigned to Group A	5	6	4	8	2
Is Result as Extreme as in Actual Study?	No	No	No	No	Yes

h. Here is a dotplot of the number of successes randomly assigned to group A:



i. The most common values for the number of successes are 5 and 6. This result makes sense because if the cards are randomly assigned to groups A and B, then half of the 11 successes should end up in group A. So, typically you expect 5 or 6 of the successes in group A.

j. Forty repetitions were performed by this class. Five of these repetitions gave a result of 3 or fewer successes in group A. This represents a proportion of .125.

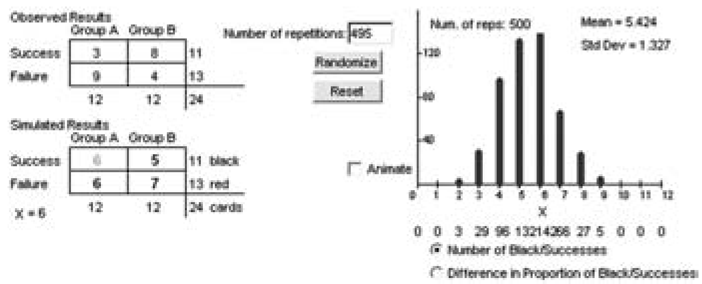
k. Based on this class's simulated results, it does not appear unlikely for random assignment to produce a result as extreme as (or more extreme than) the actual sample when the observer has no effect on subjects' performance.

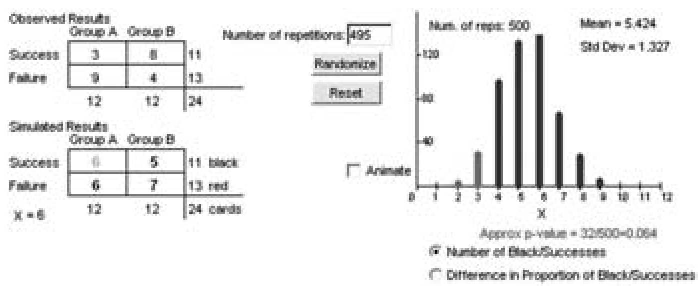
l. No, in light of your answer to the previous question, the data do not provide reasonably strong evidence in support of the researchers' conjecture; this type of sample result appears to happen in about 12–13% of random assignments when the observer has no effect on subjects' performance.

m. Yes, the number of successes randomly assigned to group A varies. The five values that occur in this running of the applet are 6, 5, 5, 8, and 7.

n. In this simulation, the resulting distribution is approximately normal, centered at about 5.4 successes, with a standard deviation of 1.327 successes.

o. Here are the simulation results:





The approximate *p*-value is 32/500, or .064.

p. Based on this approximate *p*-value from the larger simulation, it does appear somewhat unlikely for random assignment to produce a result at least as extreme as the actual sample when the observer has no effect on subjects' performance. The data provide moderate evidence in support of the researchers' conjecture because this type of sample result appears to happen in no more than about 6% of random assignments when the observer has no effect on subjects' performance.

Activity 21-2: Back to Sleep

a. This study involves random sampling from populations.

b. The null hypothesis is the proportion of all infants who sleep on their stomachs is the same in 1996 as it was in 1992. In symbols, the null hypothesis is H_0: $\pi_{1996} = \pi_{1992}$ or H_0: $\pi_{1996} - \pi_{1992} = 0$.

The alternative hypothesis is the proportion of all infants who sleep on their stomachs in 1996 is less than it was in 1992. In symbols, the alternative hypothesis is H_a: $\pi_{1996} < \pi_{1992}$ or H_a: $\pi_{1996} - \pi_{1992} < 0$.

c. The combined sample proportion of infants who slept on their stomachs is

$$\hat{p}_c = \frac{700 + 240}{2000} = .47$$

d. The test statistic is $z = \dfrac{.24 - .70}{\sqrt{(.47)(.53)\left(\frac{1}{1000} + \frac{1}{1000}\right)}} = -20.61.$

Yes, this is a very large *z*-score (in absolute value).

e. p-value $= \Pr(Z < -20.61) \approx .000$

f. If the proportion of infants who sleep on their stomachs is the same in 1996 as it was in 1992, the probability that you would find a difference in sample results as or more extreme than this study by random sampling alone is essentially 0, which means it would just about never happen. Because you did find this sample difference, you have very strong evidence that the null hypothesis was an incorrect conjecture, and the proportion of infants who sleep on their stomachs in 1996 was less than it was in 1992.

g. Reject the null hypothesis at the $\alpha = .01$ significance level because the p-value is smaller than .01.

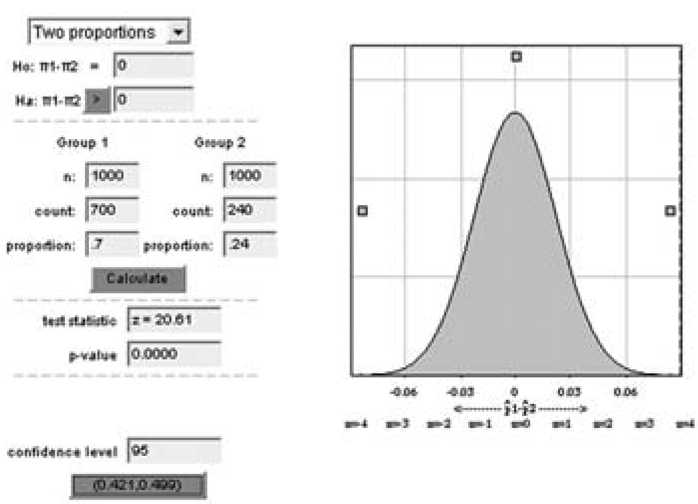
h. This sample difference provides extremely strong evidence that the proportion of all households that place infants to sleep on their stomachs decreased between 1992 and 1996.

i. For a 95% CI, you calculate $(.70 - .24) \pm (1.96)\sqrt{\frac{(.70)(.30)}{1000} + \frac{(.24)(.76)}{1000}}$,

which is $.46 \pm (1.96)\ (.0198)$, which is $.46 + .0388 = (.4212, .4988)$.

j. You are 95% confident the difference in population proportions is between .420 and .499. Because this interval does not contain 0, you are 95% confident the population proportions are not the same. Because the values in the interval are strictly positive, you are 95% confident that the percentage of infants who sleep on their stomachs has *decreased* (between 42 and 50 percentage points) in the years from 1992 to 1996.

k. Here are the applet results:



l. Answers will vary by student expectation.

m. For a 95% CI, you calculate $(.24 - .70) \pm (1.96)\sqrt{\frac{(.70)(.30)}{1000} + \frac{(.24)(.76)}{1000}} =$ $(-.4988, -.4212)$. The endpoints are reversed and multiplied by -1 in this interval. The interval has the same width as the previous interval, however, and its midpoint is the negative of that of the previous one.

Activity 21-3: Preventing Breast Cancer

a. This is an experiment because the women were randomly assigned to receive one of the treatments.

b. Explanatory: *which drug (tamoxifen or raloxifene) the woman received*

Response: *whether the woman developed invasive breast cancer*

c. H_0: $\pi_T = \pi_R$ (drugs are equally effective)

H_a: $\pi_T \neq \pi_R$ (there is a difference in the breast cancer rate between the two drugs)

d. Here is the 2 × 2 table:

	Taxmoxifen	Raloxifene	Total
Developed Invasive Breast Cancer	163	167	330
Did Not Develop Breast Cancer	9,563	9,578	19,141
Total	9,726	9.745	19,471

e. The sample proportions are $\hat{p}_T = .01676$ and $\hat{p}_R = .01714$.

f. The test statistic is $z = \dfrac{.01676 - .01714}{\sqrt{(.01695(1 - .01695)\left(\frac{1}{9726} + \frac{1}{9745}\right)}} = -0.20.$

g. No, this is not a large z-score. This is not surprising in light of the values calculated in part e, because the two proportions calculated for the tamoxifen and raloxifene groups are quite similar.

h. Using Table II, p-value $= 2 \times \Pr(Z < -0.20) = 2 \times .4207 = .8414$. Using Minitab, the p-value is .838.

i. Because the p-value is not small, do not reject H_0. You do not have statistical evidence that there is any difference between these two drugs in terms of their effectiveness in preventing breast cancer for five years. If you had found a difference, you could safely conclude it was caused by a difference in the drugs because this was a well-designed experiment. Because the study did not select a random sample, you should be cautious in generalizing these results to a larger population. In particular, you should not extend any of your conclusions beyond postmenopausal women who are at increased risk for breast cancer.

Activity 21-4: Perceptions of Self-Attractiveness

a. You need to know how many men and women were surveyed.

b. Answers will vary, but one example is if only 100 men and 100 women were surveyed because this would mean that 71 women and 81 men were satisfied with their appearance, and this difference of 10 people would not seem to be significant (it could have arisen simply from random sampling variability).

c. Answers will vary, but one example is if 1000 men and 1000 women were surveyed. In this case, you would not expect much sampling variability by chance, and the observed difference would seem much more convincing of a difference in the populations.

d. Here is the completed table:

Sample Size (each group)	Satisfied Women	Satisfied Men	Test Statistic	*p*-value	Significant at: $\alpha = .10$?	$\alpha = .05$?	$\alpha = .01$?
100	71	81	−1.67	.095	Yes	No	No
200	142	162	−2.36	.018	Yes	Yes	No
500	355	405	−3.73	.000	Yes	Yes	Yes

e. The larger the sample size, the more statistically significant a difference in proportions will be. A large difference may not be significant with a small sample size, but even a small difference may be statistically significant when the sample size is large.

Activity 21-5: Graduate Admissions Discrimination

a. The null hypothesis is that the probability of admission to the graduate programs at the University of California at Berkeley among male applicants is the same as the probability of admission among female applicants. In symbols, H_0: $\pi_m = \pi_f$.

The alternative hypothesis is that the probability of admission to the graduate programs differs for male and female applicants. In symbols, H_a: $\pi_m \neq \pi_f$.

The sample proportions are $\hat{p}_m = .4457$ and $\hat{p}_f = .3046$.

The test statistic is $z = \dfrac{.4457 - .3046}{\sqrt{(.388)(1 - .388)\left(\frac{1}{2681} + \frac{1}{1835}\right)}} = 9.56.$

This test statistic is very large (the difference between the acceptance rates is more than 9 standard deviations), so the *p*-value is $2 \times \Pr(Z > 9.56)$, which is very close to 0.

Because the *p*-value is small, reject H_0 at any commonly used level of significance.

You have extremely strong statistical evidence that the probability of admission for males differs from the probability of admission for females, much more than can be accounted for by random chance alone.

b. No, this is not evidence of discrimination. This sample was not a randomized experiment, so there could be other explanations for the observed difference. Recall that this is an example of Simpson's paradox: The women tended to apply to the graduate programs with low acceptance rates, whereas the men tended to apply to the graduate programs with high acceptance rates. Within each individual program, women were accepted for admission at roughly the same or higher rates than the men were.

Activity 21-6: Nicotine Lozenge

a. The explanatory variable is the *type of lozenge* (nicotine or placebo). The response variable is *whether the smoker successfully abstains from smoking for the year.* Both variables are categorical and binary.

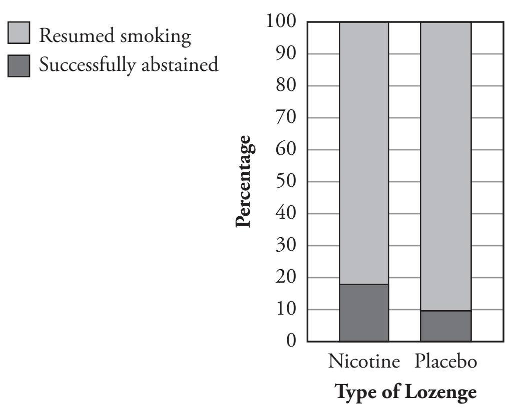
b. This is an experiment because the researchers *assigned* the subjects to take a particular kind of lozenge (nicotine or placebo).

c. The null hypothesis is that the nicotine lozenge is no more (or less) effective than the placebo, where effectiveness is measured by the proportion of smokers who successfully abstain from smoking for a year. The alternative hypothesis is that the nicotine lozenge is *more* effective than the placebo, meaning a higher proportion of smokers (who are interested in and might potentially use such a product) would successfully quit with a nicotine lozenge than with a placebo lozenge. In symbols, the hypotheses are H_0: $\pi_{\text{nicotine}} = \pi_{\text{placebo}}$ vs. H_a: $\pi_{\text{nicotine}} > \pi_{\text{placebo}}$, where π represents the population proportion of smokers who successfully abstain for a year if given the nicotine lozenge or the placebo.

d. The 2 × 2 table is shown here:

	Nicotine Lozenge	Placebo Lozenge	Total
Successfully Abstained	82	44	126
Resumed Smoking	377	414	791
Total	459	458	917

The following segmented bar graph shows those smokers taking the nicotine lozenge had a higher success rate (proportion) in this study than those smokers taking the placebo lozenge, almost twice as high (.179 vs. .096). But the graph also reveals that in both groups, many more smokers resumed smoking than were able to abstain successfully.



e. The sample proportions of smokers who successfully abstained from smoking in each group are

$$\hat{p}_{\text{nicotine}} = \frac{82}{459} \approx .179, \quad \hat{p}_{\text{placebo}} = \frac{44}{458} \approx .096$$

The combined sample proportion of smokers who abstained is

$$\hat{p}_c = \frac{82 + 44}{459 + 458} = \frac{126}{917} \approx .137$$

The test statistic is

$$z = \frac{.179 - .096}{\sqrt{(.137)(1 - .137)\left(\frac{1}{459} + \frac{1}{458}\right)}} \approx 3.65$$

The p-value is the area to the right of 3.65 under the standard normal curve, which Table II reveals to be .0001. This test is valid because random assignment was used to put subjects in groups and because the sample size condition is also met: 459(.137) = 62.88 is greater than 5, as are 458(.137) = 62.75, 459(1 −.137) = 396.12, and 458(1 − .137) = 395.25.

This very small p-value indicates the experimental data provide very strong evidence against the null hypothesis, which means very strong evidence that the population proportion of smokers who would successfully abstain from smoking is higher with the nicotine lozenge than with the placebo lozenge. Because this was a randomized experiment, you can conclude the nicotine lozenge *causes* an increase in the proportion of smokers who would successfully abstain as compared to the placebo group. You are not told how the subjects were selected for the study, but you do know that they were hoping to quit smoking, so this conclusion should be limited to smokers hoping to quit. They were probably chosen from a particular geographic area, so you might not want to generalize beyond that area, but it is hard to imagine why smokers in one area would respond differently to these lozenges than smokers in another area.

f. A 95% confidence interval for the difference $\pi_{\text{nicotine}} - \pi_{\text{placebo}}$ is

$$(.179 - .096) \pm 1.96\sqrt{\frac{(.179)(1 - .179)}{459} + \frac{(.096)(1 - .096)}{458}}$$

which is .083 ± 1.96(.023), which is .083 ± .044, which is the interval from .039 through .127. The researchers are 95% confident that the proportion of smokers using a nicotine lozenge who successfully quit would be higher than the successful proportion of smokers using a placebo lozenge by somewhere between .039 and .127. This interval procedure is valid because of random assignment and because the number of successes and failures in both groups exceeds 5 (the smallest of these numbers is 44 successes in the placebo group) with the sampling issues cautioned about in part e.

g. Notice this question calls for a confidence interval for a single proportion: π_{nicotine}, so you need to use the procedure from Topic 16:

$$\hat{p}_{\text{nicotine}} \pm z^* \sqrt{\frac{\hat{p}_{\text{nicotine}}(1 - \hat{p}_{\text{nicotine}})}{n_{\text{nicotine}}}}$$

which is .179 ± 1.96(.018), which is .179 ± .035, which is the interval from .144 through .214. The researchers can be 95% confident that if the population of all smokers who wanted to quit were to use the nicotine lozenge, the proportion who would successfully abstain from smoking for one year would be between .144 and .214. So, even though the experiment provides strong evidence that the nicotine lozenge works better than a placebo, most smokers (more than 3/4) would be unable to quit even with the nicotine lozenge.

Solutions

In-Class Activities

Activity 22-1: Close Friends

a. This is an observational study because the researcher simply observed the gender of each subject—he/she did not randomly assign gender to subjects.

b. Explanatory: *gender* Type: binary categorical

Response: *number of "close friends"* Type: quantitative

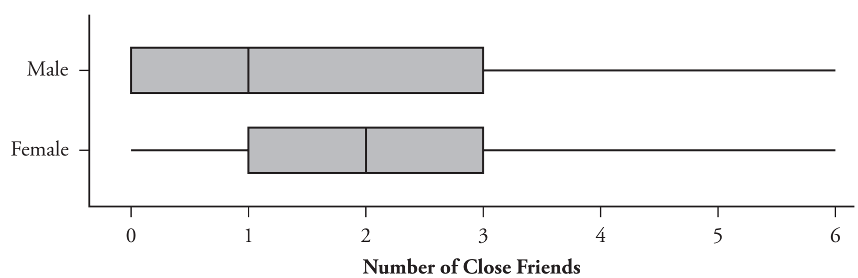
c. The null hypothesis is men and women tend to mention the same average number of close friends in response to this question. In symbols, H_0: $\mu_m = \mu_f$.

The alternative hypothesis is men and women differ in the average number of close friends they tend to mention in response to this question. In symbols, H_a: $\mu_m \neq \mu_f$.

d. The two-sample *z*-test procedure from Topic 21 does not apply in this situation because you are testing hypotheses about population means rather than population proportions. The one-sample *t*-test procedure from Topic 20 does not apply because you are comparing two means in this situation, not testing a claim about a single mean.

e. These values are statistics because they describe samples, not populations.

f. The following boxplots compare the distributions of the number of close friends:



g. There are very few differences in the distributions of the number of close friends mentioned between males and females. The males have a slightly lower mean, median, and lower quartile, but the upper quartiles and maximums are identical to those of females.

h. Yes, it would be possible to obtain sample means this far apart even if the population means were equal.

i. The test statistic is $t = \dfrac{1.861 - 2.089}{\sqrt{\dfrac{1.777^2}{654} + \dfrac{1.76^2}{813}}} = -2.45.$

j. Using 500 degrees of freedom, $-2.586 < -2.45 < -2.334$, so $2 \times .005 <$ *p*-value $< 2 \times .01$, which means $.01 <$ *p*-value $< .02$.

k. The correct interpretation of the *p*-value is "The *p*-value is the probability of getting sample data so extreme if, in fact, males and females have the same mean number of close friends in the populations."

l. Yes, this *p*-value is small enough to reject the null hypothesis at the $\alpha = .05$ significance level (*p*-value $< .02 < .05$).

m. No, the observed difference in sample means is not statistically significant at the $\alpha = .01$ significance level (p-value $> .01$).

n. Technical conditions:

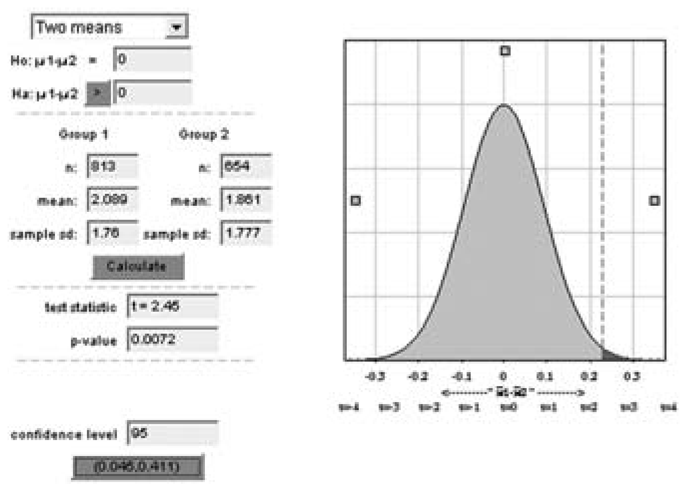
i. The data are a random sample broken into two distinct groups (see page 418).

ii. Both sample size are large (greater than 30).

Because both sample sizes are large, you do not need to worry about the strong skewness in the sample data. It does not provide any reason to doubt the validity of this test.

o. For a 95% CI for $\mu_f - \mu_m$ with 500 degrees of freedom, you calculate $(2.089 - 1.861) \pm (1.965)(0.0929) = (0.045, 0.411)$. This interval is entirely positive (and does not include zero), which means you can be 95% confident that the mean number of close friends that women have is between .045 and .411 greater than the mean number of close friends that men have.

p. Here are the applet results:



Activity 22-2: Hypothetical Commuting Times

a. No, one route does not *always* get Alex to school more quickly than the other.

b. Yes, the data suggest that Route 1 tends to get Alex to school more quickly than Route 2.

c. Here is the completed table:

	Sample Size	Sample Mean	Sample SD	*p*-value
Alex: Route 1	10	28	6	.155
Alex: Route 2	10	32	6	

d. $H_0: \mu_1 = \mu_2$ $\qquad H_a: \mu_1 \neq \mu_2$

The test statistic is $t = \dfrac{28 - 32}{\sqrt{\dfrac{6^2}{10} + \dfrac{6^2}{10}}} = -1.49.$

Using Table III with 9 degrees of freedom, $-1.833 < -1.49 < -1.383$, so $2 \times .05 < p$-value $< 2 \times .1$, which means $.1 < p$-value $< .2$. Using Minitab with 18 degrees of freedom, the p-value is .154. Using the applet (which uses 9 degrees of freedom), the p-value is .1702.

e. No, Alex's data are not statistically significant at any of the commonly used significance levels. Alex cannot reasonably conclude that one route is faster than the other route for getting to school because his p-value is not small and will not allow you to reject the null hypothesis that the average time required for both routes is the same.

f. Using Minitab, $(28 - 32) \pm t_{18}^*(2.683) = (-8.65, 0.65)$. Using the applet or Table III, t^* with 9 degrees of freedom is 1.833, and the resulting confidence interval is $(-8.92, 0.92)$.

g. Yes, this interval includes the value zero. This means that zero is a plausible value for the difference in the population means or that you cannot conclude there is a difference in the population mean travel times (with 90% confidence).

h. The most important difference between Alex's and Barb's results appears to be their mean travel times. The difference between the mean travel times on Barb's two routes appears to be significantly greater than that of Alex.

i. Carl's travel times appear to have a much smaller standard deviation than Alex's.

j. Donna appears to have used a much larger sample size than Alex did.

k. Here is the completed table:

	Sample Size	Sample Mean	Sample SD	*p*-value
Barb: Route 1	10	25	6	.002 (Minitab) .0047 (applet)
Barb: Route 2	10	35	6	
Carl: Route 1	10	28	3	.008 (Minitab) .0154 (applet)
Carl: Route 2	10	32	3	
Donna: Route 1	40	28	6	.004 (Minitab) .0049 (applet)
Donna: Route 2	40	32	6	

l. Barb: The difference in the sample means $(35 - 25)$ is large.

Carl: The sample standard deviations (3) are small.

Donna: The sample sizes are large (40).

Activity 22-3: Memorizing Letters

a. The observational units are the students in your class.

b. Explanatory: *grouping of letters* Type: binary categorical

Response: *number of letters correctly memorized* Type: quantitative

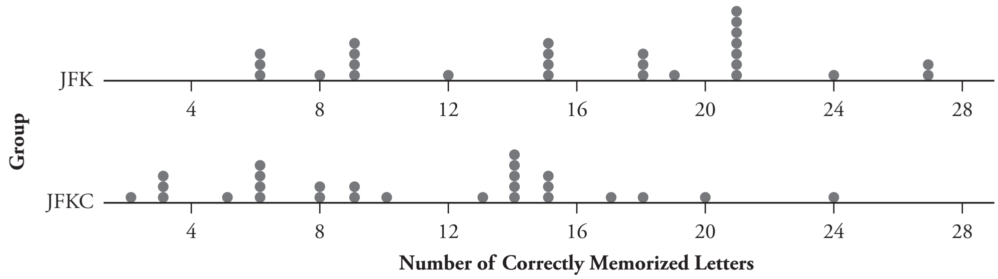
c. This study involves random assignment to two experimental groups. The students were randomly assigned to the convenient three-letter chunk group or to the inconvenient grouping group.

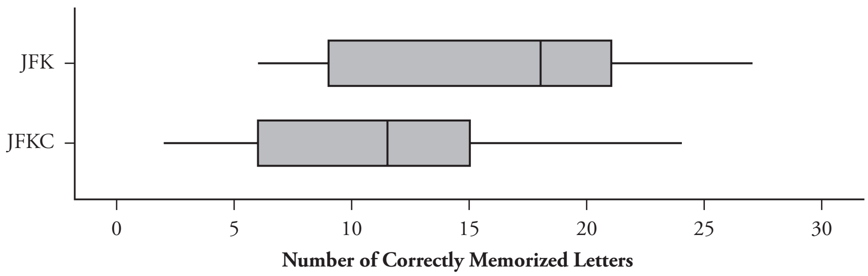
d. The null hypothesis is that the mean number of letters correctly memorized under both conditions is the same. In symbols, H_0: $\mu_{JFK} = \mu_{JFKC}$.

The alternative hypothesis is that mean number of letters correctly memorized under the three-letter grouping condition is greater than the mean number of letters correctly memorized under the inconvenient grouping condition. In symbols, H_a: $\mu_{JFK} > \mu_{JFKC}$.

Answers will vary by class. The following is one representative set of answers.

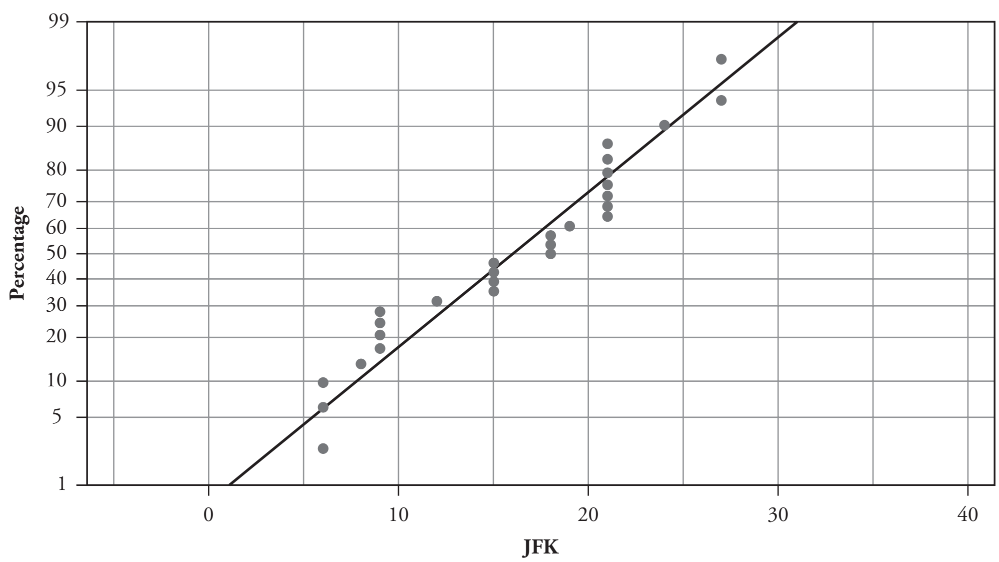
e. The following visual displays display the data:

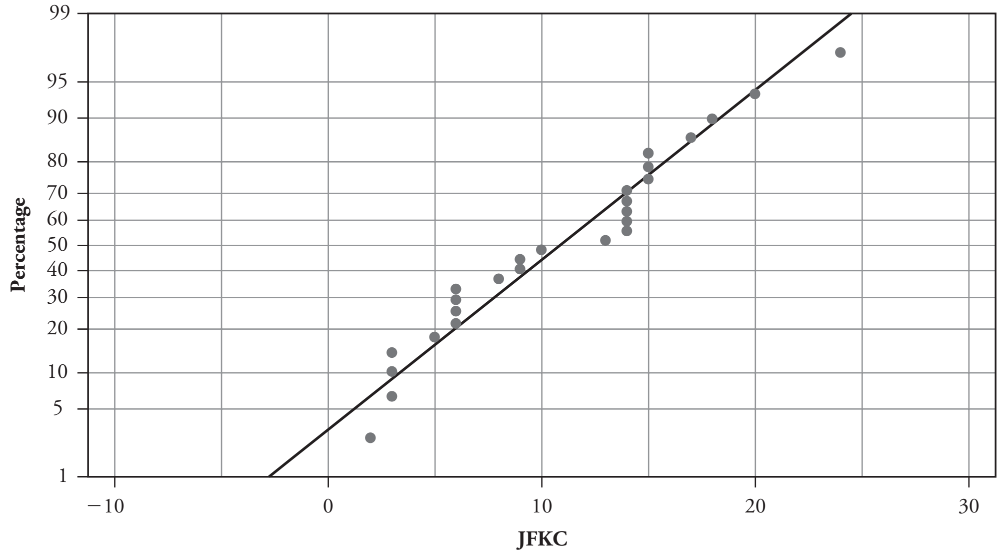




Yes, these data appear to support the conjecture that those who receive the letters in convenient three-letter chunks tend to correctly memorize more letters. The center of this plot is about six letters higher than for the JFKC plot, whereas the spreads are not very different.

f. The data arise from random assignment of subjects to two treatment groups, but the sample sizes are not quite large enough ($n = 27$ and 26, respectively). Probability plots indicate that the JFKC (inconvenient chunks) distribution is likely to be normal, but the JFK (three-letter chunk) distribution is skewed left and not normal.





Therefore, for these data, the results of the test of significance should be interpreted with caution.

g. The test statistic is $t = \dfrac{16 - 10.88}{\sqrt{\dfrac{6.45^2}{27} + \dfrac{5.86^2}{26}}} = 3.03.$

Using Table III with 25 degrees of freedom, $2.787 < 3.03 < 3.450$, so $.001 < p\text{-value} < .005$. Using Minitab, the p-value is .002. Using the applet, the p-value is .0028.

With a p-value less than .05, reject H_0. You have strong statistical evidence that the average number of letters correctly memorized is higher under the convenient three-letter condition than under the inconvenient condition. If there were no difference in the average number of letters correctly memorized under these two conditions, you would see a result as extreme or more extreme than in this study by chance alone in only about .2% of random assignments.

h. The confidence interval formula is: $(16 - 10.88) \pm t^*(1.69)$. Using Table III with 25 degrees of freedom, $t^* = 1.708$, so the confidence interval is (2.23, 8.01). Using Minitab, the confidence interval is (2.28, 7.95). Using the applet, the interval is (2.23, 8.01). You are 90% confident that the convenient three-letter grouping helps individuals memorize between 2.3 and 8.0 more letters, on average, than the inconvenient grouping.

i. Because this is a well-designed experiment in which the subjects were randomly assigned to the two groups, you can conclude that the convenient three-letter grouping did cause higher memory scores on average.

Activity 22-4: Got a Tip?

a. This is a randomized experiment because the waitress randomly assigned the two-person parties either to be told her name or not as part of her greeting. You do have to consider that the waitress was not blind to the treatment condition and may have subconsciously provided better service for the customers she expected to give her a larger tip.

b. The explanatory variable is *whether the waitress included her name as part of her greeting;* this variable is categorical and binary. The response variable is the *amount of the tip,* which is quantitative.

c. The null hypothesis is that there is no effect of the waitress using her name in her greeting. In other words, the null hypothesis says that the population mean tip amount would be the same when she uses her name as when she does not. The alternative hypothesis is that using her name has a positive effect, that the population mean tip amount would be greater when she uses her name than when she does not. In symbols: H_0: $\mu_{\text{name}} = \mu_{\text{no name}}$ vs. H_a: $\mu_{\text{name}} > \mu_{\text{no name}}$.

d. A Type I error would mean that the waitress decides that using her name helps when it really doesn't, so she would waste the minimal effort of giving her name and reap no benefit from it. A Type II error means the waitress decides using her name is not helpful even though it actually is helpful, so she would not bother to give customers her name and would lose out on that benefit. Because the cost of giving her name is minimal, losing out on potential tips is probably more of a concern.

e. The test statistic is $t = \dfrac{5.44 - 3.49}{\sqrt{\dfrac{(1.75)^2}{20} + \dfrac{(1.13)^2}{20}}} \approx \dfrac{1.95}{0.466} \approx 4.19.$

Using Table III with 19 degrees of freedom reveals that this test statistic is off the chart, so the p-value is less than .0005.

f. Because the p-value is less than .05, reject the null hypothesis at the $\alpha = .05$ level. Indeed, you would also reject the null hypothesis at the $\alpha = .01$ and even at the $\alpha = .001$ levels. The data provide very strong evidence that giving her name as part of her greeting does lead to higher tips on average.

g. You do not have enough information to check the technical conditions thoroughly. You do know that the parties were randomly assigned to one group or the other. But the sample sizes are not large, so you should check whether the tip data could reasonably have come from normal distributions; however, you only have the summary statistics, not the actual tip amounts from each party, so you cannot check this condition. You should ask the waitress to provide the actual party-by-party tip amounts to help you assess the shape of the distribution of tip amounts.

h. A 95% confidence interval for $\mu_{\text{name}} - \mu_{\text{no name}}$ is

$$(5.44 - 3.49) \pm 2.093\sqrt{\frac{(1.75)^2}{20} + \frac{(1.13)^2}{20}}$$

which is 1.95 ± 2.093(0.466), which is 1.95 ± 0.97, which is the interval (0.98, 2.92). You can be 95% confident that the waitress would earn between \$0.98 and \$2.92 more per party, on average, with a \$23.21 check, by giving her name as part of her greeting (assuming that the tip amounts are roughly normally distributed).

i. Conclude a causal link between the waitress giving her name and receiving higher tips on average. Random assignment would have assured that the only difference between the groups was whether the party was given the waitress's name. Because the group who was told her name gave significantly higher tips on average (p-value $< .0005$), you can attribute that to being told her name in greeting unless the waitress gave better service to customers to whom she gave her name. The confidence interval enables you to say more: that giving her name to customers increases the waitress' tips by an average of about 1–3 dollars per dining party at Sunday brunch in this restaurant. But you must be cautious about generalizing this result to other waitresses because only one waitress participated in this study. Even for this particular waitress, you should be cautious about generalizing the results to customers beyond those who partake of Sunday brunch at that particular Charlie Brown's restaurant in southern California. You should also remember that these *p*-value and confidences interval calculations are only valid if the tip amounts roughly follow a normal distribution.

Solutions

In-Class Activities

Activity 23-1: Marriage Ages

a. The null hypothesis is that the population mean marriage age for husbands is the same as the mean marriage age for their wives. In symbols, H_0: $\mu_{\text{husbands}} = \mu_{\text{wives}}$.

The alternative hypothesis is that the population mean marriage age for husbands is greater than the mean marriage age for their wives. In symbols, H_a: $\mu_{\text{husbands}} > \mu_{\text{wives}}$.

The test statistic is $t = \dfrac{35.71 - 33.83}{\sqrt{\dfrac{14.56^2}{24} + \dfrac{13.56^2}{24}}} = 0.46$.

Using Table III with 23 degrees of freedom, $0.46 < 0.858$, so the *p*-value is off the chart. This means the *p*-value $> .20$. Using Minitab, the *p*-value is .323. Using the applet with 23 degrees of freedom, the *p*-value is .3226.

Do not reject H_0 at the $\alpha = .05$ significance level ($.32 > .05$). You do not have statistical evidence that the population mean marriage age for husbands is greater than for wives.

b. For a 90% confidence interval: $(35.71 - 33.83) \pm t^*(4.06)$. Using Table III with 23 degrees of freedom, $t^* = 1.714$, so the confidence interval is $(-5.08, 8.84)$. Using Minitab, the interval is $(-4.95, 8.70)$. Using the applet, the interval is $(-5.08, 8.84)$.

c. Most of the values are above the $y = x$ line, which indicates a clear tendency for husbands to be older than their wives.

d. The husband's and wife's ages appear to be related. People tend to marry people in the same age group: younger people tend to marry younger people, and older people tend to marry older people.

e. For a 90% CI with 23 degrees of freedom, you calculate $1.875 \pm (1.714)\left(4.812/\sqrt{24}\right) = 1.875 \pm .98225 = (.1914, 3.5586)$. You are 90% confident the population mean difference in ages between husbands and their wives is between .19 and 3.6 years.

f. This interval is entirely positive, indicating that there is a difference in the population mean ages of husbands and their wives. The incorrect interval in part b contained both positive and negative values (and zero). The midpoint of this interval is 1.875 years and its width is 3.367 years. The midpoint of the previous (incorrect) interval was also 1.875 years, but the width was 13.65 years.

g. The null hypothesis is that the population mean difference in ages between husbands and their wives is zero. In symbols, H_0: $\mu_d = 0$. The alternative hypothesis is that the population mean difference in ages between husbands and their wives is greater than 0. In symbols, H_a: $\mu_d > 0$.

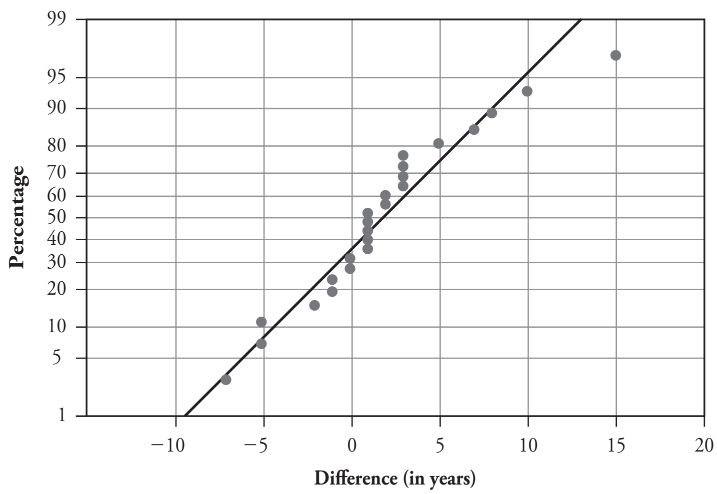
The test statistic is $t = \dfrac{1.875}{4.812/\sqrt{24}} = 1.91$.

Using Table III with 23 degrees of freedom, $1.714 < 1.91 < 2.069$, so $.025 < p\text{-value} < .05$. Using the applet with 23 degrees of freedom, the *p*-value is .0344.

Reject H_0 at the $\alpha = .05$ significance level ($.0344 < .05$). You have moderate statistical evidence the population mean difference in ages between husbands and their wives is greater than zero.

h. This is the opposite of the (incorrect) conclusion you drew in part a.

i. Technical conditions: This is not a simple random sample, but you have no reason to suspect that this sample is not representative of at least marriages in Cumberland County, Pennsylvania. The sample size of pairs is not large ($n = 24 < 30$), but a probability plot of the sample differences indicates that it is plausible the sample comes from a normally distributed population.



j. The paired analysis produces such a different conclusion from the independent-samples analysis because it reduces the variability so much. The standard deviation of the differences in ages is much smaller than the standard deviations of the ages. This makes the observed difference in the mean ages seem larger (less likely to happen by chance). You can also see that this reduces the denominator of the test statistic creating a larger value and lowering the *p*-value.

k. Yes, the researcher was wise to gather paired data. Because the husband's age and the wife's age were related, pairing was very helpful for estimating the population mean age difference among married couples.

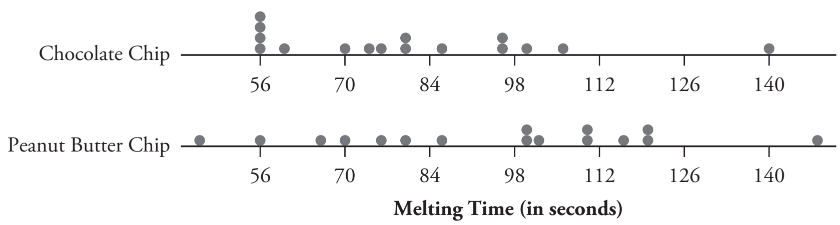
Activity 23-2: Chip Melting

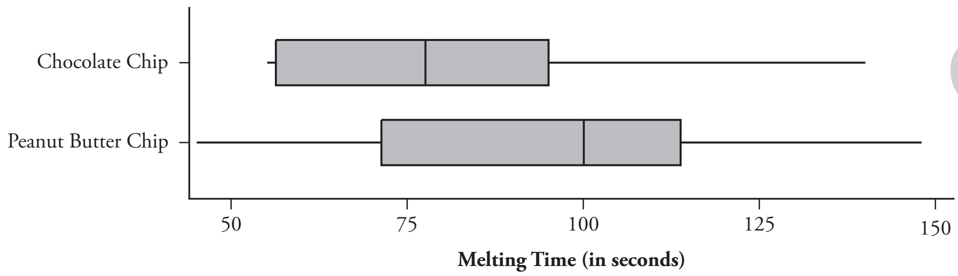
a. Explanatory: *type of chip* (chocolate or peanut butter) Type: binary categorical

Response: *time (in seconds) it takes the chip to melt* Type: quantitative

b. This is an experiment. The researchers are imposing the chips and randomly deciding the order of the chips.

c. Answers will vary by class. The following is one representative set:



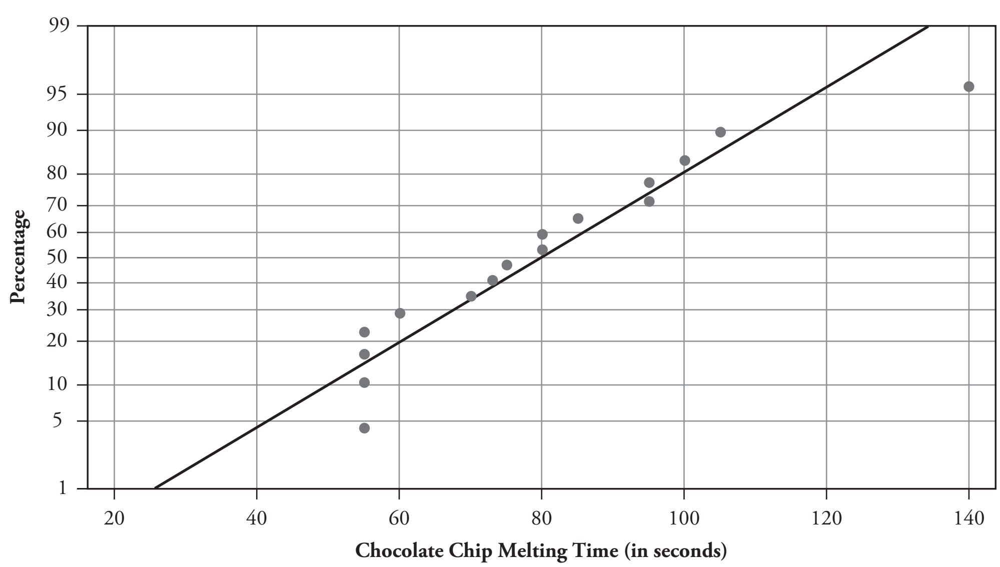


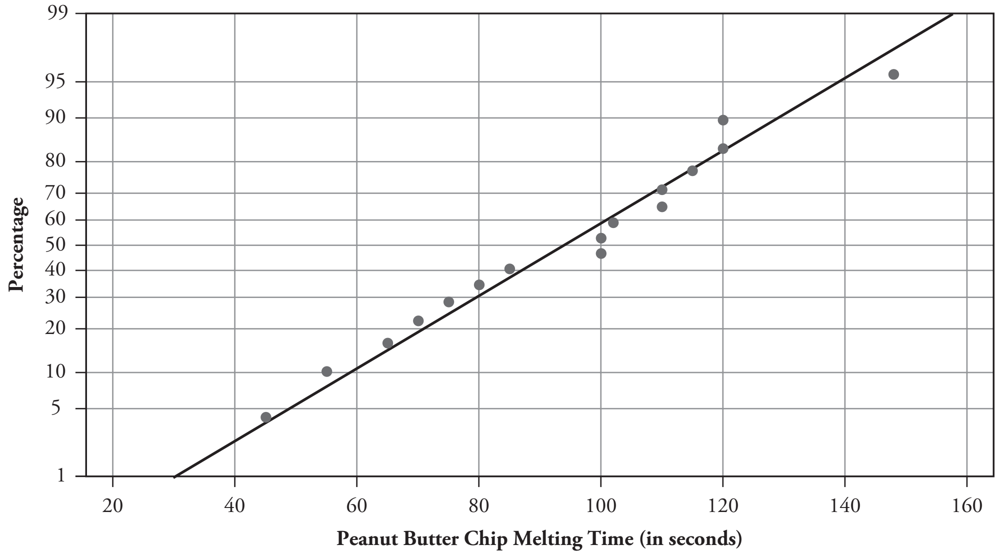
	n	$\bar{x}$	s	Min	Q_L	Median	Q_u	Max
Chocolate Chip	16	79.9	23.3	55.0	56.3	77.5	95.0	140
Peanut Butter Chip	16	93.8	27.5	45.0	71.3	100.0	113.8	148

It appears that for these students, peanut butter chips tended to take a little longer than chocolate chips to melt (mean 93.8 vs. 79.9 seconds; almost all five numbers in five number summary are larger).

d. The null hypothesis is that the population mean melting time for both chips is the same. In symbols, H_0: $\mu_{\text{chocolate}} = \mu_{\text{peanut butter}}$. The alternative hypothesis is that the population mean melting time for both chips is not the same. In symbols, H_a: $\mu_{\text{chocolate}} \neq \mu_{\text{peanut butter}}$.

Technical conditions: The data are from a completely randomized experiment. However, the sample sizes are not large (16 and 16), but probability plots do not provide strong evidence that the data come from nonnormal distributions (though there are some outliers with the chocolate chip melting times).





The test statistic is $t = \dfrac{79.9 - 93.8}{\sqrt{\dfrac{23.3^2}{16} + \dfrac{27.5^2}{16}}} = -1.54.$

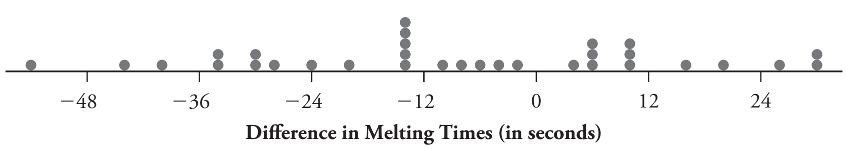
Using Table III with 15 degrees of freedom, $-1.341 < -1.54 < -1.753$, so $2 \times .05 < p$-value $< 2 \times .1$. This means $.1 < p$-value $< .2$. Using Minitab, the p-value is .135. Using the applet with 15 degrees of freedom, the p-value is .1438.

Because the p-value is not small, do not reject H_0 at any common significance level. You do not have statistical evidence that the population mean melting time differs between the two kinds of chips.

e. These data call for a paired t-test because the chips are paired-up—one chip of each type was given to each student, and each student tested the melting time of both chips. The observations in each row are clearly related—they are from the same individual!

f. Here is the numerical summary and graphical display:

	n	$\bar{x}$	s	Min.	Q_L	Median	Q_U	Max.
Differences (Chocolate – Peanut Butter)	32	−8.8	22.0	−55.0	−28.0	−9.0	8.8	30.0



The null hypothesis is that the population mean difference in melting times between chocolate and peanut butter chips is 0. In symbols, H_0: $\mu_d = 0$. The alternative hypothesis is that the population mean difference in melting times between chocolate and peanut butter chips is not 0. In symbols, H_a: $\mu_d \neq 0$.

The test statistic is $t = \dfrac{-8.8}{22.0/\sqrt{32}} = -2.27.$

Using Table III with 31 degrees of freedom, $-2.040 < -2.27 < -2.453$, so $2 \times .01 < p\text{-value} < 2 \times 0.025$. This means $.02 < p\text{-value} < .05$. Using Minitab, the *p*-value is .030. Using the applet with 31 degrees of freedom, the *p*-value is .0311.

Because the *p*-value .030 is less than .05, reject H_0 at the $\alpha = .05$ significance level.

A 90% confidence interval for the mean difference in the melting time is

$$-8.8 \pm (1.696)\left(22.0/\sqrt{32}\right) = (-15.4, -2.2)$$

Because the *p*-value of .03 is less than .05, you have some statistical evidence that there is a difference in the mean melting time between the two kinds of chips. You are 90% confident the melting time of the peanut butter chips is between 2.2 and 15.4 seconds longer than that of the chocolate chips, on average.

g. The conclusions for these two studies were quite different. In the first study, you had no statistical evidence of a difference in the chip melting time (*p*-value = .135), and in the paired data study, you had some evidence of a difference (*p*-value = .03). Pairing the data reduced the variability in the data, as the standard deviation of the differences was 22.0 seconds, whereas the standard deviations of the individual melting times were both more than 23 seconds.

Activity 23-3: Body Temperatures, Natural Selection, and Memorizing Letters

a. No, it would not be appropriate to analyze these data with a paired *t*-test because there is no link between the men and women in this study. You could change the order of the men or the women (or both) without creating a problem in your data.

b. It would be impossible to analyze these data with a paired *t*-test because the sample sizes for the two groups are not the same. You could not pair up each surviving sparrow with a dead sparrow because more sparrows survived than died.

c. No, it would not be appropriate to analyze these data on memorizing letters with a paired *t*-test because there is no link between the students who memorized the JFK chunks and the students who memorized the JFKC chunks. Randomization was used to assign subjects to the two separate groups, so you should use the two-sample *t*-procedures instead.

d. You could alter this experiment by having each student memorize the letters twice—once in the convenient three-letter (JFK) chunks and once in the inconvenient (JFKC) chunks. Randomization would still be important in this experiment because you would need to determine the order in which the student received/memorized the type of chunk randomly.

Activity 23-4: Alarming Wake-Up

a. The explanatory variable is the *sound used to wake the child*.

b. Use randomization to divide the children into two groups—one group to be awakened by recordings of their mothers' voices (personalized alarm) and the other group to be awakened by a conventional smoke alarm. Record the time it takes to wake each child. Use a two-sample *t*-test to determine whether there is a statistically significant difference in the average times of the two groups.

c. The null hypothesis is that the population mean time required to wake up to the conventional smoke alarm is the same as the population mean time required to wake up to the personalized alarm. In symbols, H_0: $\mu_{\text{conventional}} = \mu_{\text{personalized}}$.

The alternative hypothesis is the population mean time required to wake up to the conventional smoke alarm is greater than the population mean time required to wake up to the personalized alarm. In symbols, H_a: $\mu_{\text{conventional}} > \mu_{\text{personalized}}$.

d. Wake each child on two different nights—once with the personalized alarm and once with the conventional smoke alarm. Use randomization to decide the order of the alarms. Record the time required for the child to wake each time, and then calculate the difference in the times (for instance, *conventional* − *personalized*) for each child. Use the one-sample *t*-test to determine whether the average of these differences is greater than zero.

e. Let μ_d represent the average of the population differences in the times required to wake a child (*conventional alarm* − *personalized alarm*).

H_0: $\mu_d = 0$ vs. H_a: $\mu_d > 0$

f. You should recommend the matched-pairs design because it would have less variability. The times required to wake the children might vary quite a bit, but the difference in the times required to wake the children using the two methods should vary much less.

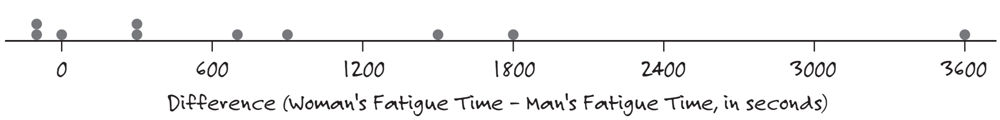
Activity 23-5: Muscle Fatigue

a. The researchers expected that strength would be related to time until fatigue; therefore, matching men and women based on strength would result in less variability in the time differences than in the individual times. This reduction in variability allows for a more powerful test, one that is better able to detect a difference in fatigue times between the genders if a difference really exists.

b. The explanatory variable is *gender,* which is categorical and binary. The response variable is *time until fatigue,* which is quantitative.

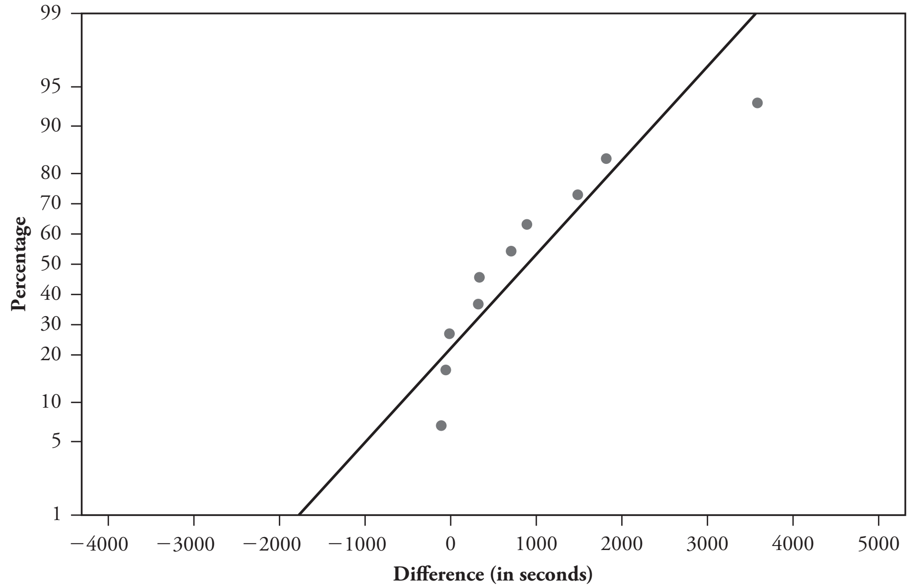
c. This is an observational study. Subjects were not assigned to a gender.

d. The differences in times until fatigue, subtracting the man's time from the woman's time in each pair, are 890, 706, 3584, −16, −58, 335, 322, 1817, 1484, −110. A dotplot follows.



Three of these differences are negative, and seven are positive. Therefore, in most pairs, the woman lasted longer until fatigue set in than the man did. The mean of these differences is 895 seconds, so the women in the sample outlasted the men by almost 15 minutes on average. The standard deviation is 1148 seconds. The distribution of differences appears to be a bit skewed to the right.

e. Because of the small sample size ($n = 10$ pairs), the distribution of differences must be approximately normal in order for the technical conditions to be satisfied. But you have already seen that the distribution is a bit skewed to the right. This skewness is also seen in a normal probability plot, which does not look very linear:



You also lack either a randomly selected sample or random assignment to treatment groups, and therefore the paired t-test conditions were not met. We should stop the analysis at this point, but we will continue with the calculations for the sake of practice, remembering that we cannot take the results very seriously.

f. You are testing H_0: $\mu_d = 0$ vs. the alternative H_a: $\mu_d \neq 0$, where μ_d represents the population mean of the differences in times until fatigue between healthy young adult men and women. The test statistic is

$$t = \frac{895}{1148/\sqrt{10}} \approx 2.465$$

The p-value, from Table III (t-Distribution Critical Values) with 9 degrees of freedom and remembering the alternative hypothesis is two-sided, is between 2(.01) and 2(.025), which means the p-value is between .02 and .05. Technology gives a p-value of .036. This p-value is fairly small, so you would reject the null hypothesis at the $\alpha = .05$ level. The sample data provide fairly strong evidence that the mean time until fatigue differs between men and women.

g. A 90% confidence interval for μ_d is

$$895 \pm 1.833 \frac{1148}{\sqrt{10}}$$

which is 895 ± 665.4, which gives the interval (229.6, 1560.4). This interval means you can be 90% confident that women take between 230 seconds and 1560 seconds longer, on average, to reach fatigue than men do (between 4−26 minutes).

h. This is not a randomized experiment, so even if the test were valid, you would not be able to attribute the longer time to fatigue for women to any particular cause. Moreover, because the subjects volunteered, you should be cautious about generalizing the results even to all healthy young adults in the area of the study. Finally, the technical conditions of the paired t-test were not met because of the small sample size and skewed distribution of differences, so you cannot take the inference results very seriously. Researchers might want to repeat the study with large sample sizes.

24

Solutions

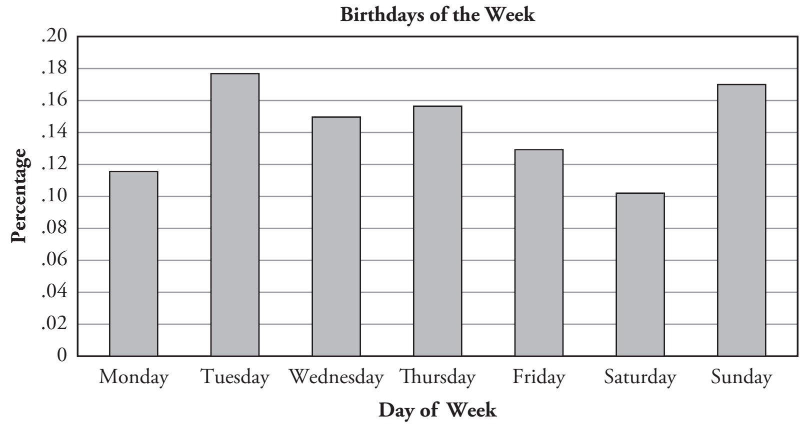
In-Class Activities

Activity 24-1: Birthdays of the Week

a. Observational units: writers

Variable: *days of the weeks on which the birthdays fall* Type: categorical

b. The following bar graph displays the data:



There does not appear to be a great difference in the proportion of people born on any given day of the week.

c. These values (π_{Tu}, π_{We}, . . . , π_{Su}) describe the proportion of all people who were born on Tuesday, Wednesday,… Sunday. The values are parameters because they describe an entire population.

d. The values are $\pi_M = \pi_{Tu} = \ldots = \pi_{Su} = 1/7 \approx .1429$.

e. You would expect the count of Monday birthdays to be $(1/7) \times 147 = 21$.

f. See the middle row of the following table.

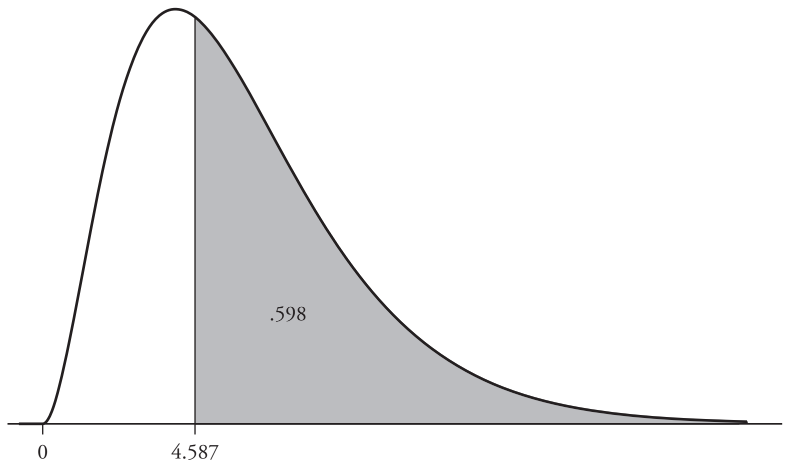
	Mon	Tues	Wed	Thu	Fri	Sat	Sun	Total
Observed Count (*O*)	17	26	22	23	19	15	25	147
Expected Count (*E*)	21	21	21	21	21	21	21	147
$\frac{(O - E)^2}{E}$	.7619	1.1905	.0476	.1905	.1905	1.7143	.7619	4.857

g. See bottom row of the previous table.

h. You calculate $X^2 = .7619 + 1.1905 + \ldots + .7619 = 4.857$.

i. Large values of the test statistic constitute evidence against the null hypothesis. Answers will vary about whether this test statistic provides convincing evidence. The additional evidence you need is the *p*-value. You need to know how likely it is that you would obtain a test statistic this large (or larger) by random chance alone if all seven days of the week are equally likely to be a person's birthday.

j. Using Table IV with 6 degrees of freedom, $4.587 < 8.56$, so the *p*-value is greater than .2.



k. If all seven days of the week are equally likely to be a person's birthday, the probability that you would obtain a test statistic this large (or larger) by random chance alone is at least .2.

l. With the larger p-value, do not reject H_0 at the $\alpha = .10$, $\alpha = .05$, or $\alpha = .01$ levels.

m. Yes; all the expected counts are 21, which is greater than 5. However, the random sampling condition is not met. These are the birthdays of "noted writers of the present," which is not a random sample of the population of U.S. citizens.

n. You have no statistical evidence against the null hypothesis that the seven days of the week are all equally likely to be a person's birthday (at least for the population of famous writers).

Activity 24-2: Birthdays of the Week

a. See middle row of the following table.

	Mon	Tues	Wed	Thu	Fri	Sat	Sun	Total
Observed Count	17	26	22	23	19	15	25	147
Expected Count	24.5	24.5	24.5	24.5	24.5	12.25	12.25	147
$\frac{(O - E)^2}{E}$	2.2959	.0918	.2551	.0918	1.235	.6173	13.270	17.857

b. The test statistic is $X^2 = 17.857$.

c. Using Table IV with 6 degrees of freedom, $16.81 < 17.857 < 18.55$, so $.005 < p\text{-value} < .01$.

d. With the smaller p-value, reject H_0 at the .01 level.

e. Sunday contributes the most to the test statistic (13.720). Its observed count is greater than its expected count, which means you have many more births on Sunday than you would expect if only 1/12 of the births occur on Sunday.

f. You have strong statistical evidence that the null hypothesis is not true; that is, that one or more of the population proportions is not equal to the values that you proposed in the null hypothesis. (So, one or more of $\pi_M = 1/6$, $\pi_{Tu} = 1/6$, $\pi_W = 1/6$, $\pi_{Th} = 1/6$, $\pi_F = 1/6$, $\pi_{Sa} = 1/12$, $\pi_{Su} = 1/12$ is not true.) If all of these were true, you would see a sample result (chi-square test statistic) as large or larger by random chance alone somewhere between .5% and 1% of the time (of random samples from such a population). Because this would occur so rarely by random chance, you conclude the null hypothesis is false.

Activity 24-3: Birthdays of the Week

a. Answers will vary, but students should expect the *p*-value to be smaller because the sample size is so much larger.

b. The sample proportions are the same as those in the actual data, so the bar graph would be identical.

c. The test statistic is $X^2 = 48.572$ (ten times larger than before).

d. Using Table IV with 6 degrees of freedom, $24.1 < 48.572$, so $p\text{-value} < .0005$. With the small *p*-value, reject H_0. You have extremely strong evidence that at least one of the seven days of the week is more or less likely to be a person's birthday (at least for the population of writers).

e. This is a different conclusion and much smaller *p*-value than in Activity 24-1. This makes sense because you have a much larger (ten times as large) sample size.

Activity 24-4: Kissing Couples

a. Observational units: 124 kissing couples

Variable: *whether the couple leaned their heads to the right when they kissed*

Type: binary categorical

b. Yes, the one-sample *z*-test from Topic 17 applies to these data.

c. Let π_{right} refer to the proportion of all kissing couples who turn their heads to the right.

H_0: $\pi_{right} = .75$, $\pi_{left} = .25$

H_a: At least one of $\pi_{right} = .75$, $\pi_{left} = .25$ is not true. Equivalently: H_a: $\pi_{right} \neq .75$ (Note that when there are only two proportions, if one proposed value is false, the other must also be false.)

d. Expected count for "leans right" $= .75 \times 124 = 93$

Expected count for "leans left" $= .25 \times 124 = 31$

The test statistic is $X^2 = (80 - 93)^2/93 + (44 - 31)^2/31 = 1.8172 + 5.4516 = 7.2688$.

Using Table IV with 1 degree of freedom, $6.63 < 7.27 < 7.88$, so $.005 < p\text{-value} < .01$. Using Minitab, the *p*-value is .007.

e. Technical conditions: The expected counts are at least five in both categories (smallest = 31). The couples selected for the sample were those who happened to be observed in public places while the researchers were watching. This is not a random sample, so you should be cautious about generalizing the results of this test to a larger population.

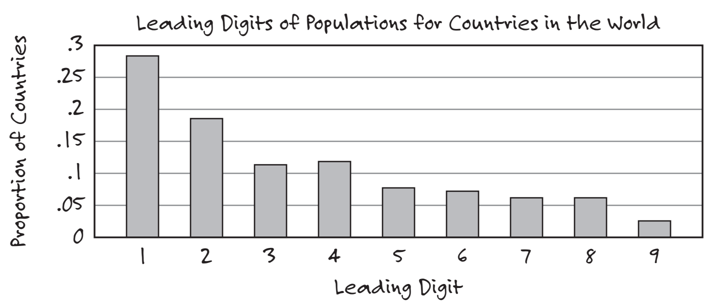
f. Considering these data to be a representative sample, you have very strong statistical evidence that H_0 is false. That is, you are confident $\pi_{right} \neq .75$.

g. You calculate $(-2.7)^2 = 7.29 \approx 7.27$. The X^2-test statistic appears to be (roughly) the square of the z-test statistic. Without rounding the z-test statistic value as much, $z = -2.696$ and $z^2 = 7.2688$, and the two test statistics are identical.

h. The p-values are roughly the same ($.0069 \approx .007$). Without rounding, the values are identical.

Activity 24-5: Leading Digits

a. A bar graph follows:



This graph does appear to be consistent with the Benford probabilities, insofar as the most common leading digit in this sample is 1 by a wide margin, followed by 2, and then gradually tapering off from there. *Note:* Even though the leading digits are numbers, you can consider this variable to be categorical, because the number simply separates the countries into groups. That's why a bar graph is appropriate rather than a histogram and also why a chi-square goodness-of-fit test is appropriate.

b. The null hypothesis is H_0: $\pi_1 = .301$, $\pi_2 = .176$, $\pi_3 = .125$, $\pi_4 = .097$, $\pi_5 = .079$, $\pi_6 = 0.067$, $\pi_7 = .058$, $\pi_8 = .051$, and $\pi_9 = .046$. In other words, the null hypothesis says the Benford probability model is correct for the data. The alternative hypothesis simply says this model is not correct, meaning at least one of the hypothesized Benford probabilities is wrong. In symbols, you can write H_a: at least one π_i differs from its hypothesized value.

c. The expected count in the "1" category is $194(.301) = 58.394$ countries. This is fairly close to the observed count of 55 countries with a leading digit of 1. Calculating the expected counts for the other categories in the same way produces these results:

Leading Digit	Observed Count	Expected Count
1	55	58.394
2	36	34.144
3	22	24.250
4	23	18.818
5	15	15.326
6	14	12.998
7	12	11.252
8	12	9.894
9	5	8.924
Total	194	194.000

All of these expected counts are greater than five, so that technical condition is satisfied. If you consider the population to be all numbers appearing in the almanac, you do not literally have a random sample of numbers from the almanac, but these are likely to be representative of population values.

d. The contribution to the test statistic from the "1" category is

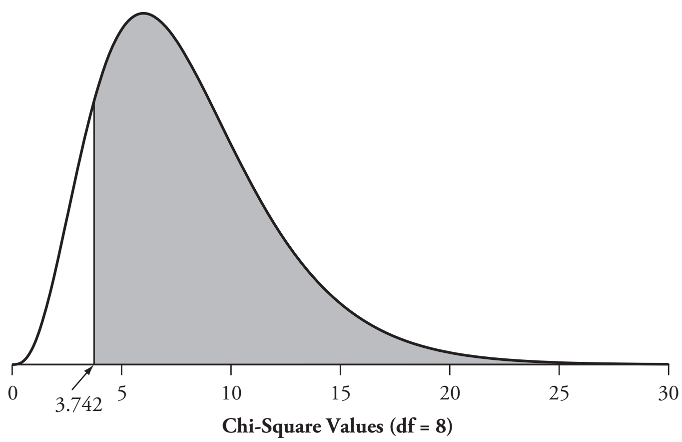
$$\frac{(55 - 58.394)^2}{58.394} \approx 0.197$$

Calculating the other contributions in the same way produces these results:

Leading Digit	Observed Count	Expected Count	$\frac{(O - E)^2}{E}$
1	55	58.394	0.197
2	36	34.144	0.100
3	22	24.250	0.209
4	23	18.818	0.929
5	15	15.326	0.007
6	14	12.998	0.077
7	12	11.252	0.050
8	12	9.894	0.448
9	5	8.924	1.725
Total	194	194.000	3.742

The test statistic equals 3.742, the sum of the nine values in the last column. (You might get a slightly different, more accurate answer if you carry more than three decimal places in your intermediate calculations.) Looking at Table IV with 8 degrees of freedom (one less than the nine categories), this test statistic is way off the chart to the left. Therefore, the p-value is much greater than .2.

Technology calculates the *p*-value to be about .88, as seen in the following graph:



e. This is a very large *p*-value, much greater than all common significance levels, so fail to reject the null hypothesis that Benford's probabilities adequately model these data.

f. The large *p*-value (based on the small value of the chi-square test statistic) from this chi-square analysis reveals the sample data are extremely consistent with the Benford probabilities. If Benford's model were correct, it would not be the least bit surprising to observe sample data as found with these nations' populations. In other words, the sample data provide no reason to doubt Benford's model describes the distribution of leading digits in the almanac.

g. An accountant or IRS agent could apply a chi-square test to see how well the numbers in a tax return follow Benford's probabilities. If the *p*-value turns out to be very small, that suggests the leading digits of numbers on that tax return differ substantially from what Benford's probabilities predict. This certainly does not prove the tax return is fraudulent, but it might suggest the numbers were made up and not legitimate. A tax return whose leading digits do not follow Benford's probabilities might be worth a closer look.

25

Solutions

In-Class Activities

Activity 25-1: Pursuit of Happiness

a. The observational units are the 4409 American adults who were surveyed.

b. Explanatory: *year* Type: categorical

Response: *happiness level* Type: categorical

c. This is an observational study because the researchers did not randomly assign the subjects to the years.

d. A segmented bar graph is the appropriate graphical display for these data.

e. The proportions of Americans who considered themselves very happy in each of these three years were very similar (about 30%).

f. The two-sample *z*-test from Topic 21 is not appropriate for testing whether these population proportions differ significantly because there are three years to consider—not just two.

g. H_0: $\pi_{1972} = \pi_{1988} = \pi_{2004}$, where π_{1972} represents the proportion of all American adults who considered themselves happy in 1972, and so on.

h. For the three years combined, the proportion of respondents who were very happy was 1403/4409 or .318.

i. For 1972, you calculate $.318 \times 1606$ or 511.05.

j. 1988: $.318 \times 1466 = 466.50$; 2004: $.318 \times 1337 = 425.45$

k. Here is the completed table:

	1972	1988	2004	Total
Very Happy	486 (511.05)	498 (466.50)	419 (425.45)	1403
Less Than Very Happy	1120 (1094.95)	968 (**999.50**)	918 (911.55)	3006
Total	1606	1466	1337	4409

l. For the "less than very happy" people in 1988, you calculate $(968 - 999.5)^2/999.5 = .993$.

m. The test statistic is $X^2 = 5.065$.

n. Large values of the test statistic would constitute evidence against the null hypothesis that the three populations have the same proportions of very happy Americans because if the proportions are the same, then the observed and expected values in each year will be very similar, and each year's contribution to the test statistic will be small, so the overall test statistic value will be small. However, if one or more of the proportions is not the same, then there will be a large difference between the observed and expected values in that year, and this difference will make a large contribution to the test statistic.

o. Using Table IV with $(2 - 1) \times (3 - 1) = 2$ degrees of freedom, $4.61 < 5.065 < 5.99$, so $.05 < p\text{-value} < .10$.

p. These sample data are a bit unlikely to have occurred by chance alone if the population proportions of very happy people had been identical for these three years, but you would fail to reject the null hypothesis at the $\alpha = .05$ level.

q. You have weak statistical evidence that the population proportions of very happy people were not identical for these three years. (Because these were random samples, you are safe in generalizing this conclusion to the populations of all American adults in each year.)

Activity 25-2: Pursuit of Happiness

a. H_0: The population distributions of happiness levels were the same in all three years.

H_a: At least one year's distribution of happiness levels differs from the others.

b. Using Minitab, the test statistic is $X^2 = 35.655$ with 4 degrees of freedom, p-value $\approx .000$.

c. Yes, the sample data provide evidence that the distributions of self-reported happiness levels differed among all adult Americans these three years, at the .05 significance level.

d. "1988/not too happy" has the largest contribution (16.927) to the test statistic. The observed count (136) is less than the expected count (193.18).

"1972/not too happy" has the next largest contribution (13.458) to the test statistic. The observed count (265) is greater than the expected count (211.63).

e. You have strong statistical evidence that the population distributions of happiness levels were not the same for all three years. In 1972, the sample proportion of those who were not too happy was substantially greater than you would expect if the population proportions were equal in all three years, and in 1988, the sample proportion of those who were not too happy was substantially less than you would expect.

Activity 25-3: Got a Tip?

a. The observational units are the customers in the coffee bar.

b. Explanatory: *type of card received* Type: categorical

Response: *whether the customer left a tip* Type: binary categorical

c. This is an experiment because the waiter randomly assigned the customers to treatment groups.

d. This study involves random assignment to groups.

e. H_0: The population proportions of potential customers who would leave a tip is the same regardless of the type of card they receive.

H_a: The population proportion of potential coffee bar customers who would leave a tip is affected by the type of card they receive.

Using Minitab, the test statistic is $X^2 = 9.953$ with 2 degrees of freedom and the *p*-value is .007.

As the *p*-value is .007, which is less than .10, reject H_0 at the .10 significance level.

The "joke/left a tip" cell makes the largest contribution to the test statistic, and the observed count is much greater than the expected count. So, you conclude that potential coffee bar customers are not equally likely to leave a tip regardless of the type of card they receive—leaving a joke card is more likely to generate a tip than leaving an advertisement card or no card.

f. Because this was an experiment where the only difference between the groups should be the type of card the customers received, you can conclude that the joke card caused customers to be more likely to leave a tip. (Note also that the technical conditions for this test to be valid were satisfied—the subjects were randomly assigned to treatment groups, and the expected counts in each cell of the table were at least five.)

g. You should not generalize your conclusion to all waiters and waitress in all food and drink establishments. This experiment was carried out in only one coffee bar with one waiter.

Activity 25-4: Pursuit of Happiness

a. Explanatory: *political inclination* Type: categorical

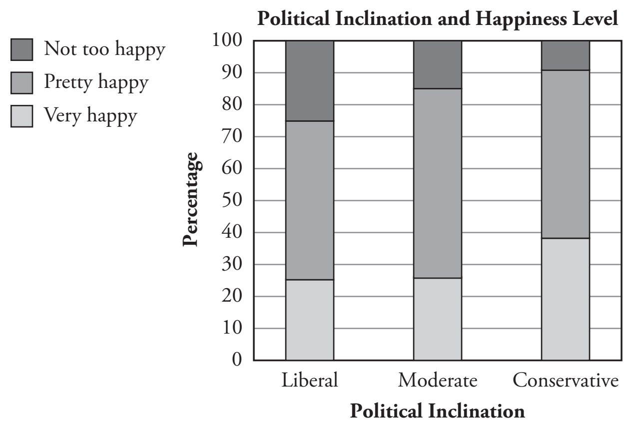
Response: *self-reported happiness level* Type: categorical

b. In Activities 25-1 and 25-2, the researchers took independent random samples from three populations (1972, 1988, and 2004). In this activity, the researchers took one random sample and then classified the subjects by *both* explanatory and response variables.

c. H_0: Political inclination and self-reported happiness level are independent in the population of adult Americans.

H_a: Political inclination is related to self-reported happiness level.

Here is a bar graph:



Using Minitab, the test statistic is $X^2 = 22.454$ with 4 degrees of freedom and p-value $\approx .000$.

With such a small p-value, reject H_0 and conclude political inclination and self-reported happiness level are related in the population of adult Americans. The largest contributions to the test statistic are made by the "conservative/very happy" cell, in which the observed count (187) is greater than the expected count (152.88), and the "conservative/not too happy" cell, in which the observed count (48) is less than the expected count (66.06).

Activity 25-5: Pets as Family

a. Here is the two-by-two table:

	Dog	Cat	Total
Feels Close to Pet	1110	577	1687
Does Not Feel Close to Pet	71	110	181
Total	1181	687	1868

H_0: Type of pet and feelings of closeness to pet are independent in the population of adult Americans.

H_a: Type of pet is related to feelings of closeness to pet.

Technical conditions: The observations arose from a random sample of the population, and the expected counts are at least five for each cell in the table.

Using Minitab, the test statistic is $X^2 = 49.63$ with 1 degree of freedom and the p-value $\approx .000$.

With such a small p-value, reject H_0 and conclude type of pet and feeling close to the pet are related in the population of adult Americans. The largest contribution to the test statistic is made by the "cat/does not feel close to pet" cell, in which the observed count (110) is greater than the expected count (66.57).

b. The null hypothesis is that the proportion of the population of all dog owners who feel close to their pets is the same as the proportion of all cat owners who feel close to their pets. In symbols, the null hypothesis is H_0: $\pi_{dog} = \pi_{cat}$.

The alternative hypothesis is the proportion of the population of all dog owners who feel close to their pets is not the same as the proportion of all cat owners who feel close to their pets. In symbols, the alternative hypothesis is H_a: $\pi_{dog} \neq \pi_{cat}$.

Technical conditions: The number of successes and failures in each group is at least five, and the data were randomly selected.

The test statistic is $z = \dfrac{.93988 - .83988}{\sqrt{.9031(1 - .9031)\left(\frac{1}{1181} + \frac{1}{687}\right)}} = 7.045.$

p-value $= 2 \times \Pr(Z > 7.045) \approx .0000$

With such a small p-value, reject H_0 at any commonly used significance level.

You have very strong statistical evidence that the population proportion of dog owners who feel close to their pets is not the same as the population proportion of cat owners who feel close to their pets.

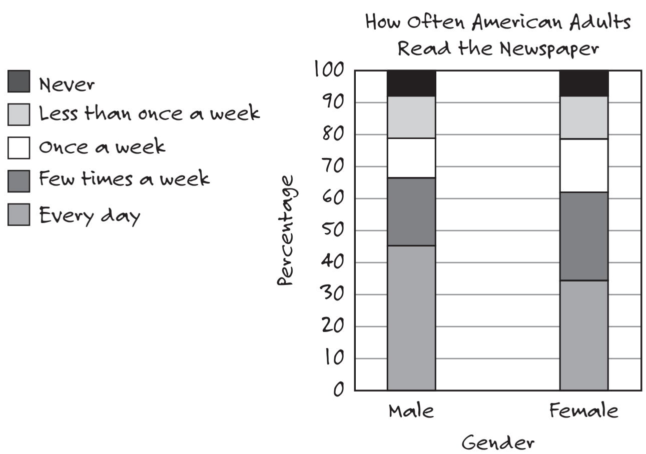
c. The p-values are the same in both cases.

d. $(7.045)^2 = 49.63$; $49.63 = X^2$

Activity 25-6: Newspaper Reading

a. The General Social Survey took *one* random sample, and this table summarizes responses on two categorical variables (*gender* and *newspaper reading frequency*), so the third scenario applies.

b. The following segmented bar graph displays the distributions (conditional proportions) of newspaper reading frequency for each gender:



This graph reveals that a higher proportion of men than women read the newspaper every day (191/422 = .453 for men vs. 167/484 = .345 for women). A higher proportion of women than men read the paper a few times a week (.211 for men vs. .275 for women) and once a week (.126 for men vs. .167 for women).

c. The null hypothesis asserts newspaper reading frequency is independent of gender. In other words, the null hypothesis says the proportional distribution of the various reading frequency categories is identical for men and women in the population of all American adults. The alternative hypothesis says newspaper reading frequency is related to gender, which means these population distributions are not the same for men and women.

The expected counts appear in parentheses in the table:

	Male	Female	Total
Every Day	191 (166.75)	167 (191.25)	358
Few Times a Week	89 (103.40)	133 (118.60)	222
Once a Week	53 (62.42)	81 (71.58)	134
Less Than Once a Week	56 (56.36)	65 (64.64)	121
Never	33 (33.07)	38 (37.93)	71
Total	422	484	906

For an example of one of these calculations, the expected count for the "every day/male" cell is found by taking (358)(422)/906 = 166.75.

Technical conditions: All of these expected counts are greater than five (smallest 33.07). The subjects were randomly selected from the population of adult Americans, so the technical conditions are satisfied.

The chi-square test statistic, computed as

$$\sum \frac{(O - E)^2}{E}$$

turns out to be

$$\begin{aligned} X^2 = {} & 3.526 + 3.075 + \\ & 2.006 + 1.749 + \\ & 1.420 + 1.238 + \\ & 0.002 + 0.002 + \\ & 0.000 + 0.000 = 13.020 \end{aligned}$$

Comparing this value to a chi-square distribution with (5 − 1)(2 − 1) = 4 degrees of freedom reveals that the p-value is slightly greater than .01 (13.02 is just below 13.28 in Table IV). This is less than .05, so the test decision is to reject the null hypothesis at the .05 significance level. The data provide fairly strong evidence that newspaper reading frequency is related to gender.

In particular, the table cells in the top row contribute the most to the test statistic calculation. This finding indicates more men than expected read the newspaper every day and fewer women than expected read the newspaper every day. The sample proportions of men and women who read the newspaper every day are $191/422 \approx .453$ for men and $167/484 \approx .345$ for women.

Solutions

In-Class Activities

Activity 26-1: House Prices

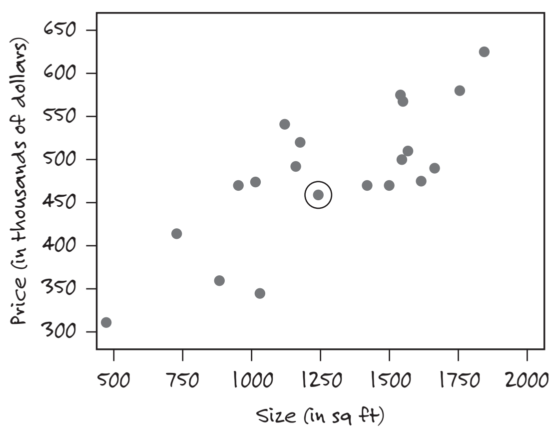
a. The observational units are the houses.

b. Two variables reported: *price* and *size.* Both are quantitative variables.

c. Yes, the data suggest that bigger houses tend to cost more than smaller ones, as the house sizes tend to increase down the list as well.

d. Bottom left: 2130 Beach St. Top right: 833 Creekside Dr.

e. The point corresponding to the house at 845 Pearl Drive is circled on the scatterplot below:



f. Yes, the scatterplot reveals a positive relationship between a house's size and price. The upward trend of the graph indicates that, in general, the larger the house, the greater the price.

g. Direction: positive

Strength: moderate

Form: linear

h. Answers will vary. One example is 2545 Lancaster Dr. (1030 sq ft, $344,720) and 415 Golden West Pl. (883 sq ft, $359,500). (The smaller house has the greater price.)

Activity 26-2: Birth and Death Rates

a. This association is weak, negative, and linear.

b. **A:** Florida; reasoning: This state has the highest death rate and lowest birth rate (of the three) because this is a state to which people like to retire.

B: Alaska; reasoning: This state has a lower than average death rate because people typically do not retire to Alaska.

C: Utah; reasoning: This state has an unusually high birth rate that may be due to the prevalence of the highly procreative Mormon (Latter Day Saints) faith in Utah.

Activity 26-3: Car Data

	Negative				None	Positive			
	Strongest			Weakest		Weakest			Strongest
Letter of Scatterplot	D	G	A	H	C	E	I	F	B

Activity 26-4: Marriage Ages

a. This scatterplot shows a strong, positive, linear association between the husband's age and the wife's age.

b. Two of the ages fall exactly on the $y = x$ line.

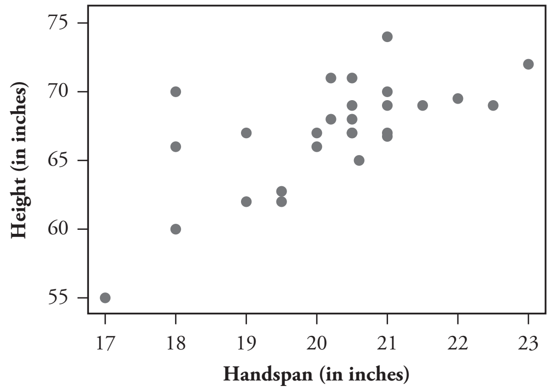
c. For 16 of the couples, the husband is older than the wife.

d. For six of the couples, the husband is younger than the wife.

e. The fact that the majority of couples produce points that lie above the $y = x$ line tells you that, in general, husbands tend to be older than their wives.

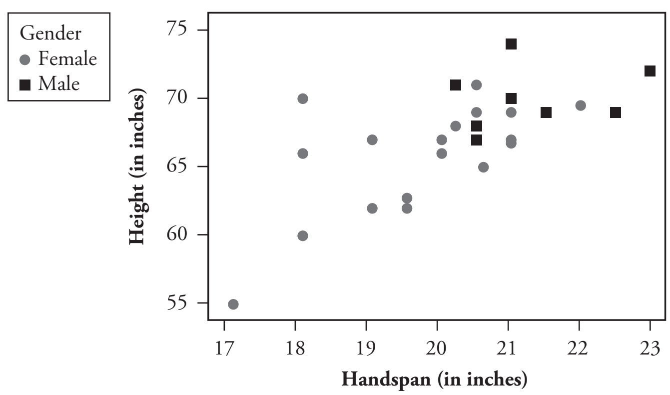
Activity 26-5: Heights, Handspans, and Foot Lengths

a. The scatterplot of *height* vs. *handspan* follows:



There is a moderate, positive, linear association between height and handspan.

b. The labeled scatterplot of *height* vs. *handspan* follows:

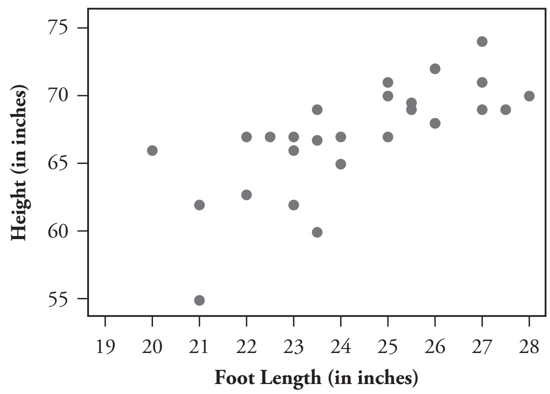


Yes, men and women tend to differ with regard to *handspan.* Men tend to have longer handspans than women and less variability in their handspans. This result is indicated by the red points clustered fairly close together on the upper end of the horizontal scale, whereas the black points are spread widely on the lower to middle end of the horizontal scale.

Yes, men and women tend to differ with regard to *height.* Most of the men in this class are taller than most of the women, although there are some fairly tall women in this class. You can tell because the black points (women) range from the low to the highest end of the vertical scale, whereas the red points are primarily clustered toward the upper end of the vertical scale.

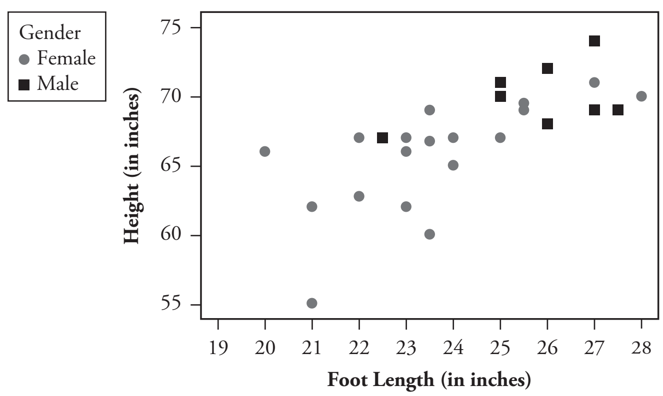
Yes, the association between *handspan* and *height* appears to be different between men and women. There is a much stronger linear relationship between these variables for the women than there is for the men. There is actually a very weak, positive association been *height* and handspan for the men.

c. The scatterplot of *height* vs. *foot length* follows:



There is a moderately strong, positive, linear relationship between *height* and *foot length.* The scatterplot shows one observation (21 in, 55 cm) that does not seem to fit the overall trend. In general, the taller the person, the longer her or his feet.

The labeled scatterplot of *height* vs. *foot length* follows:

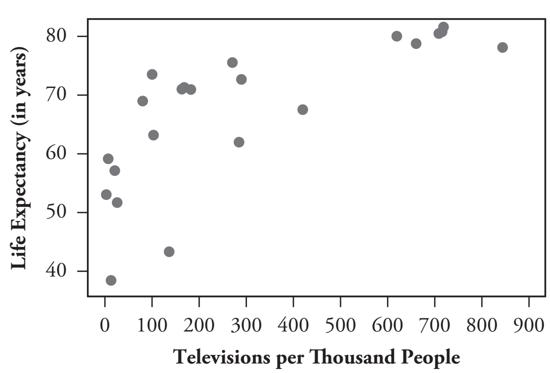


As with handspans, most of the men's feet are longer than most of the women's feet. However, there is one man of medium height with (relatively) short feet. Both genders display a moderate, positive, linear association between *height* and *foot length,* although the association is somewhat stronger among the women.

Activity 26-6: Televisions and Life Expectancy

a. Fewest: Haiti (5) Most: United States (844)

b. The scatterplot of life expectancy vs. televisions per thousand people follows:



Yes, there is a moderate, positive, curved association.

c. This is an absurd argument. Simply sending televisions to countries such as Haiti will not lead to an increase in the life expectancies of inhabitants.

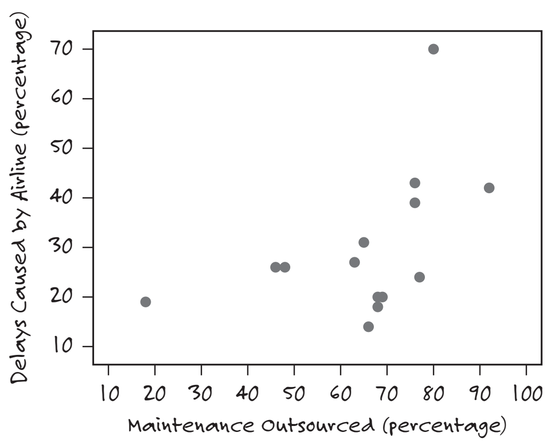
d. No; this is a very good example of a situation where two variables are strongly associated, but there is no cause-and-effect relationship between them. Having lots of televisions does not cause life expectancy to increase. There are several other variables that affect both life expectancy and number of televisions per 1000 people.

e. Many answers are possible. One possibility is the wealth (GNP) of the nation. Nations with high GNPs will also tend to have a large number of televisions per capita and will have good medical care and thus long life expectancies. Similarly, countries with low GNPs will tend to have few televisions per capita, poor national medical care, and shorter life expectancies.

Activity 26-7: Airline Maintenance

a. The observational units are the airlines.

b. The scatterplot follows:



c. The scatterplot reveals a fairly strong positive association between these variables. Airlines that outsource a higher percentage of their maintenance tend to have a greater percentage of delays caused by the airline, as compared to airlines that outsource a smaller percentage of their maintenance. The form of the association is clearly nonlinear, as the relationship appears to follow a curved pattern.

d. No, you cannot conclude from these data that outsourcing maintenance *causes* airlines to experience more delays. This is an observational study, not an experiment. Many other variables could explain the observed association between outsourcing and delays. For example, perhaps less financially successful airlines tend to outsource more of their maintenance and also to have more delays.

Solutions

In-Class Activities

Activity 27-1: Car Data

	Negative				None	Positive			
	Strongest			Weakest		Weakest			Strongest
Letter of Scatterplot	D	G	A	H	C	E	I	F	B
Correlation Coefficient	−.907	−.721	−.450	−.244	−.081	.235	.510	.889	.994

a. The correlation coefficient is −.450.

b. See table preceding part a.

c. Largest: 1.0   Smallest: −1.0

d. The correlation coefficient will assume its largest or smallest value when the scatterplot displays a perfect linear relationship (positive or negative, respectively).

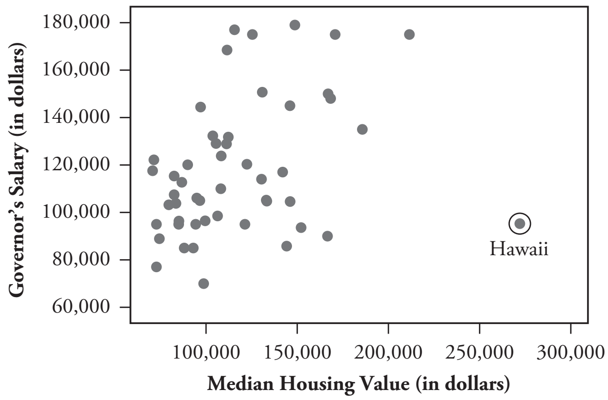
e. Positive correlation coefficients correspond to positive relationships.

f. The stronger the linear relationship, the closer the absolute value of the correlation coefficient to one.

Activity 27-2: Governors' Salaries

a. The 50 U.S. states are the observational units.

b. The scatterplot of governor's salary vs. median housing price follows:



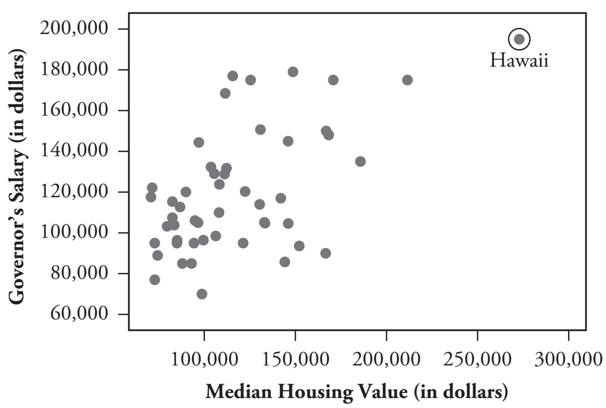
The scatterplot shows a weak positive linear association with one outlier (Hawaii).

c. Answers will vary by student guess.

d. The correlation coefficient is .334.

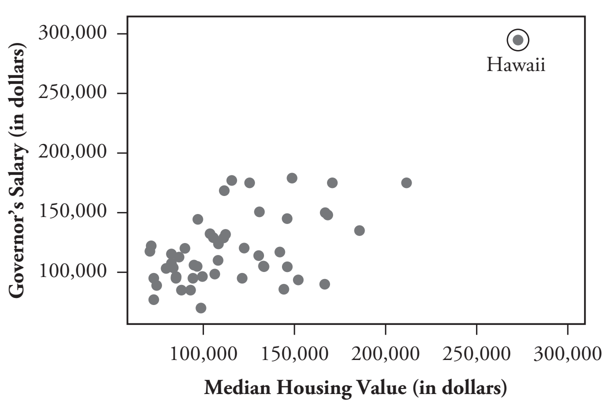
e. Yes, one of the states appears to be unusual. Hawaii has an unusually large median housing value ($272,700) *and* the salary of Hawaii's governor is in the lower half of the distribution.

f. The new correlation coefficient is .576. The revised scatterplot follows:



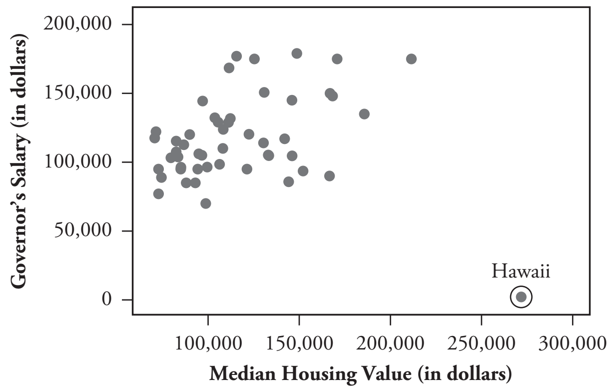
The correlation coefficient is now a bit stronger (.334 vs. .576).

g. The correlation is now .673. The revised scatterplot follows:



The correlation coefficient has again increased, but not as substantially.

h. The correlation coefficient plummets to .061. The revised scatterplot follows:



i. No, the correlation coefficient is not a resistant measure of association.

j. The government agencies tend to report the median housing price because housing prices are usually quite skewed to the right with several high outliers, and so the mean (which is not resistant) would not be an appropriate measure of center to report with these data.

Activity 27-3: Televisions and Life Expectancy

a. This scatterplot shows a moderately strong positive association. The association does not follow a linear form; it appears to be curved.

b. Answers will vary by student guess.

c. The correlation coefficient is .743.

d. Yes, the value of the correlation coefficient is fairly high even though the association between the variables is not linear.

e. No, the fairly high value of the correlation coefficient is not evidence of a cause-and-effect relationship between the variables. There are many confounding variables that could explain the association, as discussed in Activity 26-6.

27

Activity 27-4: Guess the Correlation

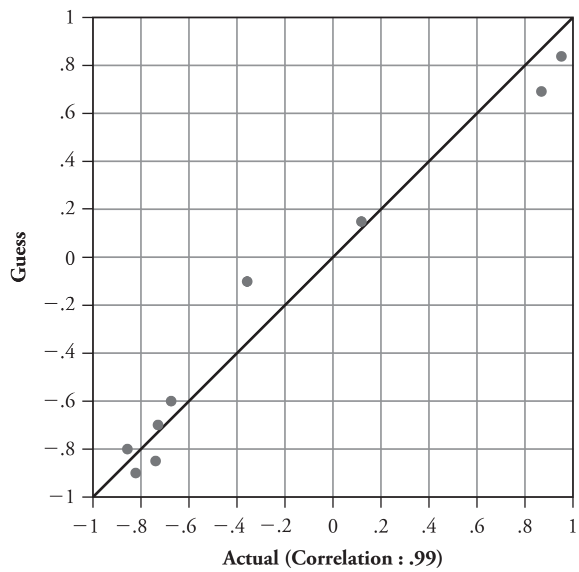
a. Answers will vary. The following is one representative set of answers.

Repetition Number	1	2	3	4	5	6	7	8	9	10
Your Guess	−.6	.7	−.7	−.8	−.85	−.7	.15	−.9	.85	−.1
Actual Value	−.69	.87	−.75	−.87	−.76	−.74	−.11	−.84	.95	−.37

b. See table in part a.

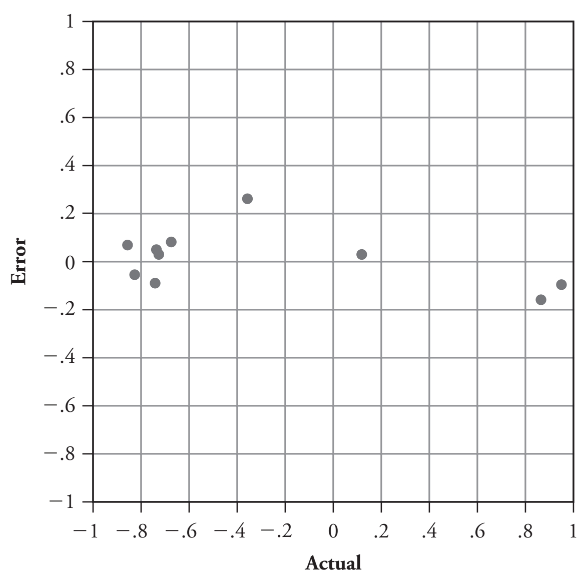
c. Guess: $r = .8$.

d. The scatterplot of *your guess* vs. *actual values* follows:



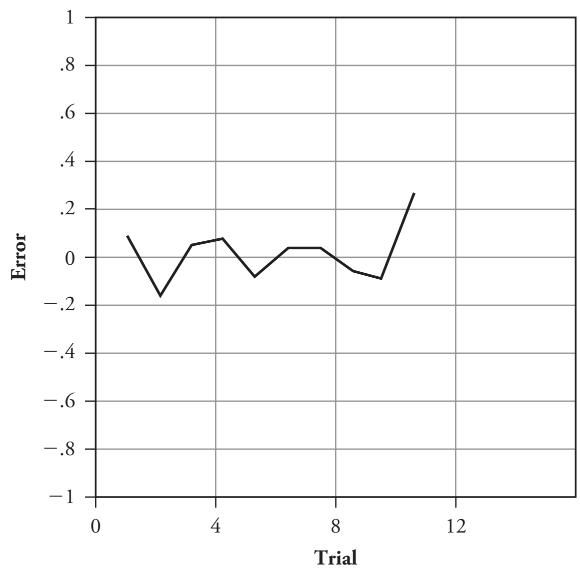
The correlation coefficient *r* is .99. Most students will be surprised that their guesses were so consistent.

e. The scatterplot of *your errors* vs. *the actual values* follows:



For most students, errors are for correlations near −1 or 1.

f. The scatterplot of your errors vs. the repetition (trial) number follows:



In this example, there is no real evidence that the accuracy of the guesses changed over time.

g. The correlation would be 1.00.

h. The correlation would be 1.00 in this case also.

i. No; if the correlation is 1.00, it does not mean that you guessed perfectly every time. It means your guesses are consistent: either consistently correct or consistently incorrect by the same amount and in the same direction. You cannot use the correlation coefficient to tell whether you guess correctly, but you can use it to tell whether your guesses change appropriately with the size and direction of the actual correlation.

Activity 27-5: House Prices

a. For the *z*-score for the price of 2130 Beach St., you calculate (311,000 − 482,386)/79601.5 = −2.148. For the product of *z*-scores, you calculate −2.148 × −2.243 = 4.817.

For the *z*-score for the size of 2461 Ocean St., you calculate (1755 − 1288.1)/369.191 = 1.265. For the product of *z*-scores, you calculate 1.223 × 1.265 = 1.547.

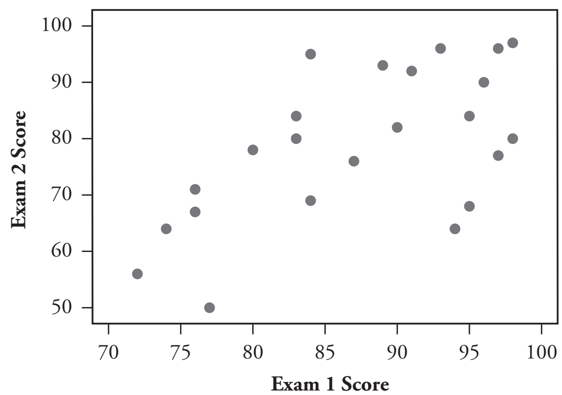
b. Correlation coefficient r = 14.819/19 or .77997 ≈ .78.

c. Most of the houses with negative price *z*-scores are paired with negative size *z*-scores, making the product of their *z*-scores positive. This is a result of the strong positive association between *house price* and *size*.

d. Using Minitab, the correlation coefficient is .780.

Activity 27-6: Exam Score Improvements

a. The scatterplot of *exam 2 score* vs. *exam 1 score* (ignoring the student who did not take exam 2) follows:

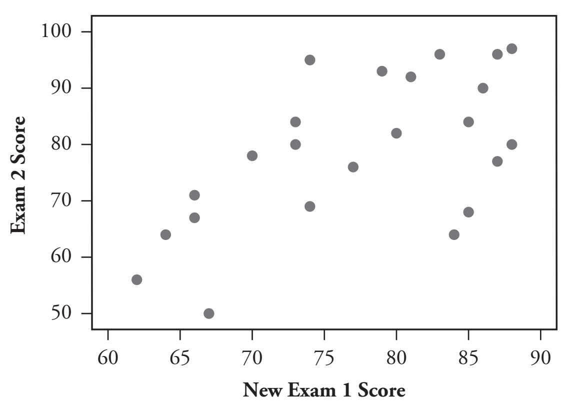


The scatterplot shows a moderate positive linear association.

b. The correlation coefficient is .602.

c. Answers will vary by student expectation.

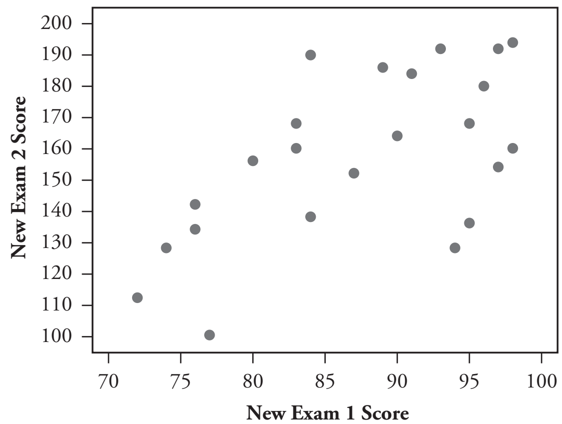
d. The scatterplot of *exam 2 score* vs. *new exam 1 score* follows:



The correlation coefficient did not change (it remains .602).

e. Answers will vary by student expectation.

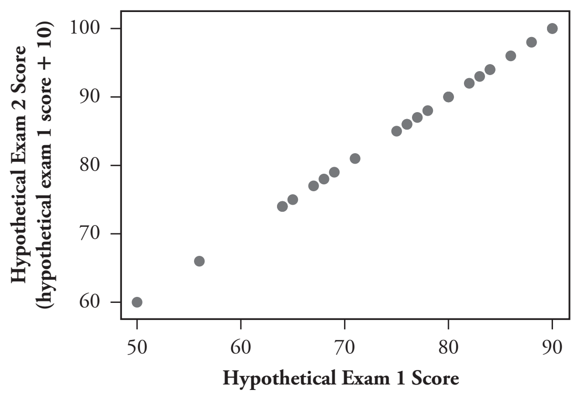
f. The scatterplot of *new exam 2 score* vs. *new exam 1 score* follows:



The correlation coefficient did not change (it is still .602).

g. Answers will vary by student expectation.

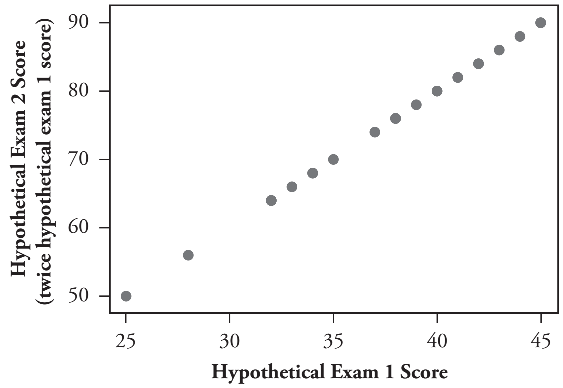
h. The hypothetical data and scatterplot will vary by student. The following is a representative example.



The correlation coefficient is 1.0.

i. Answers will vary by student expectation.

j. The hypothetical data and scatterplot will vary by student. The following is a representative example.



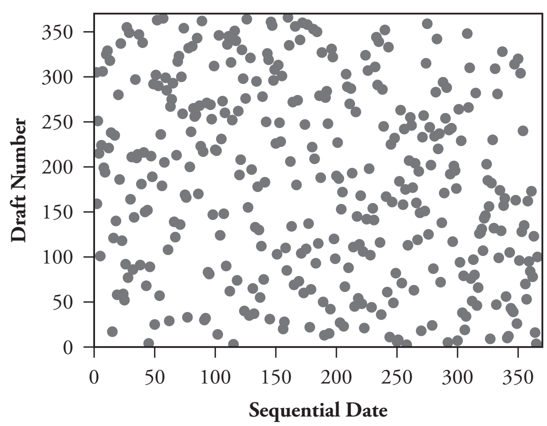
The correlation coefficient is 1.0.

Activity 27-7: Draft Lottery

a. Answers will vary.

b. With a perfectly fair, random lottery, there should be no association between *draft number* and *sequential date for the birthday.* In other words, these variables should be independent, so the correlation coefficient would equal zero. With an actual lottery, you would not expect the correlation coefficient to equal exactly zero, but it should be close to zero.

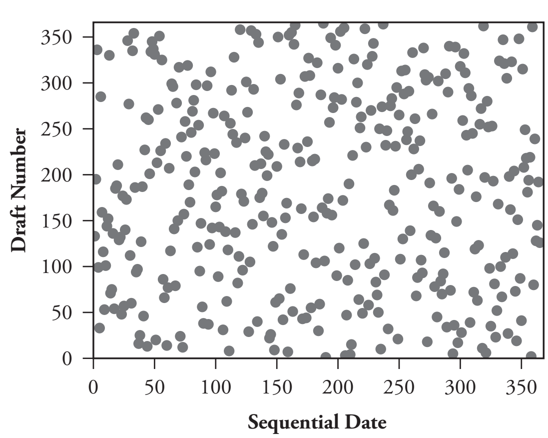
c. The scatterplot is shown here:



It's hard to see a relationship between the variables in this scatterplot, so a reasonable guess for the value of the correlation coefficient would be close to zero.

d. Technology reveals the correlation coefficient to equal −0.226. This indicates a weak negative association between draft number and sequential date. Although not large, this correlation value is farther from zero than most people expect. Looking at the scatterplot more closely, you can see there are few points in the top right and bottom left of the graph. This result suggests few birthdays late in the year were assigned high draft numbers, and few birthdays early in the year were assigned low draft numbers, which means young men born late in the year were at a disadvantage and had a better chance of getting a low draft number. Birthdays late in the year were not mixed as thoroughly as those earlier in the year, so they tended to be selected early in the process and thereby assigned a low draft number.

e. The scatterplot for the 1971 draft lottery data is shown here:



The correlation coefficient is .014, which is very close to 0. This value indicates there is no evidence of association between draft number and sequential date, suggesting the lottery process was fair and random in 1971. The mixing mechanism was greatly improved after the anomaly with the 1970 results was spotted.

Solutions

In-Class Activities

Activity 28-1: Heights, Handspans, and Foot Lengths

a. There is a moderate positive linear association between *height* and *foot length* in this scatterplot. Answers will vary by student.

The following is one representative set of answers.

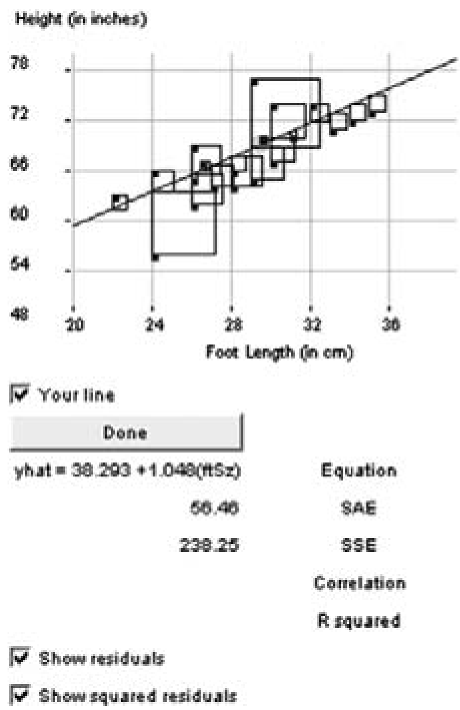
b. The equation reported is (predicted) $\widehat{height} = 38.293 + 1.048$ *foot size*.

c. No, everyone in the class did not obtain this same line.

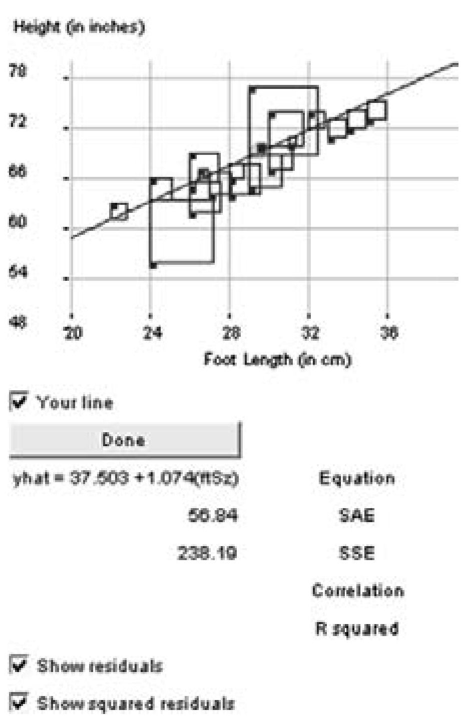
d. Answers will vary.

e. The SAE value is 56.46. Yes, someone had a smaller value (55.11).

f. The SSE value is 238.25. A graph follows:



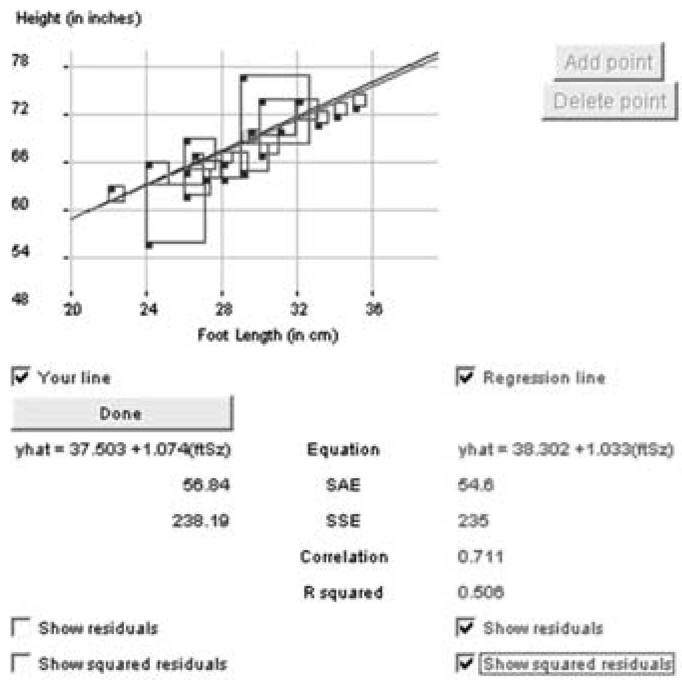
g. The equation is $\widehat{height} = 37.503 + 1.074$ *foot size;* the SSE is 238.19. Yes, someone had a smaller value (235.28). A graph follows:



h. Regression line: $\widehat{height} = 38.302 + 1.033\ \textit{foot size}$

SSE (least squares): 235

A graph follows.



i. Using the least squares regression line, predicted height is $38.302 + 1.033 \times 28$ or 67.226 inches. Yes, this prediction seems reasonable based on the scatterplot.

j. Using the least squares regression line, predicted height is $38.302 + 1.033 \times 29$ or 68.259 inches.

k. These predictions differ by 1.033 inches, which is the slope of the least squares line.

l. The least squares line would predict a height of 38.302 inches for a person with a 0 cm foot length. No, this does not make any sense because a foot length of 0 cm is outside the range of plausible values.

m. Using the least squares regression line, predicted height is $38.302 + 1.033 \times 45$ or 84.79 inches, which is more than 7 feet tall. You should not consider this prediction as reliable as the one for a person with a 28 cm foot length, because for this prediction (84.79 in) you do not have much data for someone more than 7 feet tall. You should not feel comfortable assuming the same relationship will hold for such extreme observations.

n. No, the least squares regression line does not move much at all when you change this student's height.

o. Changing the height of the student with the shortest or longest foot has a much more drastic effect on the regression line. Points on the end of the regression line are clearly *more influential* than points near the middle.

p. SSE $(\bar{y}) = 475.75$

q. Here is the percentage change in the SSE:

$$100\%\left[\frac{475.75 - 235}{475.75}\right] = 50.6\%$$

Activity 28-2: House Prices

a. Yes, the least squares line appears to provide a reasonable model for predicting the price of a house based on its size (sq ft, or ft^2).

b. Slope b =

$$.780 \times \frac{79{,}802}{369.2} = \$168.6/\text{ft}^2$$

Intercept a =

$$482{,}386 - \frac{168.6}{\text{ft}^2} \times (1288.1) = \$265{,}212.34$$

Least squares line equation: $\widehat{price} = 265{,}212.34 + 168.6/\text{ft}^2 \times size$

c. Using Minitab, *predicted price* = \$265,222 + \$169 × *size*.

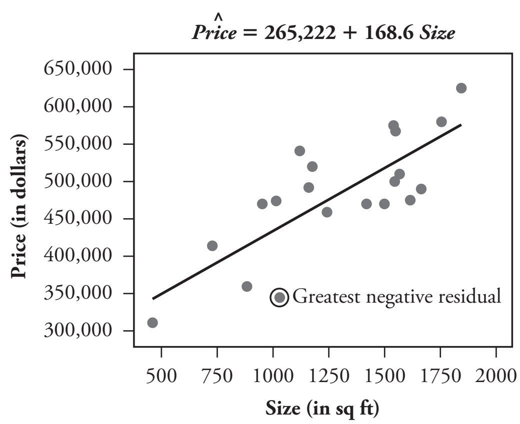
d. Predicted price ≈ \$265,222 + \$169 × (1242) = \$475,120. Yes, this answer seems reasonable.

e. The prediction was too high by \$475,120 − \$459,000 or \$16,120.

f. Fitted value: \$475,120 Residual: \$16,120

g. Points that lie above the line will have positive residuals.

h. The scatterplot, with the point circled, follows:



The house with the greatest negative residual is located at 2545 Lancaster Drive.

i. Each increase of one square foot in the house size increases the predicted house price by \$168.60.

j. You calculate 100 × \$168.60 or \$16,860.

k. The predicted price for a house with an area of 0 square feet is \$265,222. This value makes no sense in this context because there is no such thing as a house with zero area.

l. The percentage of variability in house prices explained by the least squares regression line with size is 60.8% as r^2 = correlation coefficient^2 = $(.78)^2$.

m. No, it would not be reasonable to use this regression line to predict the price of a 3500 ft^2 house because the given house sizes are from about 500–2000 ft^2. This would be extrapolation; you have no idea whether the given association continues beyond this range of house sizes.

n. No, it would not be reasonable to use this regression line to predict the price of a 2000 ft^2 house in Canton, New York. This regression line was created with data collected solely from one region in California. The housing market in New York is very different, and you cannot expect this regression line to make accurate predictions for any market other than Arroyo Grande, California.

Activity 28-3: Animal Trotting Speeds

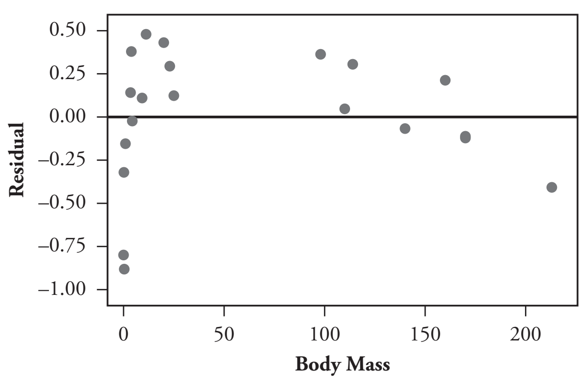
a. Yes, it appears that larger animals tend to break into trot at higher speeds than smaller animals. The scatterplot reveals a moderate positive nonlinear association.

b. No, the least squares line does not appear to provide a reasonable model for summarizing the relationship between *trot speed* and *body mass* because the relationship between these variables is nonlinear.

c. Small animals: tend to have negative residuals

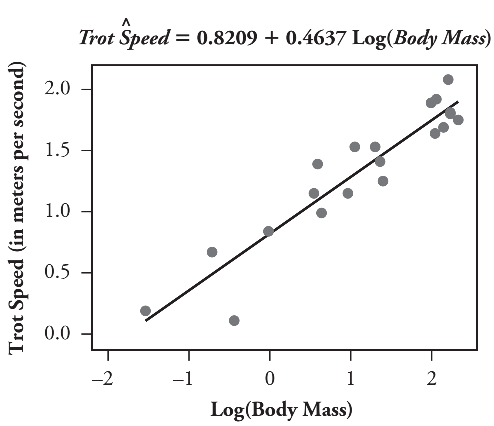
Medium-sized animals: tend to have positive residuals

d. The residual plot of *residual* vs. *body mass* follows:



Yes, the residual plot reveals a pattern that suggests the linear model is not appropriate.

e. The scatterplot of *trot speed* vs. log(*body mass*) follows:

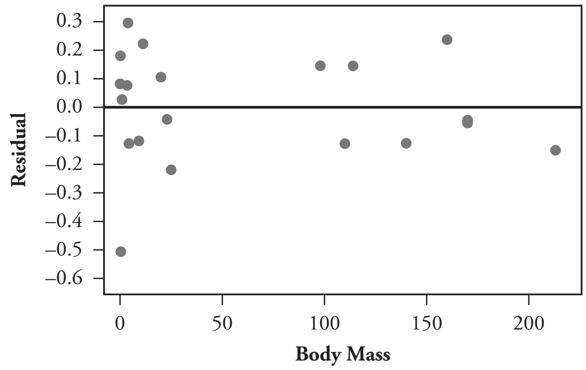


Yes, the association between *trot speed* and log10(*body mass*) appears to be fairly linear.

f. Here is the least squares line:

$$\widehat{trotspeed} = 0.8209 + 0.4637 \times \log(body\ mass);\ r^2 = 87.5\%$$

g. The residual plot for the least squares line follows:



Yes, the points in this plot are fairly randomly scattered, indicating the linear model is reasonable for the transformed data.

h. 10 kilograms: predicted trot speed = 0.8209 + 0.4637 × log(10) = 1.2846 m/sec

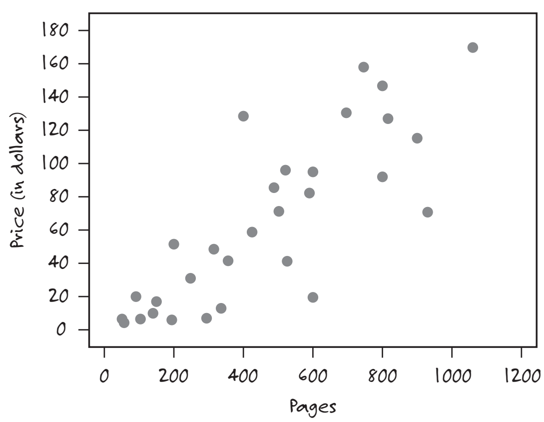
100 kilograms: predicted trot speed = 0.8209 + 0.4637 × log(100) = 1.7483 m/sec

i. The difference between these predictions is 0.4637 m/sec = slope of regression line.

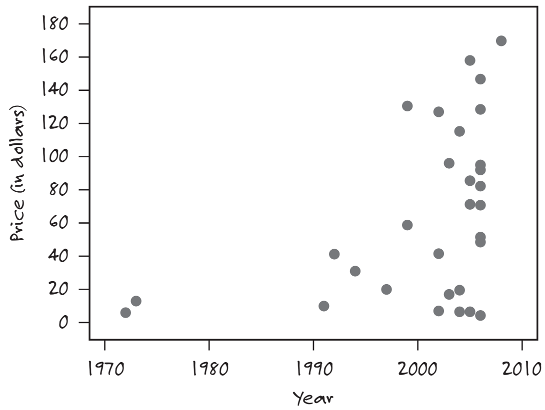
Activity 28-4: Textbook Prices

a. It seems reasonable to regard *price* as the response variable because it is natural to take an interest in predicting a textbook's price from other variables.

b. These scatterplots follow:



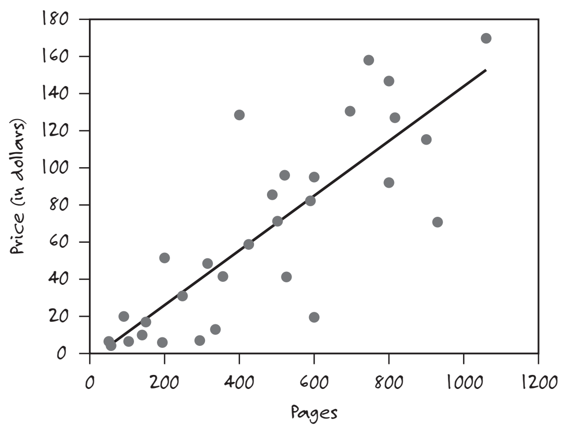
28



The scatterplot of *price* vs. *pages* reveals a fairly strong positive linear association. The scatterplot of *price* vs. *year* also indicates a positive association, but the association is much weaker and not very linear. Two unusual textbooks from the early 1970s, with very low prices, appear to be outliers, because they differ substantially from the pattern of the other textbooks, and potentially influential observations.

c. Number of pages appears to be a much better predictor of price than year. The relationship is much stronger and also more linear.

d. The equation of this least squares line is *predicted price* = −3.42 + 0.147 *pages*. It is shown on the scatterplot here:

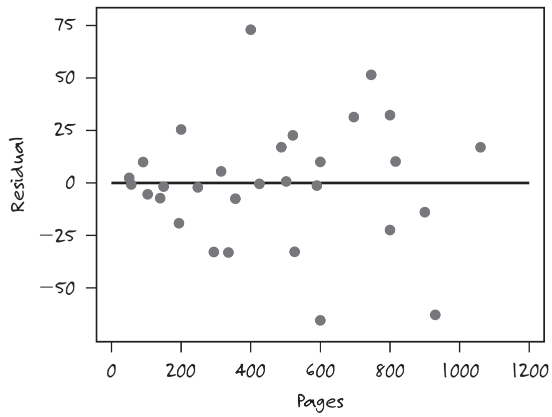


e. The predicted price for a 500-page textbook is predicted price = −3.42 + 0.147(500) = $70.08.

f. The slope coefficient is $0.147/page, which indicates the predicted price of a textbook increases by $0.147 (about 15 cents) for each additional page.

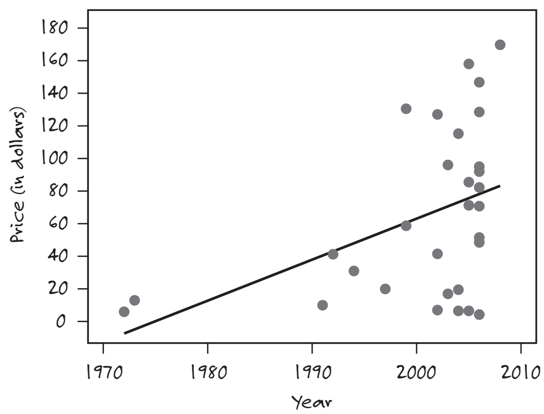
g. The value of the correlation coefficient between *price* and *pages* is $r = 0.823$, so $r^2 = (0.823)^2 = 0.677$. This coefficient says that 67.7% of the variability in textbook prices is explained by the least squares line with number of pages. The other 32.3% is explained by other factors, which could include random variation.

h. A plot of *residual* vs. *pages* follows:

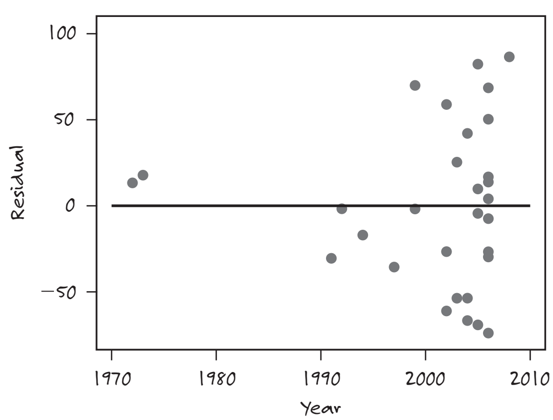


This residual plot reveals no obvious pattern, which suggests the least squares line is a reasonable model for the relationship between *price* and *pages*.

i. The equation of this least squares line is *predicted price* $= -4969 + 2.516$ *year*, with $r^2 = .186$. The least squares line is shown on the scatterplot here:



j. The residual plot is shown here:



This plot reveals that the middle years (1990−2000) tend to have negative residuals. This pattern suggests a linear model is not very appropriate for predicting price from year of publication. A transformation might lead to a more appropriate linear model.

Solutions

In-Class Activities

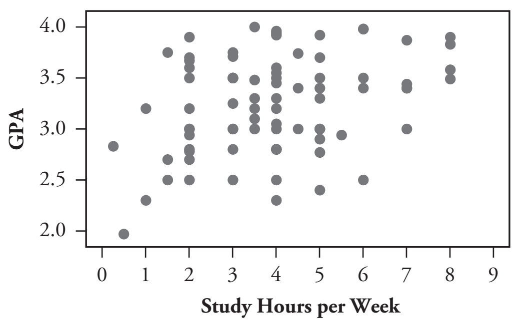
Activity 29-1: Studying and Grades

a. Observational units: students at UOP

Explanatory variable: *study hours per week* Type: quantitative

Response variable: *GPA* Type: quantitative

b. The scatterplot of *GPA* vs. *study hours* follows:



There is a weak positive association between these variables.

c. $\widehat{GPA} = 2.89 + .0894 \times (\textit{study hours})$

d. The value of r^2 is 11.8%, which means that only 11.8% of the variation in the GPAs is explained by the least squares line with study hours per week. The remaining 89.2% is due to other factors, including random variation.

e. This equation predicts the GPA will rise by 0.0894 points on average for each additional hour of study. This number is the slope in the regression equation.

f. This least squares line is based on a sample. Its coefficients are statistics because they are based on a sample.

g. No; if the student researchers took another sample of 80 students, they probably would not get the exact same regression equation from this new sample. They should get a similar equation, but sampling variability would most likely cause the equation to be somewhat different.

h. Answers will vary with each running of the applet. The following is one representative set:

$$\hat{y} = -0.023x + 3.38$$

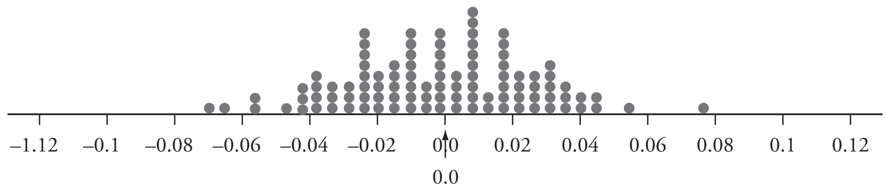
No, this is not the same equation as in part c.

i. $\hat{y} = -0.031x + 3.348$ $\hat{y} = -0.032x + 3.339$

Yes, you get a different sample line each time.

j. The generated regression lines appear to rotate about the middle, sometimes with a positive slope and sometimes with a negative slope.

k. The following graph displays the results:



The distribution of slope coefficients is roughly normal, centered at zero, with a standard deviation of 0.028.

l. None of the 100 simulated sample slopes are ≥ 0.0894. This suggests that the *p*-value for testing whether the population slope is zero against the alternative hypothesis that the population slope is positive is roughly zero. You would therefore conclude that the population slope is positive.

m. The null hypothesis is that there is no relationship between the population hours studied and GPA, or the slope of the regression line between these two variables is zero. In symbols, H_0: $\beta = 0$.

The alternative hypothesis is that students who spend more hours studying tend to have higher GPAs, or the slope of the regression line between these two variables is positive. In symbols, H_a: $\beta > 0$.

n. Using Minitab, SE(*b*) is 0.02771. Yes, this is very close to the standard deviation of the 100 simulated sample slope coefficients.

o. The test statistic is $t = \dfrac{0.0894}{0.02771} = 3.23$.

p. You calculate df $= 80 - 2 = 78$. Using Table III with 60 degrees of freedom, $.001 < p\text{-value} < .005$.

Yes, this *p*-value is consistent with the simulation results, which found a *p*-value of roughly zero.

q. Yes, this *p*-value suggests the association found in the sample of 80 students would happen by random chance alone less than .1% of the time if there were no association between *GPA* and *hours studied* in the population.

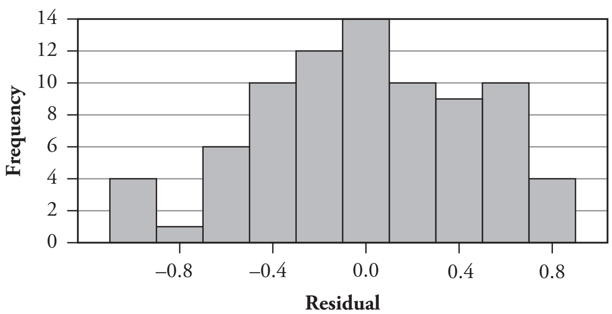
r. Using $t^* = 2.000$ with 60 degrees of freedom, a 95% CI for β is $0.0894 \pm 2 \times (0.02771)$, which is (0.034, 0.145).

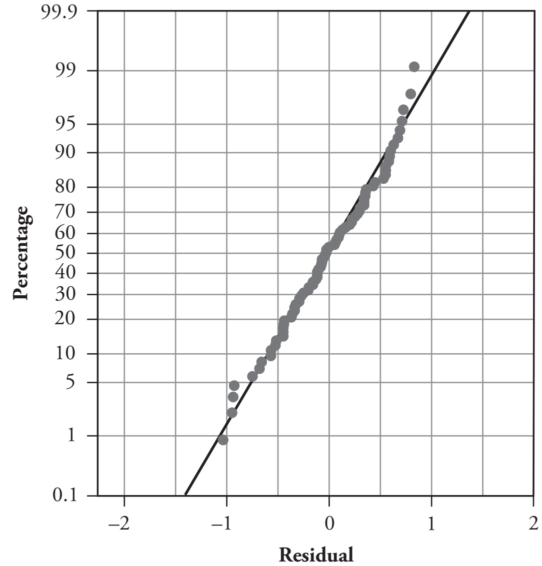
s. You are 95% confident the slope of the population regression line between *GPA* and *hours studied* is somewhere between 0.034 and 0.145 points/hour.

t. You have strong statistical evidence of a positive association between *GPA* and *hours studied* at UOP. You are 95% confident that an additional hour of study corresponds to an average increase of 0.034 to 0.145 points in overall student GPA. You are not necessarily drawing a cause-and-effect relationship, however, as this is an observational study and not an experiment.

Activity 29-2: Studying and Grades

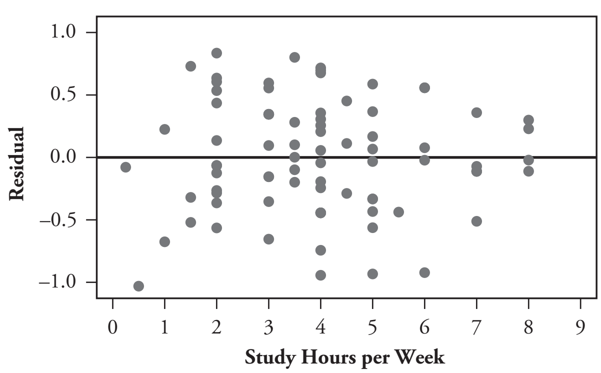
a. The following histogram and normal probability plot display the distribution of the residuals.





The distribution of residuals appears roughly normal. These plots do not reveal any marked features suggesting nonnormality.

b. The scatterplot of *residual* vs. *study hours* follows:



This residual plot does not reveal a strong pattern (curvature). The variability of the residuals does not differ substantially at various *x*-values (although it is a little different at the lowest and highest *x*-values than it is at the middle *x*-values).

Activity 29-3: House Prices

a. The equation is $\widehat{price} = \$265{,}222 + 168.6 \times (size)$.

SE(b) = 31.88 ($\$/ft^2$)

b. The null hypothesis is that there is no relationship in the population between *house size* and *price,* or the slope of the population regression line between these two variables is zero. In symbols, $H_0: \beta = 0$.

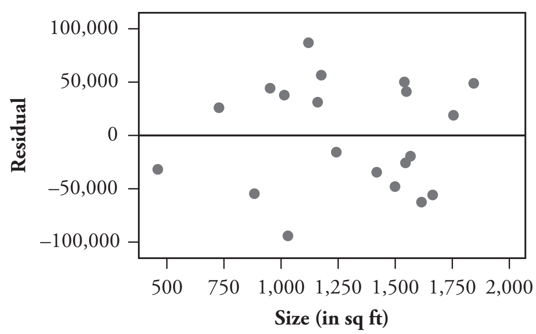
The alternative hypothesis is that larger houses in this population tend to have greater purchase prices, or the slope of the population regression line between these two variables is positive. In symbols, H_a: $\beta > 0$.

The test statistic is $t = \dfrac{168.6}{31.88} = 5.29$.

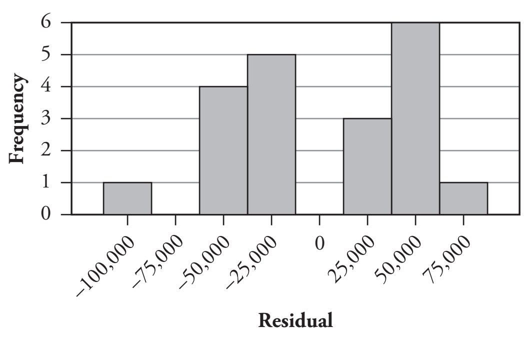
Using Table III with 18 degrees of freedom, p-value $< .0005$. Using Minitab, the p-value is .000.

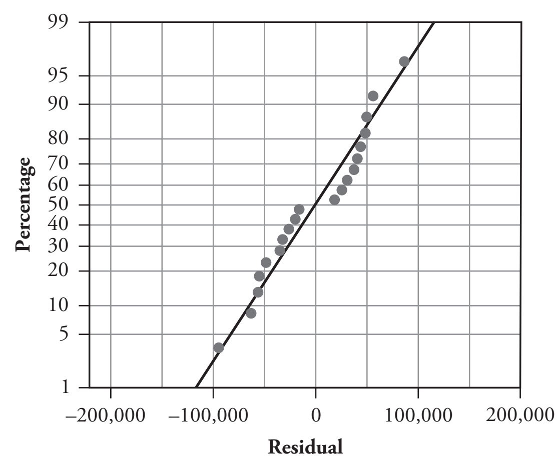
With the small p-value, reject H_0 at any commonly used significance level, and conclude there is very strong statistical evidence of a positive linear association between *house size* and *price* in this population.

The residual plot follows:



A residual plot does not reveal any strong curvature and shows that the standard deviations of the y-values can be reasonably considered the same at each x-value. The data are from a simple random sample, so technical conditions 1, 2, and 4 are met. A histogram and probability plot indicate the residuals are not beautifully normal, but the nonnormality is probably not strong enough to convince you that technical condition 3 has been violated.





c. A 90% CI for β is $168.6 \pm (1.734) \times (31.88) = (113.32, 223.88)$.

You are 90% confident the slope of the population regression line between *house price* and *size* is between 113.22 and 223.88 \$/ft^2. Therefore, you are 90% confident the average increase in the price of a house is between \$113.22 and \$223.88 for each additional square foot of area in a house in Arroyo Grande, California.

The value zero is not in this interval, which is consistent with the test result in part b where you found that zero is not a plausible value for the population slope coefficient.

Activity 29-4: Plasma and Romance

a. The scatterplot and correlation coefficient indicate there is a weak positive linear association between *plasma levels* and *passionate feelings.*

b. The correlation coefficient is a statistic because it is computed from a sample.

c. Yes, it is possible to obtain such a large correlation coefficient by random chance, even if there is no correlation between these variables in the population.

d. H_0: There is no correlation between *plasma levels* and *strength of romantic feelings* in the population, $\rho = 0$.

H_a: There is a positive correlation between *plasma levels* and *strength of romantic feelings,* $\rho > 0$.

e. The test statistic is $t = \dfrac{.347 \times \sqrt{56}}{\sqrt{1-(.347)^2}} \approx \dfrac{2.597}{\sqrt{1-(.1204)}} = 2.77$.

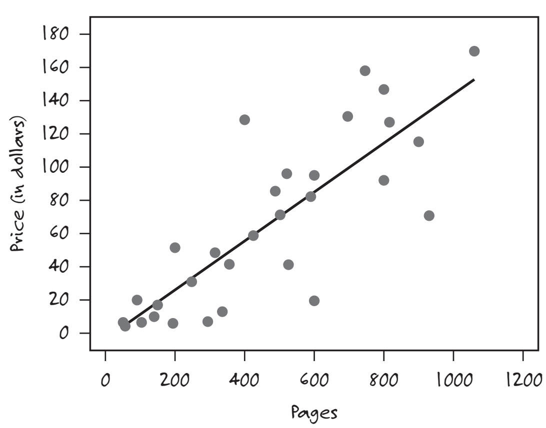
f. You calculate df = 58 − 2 = 56. Using Table III with 50 degrees of freedom, $.001 < p\text{-value} < .005$.

g. With the small *p*-value, you have strong statistical evidence there is a positive correlation between plasmas levels and strength of romantic feelings in this population. If there were no correlation in the population, you would obtain a sample result (a correlation coefficient as great or greater than .347) less than .1% of the time by random sampling alone. Because this is so unlikely to have occurred by random chance, but did happen, you conclude there must be a positive correlation between these variables in the population.

Technical conditions: Neither has been assumed met. The sample was not a simple random sample, and you have no way of testing whether the variables are normally distributed because you only have the summary statistic *r* and the sample size *n* (which is large).

Activity 29-5: Textbook Prices

a. The scatterplot with least squares line follows:



The equation of this line is *predicted price* = −3.42 + 0.147 × (*pages*). The value of r^2 is .677, indicating that 67.7% of the variability in textbook prices can be explained by the number of pages in the texts.

b. The slope coefficient is $b = 0.147$, and technology reports the standard error to be 0.019. The slope indicates that the predicted price increases by \$0.147 on average for each additional page in the book. The standard error measures the variability in the sample slopes from repeated random samples of 30 textbooks from this population.

c. The population of interest in this study is all textbooks in the Cal Poly Bookstore in November 2006.

d. Let β represent the slope of the least squares line for predicting *textbook price* from *number of pages* in the population. Then the null hypothesis is H_0: $\beta = 0$, meaning there is *no* association between *textbook price* and *number of pages* in the population. The alternative is H_a: $\beta > 0$, meaning there is a positive association between these variables in the population. The test statistic is

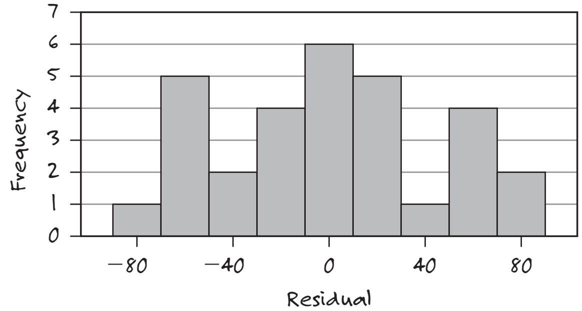
$$t = \frac{b}{SE(b)} = \frac{0.147}{0.019} \approx 7.74$$

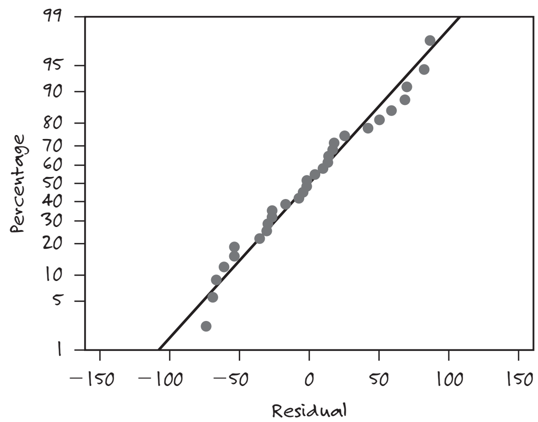
(Without rounding, technology reports the test statistic to be 7.65.) Comparing this to the t-distribution using Table III with 30 − 2 or 28 degrees of freedom reveals that the p-value is much less than .0005. Such a small p-value provides extremely strong evidence of a positive relationship between a textbook's price and its number of pages in the population of all textbooks in that bookstore at that time.

e. Using Table III, the t^* critical value for 90% confidence with 28 degrees of freedom is 1.701. A 90% confidence interval for the population slope β is $0.147 \pm 1.701(0.019)$, which is 0.147 ± 0.032, which is the interval from 0.115 through 0.179. You can be 95% confident that in the population of all textbooks in that bookstore, the predicted price of a textbook increases between 11.5 and 17.9 cents for each additional page.

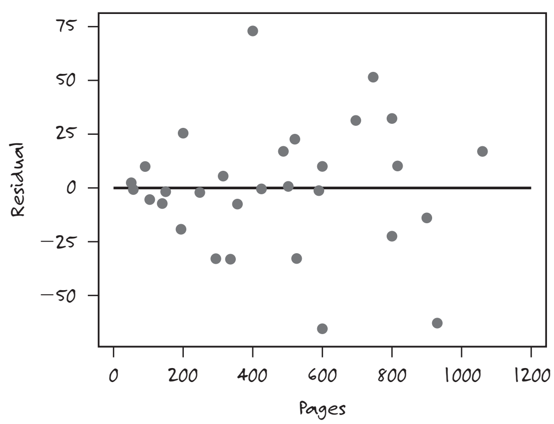
f. A 99% confidence interval for the population slope β has the same midpoint, namely the sample slope 0.147. But the 99% interval is wider than the 90% interval in order to achieve higher confidence of capturing the population slope coefficient.

g. Technical conditions: First, Shaffer and Kaplan did take a random sample of textbooks, so that condition is satisfied as long as you restrict your population to the Cal Poly Bookstore in November 2006. To check the normality condition, consider the following histogram and normal probability plot of the residuals.





Both plots indicate the distribution of the residuals is approximately normal, so that condition is satisfied. For the conditions regarding linearity and equal variability, consider a plot of *residual* vs. *pages:*



There is no obvious pattern to the residuals in this graph, so the linearity condition is met. The variability in residuals appears to be similar across all values of number of pages, although there might be a bit more variability in the residuals for larger numbers of pages. All technical conditions are met, so the significance test and confidence interval are valid.