APPENDIX C

Answers

Here are the "answers" to the exercises for the chapters and the unit reviews. As we said in Chapter 1, the answers are outlines of the complete solution. Your solution should follow the model of the Step-By-Step examples, where appropriate. You should explain the context, show your reasoning and calculations, and draw conclusions. For some problems, what you decide to include in an argument may differ somewhat from the answers here. But, of course, the numerical part of your answer should match the numbers in the answers shown.

CHAPTER 2

1. Categorical
3. Quantitative
5. Answers will vary.
7. *Who*—2500 cars
What—Distance from car to bicycle
Population—All cars passing bicyclists
9. *Who*—Coffee drinkers at a Newcastle University coffee station
What—Amount of money contributed
Population—All people in honor system payment situations
11. *Who*—25,892 men aged 30 to 87
What—Fitness level and cause of death
Population—All men
13. *Who*—54 bears
Cases—Each bear is a case.
What—Weight, neck size, length, and sex
When—Not specified
Where—Not specified
Why—To estimate weight from easier-to-measure variables
How—Researchers collected data on 54 bears they were able to catch.
Variable—Weight
Type—Quantitative
Units—Not specified
Variable—Neck size
Type—Quantitative
Units—Not specified
Variable—Length
Type—Quantitative
Units—Not specified
Variable—Sex
Type—Categorical
15. *Who*—Arby's sandwiches
Cases—Each sandwich is a case.
What—Type of meat, number of calories, and serving size
When—Not specified
Where—Arby's restaurants
Why—To assess nutritional value of sandwiches
How—Report by Arby's restaurants
Variable—Type of meat
Type—Categorical
Variable—Number of calories
Type—Quantitative
Units—Calories
Variable—Serving size
Type—Quantitative
Units—Ounces
17. *Who*—882 births
Cases—Each of the 882 births is a case.
What—Mother's age, length of pregnancy, type of birth, level of prenatal care, birth weight of baby, sex of baby, and baby's health problems
When—1998–2000
Where—Large city hospital
Why—Researchers were investigating the impact of prenatal care on newborn health.
How—Not specified exactly, but probably from hospital records
Variable—Mother's age
Type—Quantitative
Units—Not specified; probably years
Variable—Length of pregnancy
Type—Quantitative
Units—Weeks
Variable—Birth weight of baby
Type—Quantitative
Units—Not specified, probably pounds and ounces
Variable—Type of birth
Type—Categorical
Variable—Level of prenatal care
Type—Categorical
Variable—Sex
Type—Categorical
Variable—Baby's health problems
Type—Categorical
19. *Who*—Experiment subjects
Cases—Each subject is an individual.
What—Treatment (herbal cold remedy or sugar solution) and cold severity
When—Not specified

Where—Not specified
Why—To test efficacy of herbal remedy on common cold
How—The scientists set up an experiment.
Variable—Treatment
Type—Categorical
Variable—Cold severity rating
Type—Quantitative (perhaps ordinal categorical)
Units—Scale from 0 to 5
Concerns—The severity of a cold seems subjective and difficult to quantify. Scientists may feel pressure to report negative findings of herbal product.

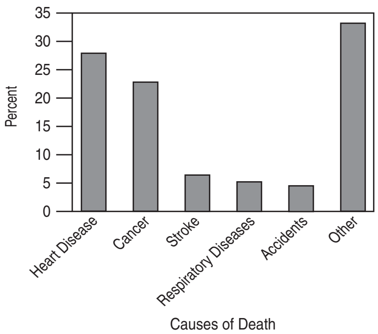
21. *Who*—Streams
Cases—Each stream is a case.
What—Name of stream, substrate of the stream, acidity of the water, temperature, BCI
When—Not specified
Where—Upstate New York
Why—To study ecology of streams
How—Not specified
Variable—Stream name
Type—Identifier
Variable—Substrate
Type—Categorical
Variable—Acidity of water
Type—Quantitative
Units—pH
Variable—Temperature
Type—Quantitative
Units—Degrees Celsius
Variable—BCI
Type—Quantitative
Units—Not specified

23. *Who*—41 refrigerator models
Cases—Each of the 41 refrigerator models is a case.
What—Brand, cost, size, type, estimated annual energy cost, overall rating, and repair history
When—2006
Where—United States
Why—To provide information to the readers of *Consumer Reports*
How—Not specified
Variable—Brand
Type—Categorical
Variable—Cost
Type—Quantitative
Units—Not specified (dollars)
Variable—Size
Type—Quantitative
Units—Cubic feet
Variable—Type
Type—Categorical
Variable—Estimated annual energy cost
Type—Quantitative
Units—Not specified (dollars)
Variable—Overall rating
Type—Categorical (ordinal)
Variable—Percent requiring repair in last 5 years
Type—Quantitative
Units—Percent

25. *Who*—Kentucky Derby races
What—Date, winner, margin, jockey, net proceed to winner, duration, track condition
When—1875 to 2008
Where—Churchill Downs, Louisville, Kentucky
Why—Not specified (To see trends in horse racing?)
How—Official statistics collected at race
Variable—Year
Type—Quantitative
Units—Day and year
Variable—Winner
Type—Identifier
Variable—Margin
Type—Quantitative
Units—Horse lengths
Variable—Jockey
Type—Categorical
Variable—Net proceeds to winner
Type—Quantitative
Units—Dollars
Variable—Duration
Type—Quantitative
Units—Minutes and seconds
Variable—Track condition
Type—Categorical

CHAPTER 3

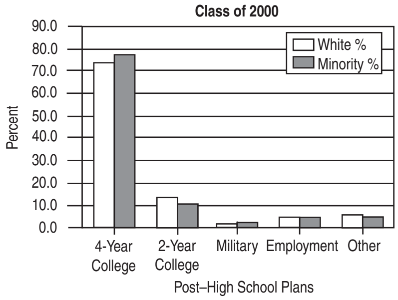
1. Answers will vary.
3. Answers will vary.
5. a) Yes; each is categorized in a single genre.
b) Thriller/Horror
7. a) Comedy
b) It is easier to tell from the bar chart; slices of the pie chart are too close in size.
9. 1755 students applied for admission to the magnet schools program. 53% were accepted, 17% were wait-listed, and the other 30% were turned away.
11. a) Yes. We can add because these categories do not overlap. (Each person is assigned only one cause of death.)
b) $100 - (27.2 + 23.1 + 6.3 + 5.1 + 4.7) = 33.6\%$
c) Either a bar chart or pie chart with "other" added would be appropriate. A bar chart is shown.



13. a) The bar chart shows that grounding and collision are the most frequent causes of oil spills. Very few have unknown causes.
b) A pie chart seems appropriate as well.
15. There's no title, the percentages total only 92%, and the three-dimensional display distorts the sizes of the regions.
17. In both the South and West, about 58% of the eighth-grade smokers preferred Marlboro. Newport was the next most popular brand, but was far more popular in the South than in the West, where Camel was cited nearly 3 times as often as in the South. Nearly twice as many smokers in the West as in the South indicated that they had no usual brand (12.9% to 6.7%).
19. a) The column totals are 100%.
b) 31.7%
c) 60%
d) i. 35.7%; ii. can't tell; iii. 0%; iv. can't tell
21. a) 82.5% b) 12.9% c) 11.1%
d) 13.4% e) 85.7%
23. a) 73.9% 4-yr college, 13.4% 2-year college, 1.5% military, 5.2% employment, 6.0% other

b) 77.2% 4-yr college, 10.5% 2-year college, 1.8% military, 5.3% employment, 5.3% other

c) Many charts are possible. Here is a side-by-side bar chart.



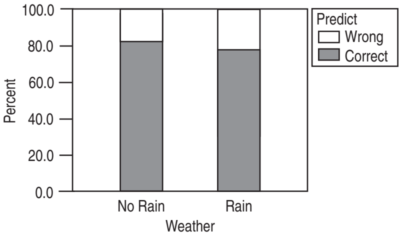
d) The white and minority students' plans are very similar. The small differences should be interpreted with caution because the total number of minority students is small. There is little evidence of an association between race and plans.

25. a) 16.6% b) 11.8% c) 37.7% d) 53.0%

27. 1755 students applied for admission to the magnet schools program: 53% were accepted, 17% were wait-listed, and the other 30% were turned away. While the overall acceptance rate was 53%, 93.8% of blacks and Hispanics were accepted, compared to only 37.7% of Asians and 35.5% of whites. Overall, 29.5% of applicants were black or Hispanic, but only 6% of those turned away were. Asians accounted for 16.6% of all applicants, but 25.4% of those turned away. Whites were 54% of the applicants and 68.5% of those who were turned away. It appears that the admissions decisions were not independent of the applicant's ethnicity.

29. a) 9.3% b) 24.7% c) 80.8%

d) No, there appears to be no association between weather and ability to forecast weather. On days it rained, his forecast was correct 79.4% of the time. When there was no rain, his forecast was correct 81.0% of the time.

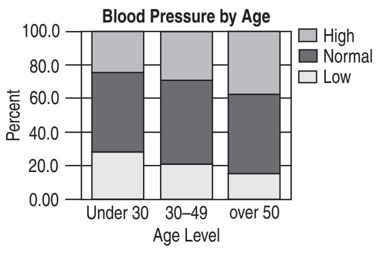


31. a) Low 20.0%, Normal 48.9%, High 31.0%

b)

Blood Pressure	Under 30	30–49	Over 50
Low	27.6%	20.7%	15.7%
Normal	49.0%	50.8%	47.2%
High	23.5%	28.5%	37.1%

c)



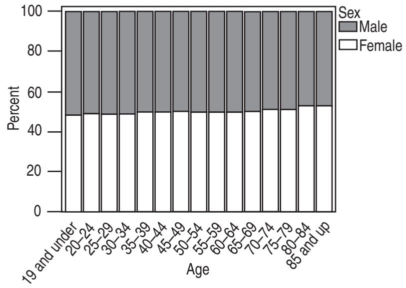
d) As age increases, the percent of adults with high blood pressure increases. By contrast, the percent of adults with low blood pressure decreases.

e) No, but it gives an indication that it might. There might be additional reasons that explain the differences in blood pressures.

33. No, there's no evidence that Prozac is effective. The relapse rates were nearly identical: 28.6% among the people treated with Prozac, compared to 27.3% among those who took the placebo.

35. a) 4.7% b) 50.0%.

c) There are about 50% of each sex in each age group, but it ranges from 48.8% female in the youngest group to 54.6% in the oldest. As the age increases, there is a slight increase in the percentage of female drivers.



d) There is a slight association. As the age increases, there is a small increase in the percentage of female drivers.

37. a) 160 of 1300, or 12.3%

b) Yes. Major surgery: 15.3% vs. minor surgery: 6.7%

c) Large hospital: 13%; small hospital: 10%

d) Large hospital: Major 15% vs. minor 5%
Small hospital: Major 20% vs. minor 8%

e) No. Smaller hospitals have a higher rate for both kinds of surgery, even though it's lower "overall."

f) The small hospital has a larger percentage of minor surgeries (83.3%) than the large hospital (20%). Minor surgeries have a lower delay rate, so the small hospital looks better "overall."

39. a) 42.6%

b) A higher percentage of males than females were admitted: Males: 47.2% to females: 30.9%

c) Program 1: Males 61.9%, females 82.4%
Program 2: Males 62.9%, females 68.0%
Program 3: Males 33.7%, females 35.2%
Program 4: Males 5.9%, females 7.0%

d) The comparisons in c) show that males have a lower admittance rate in every program, even though the overall rate shows males with a higher rate of admittance. This is an example of Simpson's paradox.

CHAPTER 4

1. Answers will vary.

3. Answers will vary.

5. a) Unimodal (near 0) and skewed to the right. Many seniors will have 0 or 1 speeding tickets. Some may have several, and a few may have more than that.

b) Probably unimodal and slightly skewed to the right. It is easier to score 15 strokes over the mean than 15 strokes under the mean.

c) Probably unimodal and symmetric. Weights may be equally likely to be over or under the average.

d) Probably bimodal. Men's and women's distributions may have different modes. It may also be skewed to the right, since it is possible to have very long hair, but hair length can't be negative.

7. a) Bimodal. Looks like two groups. Modes are near 6% and 46%. No real outliers.
b) Looks like two groups of cereals, a low-sugar and a high-sugar group.

9. a) 78%
b) Skewed to the right with at least one high outlier. Most of the vineyards are less than 90 acres with a few high ones. The mode is between 0 and 30 acres.

11. a) Because the distribution is skewed to the right, we expect the mean to be larger.
b) Bimodal and skewed to the right. Center mode near 8 days. Another mode at 1 day (may represent patients who didn't survive). Most of the patients stay between 1 and 15 days. There are some extremely high values above 25 days.
c) The median and IQR, because the distribution is strongly skewed.

13. a) 45 points b) 37 points and 54 (or 55) points
c) In the Super Bowl teams typically score a total of about 45 points, with half the games totaling between 37 and 55 points. In only one fourth of the games have the teams scored fewer than 27 points, and they once totaled 75.

15. a) The standard deviation will be larger for set 2, since the values are more spread out. SD(set 1) = 2.2, SD(set 2) = 3.2.
b) The standard deviation will be larger for set 2, since 11 and 19 are farther from 15 than are 14 and 16. Other numbers are the same. SD(set 1) = 3.6, SD(set 2) = 4.5.
c) The standard deviation will be the same for both sets, since the values in the second data set are just the values in the first data set + 80. The spread has not changed. SD(set 1) = 4.2, SD(set 2) = 4.2.

17. The mean and standard deviation because the distribution is unimodal and symmetric.

19. a) The mean is closest to $2.60 because that's the balancing point of the histogram.
b) The standard deviation is closest to $0.15 since that's a typical distance from the mean. There are no prices as far as $0.50 or $1.00 from the mean.

21. a) About 100 minutes
b) Yes, only 4 of these movies run that long.
c) The mean would be higher. The distribution is skewed high.

23. a) i. The middle 50% of movies ran between 97 and 119 minutes.
ii. On average, movie lengths varied from the mean run time by 19.6 minutes.
b) We should be cautious in using the standard deviation because the distribution of run times is skewed to the right.

25. a) The median will probably be unaffected. The mean will be larger.
b) The range and standard deviation will increase; the IQR will be unaffected.

27. The publication is using the median; the watchdog group is using the mean, pulled higher by the several very expensive movies in the long right tail.

29. a) Mean $525, median $450
b) 2 employees earn more than the mean.
c) The median because of the outlier.
d) The IQR will be least sensitive to the outlier of $1200, so it would be the best to report.

31. a)

Stem	Leaf
25	
25	
24	56
24	
23	68
23	23
22	677789
22	1234

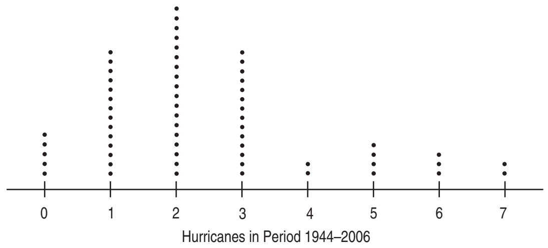
22|1 = $2.21/gallon

b) The distribution of gas prices is unimodal and skewed to the right (upward), centered around $2.27, with most stations charging between $2.26 and $2.33 per gallon. The lowest and highest prices were $2.21 and $2.46.
c) There are two high prices separated from the other gas stations by a gap.

33. a) Since these data are strongly skewed to the right, the median and IQR are the best statistics to report.
b) The mean will be larger than the median because the data are skewed to the right.
c) The median is 4 million. The IQR is 4.5 million (Q3 = 6 million, Q1 = 1.5 million).
d) The distribution of populations of the states and Washington, DC, is unimodal and skewed to the right. The median population is 4 million. One state is an outlier, with a population of 34 million.

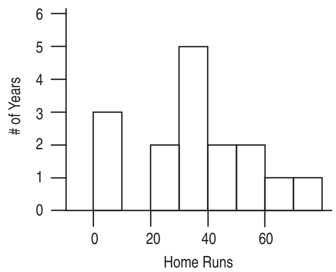
35. Skewed to the right, median at 36. Three low outliers, then a gap from 9 to 22.

37. a)



b) Slightly skewed to the right. Unimodal, mode near 2. Possibly a second mode near 5. No outliers.

39. a) This is not a histogram. The horizontal axis should split the number of home runs hit in each year into bins. The vertical axis should show the number of years in each bin.
b)



41. Skewed to the right, possibly bimodal with one fairly symmetric group near 4.4, another at 5.6. Two outliers in middle seem not to belong to either group.

Stem	Leaf
57	8
56	27
55	1
54	
53	
52	9
51	
50	8
49	
48	2
47	3
46	034
45	267
44	015
43	0199
42	669
41	22

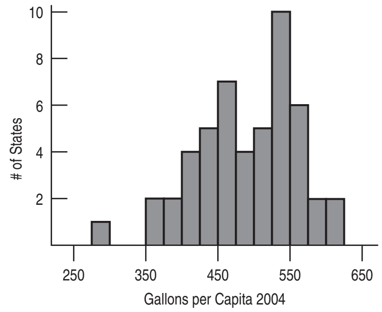
41|2 = 4.12 pH

43. Histogram bins are too wide to be useful.

45. Neither appropriate nor useful. Zip codes are categorical data, not quantitative. But they do contain *some* information. The leading digit gives a rough East-to-West placement in the United States. So, we see that they have almost no customers in the Northeast, but a bar chart by leading digit would be more appropriate.

47) a) Median 239, IQR 9, Mean 237.6, SD 5.7
b) Because it's skewed to the left, probably better to report Median and IQR.
c) Skewed to the left; may be bimodal. The center is around 239. The middle 50% of states scored between 233 and 242. Alabama, Mississippi, and New Mexico scores were much lower than other states' scores.

49. In the year 2004, per capita gasoline use by state in the United States averaged around 500 gallons per person (mean 488.8, median 500.5). States varied in per capita consumption, with a standard deviation of 68.7 gallons. The only outlier is New York. The IQR of 96.9 gallons shows that 50% of the states had per capita consumption of between 447.5 and 544.4 gallons. The data appear to be bimodal, so the median and IQR are better choices of summary statistics.



CHAPTER 5

1. Answers will vary.

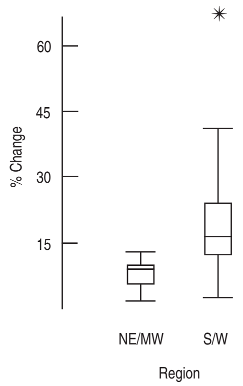
3. Answers will vary.

5. a) Prices appear to be both higher on average and more variable in Baltimore than in the other three cities. Prices in Chicago may be slightly higher than in Dallas and Denver, but the difference is very small.
b) There are outliers on the low end in Baltimore and Chicago and one high outlier in Dallas, but these do not affect the overall conclusions reached in part a).

7. a) Essentially symmetric, very slightly skewed to the right with two high outliers at 36 and 48. Most victims are between the ages of 16 and 24.
b) The slight increase between ages 22 and 24 is apparent in the histogram but not in the boxplot. It may be a second mode.
c) The median would be the most appropriate measure of center because of the slight skew and the extreme outliers.
d) The IQR would be the most appropriate measure of spread because of the slight skew and the extreme outliers.

9. a) About 59% b) Bimodal
c) Some cereals are very sugary; others are healthier low-sugar brands.
d) Yes
e) Although the ranges appear to be comparable for both groups (about 28%), the IQR is larger for the adult cereals, indicating that there's more variability in the sugar content of the middle 50% of adult cereals.

11. a)



b) Growth rates in NE/MW states are tightly clustered near 5%. S/W states are more variable, and bimodal with modes near 14 and 22. The S/W states have an outlier as well. Around all the modes, the distributions are fairly symmetric.

13. a) They should be put on the same scale, from 0 to 20 days.
b) Lengths of men's stays appear to vary more than for women. Men have a mode at 1 day and then taper off from there. Women have a mode near 5 days, with a sharp drop afterward.
c) A possible reason is childbirth.

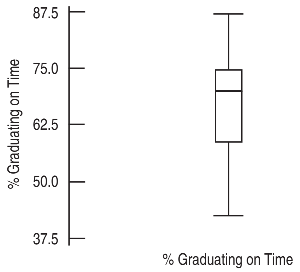
15. a) Both girls have a median score of about 17 points per game, but Scyrine is much more consistent. Her IQR is about 2 points, while Alexandra's is over 10.
b) If the coach wants a consistent performer, she should take Scyrine. She'll almost certainly deliver somewhere between 15 and 20 points. But if she wants to take a chance and needs a "big game," she should take Alexandra. Alex scores over 24 points about a quarter of the time. (On the other hand, she scores under 11 points as often.)

17. Women appear to marry about 3 years younger than men, but the two distributions are very similar in shape and spread.

19. (*Note:* Numerical details may vary.) In general, fuel economy is higher in cars than in either SUVs or vans. There are numerous outliers on both ends for cars and a few high outliers for SUVs. The top 50% of cars gets higher fuel economy than 75% of SUVs and nearly all vans. On average, SUVs and vans get about the same fuel economy, although the distribution for vans shows less spread. The range for vans is about 40 mpg, while for SUVs it is nearly 30 mpg.

21. The class A is 1, class B is 2, and class C is 3.

23. a) Probably slightly left skewed. The mean is slightly below the median, and the 25th percentile is farther from the median than the 75th percentile.
b) No, all data are within the fences.
c)



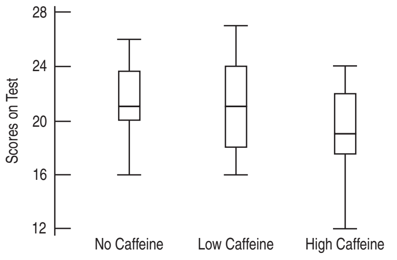
d) The 48 universities graduate, on average, about 68% of freshmen "on time," with percents ranging from 43% to 87%. The middle 50% of these universities graduate between 59% and 75% of their freshmen in 4 years.

25. a) *Who:* Student volunteers
What: Memory test
Where, when: Not specified

How: Students took memory test 2 hours after drinking caffeine-free, half-dose caffeine, or high-caffeine soda.
Why: To see if caffeine makes you more alert and aids memory retention.

b) Drink: categorical; Test score: quantitative.

c)



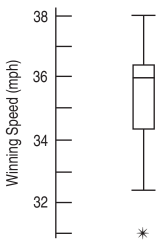
d) The participants scored about the same with no caffeine and low caffeine. The medians for both were 21 points, with slightly more variation for the low-caffeine group. The high-caffeine group generally scored lower than the other two groups on all measures of the 5-number summary: min, lower quartile, median, upper quartile, and max.

27. a) About 36 mph

b) Q_1 about 35 mph and Q_3 about 37 mph

c) The range appears to be about 7 mph, from about 31 to 38 mph. The IQR is about 2 mph.

d) We can't know exactly, but the boxplot may look something like this:



e) The median winning speed has been about 36 mph, with a max of about 38 and a min of about 31 mph. Half have run between about 35 and 37 mph, for an IQR of 2 mph.

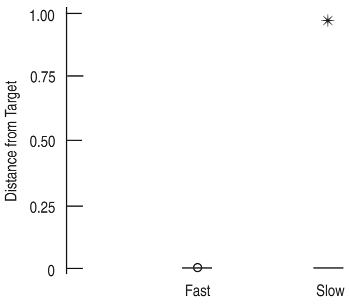
29. a) Boys b) Boys c) Girls

d) The boys appeared to have more skew, as their scores were less symmetric between quartiles. The girls' quartiles are the same distance from the median, although the left tail stretches a bit farther to the left.

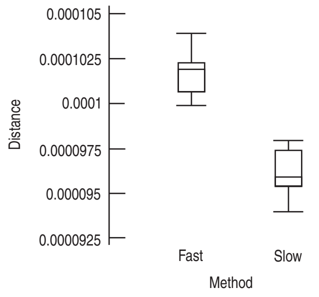
e) Girls. Their median and upper quartiles are larger. The lower quartile is slightly lower, but close.

f) $[14(4.2) + 11(4.6)]/25 = 4.38$

31.

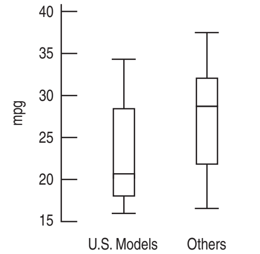


There appears to be an outlier! This point should be investigated. We'll proceed by redoing the plots with the outlier omitted:



It appears that slow speed provides much greater accuracy. But the outlier should be investigated. It is possible that slow speed can induce an infrequent very large distance.

33. a)



b) Mileage for U.S. models is typically lower, although the variability is about the same as for cars made elsewhere. The median for U.S. models is around 21 mpg, compared to 28 for the others. Half of U.S. models fall below the first quartile of others. (Other answers possible.)

35. a) Day 16 (but any estimate near 20 is okay).

b) Day 65 (but anything around 60 is okay).

c) Around day 50

37. a) Most of the data are found in the far left of this histogram. The distribution is very skewed to the right.

b) Re-expressing the data by, for example, logs or square roots might help make the distribution more nearly symmetric.

39. a) The logarithm makes the histogram more symmetric. It is easy to see that the center is around 3.5 in log assets.

b) That has a value of around 2,500 million dollars.

c) That has a value of around 1,000 million dollars.

41. a) Fusion time and group.

b) Fusion time is quantitative (units = seconds). Group is categorical.

c) Both distributions are skewed to the right with high outliers. The boxplot indicates that visual information may reduce fusion time. The median for the Verbal/Visual group seems to be about the same as the lower quartile of the No/Verbal group.

CHAPTER 6

1. a) 72 oz., 40 oz. b) 4.5 lb, 2.5 lb

3. a) Skewed to the right; mean is higher than median.

b) \$350 and \$950.

c) Minimum \$350. Mean \$750. Median \$550. Range \$1200. IQR \$600. Q1 \$400. SD \$400.

d) Minimum \$330. Mean \$770. Median \$550. Range \$1320. IQR \$660. Q1 \$385. SD \$440.

5. Lowest score = 910. Mean = 1230. SD = 120. Q3 = 1350. Median = 1270. IQR = 240.

7. Your score was 2.2 standard deviations higher than the mean score in the class.

9. 65

11. In January, a high of 55 is not quite 2 standard deviations above the mean, whereas in July a high of 55 is more than 2 standard deviations lower than the mean. So it's less likely to happen in July.

13. The z-scores, which account for the difference in the distributions of the two tests, are 1.5 and 0 for Derrick and 0.5 and 2 for Julie. Derrick's total is 1.5, which is less than Julie's 2.5.

15. a) Megan b) Anna

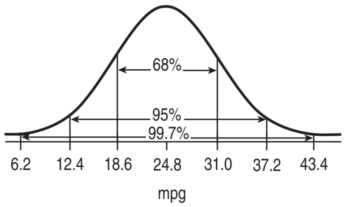
17. a) About 1.81 standard deviations below the mean.
b) 1000 ($z = 1.81$) is more unusual than 1250 ($z = 1.17$).

19. a) Mean $= 1152 - 1000 = 152$ pounds; SD is unchanged at 84 pounds.
b) Mean $= 0.40(1152) = \$460.80$; SD $= 0.40(84) = \$33.60$.

21. Min $= 0.40(980) - 20 = \$372$;
median $= 0.40(1140) - 20 = \$436$;
SD $= 0.40(84) = \$33.60$; IQR $= 0.40(102) = \$40.80$.

23. College professors can have between 0 and maybe 40 (or possibly 50) years' experience. A standard deviation of 1/2 year is impossible, because many professors would be 10 or 20 SDs away from the mean, whatever it is. An SD of 16 years would mean that 2 SDs on either side of the mean is plus or minus 32, for a range of 64 years. That's too high. So, the SD must be 6 years.

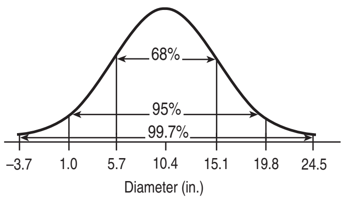
25. a)



b) 18.6 to 31.0 mpg c) 16%
d) 13.5% e) less than 12.4 mpg

27. Any weight more than 2 standard deviations below the mean, or less than $1152 - 2(84) = 984$ pounds, is unusually low. We expect to see a steer below $1152 - 3(84) = 900$ pounds only rarely.

29. a)



b) Between 1.0 and 19.8 inches c) 2.5%
d) 34% e) 16%

31. Since the histogram is not unimodal and symmetric, it is not wise to have faith in numbers from the Normal model.

33. a) 16% b) 3.8%
c) Because the Normal model doesn't fit well.
d) Distribution is skewed to the right.

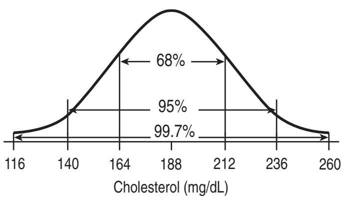
35. a) 2.5%
b) 2.5% of the receivers should gain less than −333 yards, but that's impossible, so the model doesn't fit well.
c) Data are strongly skewed to the right, not symmetric.

37. a) 12.2% b) 71.6% c) 23.3%

39. a) 1259.7 lb b) 1081.3 lb c) 1108 lb to 1196 lb

41. a) 1130.7 lb b) 1347.4 lb c) 113.3 lb

43. a)



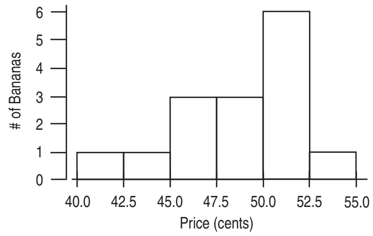
b) 30.85% c) 17.00% d) 32 points e) 212.9 points

45. a) 11.1% b) (35.9, 40.5) inches c) 40.5 inches

47. a) 5.3 grams b) 6.4 grams
c) Younger because SD is smaller.

PART I REVIEW

1. a)



b) Median 49 cents, IQR 6 cents.
c) The distribution is unimodal and left skewed. The center is near 50 cents; values range from 42 cents to 53 cents.

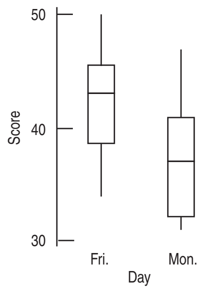
3. a) If enough sopranos have a height of 65 inches, this can happen.
b) The distribution of heights for each voice part is roughly symmetric. The basses are slightly taller than the tenors. The sopranos and altos have about the same median height. Heights of basses and sopranos are more consistent than those of altos and tenors.

5. a) It means their heights are also more variable.
b) The z-score for women to qualify is 2.40, compared with 1.75 for men, so it is harder for women to qualify.

7. a) *Who*—People who live near State University
What—Age, attended college? Favorable opinion of State?
When—Not stated
Where—Region around State U.
Why—To report to the university's directors
How—Sampled and phoned 850 local residents
b) Age—Quantitative (years); attended college?—categorical; favorable opinion?—categorical.
c) The fact that the respondents know they are being interviewed by the university's staff may influence answers.

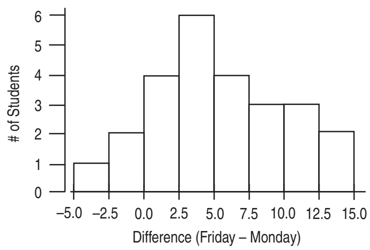
9. a) These are categorical data, so mean and standard deviation are meaningless.
b) Not appropriate. Even if it fits well, the Normal model is meaningless for categorical data.

11. a)



b) The scores on Friday were higher by about 5 points on average. This is a drop of more than 10% off the average score and shows that students fared worse on Monday after preparing for the test on Friday. The spreads are about the same, but the scores on Monday are a bit skewed to the right.

c)



d) The changes (Friday–Monday) are unimodal and centered near 4 points, with a spread of about 5 (SD). They are fairly symmetric, but slightly skewed to the right. Only 3 students did better on Monday (had a negative difference).

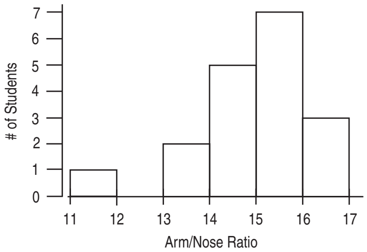
13. a) Categorical
b) Go fish. All you need to do is match the denomination. The denominations are not ordered. (Answers will vary.)
c) Gin rummy. All cards are worth their value in points (face cards are 10 points). (Answers will vary.)

15. a) Annual mortality rate for males (quantitative) in deaths per 100,000 and water hardness (quantitative) in parts per million.
b) Calcium is skewed right, possibly bimodal. There looks to be a mode down near 12 ppm that is the center of a fairly tight symmetric distribution and another mode near 62.5 ppm that is the center of a much more spread out, symmetric (almost uniform) distribution. Mortality, however, appears unimodal and symmetric with the mode near 1500 deaths per 100,000.

17. a) They are on different scales.
b) January's values are lower and more spread out.
c) Roughly symmetric but slightly skewed to the left. There are more low outliers than high ones. Center is around 40 degrees with an IQR of around 7.5 degrees.

19. a) Bimodal with modes near 2 and 4.5 minutes. Fairly symmetric around each mode.
b) Because there are two modes, which probably correspond to two different groups of eruptions, an average might not make sense.
c) The intervals between eruptions are longer for long eruptions. There is very little overlap. More than 75% of the short eruptions had intervals less than about an hour (62.5 minutes), while more than 75% of the long eruptions had intervals longer than about 75 minutes. Perhaps the interval could even be used to predict whether the next eruption will be long or short.

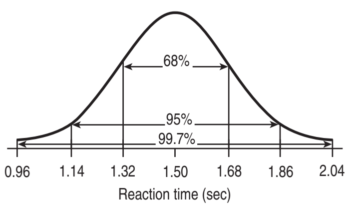
21. a)



The distribution is left skewed with a center of about 15. It has an outlier between 11 and 12.
b) Even though the distribution is somewhat skewed, the mean and median are close. The mean is 15.0 and the SD is 1.25.
c) Yes. 11.8 is already an outlier. 9.3 is more than 4.5 SDs below the mean. It is a very low outlier.

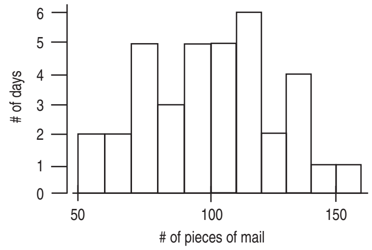
23. If we look only at the overall statistics, it appears that the follow-up group is insured at a much lower rate than those not traced (11.1% of the time compared with 16.6%). But most of the follow-up group were black, who have a lower rate of being insured. When broken down by race, the follow-up group actually has a higher rate of being insured for both blacks and whites. So the overall statistic is misleading and is attributable to the difference in race makeup of the two groups.

25. a)



b) According to the model, reaction times are symmetric with center at 1.5 seconds. About 95% of all reaction times are between 1.14 and 1.86 seconds.
c) 8.2% d) 24.1%
e) Quartiles are 1.38 and 1.62 seconds, so the IQR is 0.24 seconds.
f) The slowest 1/3 of all drivers have reaction times of 1.58 seconds or more.

27. a)



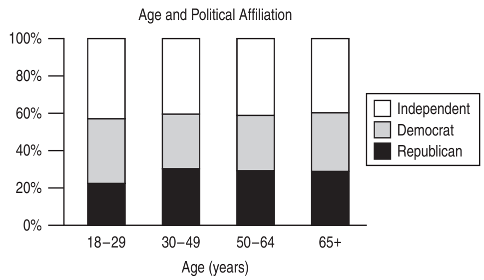
b) Mean 100.25, SD 25.54 pieces of mail.
c) The distribution is somewhat symmetric and unimodal, but the center is rather flat, almost uniform.
d) 64%. The Normal model seems to work reasonably well, since it predicts 68%.

29. a) *Who*—100 health food store customers
What—Have you taken a cold remedy?, and Effectiveness (scale 1 to 10)
When—Not stated
Where—Not stated
Why—Promotion of herbal medicine
How—In-person interviews
b) Have you taken a cold remedy?—categorical. Effectiveness—categorical or ordinal.
c) No. Customers are not necessarily representative, and the Council had an interest in promoting the herbal remedy.

31. a) 38 cars
b) Possibly because the distribution is skewed to the right.
c) Center—median is 148.5 cubic inches. Spread—IQR is 126 cubic inches.
d) No. It's bigger than average, but smaller than more than 25% of cars. The upper quartile is at 231 inches.
e) No. 1.5 IQR is 189, and $105 - 189$ is negative, so there can't be any low outliers. $231 + 189 = 420$. There aren't any cars with engines bigger than this, since the maximum has to be at most 105 (the lower quartile) + 275 (the range) = 380.
f) Because the distribution is skewed to the right, this is probably not a good approximation.
g) Mean, median, range, quartiles, IQR, and SD all get multiplied by 16.4.

33. a) 30.4%
b) If this were a random sample of all voters, yes.
c) 36.6% d) 8.8%
e) 23.1% f) 47.0%

35. a) Republican—16,535, Democrat—17,183, Other— 20,666; or Republican—30.4%, Democrat—31.6%, Other—38.0%.
b)



c) Among voters over 30, political affiliation appears to be largely unrelated to age. However there is some evidence that younger voters are less likely to be Republican
d) Voters who identified themselves as "Other" seem to be generally younger than Democrats or Republicans.

37. a) 0.43 hours. b) 1.4 hours.
c) 0.89 hours (or 53.4 minutes).
d) Survey results vary, and the mean and the SD may have changed.

CHAPTER 7

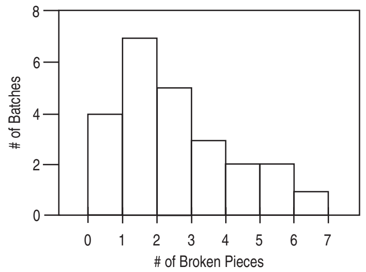
1. a) Weight in ounces: explanatory; Weight in grams: response. (Could be other way around.) To predict the weight in grams based on ounces. Scatterplot: positive, straight, strong (perfectly linear relationship).
b) Circumference: explanatory. Weight: response. To predict the weight based on the circumference. Scatterplot: positive, linear, moderately strong.
c) Shoe size: explanatory; GPA: response. To try to predict GPA from shoe size. Scatterplot: no direction, no form, very weak.
d) Miles driven: explanatory; Gallons remaining: response. To predict the gallons remaining in the tank based on the miles driven since filling up. Scatterplot: negative, straight, moderate.

3. a) Altitude: explanatory; Temperature: response. (Other way around possible as well.) To predict the temperature based on the altitude. Scatterplot: negative, possibly straight, weak to moderate.
b) Ice cream cone sales: explanatory. Air-conditioner sales: response—although the other direction would work as well. To predict one from the other. Scatterplot: positive, straight, moderate.
c) Age: explanatory; Grip strength: response. To predict the grip strength based on age. Scatterplot: curved down, moderate. Very young and elderly would have grip strength less than that of adults.
d) Reaction time: explanatory; Blood alcohol level: response. To predict blood alcohol level from reaction time test. (Other way around is possible.) Scatterplot: positive, nonlinear, moderately strong.

5. a) None b) 3 and 4 c) 2, 3, and 4
d) 1 and 2 e) 3 and possibly 1

7. There seems to be a very weak—or possibly no—relation between brain size and performance IQ.

9. a)



b) Unimodal, skewed to the right. The skew.
c) The positive, somewhat linear relation between batch number and broken pieces.

11. a) 0.006 b) 0.777 c) −0.923 d) −0.487

13. There may be an association, but not a correlation unless the variables are quantitative. There could be a correlation between average number of hours of TV watched per week per person and number of crimes committed per year. Even if there is a relationship, it doesn't mean one causes the other.

15. a) Yes. It shows a linear form and no outliers.
b) There is a strong, positive, linear association between drop and speed; the greater the coaster's initial drop, the higher the top speed.

17. The scatterplot is not linear; correlation is not appropriate.

19. The correlation may be near 0. We expect nighttime temperatures to be low in January, increase through spring and into the summer months, then decrease again in the fall and winter. The relationship is not linear.

21. The correlation coefficient won't change, because it's based on *z*-scores. The *z*-scores of the prediction errors are the same whether they are expressed in nautical miles or miles.

23. a) Assuming the relation is linear, a correlation of −0.772 shows a strong relation in a negative direction.
b) Continent is a categorical variable. Correlation does not apply.

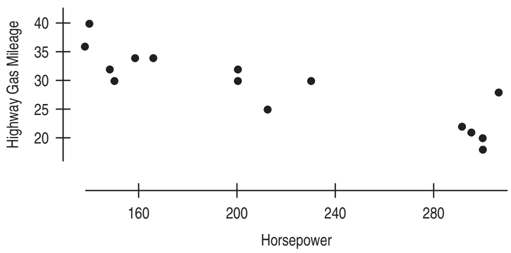
25. a) Actually, yes, taller children will tend to have higher reading scores, but this doesn't imply causation.
b) Older children are generally both taller and are better readers. Age is the lurking variable.

27. a) No. We don't know this from the correlation alone. There may be a nonlinear relationship or outliers.
b) No. We can't tell from the correlation what the form of the relationship is.
c) No. We don't know from the correlation coefficient.
d) Yes, the correlation doesn't depend on the units used to measure the variables.

29. This is categorical data even though it is represented by numbers. The correlation is meaningless.

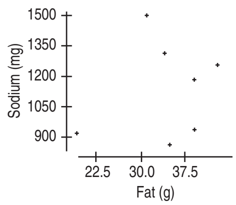
31. a) The association is positive, moderately strong, and roughly straight, with several states whose HCI seems high for their median income and one state whose HCI appears low given its median income.
b) The correlation would still be 0.65.
c) The correlation wouldn't change.
d) DC would be a moderate outlier whose HCI is high for its median income. It would lower the correlation slightly.
e) No. We can only say that higher median incomes are associated with higher housing costs, but we don't know why. There may be other economic variables at work.

33. a)



b) Negative, linear, strong. c) −0.869
d) There is a strong linear relation in a negative direction between horsepower and highway gas mileage. Lower fuel efficiency is associated with higher horsepower.

35.

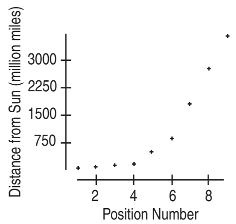


(Plot could have explanatory and predictor variables swapped.) Correlation is 0.199. There does not appear to be a relation between sodium and fat content in burgers, especially without the low-fat, low-sodium item. The correlation of 0.199 shows a weak relationship, even with the outlier included.

37. a) Yes, the scatterplot appears to be somewhat linear.
b) As the number of runs increases, the attendance also increases.
c) There is a positive association, but it does not *prove* that more fans will come if the number of runs increases. Association does not indicate causality.

39. A scatterplot shows a generally straight scattered pattern with no outliers. The correlation between *Drop* and *Duration* is 0.35, indicating that rides on coasters with greater initial drops generally last somewhat longer, but the association is weak.

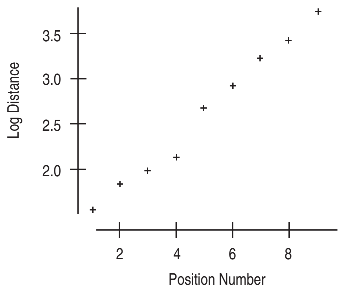
41. a)



The relation between position and distance is nonlinear, with a positive direction. There is very little scatter from the trend.

b) The relation is not linear.

c)



The relation between position number and log of distance appears to be roughly linear.

CHAPTER 8

1. 281 milligrams

3. The potassium content is actually lower than the model predicts for a cereal with that much fiber.

5. The model predicts that cereals will have approximately 27 more milligrams of potassium for every additional gram of fiber.

7. 81.5%

9. The true potassium contents of cereals vary from the predicted amounts with a standard deviation of 30.77 milligrams.

11. a) Model is appropriate.
b) Model is not appropriate. Relationship is nonlinear.
c) Model may not be appropriate. Spread is changing.

13. 300 pounds/foot. It's ridiculous to suggest an extra foot in length would add 3, 30, or 3000 pounds to a car's weight.

15. a) *Price* (in thousands of dollars) is y and *Size* (in square feet) is x.
b) Slope is thousands of $ per square foot.
c) Positive. Larger homes should cost more.

17. A linear model on *Size* accounts for 71.4% of the variation in home *Price*.

19. a) 0.845; + because larger homes cost more.
b) Price should be 0.845 SDs above the mean in price.
c) Price should be 1.690 SDs below the mean in price.

21. a) *Price* increases by about $0.061 × 1000, or $61.00, per additional sq ft.
b) 230.82 thousand, or $230,820.
c) $115,020; $6000 is the residual.

23. a) R^2 does not tell whether the model is appropriate, but measures the strength of the linear relationship. High R^2 could also be due to an outlier.
b) Predictions based on a regression line are estimates of average values of y for a given x. The actual wingspan will vary around the prediction.

25. a) Probably not. Your score is better than about 97.5% of people, assuming scores follow the Normal model. Your next score is likely to be closer to the mean.
b) The friend should probably retake the test. His score is better than only about 16% of people. His score is likely to be closer to the mean.

27. a) Probably. The residuals show some initially low points, but there is no clear curvature.
b) The linear model on *Tar* content accounts for 92.4% of the variability in *Nicotine*.

29. a) $r = 0.961$
b) Nicotine should be 1.922 SDs below average.
c) Tar should be 0.961 SDs above average.

31. a) $\widehat{Nicotine} = 0.15403 + 0.065052\ Tar$
b) 0.414 mg
c) Predicted nicotine content increases by 0.065 mg of nicotine per additional milligram of tar.
d) We'd expect a cigarette with no tar to have 0.154 mg of nicotine.
e) 0.1094 mg

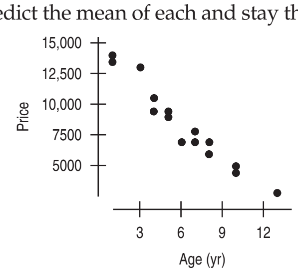
33. a) Yes. The relationship is straight enough, with a few outliers. The spread increases a bit for states with large median incomes, but we can still fit a regression line.
b) From summary statistics: $\widehat{HCI} = -156.50 + 0.0107\ MFI$; from original data: $\widehat{HCI} = -157.64 + 0.0107\ MFI$
c) From summary statistics: predicted HCI = 324.93; from original data: 324.87.
d) 223.09 e) $\widehat{z_{HCI}} = 0.65 z_{MFI}$ f) $\widehat{z_{MFI}} = 0.65 z_{HCI}$

35. a) $\widehat{Total} = 539.803 + 1.103 Age$
b) Yes. Both variables are quantitative; the plot is straight (although flat); there are no apparent outliers; the plot does not appear to change spread throughout the range of *Age*.
c) $559.65; $594.94
d) 0.14%
e) No. The plot is nearly flat. The model explains almost none of the variation in *Total Yearly Purchases*.

37. a) Moderately strong, fairly straight, and positive. Possibly some outliers (higher-than-expected math scores).
b) The student with 500 verbal and 800 math.
c) Positive, fairly strong linear relationship. 46.9% of variation in math scores is explained by verbal scores.
d) $\widehat{Math} = 217.7 + 0.662 \times Verbal$.
e) Every point of verbal score adds 0.662 points to the predicted average math score.
f) 548.5 points g) 53.0 points

39. a) 0.685 b) $\widehat{Verbal} = 162.1 + 0.71 \times Math$.
c) The observed verbal score is higher than predicted from the math score
d) 516.7 points. e) 559.6 points
f) Regression to the mean. Someone whose math score is below average is predicted to have a verbal score below average, but not as far (in SDs). So if we use *that* verbal score to predict math, they will be even closer to the mean in predicted math score than their observed math score. If we kept cycling back and forth, eventually we would predict the mean of each and stay there.

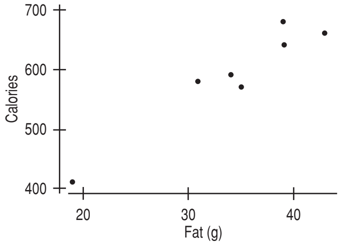
41. a)



b) Negative, linear, strong. c) Yes. d) −0.972
e) *Age* accounts for 94.4% of the variation in *Advertised Price*.
f) Other factors contribute—options, condition, mileage, etc.

43. a) $\widehat{Price} = 14{,}286 - 959 \times \text{Years}$.
b) Every extra year of age decreases average value by $959.
c) The average new Corolla costs a predicted $14,286.
d) $7573
e) Negative residual. Its price is below the predicted value for its age.
f) −$1195
g) No. After age 14, the model predicts negative prices. The relationship is no longer linear.

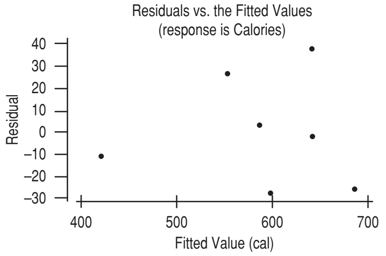
45. a)



b) 92.3% of the variation in calories can be accounted for by the fat content.

c) $\widehat{Calories} = 211.0 + 11.06 \times Fat$.

d)



Residuals show no clear pattern, so the model seems appropriate.

e) Could say a fat-free burger still has 211.0 calories, but this is extrapolation (no data close to 0).

f) Every gram of fat adds 11.06 calories, on average.

g) 553.5 calories.

47. a) The regression was for predicting calories from fat, not the other way around.

b) $\widehat{Fat} = -15.0 + 0.083 \times Calories$. Predict 34.8 grams of fat.

49. a) $\%\ Body\ Fat = -27.4 + 0.25 \times Weight$.

b) Residuals look randomly scattered around 0, so conditions are satisfied.

c) *% Body Fat* increases, on average, by 0.25 percent per pound of *Weight*.

d) Reliable is relative. R^2 is 48.5%, but residuals have a standard deviation of 7%, so variation around the line is large.

e) 0.9 percent.

51. a) $\widehat{HighJump} = 2.681 - 0.00671 \times 800mTime$. High-jump height is lower, on average, by 0.00671 meters per additional second of 800-m race time.

b) 16.4%

c) Yes, the slope is negative. Faster runners tend to jump higher.

d) There is a slight tendency for less variation in high-jump height among the slower runners than among the faster ones.

e) Not especially. The residual standard deviation is 0.060 meters, which is not much smaller than the SD of all high jumps (0.066 meters). The model doesn't appear to do a very good job of predicting.

53. The sum of the squared vertical distances to any other line would be greater than 1790.

CHAPTER 9

1. a) The trend appears to be somewhat linear up to about 1940, but from 1940 to about 1970 the trend appears to be nonlinear. From 1975 or so to the present, the trend appears to be linear.

b) Relatively strong for certain periods.

c) No, as a whole the graph is clearly nonlinear. Within certain periods (ex: 1975 to the present) the correlation is high.

d) Overall, no. You could fit a linear model to the period from 1975 to 2003, but why? You don't need to interpolate, since every year is reported, and extrapolation seems dangerous.

3. a) The relationship is not straight.

b) It will be curved downward.

c) No. The relationship will still be curved.

5. a) No. We need to see the scatterplot first to see if the conditions are satisfied, and models are always wrong.

b) No, the linear model might not fit the data everywhere.

7. a) Millions of dollars per minute of run time.

b) Costs for movies increase at the same rate per minute.

c) On average dramas cost about $20 million less for the same runtime.

9. a) The use of the Oakland airport has been growing at about 59,700 passengers/year, starting from about 282,000 in 1990.

b) 71% of the variation in passengers is accounted for by this model.

c) Errors in predictions based on this model have a standard deviation of 104,330 passengers.

d) No, that would extrapolate too far from the years we've observed.

e) The negative residual is September 2001. Air traffic was artificially low following the attacks on 9/11.

11. a) 1) High leverage, small residual.
2) No, not influential for the slope.
3) Correlation would decrease because outlier has large z_x and z_y, increasing correlation.
4) Slope wouldn't change much because the outlier is in line with other points.

b) 1) High leverage, probably small residual.
2) Yes, influential.
3) Correlation would weaken, increasing toward zero.
4) Slope would increase toward 0, since outlier makes it negative.

c) 1) Some leverage, large residual.
2) Yes, somewhat influential.
3) Correlation would increase, since scatter would decrease.
4) Slope would increase slightly.

d) 1) Little leverage, large residual.
2) No, not influential.
3) Correlation would become stronger and become more negative because scatter would decrease.
4) Slope would change very little.

13. 1) e 2) d 3) c 4) b 5) a

15. Perhaps high blood pressure causes high body fat, high body fat causes high blood pressure, or both could be caused by a lurking variable such as a genetic or lifestyle issue.

17. a) The graph shows that, on average, students progress at about one reading level per year. This graph shows averages for each grade. The linear trend has been enhanced by using averages.

b) Very close to 1.

c) The individual data points would show much more scatter, and the correlation would be lower.

d) A slope of 1 would indicate that for each 1-year grade level increase, the average reading level is increasing by 1 year.

19. a) *Cost* decreases by $2.13 per degree of average daily *Temp.* So warmer temperatures indicate lower costs.

b) For an avg. monthly temperature of 0°F, the cost is predicted to be $133.

c) Too high; the residuals (observed − predicted) around 32°F are negative, showing that the model overestimates the costs.

d) $111.70 e) About $105.70

f) No, the residuals show a definite curved pattern. The data are probably not linear.

g) No, there would be no difference. The relationship does not depend on the units.

21. a) 0.88

b) Interest rates during this period grew at about 0.25% per year, starting from an interest rate of about 0.64%.

c) Substituting 50 in the model yields a prediction of about 13%.

d) Not really. Extrapolating 20 years beyond the end of these data would be dangerous and unlikely to be accurate.

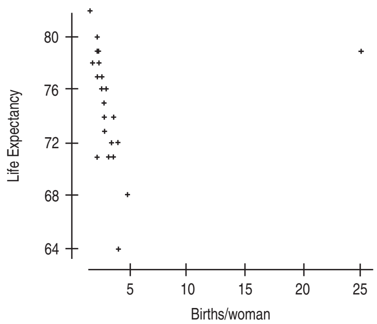
23. a) The two models fit comparably well, but they have very different slopes.
b) This model predicts the interest rate in 2000 to be 3.24%, much lower than the other model predicts.
c) We can trust the new predicted value because it is in the middle of the data used for the regression.
d) The best answer is "I can't predict that."

25. a) Stronger. Both slope and correlation would increase.
b) Restricting the study to nonhuman animals would justify it.
c) Moderately strong.
d) For every year increase in life expectancy, the gestation period increases by about 15.5 days, on average.
e) About 270.5 days.

27. a) Removing hippos would make the association stronger, since hippos are more of a departure from the pattern.
b) Increase.
c) No, there must be a good reason for removing data points.
d) Yes, removing it lowered the slope from 15.5 to 11.6 days per year.

29. a) Answers may vary. Using the data for 1955–2000 results in a scatterplot that is relatively linear with some curvature. The residuals plot shows a definite trend, indicating that the data are not linear. If you used the line, for 2010 the predicted age is 26.07 years.
b) Not much, since the data are not truly linear and 2010 is 10 years from the last data point (extrapolating is risky).
c) No, that extrapolation of more than 50 years would be absurd. There's no reason to believe the trend from 1955 to 2000 will continue.

31.



a) Except for the outlier, Costa Rica, the data appear to have a linear form in a negative direction.
b) The outlier is Costa Rica, whose data appear to be wrong, with 25 births per woman. That's impossible.
c) With Costa Rica, $r = 0.168$ and R-squared $= 2.8\%$, indicating that 2.8% of the variation in *Life Expectancy* is explained by the variation in *Births per Woman.* Without Costa Rica, $r = -0.796$ and R-squared $= 63.3\%$, indicating that 63.3% of the variation in *Life Expectancy* is explained by the variation in *Births/Woman.*
d) With Costa Rica, $\widehat{Life\ Expectancy} = 72.6 + 0.15\ Births$; without Costa Rica, $\widehat{Life\ Expectancy} = 84.5 - 4.44\ Births$.
e) The model with Costa Rica is not appropriate. The residuals plot shows a distinct outlier, which is Costa Rica. Removing Costa Rica gives a better residuals plot, suggesting that the linear equation is more appropriate.
f) With Costa Rica, the slope is near 0, suggesting that the linear model is not very useful. The y-intercept suggests that with no births, the life expectancy is about 72.6 years. Without Costa Rica, the slope is -4.44, indicating that an average increase of one child per woman predicts a lower life expectancy of 4.44 years, on average. The y-intercept indicates that a country with a birth rate of zero would have a life expectancy of 84.5 years. This is extrapolation.
g) While there is an association, there is no reason to expect causality. Lurking variables may be involved.

33. a) The scatterplot is clearly nonlinear; however, the last few years—say, from 1970 on—do appear to be linear.
b) Using the data from 1970 to 2006 gives $r = 0.997$ and $\widehat{CPI} = -9052.42 + 4.61\ Year$. Predicted CPI in 2016 $= 241.34$ (an extrapolation of doubtful accuracy).

CHAPTER 10

1. a) No re-expression needed.
b) Re-express to straighten the relationship.
c) Re-express to equalize spread.

3. a) There's an annual pattern in when people fly, so the residuals cycle up and down.
b) No, this kind of pattern can't be helped by re-expression.

5. a) 16.44 b) 7.84 c) 0.36 d) 1.75 e) 27.59

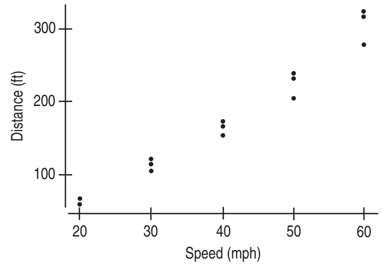
7. a) Fairly linear, negative, strong.
b) Gas mileage decreases an average 7.652 mpg for each thousand pounds of weight.
c) No. Residuals show a curved pattern.

9. a) Residuals are more randomly spread around 0, with some low outliers.
b) $\widehat{Fuel\ Consumption} = 0.625 + 1.178 \times Weight$.
c) For each additional 1000 pounds of *Weight*, an additional 1.178 gallons will be needed to drive 100 miles.
d) 21.06 miles per gallon.

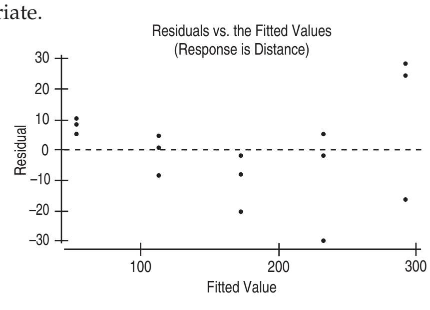
11. a) Although more than 97% of the variation in GDP can be accounted for by this model, we should examine a scatterplot of the residuals to see if it's appropriate.
b) No. The residuals show clear curvature.

13. Yes, the pattern in the residuals is somewhat weaker.

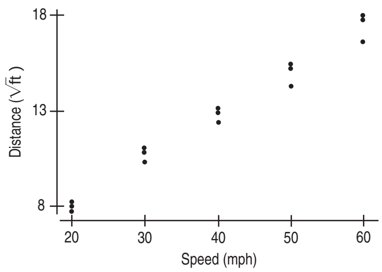
15. a)



$\widehat{Distance} = -65.9 + 5.98\ Speed$.
But residuals have a curved shape, so linear model is not appropriate.



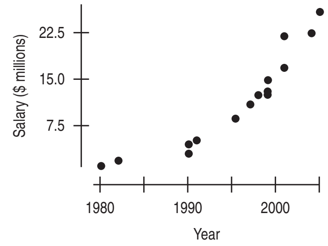
b)



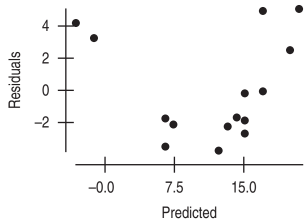
$\sqrt{Distance}$ linearizes the plot.
c) *Predicted* $\sqrt{Distance} = 3.30 + 0.235 \times Speed$.
d) 263.4 feet. e) 390.2 feet (an extrapolation)

f) Fairly confident, since $R^2 = 98.4\%$, and s is small.

17. a) The plot looks fairly straight. (It is okay to see a bend in the plot; there's one there.)



b) $\widehat{Salary} = -1913.88 + 0.965\,Year$

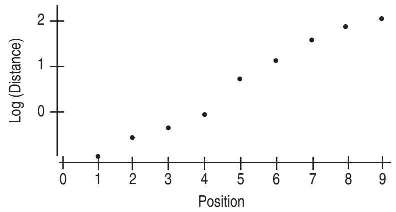


The residuals plot shows a strong bend.

c) *log(Salary)* works well.

d) $\log\left(\widehat{Salary}\right) = -109.133 + 0.05516\,Year$

19. a)



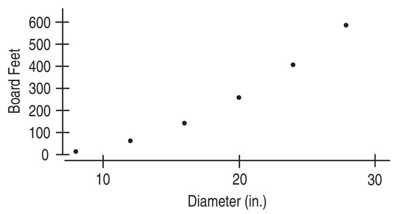
Log(*Distance*) against position works pretty well.

$\widehat{\log(Distance)} = 1.245 + 0.271 \times Position\ number.$

b) Pluto's residual is not especially larger in the log scale. However, a model without Pluto predicts the 9th planet should be 5741 million miles. Pluto, at "only" 3707 million miles, doesn't fit very well, giving support to the argument that Pluto doesn't behave like a planet.

21. The predicted log(*Distance*) of Eris is 3.685, corresponding to a distance of 4841 million miles. That's short of the actual average distance of 6300 million miles.

23. a)

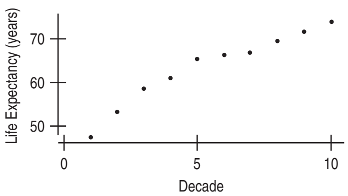


$\widehat{\sqrt{Bdft}} = -4 + diam$

The model is exact.

b) 36 board feet. c) 1024 board feet.

25.



$\widehat{\log Life} = 1.685 + 0.18497 \log Decade$

27. The relationship cannot be made straight by the methods of this chapter.

29. a) $\widehat{\sqrt{Left}} = 8.465 - 0.06926(Age)$ b) 52.10 years

c) No; the residuals plot still shows a pattern.

PART II REVIEW

1. % over 50, 0.69.
% under 20, −0.71.
% Graduating on time, −0.51.
% Full-time Faculty, 0.09

3. a) There does not appear to be a linear relationship.

b) Nothing, there is no reason to believe that the results for the Finger Lakes region are representative of the vineyards of the world.

c) $\widehat{CasePrice} = 92.77 + 0.567 \times Years.$

d) Only 2.7 % of the variation in case price is accounted for by the ages of vineyards. Most of that is due to two outliers. We are better off using the mean price rather than this model.

5. a) $\widehat{TwinBirths} = -5119590 + 2618.25 \times Year.$

b) Each year, the number of twins born in a year increases, on average, by approximately 2618.25.

c) 143,092.5 births. The scatterplot appears to be somewhat linear, but there is some curvature in the pattern. There is no reason to believe that the increase will continue to be linear 5 years beyond the data.

d) The residuals plot shows a definite curved pattern, so the relation is not linear.

7. a) −0.520

b) Negative, not strong, somewhat linear, but with more variation as pH increases.

c) The BCI would also be average.

d) The predicted *BCI* will be 1.56 SDs of *BCI* below the mean *BCI*.

9. a) $\widehat{Manatee\,Deaths} = -45.67 \times 0.1315\ Powerboat\,Registrations$ (in 1000s).

b) According to the model, for each increase of 10,000 motorboat registrations, the number of manatees killed increases by approximately 1.315.

c) If there were 0 motorboat registrations, the number of manatee deaths would be −45.67. This is obviously a silly extrapolation.

d) The predicted number is 82.41 deaths. The actual number of deaths was 79. The residual is $79 - 82.41 = -3.41$. The model overestimated the number of deaths by 3.41.

e) Negative residuals would suggest that the actual number of deaths was lower than the predicted number.

f) Over time, the number of motorboat registrations has increased and the number of manatee kills has increased. The trend may continue. Extrapolation is risky, however, because the government may enact legislation to protect the manatee.

11. a) −0.984 b) 96.9% c) 32.95 mph d) 1.66 mph

e) Slope will increase.

f) Correlation will weaken (become less negative).

g) Correlation is the same, regardless of units.

13. a) Weight (but unable to verify linearity).

b) As weight increases, mileage decreases.

c) *Weight* accounts for 81.5% of the variation in *Fuel Efficiency.*

15. a) $\widehat{Horsepower} = 3.50 + 34.314 \times Weight.$

b) Thousands. For the equation to have predicted values between 60 and 160, the *X* values would have to be in thousands of pounds.

c) Yes. The residual plot does not show any pattern.

d) 115.0 horsepower.

17. a) The scatterplot shows a fairly strong linear relation in a positive direction. There seem to be two distinct clusters of data.

b) $\widehat{Interval} = 33.967 \div 10.358 \times Duration.$

c) The time between eruptions increases by about 10.4 minutes per minute of *Duration* on average.

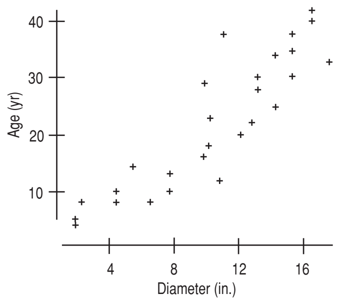
d) Since 77% of the variation in *Interval* is accounted for by *Duration* and the error standard deviation is 6.16 minutes, the prediction will be relatively accurate.

e) 75.4 minutes.

f) A residual is the observed value minus the predicted value. So the residual $= 79 - 75.4 = 3.6$ minutes, indicating that the model underestimated the interval in this case.

19. a) $r = 0.888$. Although r is high, you must look at the scatterplot and verify that the relation is linear in form.

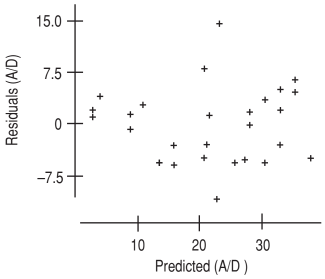
b)



The association between diameter and age appears to be strong, somewhat linear, and positive.

c) $\widehat{Age} = -0.97 + 2.21 \times Diameter$.

d)



The residuals show a curved pattern (and two outliers).

e) The residuals for five of the seven largest trees (15 in. or larger) are positive, indicating that the predicted values underestimate the age.

21. Most houses have areas between 1000 and 5000 square feet. Increasing 1000 square feet would result in either $1000(.008) =$ 8 thousand dollars, $1000(.08) = 80$ thousand dollars, $1000(.8) =$ 800 thousand dollars, or $1000(8) = 8000$ thousand dollars. Only $80,000 is reasonable, so the slope must be 0.08.

23. a) The model predicts % smoking from year, not the other way around.

b) $\widehat{Year} = 2027.91 - 202.74 \times \%\ Smoking$.

c) The smallest % smoking given is 12.7, and an extrapolation to $x = 0$ is probably too far from the given data. The prediction is not very reliable in spite of the strong correlation.

25. The relation shows a negative direction, with a somewhat linear form, but perhaps with some slight curvature. There are several model outliers.

27. a) 71.9%

b) As latitude increases, the January temperature decreases.

c) $\widehat{January\,Temperature} = 108.80 - 2.111 \times Latitude$.

d) As the latitude increases by 1 degree, the average January temperature drops by about 2.11 degrees, on average.

e) The y-intercept would indicate that the average January temperature is 108.8 when the latitude is 0. However, this is extrapolation and may not be meaningful.

f) 24.4 degrees.

g) The equation underestimates the average January temperature.

29. a) The scatterplot shows a strong, linear, positive association.

b) There is an association, but it is likely that training and technique have increased over time and affected both jump performances.

c) Neither; the change in units does not affect the correlation.

d) The long-jumper would jump 0.925 SDs above the mean long jump, on average.

31. a) No relation; the correlation would probably be close to 0.

b) The relation would have a positive direction and the correlation would be strong, assuming that students were studying French in each grade level. Otherwise, no correlation.

c) No relation; correlation close to 0.

d) The relation would have a positive direction and the correlation would be strong, since vocabulary would increase with each grade level.

33. $\widehat{Calories} = 560.7 - 3.08 \times Time$.
Each minute extra at the table results in 3.08 fewer calories being consumed, on average. Perhaps the hungry children eat fast and eat more.

35. There seems to be a strong, positive, linear relationship with one high-leverage point (Northern Ireland) that makes the overall R^2 quite low. Without that point, the R^2 increases to 61.5%. Of course, these data are averaged across thousands of households, so the correlation appears to be higher than it would be for individuals. Any conclusions about individuals would be suspect.

37. a) 3.842 b) 501.187 c) 4.0

39. a) 30,818 pounds.

b) 1302 pounds.

c) 31,187.6 pounds.

d) I would be concerned about using this relation if we needed accuracy closer than 1000 pounds or so, as the residuals are more than ± 1000 pounds.

e) Negative residuals will be more of a problem, as the predicted weight would overestimate the weight of the truck; trucking companies might be inclined to take the ticket to court.

41. The original data are nonlinear, with a significant curvature. Using reciprocal square root of diameter gave a scatterplot that is nearly linear:

$$\widehat{1/\sqrt{Drain\ Time}} = 0.0024 + 0.219\ Diameter.$$

CHAPTER 11

1. Yes. You cannot predict the outcome beforehand.

3. A machine pops up numbered balls. If it were truly random, the outcome could not be predicted and the outcomes would be equally likely. It is random only if the balls generate numbers in equal frequencies.

5. Use two-digit numbers 00–99; let 00–02 = defect, 03–99 = no defect

7. a) 45, 10 b) 17, 22

9. If the lottery is random, it doesn't matter which number you play; all are equally likely to win.

11. a) The outcomes are not equally likely; for example, tossing 5 heads does not have the same probability as tossing 0 or 9 heads, but the simulation assumes they are equally likely.

b) The even-odd assignment assumes that the player is equally likely to score or miss the shot. In reality, the likelihood of making the shot depends on the player's skill.

c) The likelihood for the first ace in the hand is not the same as for the second or third or fourth. But with this simulation, the likelihood is the same for each. (And it allows you to get 5 aces, which could get you in trouble in a real poker game!)

13. The conclusion should indicate that the simulation *suggests* that the average length of the line would be 3.2 people. Future results might not match the simulated results exactly.

15. a) The component is one voter voting. An outcome is a vote for our candidate or not. Use two random digits, giving 00–54 a vote for your candidate and 55–99 for the underdog.

b) A trial is 100 votes. Examine 100 two-digit random numbers, and count how many people voted for each candidate. Whoever gets the majority of votes wins that trial.

c) The response variable is whether the underdog wins or not.

17. Answers will vary, but average answer will be about 51%.
19. Answers will vary, but average answer will be about 26%.
21. a) Answers will vary, but you should win about 10% of the time.
b) You should win at the same rate with any number.
23. Answers will vary, but you should win about 10% of the time.
25. Answers will vary, but average answer will be about 1.9 tests.
27. Answers will vary, but average answer will be about 1.24 points.
29. Do the simulation in two steps. First simulate the payoffs. Then count until $500 is reached. Answers will vary, but average should be near 10.2 customers.
31. Answers will vary, but average answer will be about 3 children.
33. Answers will vary, but average answer will be about 7.5 rolls.
35. No, it will happen about 40% of the time.
37. Answers will vary, but average answer will be about 37.5%.
39. Three women will be selected about 7.8% of the time.

CHAPTER 12

1. a) No. It would be nearly impossible to get exactly 500 males and 500 females from every country by random chance.
b) A stratified sample, stratified by whether the respondent is male or female.
3. a) Voluntary response.
b) We have no confidence at all in estimates from such studies.
5. a) The population of interest is all adults in the United States aged 18 and older.
b) The sampling frame is U.S. adults with telephones.
c) Some members of the population (e.g, many college students) don't have landline phones, which could create a bias.
7. a) Population—All U.S. adults.
b) Parameter—Proportion who have used and benefited from alternative medicine.
c) Sampling Frame—All Consumers Union subscribers.
d) Sample—Those who responded.
e) Method—Questionnaire to all (nonrandom).
f) Bias—Nonresponse. Those who respond may have strong feelings one way or another.
9. a) Population—Adults.
b) Parameter—Proportion who think drinking and driving is a serious problem.
c) Sampling Frame—Bar patrons.
d) Sample—Every 10th person leaving the bar.
e) Method—Systematic sampling (may be random).
f) Bias—Those interviewed had just left a bar. They may think drinking and driving is less of a problem than do other adults.
11. a) Population—Soil around a former waste dump.
b) Parameter—Concentrations of toxic chemicals.
c) Sampling Frame—Accessible soil around the dump.
d) Sample—16 soil samples.
e) Method—Not clear.
f) Bias—Don't know if soil samples were randomly chosen. If not, may be biased toward more or less polluted soil.
13. a) Population—Snack food bags.
b) Parameter—Weight of bags, proportion passing inspection.
c) Sampling Frame—All bags produced each day.
d) Sample—Bags in 10 randomly selected cases, 1 bag from each case for inspection.
e) Method—Multistage random sampling.
f) Bias—Should be unbiased.
15. Bias. Only people watching the news will respond, and their preference may differ from that of other voters. The sampling method may systematically produce samples that don't represent the population of interest.
17. a) Voluntary response. Only those who see the ad, have Internet access, *and* feel strongly enough will respond.
b) Cluster sampling. One school may not be typical of all.
c) Attempted census. Will have nonresponse bias.
d) Stratified sampling with follow-up. Should be unbiased.
19. a) This is a multistage design, with a cluster sample at the first stage and a simple random sample for each cluster.
b) If any of the three churches you pick at random is not representative of all churches, then you'll introduce sampling error by the choice of that church.
21. a) This is a systematic sample.
b) The sampling frame is patrons willing to wait for the roller coaster on that day at that time. It should be representative of the people in line, but not of all people at the amusement park.
c) It is likely to be representative of those waiting for the roller coaster. Indeed, it may do quite well if those at the front of the line respond differently (after their long wait) than those at the back of the line.
23. a) Answers will definitely differ. Question 1 will probably get many "No" answers, while Question 2 will get many "Yes" answers. This is response bias.
b) "Do you think standardized tests are appropriate for deciding whether a student should be promoted to the next grade?" (Other answers will vary.)
25. a) Biased toward yes because of "pollute." "Should companies be responsible for any costs of environmental cleanup?"
b) Biased toward no because of "old enough to serve in the military." "Do you think the drinking age should be lowered from 21?"
27. a) Not everyone has an equal chance. Misses people with unlisted numbers, or without landline phones, or at work.
b) Generate random numbers and call at random times.
c) Under the original plan, those families in which one person stays home are more likely to be included. Under the second plan, many more are included. People without landline phones are still excluded.
d) It improves the chance of selected households being included.
e) This takes care of phone numbers. Time of day may be an issue. People without landline phones are still excluded.
29. a) Answers will vary.
b) Your own arm length. Parameter is your own arm length; population is all possible measurements of it.
c) Population is now the arm lengths of you and your friends. The average estimates the mean of these lengths.
d) Probably not. Friends are likely to be of the same age and not very diverse or representative of the larger population.
31. a) Assign numbers 001 to 120 to each order. Use random numbers to select 10 transactions to examine.
b) Sample proportionately within each type. (Do a stratified random sample.)
33. a) Select three cases at random; then select one jar randomly from each case.
b) Use random numbers to choose 3 cases from numbers 61 through 80; then use random numbers between 1 and 12 to select the jar from each case.
c) No. Multistage sampling.
35. a) Depends on the Yellow Page listings used. If from regular (line) listings, this is fair if all doctors are listed. If from ads, probably not, as those doctors may not be typical.
b) Not appropriate. This cluster sample will probably contain listings for only one or two business types.

CHAPTER 13

1. a) No. There are no manipulated factors. Observational study.
b) There may be lurking variables that are associated with both parental income and performance on the SAT.
3. a) This is a retrospective observational study.
b) That's appropriate because MS is a relatively rare disease.

c) The subjects were U.S. military personnel, some of whom had developed MS.
d) The variables were the vitamin D blood levels and whether or not the subject developed MS.

5. a) This was a randomized, placebo-controlled experiment.
b) Yes, such an experiment is the right way to determine whether black cohosh has an effect.
c) 351 women aged 45 to 55 who reported at least two hot flashes a day.
d) The treatments were black cohosh, a multiherb supplement, a multiherb supplement plus advice, estrogen, and a placebo. The response was the women's symptoms (presumably frequency of hot flashes).

7. a) Experiment.
b) Bipolar disorder patients.
c) Omega-3 fats from fish oil, two levels.
d) 2 treatments.
e) Improvement (fewer symptoms?).
f) Design not specified.
g) Blind (due to placebo), unknown if double-blind.
h) Individuals with bipolar disease improve with high-dose omega-3 fats from fish oil.

9. a) Observational study.
b) Prospective.
c) Men and women with moderately high blood pressure and normal blood pressure, unknown selection process.
d) Memory and reaction time.
e) As there is no random assignment, there is no way to know that high blood pressure *caused* subjects to do worse on memory and reaction-time tests. A lurking variable may also be the cause.

11. a) Experiment.
b) Postmenopausal women.
c) Alcohol—2 levels; blocking variable—estrogen supplements (2 levels).
d) 1 factor (alcohol) at 2 levels = 2 treatments.
e) Increase in estrogen levels.
f) Blocked.
g) Not blind.
h) Indicates that alcohol consumption *for those taking estrogen supplements* may increase estrogen levels.

13. a) Observational study.
b) Retrospective.
c) Women in Finland, unknown selection process with data from church records.
d) Women's lifespans.
e) As there is no random assignment, there is no way to know that having sons or daughters shortens or lengthens the lifespan of mothers.

15. a) Observational study.
b) Prospective.
c) People with or without depression, unknown selection process.
d) Frequency of crying in response to sad situations.
e) There is no apparent difference in crying response (to sad movies) for depressed and nondepressed groups.

17. a) Experiment.
b) People experiencing migraines.
c) 2 factors (pain reliever and water temperature), 2 levels each.
d) 4 treatments.
e) Level of pain relief.
f) Completely randomized over 2 factors.
g) Blind, as subjects did not know if they received the pain medication or the placebo, but not blind, as the subjects will know if they are drinking regular or ice water.
h) It may indicate whether pain reliever alone or in combination with ice water gives pain relief, but patients are not blinded to ice water, so placebo effect may also be the cause of any relief seen caused by ice water.

19. a) Experiment.
b) Athletes with hamstring injuries.
c) 1 factor: type of exercise program (2 levels).
d) 2 treatments.
e) Time to return to sports.
f) Completely randomized.
g) No blinding—subjects must know what kind of exercise they do.
h) Can determine which of the two exercise programs is more effective.

21. They need to compare omega-3 results to something. Perhaps bipolarity is seasonal and would have improved during the experiment anyway.

23. a) Subjects' responses might be related to many other factors (diet, exercise, genetics, etc). Randomization should equalize the two groups with respect to unknown factors.
b) More subjects would minimize the impact of individual variability in the responses, but the experiment would become more costly and time consuming.

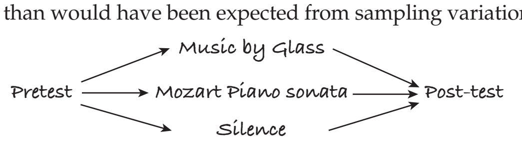
25. People who engage in regular exercise might differ from others with respect to bipolar disorder, and that additional variability could obscure the effectiveness of this treatment.

27. Answers may vary. Use a random-number generator to randomly select 24 numbers from 01 to 24 without replication. Assign the first 8 numbers to the first group, the second 8 numbers to the second group, and the third 8 numbers to the third group.

29. a) First, they are using athletes who have a vested interest in the success of the shoe by virtue of their sponsorship. They should choose other athletes. Second, they should randomize the order of the runs, not run all the races with their shoes second. They should blind the athletes by disguising the shoes if possible, so they don't know which is which. The timers shouldn't know which athletes are running with which shoes, either. Finally, they should replicate several times, since times will vary under both shoe conditions.
b) Because of the problems in (a), the results they obtain may favor their shoes. In addition, the results obtained for Olympic athletes may not be the same as for the general runner.

31. a) Allowing athletes to self-select treatments could confound the results. Other issues such as severity of injury, diet, age, etc., could also affect time to heal; randomization should equalize the treatment groups with respect to any such variables.
b) A control group could have revealed whether either exercise program was better (or worse) than just letting the injury heal.
c) Doctors who evaluated the athletes to approve their return to sports should not know which treatment the subject had.
d) It's hard to tell. The difference of 15 days seems large, but the standard deviations indicate that there was a great deal of variability in the times.

33. a) The differences among the Mozart and quiet groups were more than would have been expected from sampling variation.
b)



c) The Mozart group seems to have the smallest median difference and thus the *least* improvement, but there does not appear to be a significant difference.
d) No, if anything, there is less improvement, but the difference does not seem significant compared with the usual variation.

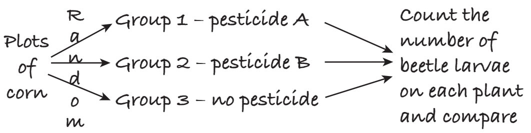
35. a) Observational, prospective study.
b) The supposed relation between health and wine consumption might be explained by the confounding variables of income and education.
c) None of these. While the variables have a relation, there is no causality indicated for the relation.

37. a) Arrange the 20 containers in 20 separate locations. Use a random-number generator to identify the 10 containers that should be filled with water.
b) Guessing, the dowser should be correct about 50% of the time. A record of 60% (12 out of 20) does not appear to be significantly different.
c) Answers may vary. You would need to see a high level of success—say, 90% to 100%, that is, 18 to 20 correct.

39. Randomly assign half the reading teachers in the district to use each method. Students should be randomly assigned to teachers as well. Make sure to block both by school and grade (or control grade by using only one grade). Construct an appropriate reading test to be used at the end of the year, and compare scores.

41. a) They mean that the difference is higher than they would expect from normal sampling variability.
b) An observational study.
c) No. Perhaps the differences are attributable to some confounding variable (e.g., people are more likely to engage in riskier behaviors on the weekend) rather than the day of admission.
d) Perhaps people have more serious accidents and traumas on weekends and are thus more likely to die as a result.

43. Answers may vary. This experiment has 1 factor (pesticide), at 3 levels (pesticide A, pesticide B, no pesticide), resulting in 3 treatments. The response variable is the number of beetle larvae found on each plant. Randomly select a third of the plots to be sprayed with pesticide A, a third with pesticide B, and a third with no pesticide (since the researcher also wants to know whether the pesticides even work at all). To control the experiment, the plots of land should be as similar as possible with regard to amount of sunlight, water, proximity to other plants, etc. If not, plots with similar characteristics should be blocked together. If possible, use some inert substance as a placebo pesticide on the control group, and do not tell the counters of the beetle larvae which plants have been treated with pesticides. After a given period of time, count the number of beetle larvae on each plant and compare the results.



45. Answers may vary. Find a group of volunteers. Each volunteer will be required to shut off the machine with his or her left hand and right hand. Randomly assign the left or right hand to be used first. Complete the first attempt for the whole group. Now repeat the experiment with the alternate hand. Check the differences in time for the left and right hands.

47. a) Jumping with or without a parachute.
b) Volunteer skydivers (the dimwitted ones).
c) A parachute that looks real but doesn't work.
d) A good parachute and a placebo parachute.
e) Whether parachutist survives the jump (or extent of injuries).
f) All should jump from the same altitude in similar weather conditions and land on similar surfaces.
g) Randomly assign people the parachutes.
h) The skydivers (and the people involved in distributing the parachute packs) shouldn't know who got a working chute. And the people evaluating the subjects after the jumps should not be told who had a real parachute either!

PART III REVIEW

1. Observational prospective study. Indications of behavior differences can be seen in the two groups. May show a link between premature birth and behavior, but there may be lurking variables involved.

3. Experiment, matched by gender and weight, randomization within blocks of two pups of same gender and weight. Factor: type of diet. Treatments: low-calorie diet and allowing the dog to eat all it wants. Response variable: length of life. Can conclude that, on average, dogs with a lower-calorie diet live longer.

5. Observational prospective study. Indicates folate *may* help in reducing colon cancer for those with family histories of the disease.

7. Sampling. Probably a simple random sample, although may be stratified by type of firework. Population is all fireworks produced each day. Parameter is proportion of duds. Can determine if the day's production is ready for sale.

9. Observational retrospective study. Living near strong electromagnetic fields may be associated with more leukemia than normal. May be lurking variables, such as socioeconomic level.

11. Experiment. Blocked by sex of rat. Randomization is not specified. Factor is type of hormone given. Treatments are leptin and insulin. Response variable is lost weight. Can conclude that hormones can help suppress appetites in rats, and the type of hormone varies by gender.

13. Experiment. Factor is gene therapy. Hamsters were randomized to treatments. Treatments were gene therapy or not. Response variable is heart muscle condition. Can conclude that gene therapy is beneficial (at least in hamsters).

15. Sampling. Population is all oranges on the truck. Parameter is proportion of unsuitable oranges. Procedure is probably simple random sampling. Can conclude whether or not to accept the truckload.

17. Observational prospective study. Physically fit men may have a lower risk of death from cancer.

19. Answers will vary. This is a simulation problem. Using a random digits table or software, call 0–4 a loss and 5–9 a win for the gambler on a game. Use blocks of 5 digits to simulate a week's pick.

21. Answers will vary.

23. a) Experiment. Actively manipulated candy giving, diners were randomly assigned treatments, control group was those with no candy, lots of dining parties.
b) It depends on when the decision was made. If early in the meal, the server may give better treatment to those who will receive candy—biasing the results.
c) A difference in response so large it cannot be attributed to natural sampling variability.

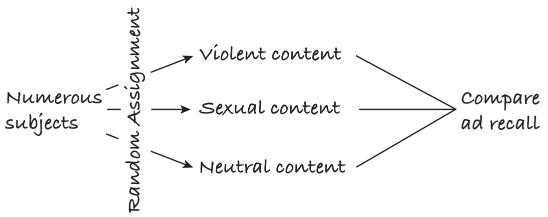
25. a) Voluntary response. Only those who feel strongly will pay for the 900 phone call.
b) "If it would help future generations live a longer, healthier life, would you be in favor of human cloning?"

27. a) Simulation results will vary. Average will be around 5.8 points.
b) Simulation results will vary. Average will also be around 5.8 points.
c) Answers will vary.

29. a) Yes.
b) No. Residences without phones are excluded. Residences with more than one phone had a higher chance.
c) No. People who respond to the survey may be of age but not registered voters.
d) No. Households who answered the phone may be more likely to have someone at home when the phone call was generated. These may not be representative of all households.

31. a) Does not prove it. There may be other confounding variables. Only way to prove this would be to do a controlled experiment.
b) Alzheimer's usually shows up late in life. Perhaps smokers have died of other causes before Alzheimer's can be seen.
c) An experiment would be unethical. One could design a prospective study in which groups of smokers and nonsmokers are followed for many years and the incidence of Alzheimer's is tracked.

33.



Numerous subjects will be randomly assigned to see shows with violent, sexual, or neutral content. They will see the same commercials. After the show, they will be interviewed for their recall of brand names in the commercials.

35. a) May have been a simple random sample, but given the relative equality in age groups, may have been stratified.
b) 35.1%.
c) We don't know. Perhaps cell phones or unlisted numbers were excluded, and Democrats have more (or fewer) of those. Probably OK, though.
d) Do party affiliations differ for different age groups?

37. The factor in the experiment will be type of bird control. I will have three treatments: scarecrow, netting, and no control. I will randomly assign several different areas in the vineyard to one of the treatments, taking care that there is sufficient separation that the possible effect of the scarecrow will not be confounded. At the end of the season, the response variable will be the proportion of bird-damaged grapes.

39. a) We want all subjects treated as alike as possible. If there were no "placebo surgery," subjects would know this and perhaps behave differently.
b) The experiment looked for a difference in the effectiveness of the two treatments. (If we wanted to generalize, we would need to assume that the results for these volunteers are the same as on all patients who might need this operation.)
c) "Not statistically significant" means the difference in results were small enough that it could be explained by natural sampling variability.

41. a) Use stratified sampling to select 2 first-class passengers and 12 from coach.
b) Number passengers alphabetically, 01 = Bergman to 20 = Testut. Read in blocks of two, ignoring any numbers more than 20. This gives 65, 43, 67, 11 (selects Fontana), 27, 04 (selects Castillo).
c) Number passengers alphabetically from 001 to 120. Use the random-number table to find three-digit numbers in this range until 12 different values have been selected.

43. Simulation results will vary.
(Use integers 00 to 99 as a basis. Use integers 00 to 69 to represent a tee shot on the fairway. If on the fairway, use digits 00 to 79 to represent on the green. If off the fairway, use 00 to 39 to represent getting on the green. If not on the green, use digits 00 to 89 to represent landing on the green. For the first putt, use digits 00 to 19 to represent making the shot. For subsequent putts, use digits 00 to 89 to represent making the shot.)

CHAPTER 14

1. a) S = {HH, HT, TH, TT}, equally likely.
b) S = {0, 1, 2, 3}, not equally likely.
c) S = {H, TH, TTH, TTT}, not equally likely.
d) S = {1, 2, 3, 4, 5, 6}, not equally likely.

3. In this context "truly random" should mean that every number is equally likely to occur.

5. There is no "Law of Averages." She would be wrong to think that they are "due" for a harsh winter.

7. There is no "Law of Averages." If at bats are independent, his chance for a hit does not change based on recent successes or failures.

9. a) There is some chance you would have to pay out much more than the $300.
b) Many customers pay for insurance. The small risk for any one customer is spread among all.

11. a) Legitimate. b) Legitimate.
c) Not legitimate (sum more than 1). d) Legitimate.
e) Not legitimate (can't have negatives or values more than 1).

13. A family may own both a car and an SUV. The events are not disjoint, so the Addition Rule does not apply.

15. When cars are traveling close together, their speeds are not independent, so the Multiplication Rule does not apply.

17. a) He has multiplied the two probabilities.
b) He assumes that being accepted at the colleges are independent events.
c) No. Colleges use similar criteria for acceptance, so the decisions are not independent.

19. a) 0.72 b) 0.89 c) 0.28

21. a) 0.5184 b) 0.0784 c) 0.4816

23. a) Repair needs for the two cars must be independent.
b) Maybe not. An owner may treat the two cars similarly, taking good (or poor) care of both. This may decrease (or increase) the likelihood that each needs to be repaired.

25. a) $342/1005 = 0.340$.
b) $30/1005 + 50/1005 = 80/1005 = 0.080$.

27. a) 0.195 b) 0.913
c) Responses are independent.
d) People were polled at random.

29. a) 0.4712 b) 0.7112
c) $(1 - 0.76) + 0.76(1 - 0.38)$ or $1 - (0.76)(0.38)$

31. a) 1) 0.30 2) 0.30 3) 0.90 4) 0.0
b) 1) 0.027 2) 0.128 3) 0.512 4) 0.271

33. a) Disjoint (can't be both red and orange).
b) Independent (unless you're drawing from a small bag).
c) No. Once you know that one of a pair of disjoint events has occurred, the other is impossible.

35. a) 0.0046 b) 0.125 c) 0.296 d) 0.421 e) 0.995

37. a) 0.027 b) 0.063 c) 0.973 d) 0.014

39. a) 0.024 b) 0.250 c) 0.543

41. 0.078.

43. a) For any day with a valid three-digit date, the chance is 0.001, or 1 in 1000. For many dates in October through December, the probability is 0. (No three digits will make 10/15, for example.)
b) There are 65 days when the chance to match is 0. (Oct. 10–31, Nov. 10–30, and Dec. 10–31.) The chance for no matches on the remaining 300 days is 0.741
c) 0.259 d) 0.049

CHAPTER 15

1. a) 0.68 b) 0.32 c) 0.04

3. a) 0.31 b) 0.48 c) 0.31

5. a) 0.2025 b) 0.6965 c) 0.2404 d) 0.0402

7. a) 0.50 b) 1.00 c) 0.077 d) 0.333

9. a) 0.11 b) 0.27 c) 0.407 d) 0.344

11. a) 0.011 b) 0.222 c) 0.054 d) 0.337 e) 0.436

13. 0.21

15. a) 0.145 b) 0.118 c) 0.414 d) 0.217

17. a) 0.318 b) 0.955 c) 0.071 d) 0.009

19. a) 32% b) 0.135
c) No, 7% of juniors have taken both.
d) No, the probability that a junior has taken a computer course is 0.23. The probability that a junior has taken a computer course *given* he or she has taken a Statistics course is 0.135.

21. a) 0.266
b) No, 26.6% of homes with garages have pools; 21% of homes overall have pools.
c) No, 17% of homes have both.

23. Yes, $P(\text{Ace}) = 4/52$. $P(\text{Ace} \mid \text{any suit}) = 1/13$.
25. a) 0.17
b) No; 13% of the chickens had both contaminants.
c) No; $P(C \mid S) = 0.87 \neq P(C)$. If a chicken is contaminated with salmonella, it's more likely also to have campylobacter.
27. No, only 32% of all men have high cholesterol, but 40.7% of those with high blood pressure do.
29. a) 95.6%
b) Probably. 95.4% of people with cell phones had landlines, and 95.6% of all people did.
31. No. Only 34% of men were Democrats, but over 41% of all voters were.
33. a) No, the probability that the luggage arrives on time depends on whether the flight is on time. The probability is 95% if the flight is on time and only 65% if not.
b) 0.695
35. 0.975
37. a) No, the probability of missing work for day-shift employees is 0.01. It is 0.02 for night-shift employees. The probability depends on whether they work day or night shift.
b) 1.4%
39. 57.1%
41. a) 0.20 b) 0.272 c) 0.353 d) 0.033
43. 0.563 **45.** Over 0.999

CHAPTER 16

1. a) 19 b) 4.2
3. a)

Amount won	\$0	\$5	\$10	\$30
P(Amount won)	$\frac{26}{52}$	$\frac{13}{52}$	$\frac{12}{52}$	$\frac{1}{52}$

b) \$4.13 c) \$4 or less (answers may vary)
5. a)

Children	1	2	3
P(Children)	0.5	0.25	0.25

b) 1.75 children c) 0.875 boys

Boys	0	1	2	3
P(Boys)	0.5	0.25	0.125	0.125

7. \$27,000
9. a) 7 b) 1.89
11. \$5.44
13. 0.83
15. a) 1.7 b) 0.9
17. $\mu = 0.64, \sigma = 0.93$
19. a) \$50 b) \$100
21. a) No. The probability of winning the second depends on the outcome of the first.
b) 0.42 c) 0.08
d)

Games won	0	1	2
P(Games won)	0.42	0.50	0.08

e) $\mu = 0.66, \sigma = 0.62$
23. a)

Number good	0	1	2
P(Number good)	0.067	0.467	0.467

b) 1.40 c) 0.61
25. a) $\mu = 30, \sigma = 6$ b) $\mu = 26, \sigma = 5$ c) $\mu = 30, \sigma = 5.39$
d) $\mu = -10, \sigma = 5.39$ e) $\mu = 20, \sigma = 2.83$
27. a) $\mu = 240, \sigma = 12.80$ b) $\mu = 140, \sigma = 24$
c) $\mu = 720, \sigma = 34.18$ d) $\mu = 60, \sigma = 39.40$
e) $\mu = 600, \sigma = 22.63$
29. a) 1.8 b) 0.87
c) Cartons are independent of each other.
31. $\mu = 13.6, \sigma = 2.55$ (assuming the hours are independent of each other).
33. a) $\mu = 23.4, \sigma = 2.97$
b) We assume each truck gets tickets independently.
35. a) There will be many gains of \$150 with a few large losses.
b) $\mu = \$300, \sigma = \$8485.28$
c) $\mu = \$1{,}500{,}000, \sigma = \$600{,}000$
d) Yes. \$0 is 2.5 SDs below the mean for 10,000 policies.
e) Losses are independent of each other. A major catastrophe with many policies in an area would violate the assumption.
37. a) 1 oz b) 0.5 oz c) 0.023
d) $\mu = 4$ oz, $\sigma = 0.5$ oz
e) 0.159
f) $\mu = 12.3$ oz, $\sigma = 0.54$ oz
39. a) 12.2 oz b) 0.51 oz c) 0.058
41. a) $\mu = 200.57$ sec, $\sigma = 0.46$ sec
b) No, $z = \frac{199.48 - 200.57}{0.461} = -2.36$. There is only 0.009 probability of swimming that fast or faster.
43. a) A = price of a pound of apples; P = price of a pound of potatoes; Profit $= 100A + 50P - 2$
b) \$63.00 c) \$20.62
d) Mean—no; SD—yes (independent sales prices).
45. a) $\mu = 1920, \sigma = 48.99$; $P(T > 2000) = 0.051$
b) $\mu = \$220, \sigma = 11.09$; No—\$300 is more than 7 SDs above the mean.
c) $P(D - \frac{1}{2}C > 0) \approx 0.26$

CHAPTER 17

1. a) No. More than two outcomes are possible.
b) Yes, assuming the people are unrelated to each other.
c) No. The chance of a heart changes as cards are dealt so the trials are not independent.
d) No, 500 is more than 10% of 3000.
e) If packages in a case are independent of each other, yes.
3. a) Use single random digits. Let 0, 1 = Tiger. Count the number of random numbers until a 0 or 1 occurs.
c) Results will vary.
d)

x	1	2	3	4	5	6	7	8	≥ 9
$P(x)$	0.2	0.16	0.128	0.102	0.082	0.066	0.052	0.042	0.168

5. a) Use single random digits. Let 0, 1 = Tiger. Examine random digits in groups of five, counting the number of 0's and 1's.
c) Results will vary.
d)

x	0	1	2	3	4	5
$P(x)$	0.33	0.41	0.20	0.05	0.01	0.0

7. Departures from the same airport during a 2-hour interval may not be independent. All could be delayed by weather, for example.
9. a) 0.0819 b) 0.0064 c) 0.992
11. 5 **13.** 20 calls
15. a) 25 b) 0.185 c) 0.217 d) 0.693
17. a) 0.0745 b) 0.502 c) 0.211
d) 0.0166 e) 0.0179 f) 0.9987
19. a) 0.65 b) 0.75 c) 7.69 picks
21. a) $\mu = 10.44, \sigma = 1.16$
b) i) 0.812 ii) 0.475 iii) 0.00193 iv) 0.998

23. $\mu = 20.28, \sigma = 4.22$

25. a) 0.118 b) 0.324 c) 0.744 d) 0.580

27. a) $\mu = 56, \sigma = 4.10$

b) Yes, $np = 56 \geq 10$, $nq = 24 \geq 10$, serves are independent.

c) In a match with 80 serves, approximately 68% of the time she will have between 51.9 and 60.1 good serves, approximately 95% of the time she will have between 47.8 and 64.2 good serves, and approximately 99.7% of the time she will have between 43.7 and 68.3 good serves.

d) Normal, approx.: 0.014; Binomial, exact: 0.016

29. a) Assuming apples fall and become blemished independently of each other, Binom(300, 0.06) is appropriate. Since $np \geq 10$ and $nq \geq 10$, $N(18, 4.11)$ is also appropriate.

b) Normal, approx.: 0.072; Binomial, exact: 0.085

c) No, 50 is 7.8 SDs above the mean.

31. Normal, approx.: 0.053; Binomial, exact: 0.061

33. The mean number of sales should be 24 with SD 4.60. Ten sales is more than 3.0 SDs below the mean. He was probably misled.

35. a) 5 b) 0.066 c) 0.107 d) $\mu = 24, \sigma = 2.19$

e) Normal, approx.: 0.819; Binomial, exact: 0.848

37. $\mu = 20, \sigma = 4$. I'd want *at least* 32 (3 SDs above the mean). (Answers will vary.)

39. Probably not. There's a more than 9% chance that he could hit 4 shots in a row, so he can expect this to happen nearly once in every 10 sets of 4 shots he takes. That does not seem unusual.

41. Yes. We'd expect him to make 22 shots, with a standard deviation of 3.15 shots. 32 shots is more than 3 standard deviations above the expected value, an unusually high rate of success.

PART IV REVIEW

1. a) 0.34 b) 0.27 c) 0.069

d) No, 2% of cars have both types of defects.

e) Of all cars with cosmetic defects, 6.9% have functional defects. Overall, 7.0% of cars have functional defects. The probabilities here are estimates, so these are probably close enough to say the defects are independent.

3. a) C = Price to China; F = Price to France; Total $= 3C + 5F$

b) $\mu = \$5500, \sigma = \$672.68$ c) $\mu = \$500, \sigma = \$180.28$

d) Means—no. Standard deviations—yes; ticket prices must be independent of each other for different countries, but all tickets to the same country are at the same price.

5. a) $\mu = -\$0.20, \sigma = \$1.89$ b) $\mu = -\$0.40, \sigma = \$2.67$

7. a) 0.106 b) 0.651 c) 0.442

9. a) 0.590 b) 0.328 c) 0.00856

11. a) $\mu = 15.2, \sigma = 3.70$ b) Yes, $np \geq 10$ and $nq \geq 10$

c) Normal, approx.: 0.080; Binomial, exact: 0.097

13. a) 0.0173 b) 0.591

c) Left: 960; right: 120; both: 120

d) $\mu = 120, \sigma = 10.39$

e) About 68% chance of between 110 and 130; about 95% between 99 and 141; about 99.7% between 89 and 151.

15. a) Men's heights are more variable than women's.

b) Men (1.75 SD vs 2.4 SD for women)

c) M = Man's height; W = Woman's height; $M - W$ is how much taller the man is.

d) 5.1" e) 3.75" f) 0.913

g) If independent, it should be about 91.3%. We are told 92%. This difference seems small and may be due to natural sampling variability.

17. a) The chance is 1.6×10^{-7}. b) 0.952 c) 0.063

19. \$240

21. a) 0.717 b) 0.588

23. a) $\mu = 100, \sigma = 8$ b) $\mu = 1000, \sigma = 60$

c) $\mu = 100, \sigma = 8.54$ d) $\mu = -50, \sigma = 10$

e) $\mu = 100, \sigma = 11.31$

25. a) Many do both, so the two categories can total more than 100%.

b) No. They can't be disjoint. If they were, the total would be 100% or less.

c) No. Probabilities are different for boys and girls.

d) 0.0524

27. a) 21 days b) 1649.73 som

c) 3300 som extra. About 157-som "cushion" each day.

29. No, you'd expect 541.2 homeowners, with an SD of 13.56. 523 is 1.34 SDs below the mean; not unusual.

31. a) 0.018 b) 0.300 c) 0.26

33. a) 6 b) 15 c) 0.402

35. a) 34% b) 35% c) 31.4%

d) 31.4% of classes that used calculators used computer assignments, while in classes that didn't use calculators, 30.6% used computer assignments. These are close enough to think the choice is probably independent.

37. a) 1/11 b) 7/22 c) 5/11 d) 0 e) 19/66

39. a) Expected number of stars with planets.

b) Expected number of planets with intelligent life.

c) Probability of a planet with a suitable environment having intelligent life.

d) f_l: If a planet has a suitable environment, the probability that life develops.

f_i: If a planet develops life, the probability that the life evolves intelligence.

f_c: If a planet has intelligent life, the probability that it develops radio communication.

41. 0.991

CHAPTER 18

1. All the histograms are centered near 0.05. As n gets larger, the histograms approach the Normal shape, and the variability in the sample proportions decreases.

3. a)

n	Observed mean	Theoretical mean	Observed st. dev.	Theoretical st. dev.
20	0.0497	0.05	0.0479	0.0487
50	0.0516	0.05	0.0309	0.0308
100	0.0497	0.05	0.0215	0.0218
200	0.0501	0.05	0.0152	0.0154

b) They are all quite close to what we expect from the theory.

c) The histogram is unimodal and symmetric for $n = 200$.

d) The success/failure condition says that np and nq should both be at least 10, which is not satisfied until $n = 200$ for $p = 0.05$. The theory predicted my choice.

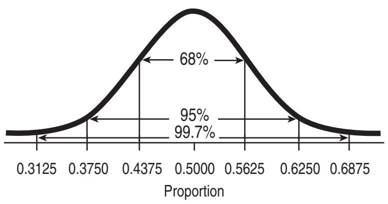
5. a) Symmetric, because probability of heads and tails is equal.

b) 0.5 c) 0.125 d) $np = 8 < 10$

7. a) About 68% should have proportions between 0.4 and 0.6, about 95% between 0.3 and 0.7, and about 99.7% between 0.2 and 0.8.

b) $np = 12.5$, $nq = 12.5$; both are ≥ 10.

c)

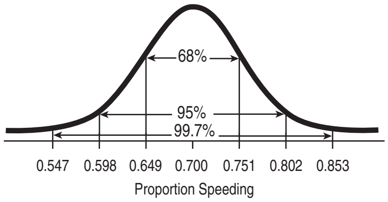


$np = nq = 32$; both are ≥ 10.

d) Becomes narrower (less spread around 0.5).

9. This is a fairly unusual result: about 2.26 SDs below the mean. The probability of that is about 0.012. So, in a class of 100 this is certainly a reasonable possibility.

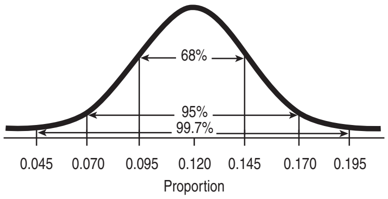
11. a)



b) Both $np = 56$ and $nq = 24 \geq 10$. Drivers *may* be independent of each other, but if flow of traffic is very fast, they may not be. Or weather conditions may affect all drivers. In these cases they may get more or fewer speeders than they expect.

13. a) Assume that these children are typical of the population. They represent fewer than 10% of all children. We expect 20.4 near-sighted and 149.6 not; both are at least 10.

b)



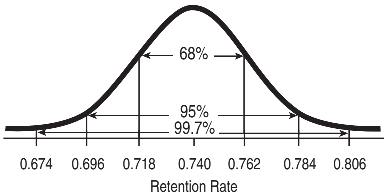
c) Probably between 12 and 29.

15. a) $\mu = 7\%$, $\sigma = 1.8\%$

b) Assume that clients pay independently of each other, that we have a random sample of all possible clients, and that these represent less than 10% of all possible clients. $np = 14$ and $nq = 186$ are both at least 10.

c) 0.048

17.



These are not random samples, and not all colleges may be typical (representative). $np = 296$, $nq = 104$ are both at least 10.

19. Yes; if their students were typical, a retention rate of $522/603 = 86.6\%$ would be over 7 standard deviations above the expected rate of 74%.

21. 0.212. Reasonable that those polled are independent of each other and represent less than 10% of all potential voters. We assume the sample was selected at random. Success/Failure Condition met: $np = 208$, $nq = 192$. Both ≥ 10.

23. 0.088 using $N(0.08, 0.022)$ model.

25. Answers will vary. Using $\mu + 3\sigma$ for "very sure," the restaurant should have 89 nonsmoking seats. Assumes customers at any time are independent of each other, a random sample, and represent less than 10% of all potential customers. $np = 72$, $nq = 48$, so Normal model is reasonable ($\mu = 0.60$, $\sigma = 0.045$).

27. a) Normal, center at μ, standard deviation $\sigma/\sqrt{n}$.

b) Standard deviation will be smaller. Center will remain the same.

29. a) The histogram is unimodal and slightly skewed to the right, centered at 36 inches with a standard deviation near 4 inches.

b) All the histograms are centered near 36 inches. As n gets larger, the histograms approach the Normal shape and the variability in the sample means decreases. The histograms are fairly normal by the time the sample reaches size 5.

31. a)

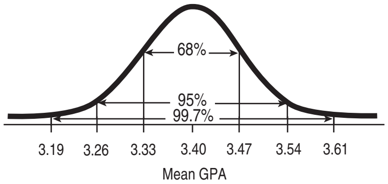
n	Observed mean	Theoretical mean	Observed st. dev.	Theoretical st. dev.
2	36.314	36.33	2.855	2.842
5	36.314	36.33	1.805	1.797
10	36.341	36.33	1.276	1.271
20	36.339	36.33	0.895	0.899

b) They are all very close to what we would expect.

c) For samples as small as 5, the sampling distribution of sample means is unimodal and very symmetric.

d) The distribution of the original data is nearly unimodal and symmetric, so it doesn't take a very large sample size for the distribution of sample means to be approximately Normal.

33.



Normal, $\mu = 3.4$, $\sigma = 0.07$. We assume that the students are randomly assigned to the seminars and represent less than 10% of all possible students, and that individual's GPAs are independent of one another.

35. a) As the CLT predicts, there is more variability in the smaller outlets.

b) If the lottery is random, all outlets are equally likely to sell winning tickets.

37. a) 21.1% b) 276.8 days or more

c) $N(266, 2.07)$ d) 0.002

39. a) There are more premature births than very long pregnancies. Modern practice of medicine stops pregnancies at about 2 weeks past normal due date.

b) Parts (a) and (b)—yes—we can't use Normal model if it's very skewed. Part (c)—no—CLT guarantees a Normal model for this large sample size.

41. a) $\mu = \$2.00$, $\sigma = \$3.61$

b) $\mu = \$4.00$, $\sigma = \$5.10$

c) 0.191. Model is $N(80, 22.83)$.

43. a) $\mu = 2.859$, $\sigma = 1.324$

b) No. The score distribution in the sample should resemble that in the population, somewhat uniform for scores 1–4 and about half as many 5's.

c) Approximately $N\left(2.859, \frac{1.324}{\sqrt{40}}\right)$.

45. About 20%, based on $N(2.859, 0.167)$.

47. a) $N(2.9, 0.045)$ b) 0.0131 c) 2.97 gm/mi

49. a) Can't use a Normal model to estimate probabilities. The distribution is skewed right—not Normal.

b) 4 is probably not a large enough sample to say the average follows the Normal model.

c) No. This is 3.16 SDs above the mean.

51. a) 0.0003. Model is $N(384, 34.15)$. b) \$427.77 or more.

53. a) 0.734

b) 0.652. Model is $N(10, 12.81)$.

c) 0.193. Model is $N(120, 5.774)$.

d) 0.751. Model is $N(10, 7.394)$.

CHAPTER 19

1. She believes the true proportion is within 4% of her estimate, with some (probably 95%) degree of confidence.

3. a) Population—all cars; sample—those actually stopped at the checkpoint; p—proportion of all cars with safety problems;

$\hat{p}$—proportion actually seen with safety problems (10.4%); if sample (a cluster sample) is representative, then the methods of this chapter will apply.
b) Population—general public; sample—those who logged onto the Web site; p—population proportion of those who favor prayer in school; $\hat{p}$—proportion of those who voted in the poll who favored prayer in school (81.1%); can't use methods of this chapter—sample is biased and nonrandom.
c) Population—parents at the school; sample—those who returned the questionnaire; p—proportion of all parents who favor uniforms; $\hat{p}$—proportion of respondents who favor uniforms (60%); should not use methods of this chapter, since not SRS (possible non-response bias).
d) Population—students at the college; sample—the 1632 students who entered that year; p—proportion of all students who will graduate on time; $\hat{p}$—proportion of that year's students who graduate on time (85.0%); can use methods of this chapter if that year's students (a cluster sample) are viewed as a representative sample of all possible students at the school.

5. a) Not correct. This implies certainty.
b) Not correct. Different samples will give different results. Many fewer than 95% will have 88% on-time orders.
c) Not correct. The interval is about the population proportion, not the sample proportion in different samples.
d) Not correct. In this sample, we *know* 88% arrived on time.
e) Not correct. The interval is about the parameter, not the days.

7. a) False b) True c) True d) False

9. On the basis of this sample, we are 90% confident that the proportion of Japanese cars is between 29.9% and 47.0%.

11. a) (0.798, 0.863)
b) We're 95% confident that between 80% and 86% of all broiler chicken sold in U.S. food stores is infected with *Campylobacter*.
c) The size of the population is irrelevant. If *Consumer Reports* had a random sample, 95% of intervals generated by studies like this will capture the true contamination level.

13. a) 0.025
b) We're 90% confident that this poll's estimate is within ±2.5% of the true proportion of people who are baseball fans.
c) Larger. To be more certain, we must be less precise.
d) 0.039 e) less confidence
f) No evidence of change; given the margin of error, 0.37 is a plausible value for 2007 as well.

15. a) (0.0465, 0.0491). The assumptions and conditions for constructing a confidence interval are satisfied.
b) The confidence interval gives the set of plausible values (with 95% confidence). Since 0.05 is outside the interval, that seems to be a bit too optimistic.

17. a) (12.7%, 18.6%)
b) We are 95% confident, based on this sample, that the proportion of all auto accidents that involve teenage drivers is between 12.7% and 18.6%.
c) About 95% of all random samples will produce confidence intervals that contain the true population proportion.
d) Contradicts. The interval is completely below 20%.

19. Probably nothing. Those who bothered to fill out the survey may be a biased sample.

21. a) Response bias (wording) b) (54%, 60%)
c) Smaller—the sample size was larger.

23. a) (18.2%, 21.8%)
b) We are 98% confident, based on the sample, that between 18.2% and 21.8% of English children are deficient in vitamin D.
c) About 98% of all random samples will produce a confidence interval that contains the true proportion of children deficient in vitamin D.

25. a) Wider. The sample size is probably about one-fourth of the sample size for all adults, so we'd expect the confidence interval to be about twice as wide.
b) Smaller. The second poll used a slightly larger sample size.

27. a) (15.5%, 26.3%) b) 612
c) Sample may not be random or representative. Deer that are legally hunted may not represent all sexes and ages.

29. a) 141 b) 318 c) 564

31. 1801 **33.** 384 total, using $p = 0.15$ **35.** 90%

CHAPTER 20

1. a) $H_0: p = 0.30; H_A: p < 0.30$
b) $H_0: p = 0.50; H_A: p \neq 0.50$
c) $H_0: p = 0.20; H_A: p > 0.20$

3. Statement d is correct.

5. No, we can say only that there is a 27% chance of seeing the observed effectiveness just from natural sampling variation. There is no *evidence* that the new formula is more effective, but we can't conclude that they are equally effective.

7. a) No. There's a 25% chance of losing twice in a row. That's not unusual.
b) 0.125 c) No, we expect that to happen 1 time in 8.
d) Maybe 5? The chance of 5 losses in a row is only 1 in 32, which seems unusual.

9. 1) Use p, not $\hat{p}$, in hypotheses.
2) The question was about failing to meet the goal, so H_A should be $p < 0.96$.
3) Did not check $0.04(200) = 8$. Since $nq < 10$, the Success/Failure Condition is violated. Didn't check 10% Condition.
4) $188/200 = 0.94; SD(\hat{p}) = \sqrt{\dfrac{(0.96)(0.04)}{200}} = 0.014$
5) z is incorrect; should be $z = \dfrac{0.94 - 0.96}{0.014} = -1.43$
6) $P = P(z < -1.43) = 0.076$
7) There is only weak evidence that the new instructions do not work.

11. a) $H_0: p = 0.30; H_A: p > 0.30$
b) Possibly an SRS; we don't know if the sample is less than 10% of his customers, but it could be viewed as less than 10% of all possible customers; $(0.3)(80) \geq 10$ and $(0.7)(80) \geq 10$. Wells are independent only if customers don't have farms on the same underground springs.
c) $z = 0.73$; P-value $= 0.232$
d) If his dowsing is no different from standard methods, there is more than a 23% chance of seeing results as good as those of the dowser's, or better, by natural sampling variation.
e) These data provide no evidence that the dowser's chance of finding water is any better than normal drilling.

13. a) $H_0: p_{2000} = 0.34; H_A: p_{2000} \neq 0.34$
b) Students were randomly sampled and should be independent. 34% and 66% of 8302 are greater than 10. 8302 students is less than 10% of the entire student population of the United States.
c) $P = 0.058$
d) With such a small P-value, I reject H_0. There has been a statistically significant change in the proportion of students who have no absences.
e) No. A difference this small, although statistically significant, is not meaningful. We might look at new data in a few years.

15. a) $H_0: p = 0.05$ vs. $H_A: p < 0.05$
b) We assume the whole mailing list has over 1,000,000 names. This is a random sample, and we expect 5000 successes and 95,000 failures.
c) $z = -3.178$; P-value $= 0.00074$, so we reject H_0; there is strong evidence that the donation rate would be below 5%.

17. a) $H_0: p = 0.63, H_A: p > 0.63$
b) The sample is representative. $240 < 10\%$ of all law school applicants. We expect $240(0.63) = 151.2$ to be admitted and $240(0.37) = 88.8$ not to be, both at least 10. $z = 1.58$; P-value = 0.057

c) Although the evidence is weak, there is some indication that the program may be successful. Candidates should decide whether they can afford the time and expense.

19. $H_0: p = 0.20; H_A: p > 0.20$. SRS (not clear from information provided); 22 is more than 10% of the population of 150; $(0.20)(22) < 10$. Do not proceed with a test.

21. $H_0: p = 0.03; p \neq 0.03$. $\hat{p} = 0.015$. One mother having twins will not affect another, so observations are independent; not an SRS; sample is less than 10% of all births. However, the mothers at this hospital may not be representative of all teenagers; $(0.03)(469) = 14.07 \geq 10; (0.97)(469) \geq 10$. $z = -1.91$; P-value $= 0.0556$. With a P-value this low, reject H_0. These data show some evidence that the rate of twins born to teenage girls at this hospital is less than the national rate of 3%. It is not clear whether this can be generalized to all teenagers.

23. $H_0: p = 0.25; H_A: p > 0.25$. SRS; sample is less than 10% of all potential subscribers; $(0.25)(500) \geq 10; (0.75)(500) \geq 10$. $z = 1.24$; P-value $= 0.1076$. The P-value is high, so do not reject H_0. These data do not show that more than 25% of current readers would subscribe; the company should not go ahead with the WebZine on the basis of these data.

25. $H_0: p = 0.40; H_A: p < 0.40$. Data are for all executives in this company and may not be able to be generalized to all companies; $(0.40)(43) \geq 10; (0.60)(43) \geq 10$. $z = -1.31$; P-value $= 0.0955$. Because the P-value is high, we fail to reject H_0. These data do not show that the proportion of women executives is less than the 40% of women in the company in general.

27. $H_0: p = 0.103; H_A: p > 0.103$. $\hat{p} = 0.118; z = 2.06$; P-value $= 0.02$. Because the P-value is low, we reject H_0. These data provide evidence that the dropout rate has increased.

29. $H_0: p = 0.90; H_A: p < 0.90$. $\hat{p} = 0.844; z = -2.05$; P-value $= 0.0201$. Because the P-value is so low, we reject H_0. There is strong evidence that the actual rate at which passengers with lost luggage are reunited with it within 24 hours is less than the 90% claimed by the airline.

31. a) Yes; assuming this sample to be a typical group of people, $P = 0.0008$. This cancer rate is very unusual.
b) No, this group of people may be atypical for reasons that have nothing to do with the radiation.

CHAPTER 21

1. a) Two sided. Let p be the percentage of students who prefer Diet Pepsi. $H_0: p = 0.5$ vs. $H_A: p \neq 0.5$
b) One sided. Let p be the percentage of teenagers who prefer the new formulation. $H_0: p = 0.5$ vs. $H_A: p > 0.5$
c) One sided. Let p be the percentage of people who intend to vote for the override. $H_0: p = 2/3$ vs. $H_A: p > 2/3$.
d) Two sided. Let p be the percentage of days that the market goes up. $H_0: p = 0.5$ vs. $H_A: p \neq 0.5$

3. If there is no difference in effectiveness, the chance of seeing an observed difference this large or larger is 4.7% by natural sampling variation.

5. $\alpha = 0.10$: Yes. The P-value is less than 0.05, so it's less than 0.10. But to reject H_0 at $\alpha = 0.01$, the P-value must be below 0.01, which isn't necessarily the case.

7. a) There is only a 1.1% chance of seeing a sample proportion as low as 89.4% vaccinated by natural sampling variation if 90% have really been vaccinated.
b) We conclude that p is below 0.9, but a 95% confidence interval would suggest that the true proportion is between (0.889, 0.899). Most likely, a decrease from 90% to 89.9% would not be considered important. On the other hand, with 1,000,000 children a year vaccinated, even 0.1% represents about 1000 kids—so this may very well be important.

9. a) (1.9%, 4.1%)
b) Because 5% is not in the interval, there is strong evidence that fewer than 5% of all men use work as their primary measure of success.
c) $\alpha = 0.01$; it's a lower-tail test based on a 98% confidence interval.

11. a) (0.274, 0.327)
b) Since 0.27 is not in the confidence interval, we reject the hypothesis that $p = 0.27$

13. a) The Success/Failure Condition is violated: only 5 pups had dysplasia.
b) We are 95% confident that between 5% and 26% of puppies will show signs of hip dysplasia at the age of 6 months.

15. a) Type II error b) Type I error
c) By making it easier to get the loan, the bank has reduced the alpha level.
d) The risk of a Type I error is decreased and the risk of a Type II error is increased.

17. a) Power is the probability that the bank denies a loan that would not have been repaid.
b) Raise the cutoff score.
c) A larger number of trustworthy people would be denied credit, and the bank would miss the opportunity to collect interest on those loans.

19. a) The null is that the level of home ownership remains the same. The alternative is that it rises.
b) The city concludes that home ownership is on the rise, but in fact the tax breaks don't help.
c) The city abandons the tax breaks, but they were helping.
d) A Type I error causes the city to forego tax revenue, while a Type II error withdraws help from those who might have otherwise been able to buy a home.
e) The power of the test is the city's ability to detect an actual increase in home ownership.

21. a) It is decided that the shop is not meeting standards when it is.
b) The shop is certified as meeting standards when it is not.
c) Type I d) Type II

23. a) The probability of detecting a shop that is not meeting standards.
b) 40 cars. Larger n. c) 10%. More chance to reject H_0.
d) A lot. Larger differences are easier to detect.

25. a) One-tailed. The company wouldn't be sued if "too many" minorities were hired.
b) Deciding the company is discriminating when it is not.
c) Deciding the company is not discriminating when it is.
d) The probability of correctly detecting actual discrimination.
e) Increases power. f) Lower, since n is smaller.

27. a) One-tailed. Software is supposed to decrease the dropout rate.
b) $H_0: p = 0.13; H_A: p < 0.13$
c) He buys the software when it doesn't help students.
d) He doesn't buy the software when it does help students.
e) The probability of correctly deciding the software is helpful.

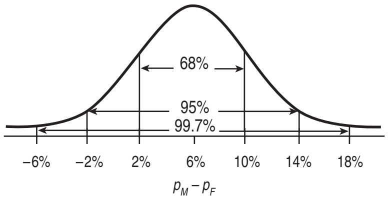
29. a) $z = -3.21, p = 0.0007$. The change is statistically significant. A 95% confidence interval is (2.3%, 8.5%). This is clearly lower than 13%. If the cost of the software justifies it, the professor should consider buying the software.
b) The chance of observing 11 or fewer dropouts in a class of 203 is only 0.07% if the dropout rate is really 13%.

31. a) $H_A: p = 0.30$, where p is the probability of heads
b) Reject the null hypothesis if the coin comes up tails—otherwise fail to reject.
c) $P(\text{tails given the null hypothesis}) = 0.1 = \alpha$.
d) $P(\text{tails given the alternative hypothesis}) = \text{power} = 0.70$
e) Spin the coin more than once and base the decision on the sample proportion of heads.

33. a) 0.0464 b) Type I c) 37.6%
d) Increase the number of shots. Or keep the number of shots at 10, but increase alpha by declaring that 8, 9, or 10 will be deemed as having improved.

CHAPTER 22

1. It's very unlikely that samples would show an observed difference this large if in fact there is no real difference in the proportions of boys and girls who have used online social networks.
3. The ads may be working. If there had been no real change in name recognition, there'd be only about a 3% chance the percentage of voters who heard of this candidate would be at least this much higher in a different sample.
5. The responses are not from two independent groups, but are from the same individuals.
7. a) Stratified b) 6% higher among males c) 4%
d)



e) Yes; a poll result showing little difference is only 1–2 standard deviations below the expected outcome.
9. a) Yes. Random sample; less than 10% of the population; samples are independent; more than 10 successes and failures in each sample.
b) (0.055, 0.140)
c) We are 95% confident, based on these samples, that the proportion of American women age 65 and older who suffer from arthritis is between 5.5% and 14.0% more than the proportion of American men of the same age who suffer from arthritis.
d) Yes; the entire interval lies above 0.
11. a) 0.035 b) (0.356, 0.495)
c) We are 95% confident, based on these data, that the proportion of pets with a malignant lymphoma in homes where herbicides are used is between 35.6% and 49.5% higher than the proportion of pets with lymphoma in homes where no pesticides are used.
13. a) Yes, subjects were randomly divided into independent groups, and more than 10 successes and failures were observed in each group.
b) (4.7%, 8.9%)
c) Yes, we're 95% confident that the rate of infection is 5–9 percentage points lower. That's a meaningful reduction, considering the 20% infection rate among the unvaccinated kids.
15. a) $H_0: p_V - p_{NV} = 0, H_A: p_V - p_{NV} < 0$.
b) Because 0 is not in the confidence interval, reject the null. There's evidence that the vaccine reduces the rate of ear infections.
c) 2.5% d) Type I
e) Babies would be given ineffective vaccinations.
17. a) Prospective study
b) $H_0: p_1 - p_2 = 0; H_A: p_1 - p_2 \neq 0$ where p_1 is the proportion of students whose parents disapproved of smoking who became smokers and p_2 is the proportion of students whose parents are lenient about smoking who became smokers.
c) Yes. We assume the students were randomly selected; they are less than 10% of the population; samples are independent; at least 10 successes and failures in each sample.
d) $z = -1.17$, P-value $= 0.2422$. These samples do not show evidence that parental attitudes influence teens' decisions to smoke.
e) If there is no difference in the proportions, there is about a 24% chance of seeing the observed difference or larger by natural sampling variation.
f) Type II
19. a) $(-0.065, 0.221)$
b) We are 95% confident that the proportion of teens whose parents disapprove of smoking who will eventually smoke is between 22.1% less and 6.5% more than for teens with parents who are lenient about smoking.
c) 95% of all random samples will produce intervals that contain the true difference.
21. a) No; subjects weren't assigned to treatment groups. It's an observational study.
b) $H_0: p_1 - p_2 = 0; H_A: p_1 - p_2 \neq 0$. $z = 3.56$, P-value $= 0.0004$. With a P-value this low, we reject H_0. There is a significant difference in the clinic's effectiveness. Younger mothers have a higher birth rate than older mothers. Note that the Success/Failure Condition is met based on the pooled estimate of p.
c) We are 95% confident, based on these data, that the proportion of successful live births at the clinic is between 10.0% and 27.8% higher for mothers under 38 than in those 38 and older. However, the Success/Failure Condition is not met for the older women, since # Successes < 10. We should be cautious in trusting this confidence interval.
23. a) $H_0: p_1 - p_2 = 0; H_A: p_1 - p_2 > 0$. $z = 1.18$, P-value $= 0.118$. With P-value this high, we fail to reject H_0. These data do not show evidence of a decrease in the voter support for the candidate.
b) Type II
25. a) $H_0: p_1 - p_2 = 0; H_A: p_1 - p_2 \neq 0$. $z = -0.39$, P-value $= 0.6951$. With a P-value this high, we fail to reject H_0. There is no evidence of racial differences in the likelihood of multiple births, based on these data.
b) Type II
27. a) We are 95% confident, that between 67.0% and 83.0% of patients with joint pain will find medication A effective.
b) We are 95% confident, that between 51.9% and 70.3% of patients with joint pain will find medication B effective.
c) Yes, they overlap. This might indicate no difference in the effectiveness of the medications. (Not a proper test.)
d) We are 95% confident that the proportion of patients with joint pain who will find medication A effective is between 1.7% and 26.1% higher than the proportion who will find medication B effective.
e) No. There is a difference in the effectiveness of the medications.
f) To estimate the variability in the difference of proportions, we must add variances. The two one-sample intervals do not. The two-sample method is the correct approach.
29. The conditions are satisfied to test $H_0: p_{young} = p_{old}$ against $H_A: p_{young} > p_{old}$. The one-sided P-value is 0.0619, so we may reject the null hypothesis. Although the evidence is not strong, *Time* may be justified in saying that younger men are more comfortable discussing personal problems.
31. Yes. With a low P-value of 0.003, reject the null hypothesis of no difference. There's evidence of an increase in the proportion of parents checking the Web sites visited by their teens.

PART V REVIEW

1. H_0: There is no difference in cancer rates, $p_1 - p_2 = 0$. H_A: The cancer rate in those who use the herb is higher, $p_1 - p_2 > 0$.
3. a) 10.29
b) Not really. The z-score is -1.11. Not any evidence to suggest that the proportion for Monday is low.
c) Yes. The z-score is 2.26 with a P-value of 0.024 (two-sided).
d) Some births are scheduled for the convenience of the doctor and/or the mother.
5. a) $H_0: p_1 = 0.40; H_A: p_1 < 0.40$
b) Random sample; less than 10% of all California gas stations, $0.4(27) = 10.8, 0.6(27) = 16.2$. Assumptions and conditions are met.
c) $z = -1.49$, P-value $= 0.0677$
d) With a P-value this high, we fail to reject H_0. These data do not provide evidence that the proportion of leaking gas tanks is less than 40% (or that the new program is effective in decreasing the proportion).

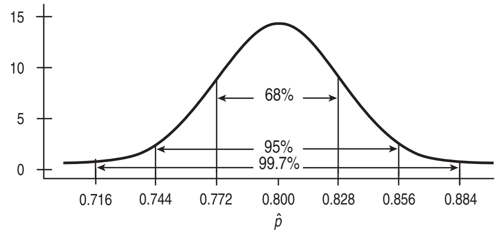
e) Yes, Type II.
f) Increase α, increase the sample size.
g) Increasing α—increases power, lowers chance of Type II error, but increases chance of Type I error.
Increasing sample size—increases power, costs more time and money.

7. a) The researcher believes that the true proportion of "A's" is within 10% of the estimated 54%, namely, between 44% and 64%.
b) Small sample c) No, 63% is contained in the interval.

9. a) Pew believes that the true proportion is within 3% of the 33% from the sample; that is, between 30% and 36%.
b) Larger, since it's a smaller sample.
c) We are 95% confident that the proportion of active traders who rely on the Internet for investment information is between 38.7% and 51.3%, based on this sample.
d) Larger, since it's a smaller sample.

11. a) Bimodal!
b) μ, the population mean. Sample size does not matter.
c) $\sigma/\sqrt{n}$; sample size does matter.
d) It becomes closer to a Normal model and narrower as the sample size increases.

13. a) $\mu = 0.80, \sigma = 0.028$
b) Yes. $0.8(200) = 160, 0.2(200) = 40$. Both ≥ 10.
c)



d) 0.039

15. H_0: There is no difference, $p_1 - p_2 = 0$. H_A: Early births have increased, $p_1 - p_2 < 0$. $z = -0.729$, P-value $= 0.2329$. Because the P-value is so high, we do not reject H_0. These data do not show an increase in the incidence of early birth of twins.

17. a) H_0: There is no difference, $p_1 - p_2 \geq 0$. H_A: Treatment prevents deaths from eclampsia, $p_1 - p_2 < 0$.
b) Samples are random and independent; less than 10% of all pregnancies (or eclampsia cases); more than 10 successes and failures in each group.
c) 0.8008
d) There is insufficient evidence to conclude that magnesium sulfide is effective in preventing eclampsia deaths.
e) Type II f) Increase the sample size, increase α.
g) Increasing sample size: decreases variation in the sampling distribution, is costly. Increasing α: Increases likelihood of rejecting H_0, increases chance of Type I error.

19. a) It is not clear what the pollster asked. Otherwise they did fine.
b) Stratified sampling. c) 4%
d) 95% e) Smaller sample size.
f) Wording and order of questions (response bias).

21. a) H_0: There is no difference, $p = 0.143$. H_A: The fatal accident rate is lower in girls, $p < 0.143$. $z = -1.67$, P-value $= 0.0479$. Because the P-value is low, we reject H_0. These data give some evidence that the fatal accident rate is lower for girls than for teens in general.
b) If the proportion is really 14.3%, we will see the observed proportion (11.3%) or lower 4.8% of the time by sampling variation.

23. a) One would expect many small fish, with a few large ones.
b) We don't know the exact distribution, but we know it's not Normal.
c) Probably not. With a skewed distribution, a sample size of five is not a large enough sample to say the sampling model for the mean is approximately Normal.
d) 0.961

25. a) Yes. $0.8(60) = 48, 0.2(60) = 12$. Both are ≥ 10.
b) 0.834
c) Higher. Bigger sample means smaller standard deviation for $\hat{p}$.
d) Answers will vary. For $n = 500$, the probability is 0.997.

27. a) 54.4 to 62.5%
b) Based on this study, with 95% confidence the proportion of Crohn's disease patients who will respond favorable to infliximab is between 54.4% and 62.5%.
c) 95% of all such random samples will produce confidence intervals that contain the true proportion of patients who respond favorably.

29. At least 423, assuming that p is near 50%.

31. a) Random sample (?); certainly less than 10% of all preemies and normal babies; more than 10 failures and successes in each group. 1.7% to 16.3% greater for normal-birth weight children.
b) Since 0 is not in the interval, there is evidence that preemies have a lower high school graduation rate than children of normal birth weight.
c) Type I, since we rejected the null hypothesis.

33. a) H_0: The computer is undamaged. H_A: The computer is damaged.
b) 20% of good PCs will be classified as damaged (bad), while all damaged PCs will be detected (good).
c) 3 or more. d) 20%
e) By switching to two or more as the rejection criterion, 7% of the good PCs will be misclassified, but only 10% of the bad ones will, increasing the power from 20% to 90%.

35. The null hypothesis is that Bush's disapproval proportion is 66%—the Nixon benchmark. The one-tailed test has a z-value of -2.00, so the P-value is 0.0228. It looks like Bush's May 2007 ratings were better than the Nixon benchmark low.

37. a) The company is interested only in confirming that the athlete is well known.
b) Type I: the company concludes that the athlete is well known, but that's not true. It offers an endorsement contract to someone who lacks name recognition. Type II: the company overlooks a well-known athlete, missing the opportunity to sign a potentially effective spokesperson.
c) Type I would be more likely, Type II less likely.

39. I am 95% confident that the proportion of U.S. adults who favor nuclear energy is between 7 and 19 percentage points higher than the proportion who would accept a nuclear plant near their area.

CHAPTER 23

1. a) 1.74 b) 2.37 c) 0.0524 d) 0.0889

3. Shape becomes closer to Normal; center does not change; spread becomes narrower.

5. a) The confidence interval is for the population mean, not the individual cows in the study.
b) The confidence interval is not for individual cows.
c) We *know* the average gain in this study was 56 pounds!
d) The average weight gain of all cows does not vary. It's what we're trying to estimate.
e) No. There is not a 95% chance for another sample to have an average weight gain between 45 and 67 pounds. There is a 95% chance that another sample will have its average weight gain within two standard errors of the true mean.

7. a) No. A confidence interval is not about individuals in the population.
b) No. It's not about individuals in the sample, either.
c) No. We know the mean cost for students in the sample was \$1196.
d) No. A confidence interval is not about other sample means.
e) Yes. A confidence interval estimates a population parameter.

9. a) Based on this sample, we can say, with 95% confidence, that the mean pulse rate of adults is between 70.9 and 74.5 beats per minute.
b) 1.8 beats per minute
c) Larger

11. The assumptions and conditions for a t-interval are not met. The distribution is highly skewed to the right and there is a large outlier.

13. a) Yes. Randomly selected group; less than 10% of the population; the histogram is not unimodal and symmetric, but it is not highly skewed and there are no outliers, so with a sample size of 52, the CLT says $\bar{y}$ is approximately Normal.
b) (98.06, 98.51) degrees F
c) We are 98% confident, based on the data, that the average body temperature for an adult is between 98.06°F and 98.51°F.
d) 98% of all such random samples will produce intervals containing the true mean temperature.
e) These data suggest that the true normal temperature is somewhat less than 98.6°F.

15. a) Narrower. A smaller margin of error, so less confident.
b) Advantage: more chance of including the true value. Disadvantage: wider interval.
c) Narrower; due to the larger sample, the SE will be smaller.
d) About 252

17. a) (709.90, 802.54)
b) With 95% confidence, based on these data, the speed of light is between 299,709.9 and 299,802.5 km/sec.
c) Normal model for the distribution, independent measurements. These seem reasonable here, but it would be nice to see if the Nearly Normal Condition held for the data.

19. a) Given no time trend, the monthly on-time departure rates should be independent. Though not a random sample, these months should be representative, and they're fewer than 10% of all months. The histogram looks unimodal, but slightly left-skewed; not a concern with this large sample.
b) $80.57 < \mu(\text{OT Departure\%}) < 81.80$
c) We can be 90% confident that the interval from 80.57% to 81.80% holds the true mean monthly percentage of on-time flight departures.

21. The 95% confidence interval lies entirely above the 0.08 ppm limit, evidence that mirex contamination is too high and consistent with rejecting the null. We used an upper-tail test, so the P-value should therefore be smaller than $\frac{1}{2}(1 - 0.95) = 0.025$, and it was.

23. If in fact the mean cholesterol of pizza eaters does not indicate a health risk, then only 7 of every 100 samples would have mean cholesterol levels as high (or higher) as observed in this sample.

25. a) Upper-tail. We want to show it will hold 500 pounds (or more) easily.
b) They will decide the stands are safe when they're not.
c) They will decide the stands are unsafe when they are in fact safe.

27. a) Decrease α. This means a smaller chance of declaring the stands safe if they are not.
b) The probability of correctly detecting that the stands are capable of holding more than 500 pounds.
c) Decrease the standard deviation—probably costly. Increase the sample size—takes more time for testing and is costly. Increase α—more Type I errors. Increase the "design load" to be well above 500 pounds—again, costly.

29. a) $H_0: \mu = 23.3; H_A: \mu > 23.3$
b) We have a random sample of the population. Population may not be normally distributed, as it would be easier to have a few much older men at their first marriage than some very young men. However, with a sample size of 40, $\bar{y}$ should be approximately Normal. We should check the histogram for severity of skewness and possible outliers.
c) $(\bar{y} - 23.3)/(s/\sqrt{40}) \sim t_{39}$ d) 0.1447
e) If the average age at first marriage is still 23.3 years, there is a 14.5% chance of getting a sample mean of 24.2 years or older simply from natural sampling variation.
f) We lack evidence that the average age at first marriage has increased from the mean of 23.3 years.

31. a) Probably a representative sample; the Nearly Normal Condition seems reasonable. (Show a Normal probability plot or histogram.) The histogram is nearly uniform, with no outliers or skewness.
b) $\bar{y} = 28.78, s = 0.40$ c) (28.36, 29.21) grams
d) Based on this sample, we are 95% confident the average weight of the content of Ruffles bags is between 28.36 and 29.21 grams.
e) The company is erring on the safe side, as it appears that, on average, it is putting in slightly more chips than stated.

33. a) Type I; he mistakenly rejected the null hypothesis that $p = 0.10$ (or worse).
b) Yes. These are a random sample of bags and the Nearly Normal Condition is met (Show a Normal probability plot or histogram.); $t = -2.51$ with 7 df for a one-sided P-value of 0.0203.

35. a) Random sample; the Nearly Normal Condition seems reasonable from a Normal probability plot. The histogram is roughly unimodal and symmetric with no outliers. (Show plot.)
b) (1187.9, 1288.4) chips
c) Based on this sample, the mean number of chips in an 18-ounce bag is between 1187.9 and 1288.4, with 95% confidence. The *mean* number of chips is clearly greater than 1000. However, if the claim is about individual bags, then it's not necessarily true. If the mean is 1188 and the SD deviation is near 94, then 2.5% of the bags will have fewer than 1000 chips, using the Normal model. If in fact the mean is 1288, the proportion below 1000 will be less than 0.1%, but the claim is still false.

37. a) The Normal probability plot is relatively straight, with one outlier at 93.8 sec. Without the outlier, the conditions seem to be met. The histogram is roughly unimodal and symmetric with no other outliers. (Show your plot.)
b) $t = -2.63$, P-value $= 0.0160$. With the outlier included, we might conclude that the mean completion time for the maze is not 60 seconds; in fact, it is less.
c) $t = -4.46$, P-value $= 0.0003$. Because the P-value is so small, we reject H_0. Without the outlier, we see strong evidence that the average completion time for the maze is less than 60 seconds. The outlier here did not change the conclusion.
d) The maze does not meet the "one-minute average" requirement. Both tests rejected a null hypothesis of a mean of 60 seconds.

39. a) $287.3 < \mu(\text{Drive Distance}) < 289.9$
b) These data are not a random sample of golfers. The top professionals are (unfortunately) not representative and were not selected at random. We might consider the 2006 data to represent the population of all professional golfers, past, present, and future.
c) The data are means for each golfer, so they are less variable than if we looked at all the separate drives.

CHAPTER 24

1. Yes. The high P-value means that we lack evidence of a difference, so 0 is a possible value for $\mu_{Meat} - \mu_{Beef}$.

3. a) Plausible values of $\mu_{Meat} - \mu_{Beef}$ are all negative, so the mean fat content is probably higher for beef hot dogs.
b) The difference is significant. c) 10%

5. a) False. The confidence interval is about means, not about individual hot dogs.
b) False. The confidence interval is about means, not about individual hot dogs.

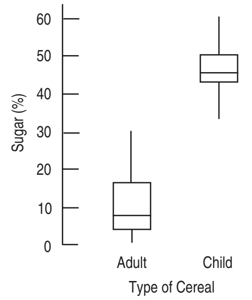
c) True.
d) False. CI's based on other samples will also try to estimate the true difference in population means; there's no reason to expect other samples to conform to this result.
e) True.

7. a) 2.927 b) Larger
c) Based on this sample, we are 95% confident that students who learn Math using the CPMP method will score, on average, between 5.57 and 11.43 points better on a test solving applied Algebra problems with a calculator than students who learn by traditional methods.
d) Yes; 0 is not in the interval.

9. a) $H_0: \mu_C - \mu_T = 0$ vs. $H_A: \mu_C - \mu_T \neq 0$
b) Yes. Groups are independent, though we don't know if students were randomly assigned to the programs. Sample sizes are large, so CLT applies.
c) If the means for the two programs are really equal, there is less than a 1 in 10,000 chance of seeing a difference as large as or larger than the observed difference just from natural sampling variation.
d) On average, students who learn with the CPMP method do significantly worse on Algebra tests that do not allow them to use calculators than students who learn by traditional methods.

11. a) (1.36, 4.64)
b) No; 5 minutes is beyond the high end of the interval.

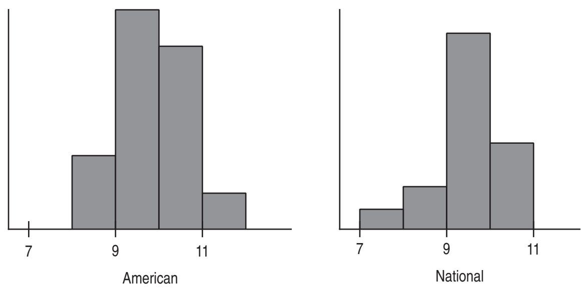
13.



Random sample—questionable, but probably representative, independent samples, less than 10% of all cereals; boxplot shows no outliers—not exactly symmetric, but these are reasonable sample sizes. Based on these samples, with 95% confidence, children's cereals average between 32.49% and 40.80% more sugar content than adult's cereals.

15. $H_0: \mu_N - \mu_C = 0$ vs. $H_A: \mu_N - \mu_C > 0$; $t = 2.207$; P-value $= 0.0168$; df $= 33.4$. Because of the small P-value, we reject H_0. These data do suggest that new activities are better. The mean reading comprehension score for the group with new activities is significantly (at $\alpha = 0.05$) higher than the mean score for the control group.

17. a)



Both are unimodal and reasonably symmetric.
b) Based on these data, the average number of runs in an American League stadium is between 9.36 and 10.23, with 95% confidence.
c) No. The boxplot indicates it isn't an outlier.
d) We want to work directly with the average difference. The two separate confidence intervals do not answer questions about the difference. The difference has a different standard deviation, found by adding variances.

19. a) $(-0.18, 0.89)$
b) Based on these data, with 95% confidence, American League stadiums average between 0.18 fewer runs and 0.89 more runs per game than National League stadiums.
c) No; 0 is in the interval.

21. These are not two independent samples. These are before and after scores for the same individuals.

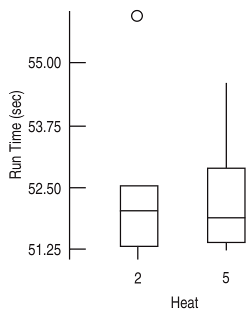
23. a) These data do not provide evidence of a difference in ad recall between shows with sexual content and violent content.
b) $H_0: \mu_S - \mu_N = 0$ vs. $H_A: \mu_S - \mu_N \neq 0$. $t = -6.08$, df $= 213.99$, P-value $= 5.5 \times 10^{-9}$. Because the P-value is low, we reject H_0. These data suggest that ad recall between shows with sexual and neutral content is different; those who saw shows with neutral content had higher average recall.

25. a) $H_0: \mu_V - \mu_N = 0$ vs. $H_A: \mu_V - \mu_N \neq 0$. $t = -7.21$, df $= 201.96$, P-value $= 1.1 \times 10^{-11}$. Because of the very small P-value, we reject H_0. There is a significant difference in mean ad recall between shows with violent content and neutral content; viewers of shows with neutral content remember more brand names, on average.
b) With 95% confidence, the average number of brand names remembered 24 hours later is between 1.45 and 2.41 higher for viewers of neutral content shows than for viewers of sexual content shows, based on these data.

27. $H_0: \mu_{big} - \mu_{small} = 0$ vs. $H_A: \mu_{big} - \mu_{small} \neq 0$; bowl size was assigned randomly; amount scooped by individuals and by the two groups should be independent. With 34.3 df, $t = 2.104$ and P-value $= 0.0428$. The low P-value leads us to reject the null hypothesis. There is evidence of a difference in the average amount of ice cream that people scoop when given a bigger bowl.

29. a) The 95% confidence interval for the difference is (0.61, 5.39). 0 is not in the interval, so scores in 1996 were significantly higher. (Or the t, with more than 7500 df, is 2.459 for a P-value of 0.0070.)
b) Since both samples were very large, there shouldn't be a difference in how certain you are, assuming conditions are met.

31. Independent Groups Assumption: The runners are different women, so the groups are independent. The Randomization Condition is satisfied since the runners are selected at random for these heats.



Nearly Normal Condition: The boxplots show an outlier, but we will proceed and then redo the analysis with the outlier deleted. When we include the outlier, $t = 0.035$ with a two-sided P-value of 0.97. With the outlier deleted, $t = -1.14$, with $P = 0.2837$. Either P-value is so large that we fail to reject the null hypothesis of equal means and conclude that there is no evidence of a difference in the mean times for runners in unseeded heats.

33. With $t = -4.57$ and a very low P-value of 0.0013, we reject the null hypothesis of equal mean velocities. There is strong evidence that golf balls hit off Stinger tees will have a higher mean initial velocity.

35. a) We can be 95% confident that the interval 74.8 ± 178.05 minutes includes the true difference in mean crossing times between men and women. Because the interval includes zero, we cannot be confident that there is any difference at all.

b) Independence Assumption: There is no reason to believe that the swims are not independent or that the two groups are not independent of each other.

Randomization Condition: The swimmers are not a random sample from any identifiable population, but they may be representative of swimmers who tackle challenges such as this.

Nearly Normal Condition: the boxplots show no outliers. The histograms are unimodal; the histogram for men is somewhat skewed to the right. (Show your graphs.)

37. a) H_0: $\mu_R - \mu_N = 0$ vs. H_A: $\mu_R - \mu_N < 0$. $t = -1.36$, df $= 20.00$, P-value $= 0.0945$. Because the P-value is large, we fail to reject H_0. These data show no evidence of a difference in mean number of objects recalled between listening to rap or no music at all.

b) Didn't conclude any difference.

CHAPTER 25

1. a) Randomly assign 50 hens to each of the two kinds of feed. Compare production at the end of the month.

b) Give all 100 hens the new feed for 2 weeks and the old food for 2 weeks, randomly selecting which feed the hens get first. Analyze the differences in production for all 100 hens.

c) Matched pairs. Because hens vary in egg production, the matched-pairs design will control for that.

3. a) Show the same people ads with and without sexual images, and record how many products they remember in each group. Randomly decide which ads a person sees first. Examine the differences for each person.

b) Randomly divide volunteers into two groups. Show one group ads with sexual images and the other group ads without. Compare how many products each group remembers.

5. a) Matched pairs—same cities in different periods.

b) There is a significant difference (P-value $= 0.0244$) in the labor force participation rate for women in these cities; women's participation seems to have increased between 1968 and 1972.

7. a) Use the paired *t*-test because we have pairs of Fridays in 5 different months. Data from adjacent Fridays within a month may be more similar than data from randomly chosen Fridays.

b) We conclude that there is evidence (P-value 0.0212) that the mean number of cars found on the M25 motorway on Friday the 13th is less than on the previous Friday.

c) We don't know if these Friday pairs were selected at random. If these are the Fridays with the largest differences, this will affect our conclusion. The Nearly Normal Condition appears to be met by the differences, but the sample size is small.

9. Adding variances requires that the variables be independent. These price quotes are for the same cars, so they are paired. Drivers quoted high insurance premiums by the local company will be likely to get a high rate from the online company, too.

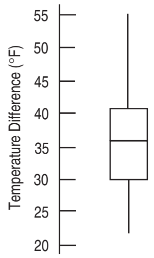
11. a) The histogram—we care about differences in price.

b) Insurance cost is based on risk, so drivers are likely to see similar quotes from each company, making the differences relatively smaller.

c) The price quotes are paired; they were for a random sample of fewer than 10% of the agent's customers; the histogram of differences looks approximately Normal.

13. H_0: $\mu(Local - Online) = 0$ vs. H_A: $\mu(Local - Online) > 0$; with 9 df, $t = 0.83$. With a high P-value of 0.215, we don't reject the null hypothesis. These data don't provide evidence that online premiums are lower, on average.

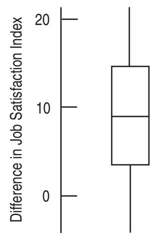
15.



Data are paired for each city; cities are independent of each other; boxplot shows the temperature differences are reasonably symmetric, with no outliers. This is probably not a random sample, so we might be wary of inferring that this difference applies to all European cities. Based on these data, we are 90% confident that the average temperature in European cities in July is between 32.3°F and 41.3°F higher than in January.

17. Based on these data, we are 90% confident that boys, on average, can do between 1.6 and 13.0 more push-ups than girls (independent samples—not paired).

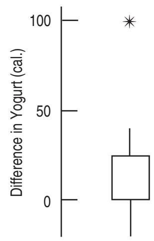
19. a) Paired sample test. Data are before/after for the same workers; workers randomly selected; assume fewer than 10% of all this company's workers; boxplot of differences shows them to be symmetric, with no outliers.



b) H_0: $\mu_D = 0$ vs. H_A: $\mu_D > 0$. $t = 3.60$, P-value $= 0.0029$. Because $P < 0.01$, reject H_0. These data show evidence that average job satisfaction has increased after implementation of the program.

c) Type I

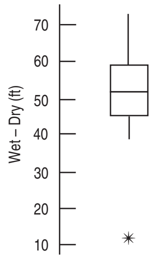
21. H_0: $\mu_D = 0$ vs. H_A: $\mu_D \neq 0$. Data are paired by brand; brands are independent of each other; fewer than 10% of all yogurts (questionable); boxplot of differences shows an outlier (100) for Great Value:



With the outlier included, the mean difference (Strawberry – Vanilla) is 12.5 calories with a *t*-stat of 1.332, with 11 df, for a P-value of 0.2098. Deleting the outlier, the difference is even smaller, 4.55 calories with a *t*-stat of only 0.833 and a P-value of 0.4241. With P-values so large, we do not reject H_0. We conclude that the data do not provide evidence of a difference in mean calories.

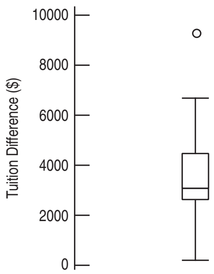
23. a) Cars were probably not a simple random sample, but may be representative in terms of stopping distance; boxplot does not show outliers, but does indicate right skewness. A 95% confidence interval for the mean stopping distance on dry pavement is (131.8, 145.6) feet.

b) Data are paired by car; cars were probably not randomly chosen, but representative; boxplot shows an outlier (car 4) with a difference of 12. With deletion of that car, a Normal probability plot of the differences is relatively straight.



Retaining the outlier, we estimate with 95% confidence that the average braking distance is between 38.8 and 62.6 feet more on wet pavement than on dry, based on this sample. (Without the outlier, the confidence interval is 47.2 to 62.8 feet.)

25. a) Paired Data Assumption: Data are paired by college. Randomization Condition: This was a random sample of public colleges and universities. 10% Condition: these are fewer than 10% of all public colleges and universities.



Normal Population Assumption: U.C. Irvine seems to be an outlier; we might consider removing it.

b) Having deleted the observation for U.C.-Irvine, whose difference of \$9300 was an outlier, we are 90% confident, based on the remaining data, that nonresidents pay, on average, between \$2615.31 and \$3918.02 more than residents. If we retain the outlier, the interval is (\$2759, \$4409).

c) Assertion is reasonable; with or without the outlier, \$3500 is in the confidence interval.

27. a) 60% is 30 strikes; H_0: $\mu = 30$ vs. H_A: $\mu > 30$. $t = 6.07$, P-value $= 3.92 \times 10^{-6}$. With a very small P-value, we reject H_0. There is very strong evidence that players can throw more than 60% strikes after training, based on this sample.

b) H_0: $\mu_D = 0$ vs. H_A: $\mu_D > 0$. $t = 0.135$, P-value $= 0.4472$. With such a high P-value, we do not reject H_0. These data provide no evidence that the program has improved pitching in these Little League players.

PART VI REVIEW

1. a) H_0: $\mu_{Jan} - \mu_{Jul} = 0$; H_A: $\mu_{Jan} - \mu_{Jul} \neq 0$. $t = -1.94$, df $= 43.68$, P-value $= 0.0590$. Since $P < 0.10$, reject the null. These data show a significant difference in mean age to crawl between January and July babies.

b) H_0: $\mu_{Apr} - \mu_{Oct} = 0$; H_A: $\mu_{Apr} - \mu_{Oct} \neq 0$. $t = -0.92$; df $= 59.40$; P-value $= 0.3610$. Since $P > 0.10$, do not reject the null; these data do not show a significant difference between April and October with regard to the mean age at which crawling begins.

c) These results are not consistent with the claim.

3. H_0: $p = 0.26$; H_A: $p \neq 0.26$. $z = 0.946$; P-value $= 0.3443$. Because the P-value is high, we do not reject H_0. These data do not show that the Denver-area rate is different from the national rate in the proportion of businesses with women owners.

5. Based on these data, we are 95% confident that the mean difference in aluminum oxide content is between −3.37 and 1.65. Since the interval contains 0, the means in aluminum oxide content of the pottery made at the two sites could reasonably be the same.

7. a) H_0: $p_{ALS} - p_{Other} = 0$; H_A: $p_{ALS} - p_{Other} > 0$. $z = 2.52$; P-value $= 0.0058$. With such a low P-value, we reject H_0. This is strong evidence that there is a higher proportion of varsity athletes among ALS patients than those with other disorders.

b) Observational retrospective study. To make the inference, one must assume the patients studied are representative.

9. H_0: $\mu = 7.41$; H_A: $\mu \neq 7.41$. $t = 2.18$; df $= 111$; P-value $= 0.0313$. With such a low P-value, we reject H_0. Assuming that Missouri babies fairly represent the United States, these data suggest that American babies are different from Australian babies in birth weight; it appears American babies are heavier, on average.

11. a) If there is no difference in the average fish sizes, the chance of seeing an observed difference this large just by natural sampling variation is less than 0.1%.

b) If cost justified, feed them a natural diet. c) Type I

13. a) Assuming the conditions are met, from these data we are 95% confident that patients with cardiac disease average between 3.39 and 5.01 years older than those without cardiac disease.

b) Older patients are at greater risk from a variety of other health issues, and perhaps more depressed.

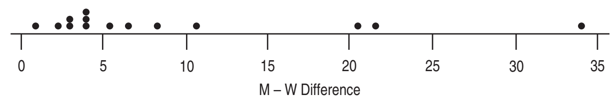
15. a) Stratified sample survey.

b) We are 95% confident that the proportion of boys who play computer games is between 7.0 and 17.0 percentage points higher than among girls.

c) Yes. The entire interval lies above 0.

17. Based on the data, we are 95% confident that the mean difference in words misunderstood is between −3.76 and 3.10. Because 0 is in the confidence interval, we would conclude that the two tapes could be equivalent.

19. a)



The countries that appear to be outliers are Spain, Italy, and Portugal. They are all Mediterranean countries.

b) H_0: $\mu_D = 0$; H_A: $\mu_D > 0$.
$t = 5.56$; df $= 10$; P-value $= 0.0001$. With such a low P-value, we reject H_0. These data show that European men are more likely than women to read newspapers.

21. We are 95% confident that the proportion of American adults who would agree with the statement is between 57.0% and 63.0%.

23. Data are matched pairs (before and after for the same rooms); less than 10% of all rooms in a large hotel; uncertain how these rooms were selected (are they representative?). Histogram shows that differences are roughly unimodal and symmetric, with no outliers. A 95% confidence interval for the difference, before – after, is (0.58, 2.65) counts. Since the entire interval is above 0, these data suggest that the new air-conditioning system was effective in reducing average bacteria counts.

25. a) We are 95% confident that between 19.77% and 38.66% of children with bipolar symptoms will be helped with medication and psychotherapy, based on this study.

b) 221 children

27. a) From this histogram, about 115 loaves or more. (Not Normal.) This assumes the last 100 days are typical.

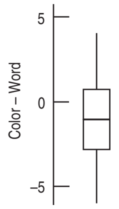
b) Large sample size; CLT says $\bar{y}$ will be approximately Normal.

c) From the data, we are 95% confident that on average the bakery will sell between 101.2 and 104.8 loaves of bread a day.
d) 25
e) Yes, 100 loaves per day is too low—the entire confidence interval is above that.

29. a) H_0: $p_{High} - p_{Low} = 0$; H_A: $p_{High} - p_{Low} \neq 0$. $z = -3.57$; P-value $= 0.0004$. Because the P-value is so low, we reject H_0. These data suggest the IRS risk is different in the two groups; it appears people who consume dairy products often have a lower risk, on average.
b) Doesn't indicate causality; this is not an experiment.

31. Based on these data, we are 95% confident that seeded clouds will produce an average of between -4.76 and 559.56 more acre-feet of rain than unseeded clouds. Since the interval contains negative values, it may be that seeding is unproductive.

33. a) Randomizing order of the tasks helps avoid bias and memory effects. Randomizing the cards helps avoid bias as well.
b) H_0: $\mu_D = 0$; H_A: $\mu_D \neq 0$



Boxplot of the differences looks symmetric with no outliers.

$t = -1.70$; P-value $= 0.0999$; do not reject H_0, because $P > 0.05$. The data do not provide evidence that the color or written word dominates.

35. a) Different samples give different means; this is a fairly small sample. The difference may be due to natural sampling variation.
b) H_0: $\mu = 100$; H_A: $\mu < 100$
c) Batteries selected are a SRS (representative); fewer than 10% of the company's batteries; lifetimes are approximately Normal.
d) $t = -1.0$; P-value $= 0.1666$; do not reject H_0. This sample does not show that the average life of the batteries is significantly less than 100 hours.
e) Type II.

CHAPTER 26

1. a) Chi-square test of independence. We have one sample and two variables. We want to see if the variable *Account Type* is independent of the variable *Trade Type*.
b) Other test. *Account Size* is quantitative, not counts.
c) Chi-square test of homogeneity. We want to see if the distribution of one variable, *Courses*, is the same for two groups (resident and nonresident students).

3. a) 10 b) Goodness-of-fit
c) H_0: The die is fair (all faces have $p = 1/6$).
H_A: The die is not fair.
d) Count data; rolls are random and independent; expected frequencies are all bigger than 5.
e) 5 f) $\chi^2 = 5.600$, P-value $= 0.3471$
g) Because the P-value is high, do not reject H_0. The data show no evidence that the die is unfair.

5. a) Weights are quantitative, not counts.
b) Count the number of each kind of nut, assuming the company's percentages are based on counts rather than weights.

7. H_0: The police force represents the population (29.2% white, 28.2% black, etc.). H_A: The police force is not representative of the population. $\chi^2 = 16516.88$, df $= 4$, P-value $= 0.0000$. Because the P-value is so low, we reject H_0. These data show that the police force is not representative of the population. In particular, there are too many white officers in relationship to their membership in the community.

9. a) $\chi^2 = 5.671$, df $= 3$, P-value $= 0.1288$. With a P-value this high, we fail to reject H_0. Yes, these data are consistent with those predicted by genetic theory.
b) $\chi^2 = 11.342$, df $= 3$, P-value $= 0.0100$. Because of the low P-value, we reject H_0. These data provide evidence that the distribution is not as specified by genetic theory.
c) With small samples, many more data sets will be consistent with the null hypothesis. With larger samples, small discrepancies will show evidence against the null hypothesis.

11. a) $96/16 = 6$ b) Goodness of Fit
c) H_0: The number of large hurricanes remains constant over decades.
H_A: The number of large hurricanes has changed.
d) 15 e) P-value $= 0.63$
f) The very high P-value means these data offer no evidence that the numbers of large hurricanes has changed.
g) The final period is only 6 years rather than 10 and already 7 large hurricanes have been observed. Perhaps this decade will have an unusually large number of such hurricanes.

13. a) Independence
b) H_0: Breastfeeding success is independent of having an epidural.
H_A: There's an association between breastfeeding success and having an epidural.

15. a) 1 b) 159.34
c) Breastfeeding behavior should be independent for these babies. They are fewer than 10% of all babies; we assume they are representative. We have counts, and all the expected counts are at least 5.

17. a) 5.90 b) P-value < 0.005
c) The P-value is very low, so reject the null. There's evidence of an association between having an epidural and subsequent success in breastfeeding.

19. a) $\dfrac{(190 - 159.34)}{\sqrt{159.34}} = 2.43$
b) It appears that babies whose mothers had epidurals during childbirth are much less likely to be breastfeeding 6 months later.

21. These factors would not be mutually exclusive. There would be yes or no responses for every baby for each.

23. a) 40.2% b) 8.1% c) 62.2% d) 285.48
e) H_0: Survival was independent of status on the ship.
H_A: Survival depended on the status.
f) 3
g) We reject the null hypothesis. Survival depended on status. We can see that first-class passengers were more likely to survive than passengers of any other class.

25. First class passengers were most likely to survive, while 3rd-class passengers and crew were under-represented among the survivors.

27. a) Experiment—actively imposed treatments (different drinks)
b) Homogeneity
c) H_0: The rate of urinary tract infection is the same for all three groups. H_A: The rate of urinary tract infection is different among the groups.
d) Count data; random assignment to treatments; all expected frequencies larger than 5.
e) 2 f) $\chi^2 = 7.776$, P-value $= 0.020$.
g) With a P-value this low, we reject H_0. These data provide reasonably strong evidence that there is a difference in urinary tract infection rates between cranberry juice drinkers, lactobacillus drinkers, and the control group.
h) The standardized residuals are

	Cranberry	Lactobacillus	Control
Infection	−1.87276	1.19176	0.68100
No Infection	1.24550	−0.79259	−0.45291

From the standardized residuals (and the sign of the residuals), it appears those who drank cranberry juice were less likely to develop urinary tract infections; those who drank lactobacillus were more likely to have infections.

29. a) Independence
b) H_0: *Political Affiliation* is independent of *Sex*.
H_A: There is a relationship between *Political Affiliation* and *Sex*.
c) Counted data; probably a random sample, but can't extend results to other states; all expected frequencies greater than 5.
d) $\chi^2 = 4.851$, df $= 2$, P-value $= 0.0884$.
e) Because of the high P-value, we do not reject H_0. These data do not provide evidence of a relationship between *Political Affiliation* and *Sex*.

31. H_0: *Political Affiliation* is independent of *Region*. H_A: There is a relationship between *Political Affiliation* and *Region*. $\chi^2 = 13.849$, df $= 4$, P-value $= 0.0078$. With a P-value this low, we reject H_0. *Political Affiliation* and *Region* are related. Examination of the residuals shows that those in the West are more likely to be Democrat than Republican; those in the Northeast are more likely to be Republican than Democrat.

33. a) Homogeneity
b) H_0: The grade distribution is the same for both professors.
H_A: The grade distributions are different.
c)

	Dr. Alpha	Dr. Beta
A	6.667	5.333
B	12.778	10.222
C	12.222	9.778
D	6.111	4.889
F	2.222	1.778

Three cells have expected frequencies less than 5.

35. a)

	Dr. Alpha	Dr. Beta
A	6.667	5.333
B	12.778	10.222
C	12.222	9.778
Below C	8.333	6.667

All expected frequencies are now larger than 5.
b) Decreased from 4 to 3.
c) $\chi^2 = 9.306$, P-value $= 0.0255$. Because the P-value is so low, we reject H_0. The grade distributions for the two professors are different. Dr. Alpha gives fewer A's and more grades below C than Dr. Beta.

37. $\chi^2 = 14.058$, df $= 1$, P-value $= 0.0002$. With a P-value this low, we reject H_0. There is evidence of racial steering. Blacks are much less likely to rent in Section A than Section B.

39. a) $z = 3.74936$, $z^2 = 14.058$.
b) P-value $(z) = 0.0002$ (same as in Exercise 25).

41. $\chi^2 = 5.89$, df $= 3$, P $= 0.117$. Because the P-value is >0.05, these data show no evidence of an association between the mother's age group and the outcome of the pregnancy.

CHAPTER 27

1. a) $\widehat{Error} = 453.22 - 8.37\ YearSince1970$; according to the model, the error made in predicting a hurricane's path was about 453 nautical miles, on average, in 1970. It has been declining at a rate of about 8.37 nautical miles per year.
b) H_0: $\beta_1 = 0$; there has been no change in prediction accuracy.
H_A: $\beta_1 \neq 0$; there has been a change in prediction accuracy.
c) With a P-value < 0.001, I reject the null hypothesis and conclude that prediction accuracies have in fact been changing during this period.
d) 58.5% of the variation in hurricane prediction accuracy is accounted for by this linear model on time.

3. a) $\widehat{Budget} = -31.387 + 0.714\ RunTime$. The model suggests that movies cost about $714,000 per minute to make.
b) A negative starting value makes no sense, but the P-value of 0.07 indicates that we can't discern a difference between our estimated value and zero. The statement that a movie of zero length should cost $0 makes sense.
c) Amounts by which movie costs differ from predictions made by this model vary, with a standard deviation of about $33 million.
d) 0.154 $m/min
e) If we constructed other models based on different samples of movies, we'd expect the slopes of the regression lines to vary, with a standard deviation of about $154,000 per minute.

5. a) The scatterplot looks straight enough, the residuals look random and nearly normal, and the residuals don't display any clear change in variability.
b) I'm 95% confident that the cost of making longer movies increases at a rate of between 0.41 and 1.02 million dollars per additional minute.

7. a) H_0: $\beta_1 = 0$; there's no association between calories and sodium content in all-beef hot dogs. H_A: $\beta_1 \neq 0$: there is an association.
b) Based on the low P-value (0.0018), I reject the null. There is evidence of an association between the number of calories in all-beef hot dogs and their sodium contents.

9. a) Among all-beef hot dogs with the same number of calories, the sodium content varies, with a standard deviation of about 60 mg.
b) 0.561 mg/cal
c) If we tested many other samples of all-beef hot dogs, the slopes of the resulting regression lines would be expected to vary, with a standard deviation of about 0.56 mg of sodium per calorie.

11. I'm 95% confident that for every additional calorie, all-beef hot dogs have, on average, between 1.07 and 3.53 mg more sodium.

13. a) H_0: Difference in age at first marriage has not been changing, $\beta_1 = 0$. H_A: Difference in age at first marriage has been changing, $\beta_1 \neq 0$.
b) Residual plot shows no obvious pattern; histogram is not particularly Normal, but shows no obvious skewness or outliers.
c) $t = -7.04$, P-value <0.0001. With such a low P-value, we reject H_0. These data show evidence that difference in age at first marriage is decreasing.

15. Based on these data, we are 95% confident that the average difference in age at first marriage is decreasing at a rate between 0.039 and 0.021 years per year.

17. a) H_0: *Fuel Economy* and *Weight* are not (linearly) related, $\beta_1 = 0$. H_A: *Fuel Economy* changes with *Weight*, $\beta_1 \neq 0$. P-value < 0.0001, indicating strong evidence of an association.
b) Yes, the conditions seem satisfied. Histogram of residuals is unimodal and symmetric; residual plot looks OK, but some "thickening" of the plot with increasing values.
c) $t = -12.2$, P-value < 0.0001.These data show evidence that *Fuel Economy* decreases with the *Weight* of the car.

19. a) $(-9.57, -6.86)$ mpg per 1000 pounds.
b) Based on these data, we are 95% confident that *Fuel Efficiency* decreases between 6.86 and 9.57 miles per gallon, on average, for each additional 1000 pounds of *Weight*.

21. a) We are 95% confident that 2500-pound cars will average between 27.34 and 29.07 miles per gallon.
b) Based on the regression, a 3450-pound car will get between 15.44 and 25.36 miles per gallon, with 95% confidence.

23. a) Yes. $t = 2.73$, P-value $= 0.0079$. With a P-value so low, we reject H_0. There is a positive relationship between *Calories* and *Sodium* content.
b) No. $R^2 = 9\%$ and s appears to be large, although without seeing the data, it is a bit hard to tell.

25. Plot of *Calories* against *Fiber* does not look linear; the residuals plot also shows increasing variance as predicted values get large. The histogram of residuals is right skewed.

27. a) H_0: No (linear) relationship between *BCI* and *pH*, $\beta_1 = 0$.
H_A: There seems to be a relationship, $\beta_1 \neq 0$.
b) $t = -7.73$ with 161 df; P-value < 0.0001
c) There seems to be a negative relationship; *BCI* decreases as *pH* increases at an average of 197.7 *BCI* units per increase of 1 *pH*.

29. a) H_0: No linear relationship between *Population* and *Ozone*, $\beta_1 = 0$. H_A: *Ozone* increases with *Population*, $\beta_1 > 0$. $t = 3.48$, P-value $= 0.0018$. With a P-value so low, we reject H_0. These data show evidence that *Ozone* increases with *Population*.
b) Yes, *Population* accounts for 84% of the variability in *Ozone* level, and s is just over 5 parts per million.

31. a) Based on this regression, each additional million residents corresponds to an increase in average ozone level of between 3.29 and 10.01 ppm, with 90% confidence.
b) The mean *Ozone* level for cities with 600,000 people is between 18.47 and 27.29 ppm, with 90% confidence.

33. a) 33 batteries.
b) Yes. The scatterplot is roughly linear with lots of scatter; plot of residuals vs. predicted values shows no overt patterns; Normal probability plot of residuals is reasonably straight.
c) H_0: No linear relationship between *Cost* and *Cranking Amps*, $\beta_1 = 0$. H_A: *Cranking Amps* increase with cost, $\beta_1 > 0$. $t = 3.23$; P-value $= \frac{1}{2}(0.0029) = 0.00145$. With a P-value so low, we reject H_0. These data provide evidence that more expensive batteries do have more cranking amps.
d) No. $R^2 = 25.2\%$ and $s = 116$ amps. Since the range of amperage is only about 400 amps, an s of 116 is not very useful.
e) $\widehat{Cranking\ amps} = 384.59 + 4.15 \times Cost$.
f) (1.97, 6.32) cold cranking amps per dollar.
g) *Cranking amps* increase, on average, between 1.97 and 6.32 per dollar of battery *Cost* increase, with 90% confidence.

35. a) H_0: No linear relationship between *Waist* size and *%Body Fat*, $\beta_1 = 0$. H_A: *%Body Fat* changes with *Waist* size, $\beta_1 \neq 0$. $t = 8.14$; P-value < 0.0001. There's evidence that *%Body Fat* seems to increase with *Waist* size.
b) With 95% confidence, mean *%Body Fat* for people with 40-inch waists is between 23.58 and 29.02, based on this regression.

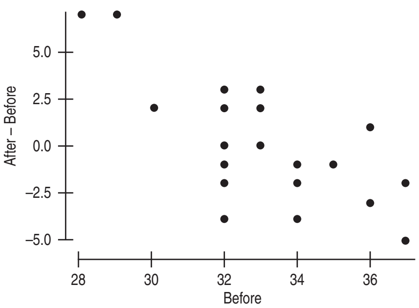
37. a) The regression model is $\widehat{Midterm2} = 12.005 + 0.721\ Midterm1$

	Estimate	Std Error	t-ratio	P-value
Intercept	12.00543	15.9553	0.752442	0.454633
Slope	0.72099	0.183716	3.924477	0.000221

RSquare	0.198982
s	16.78107
n	64

b) The scatterplot shows a weak, somewhat linear, positive relationship. There are several outlying points, but removing them only makes the relationship slightly stronger. There is no obvious pattern in the residual plot. The regression model appears appropriate. The small P-value for the slope shows that the slope is statistically distinguishable from 0 even though the R^2 value of 0.199 suggests that the overall relationship is weak.
c) No. The R^2 value is only 0.199 and the value of s of 16.8 points indicates that she would not be able to predict performance on *Midterm2* very accurately.

39. H_0: Slope of *Effectiveness* vs *Initial Ability* $= 0$; H_A: Slope $\neq 0$



Scatterplot is straight enough. Regression conditions appear to be met. $t = -4.34$, df $= 19$, P-value $= 0.004$. With a P-value this small, we reject the null hypothesis. There is strong evidence that the effectiveness of the video depends on the player's initial ability. The negative slope observed that the method is more effective for those whose initial performance was poorest and less so for those whose initial performance was better. This looks like a case of regression to the mean. Those who were above average initially tended to be worse after training. Those who were below average initially tended to improve.

41. a) Data plot looks linear; no overt pattern in residuals; histogram of residuals roughly symmetric and unimodal.
b) H_0: No linear relationship between *Education* and *Mortality*, $\beta_1 = 0$. H_A: $\beta_1 \neq 0$. $t = -6.24$; P-value < 0.001. There is evidence that cities in which the mean education level is higher also tend to have a lower mortality rate.
c) No. Data are on cities, not individuals. Also, these are observational data. We cannot predict causal consequences from them.
d) $(-65.95, -33.89)$ deaths per 100,000 people.
e) *Mortality* decreases, on average, between 33.89 and 65.95 deaths per 100,000 for each extra year of average *Education*.
f) Based on the regression, the average *Mortality* for cities with an average of 12 years of *Education* will be between 874.239 and 914.196 deaths per 100,000 people.

PART VII REVIEW

1. H_0: The proportions are as specified by the ratio 1:3:3:9; H_A: The proportions are not as stated. $\chi^2 = 5.01$; df $= 3$; P-value $= 0.1711$. Since $P > 0.05$, we fail to reject H_0. These data do not provide evidence to indicate that the proportions are other than 1:3:3:9.

3. a) H_0: *Mortality* and *calcium concentration* in water are not linearly related, $\beta_1 = 0$; H_A: They are linearly related, $\beta_1 \neq 0$.
b) $t = -6.73$; P-value < 0.0001. There is a significant negative relationship between calcium in drinking water and mortality.
c) $(-4.19, -2.27)$ deaths per 100,000 for each ppm calcium.
d) Based on the regression, we are 95% confident that mortality (deaths per 100,000) decreases, on average, between 2.27 and 4.19 for each part per million of calcium in drinking water.

5. 404 checks

7. H_0: *Income* and *Party* are independent. H_A: *Income* and *Party* are not independent. $\chi^2 = 17.19$; P-value $= 0.0018$. With such a small P-value, we reject H_0. These data show evidence that income level and party are not independent. Examination of components suggests Democrats are most likely to have low incomes; Independents are most likely to have middle incomes, and Republicans are most likely to have high incomes.

9. H_0: $p_L - p_R = 0$; H_A: $p_L - p_R \neq 0$. $z = 1.38$; P-value $= 0.1683$. Since $P > 0.05$, we do not reject H_0. These data do not provide evidence of a difference in musical abilities between right- and left-handed people.

11. a) H_0: $\mu_D = 0$; H_A: $\mu_D \neq 0$.
Boxplot of the differences indicates a strong outlier (1958). With the outlier kept in, the t-stat is 0, with a P-value of 1.00 (two sided). There is no evidence of a difference (on averag of actual and that predicted by Gallup. With the outlier taken out, the t-stat is still only -0.8525 with a P-value of 0.4106, so the conclusion is the same.
b) H_0: There is no (linear) relationship between predicted and actual number of Democratic seats won ($\beta_1 = 0$). H_A: There is a relationship ($\beta_1 \neq 0$). The relationship is very strong, with an R^2 of 97.7%. The t-stat is 22.56. Even with only 12 df, this is clearly significant (P-value < 0.0001). There is an outlying residual (1958), but without it, the regression is even stronger.

13. Conditions are met; df $= 4$; $\chi^2 = 0.69$; P-value $= 0.9526$. Since $P > 0.05$, we do not reject H_0. We do not have evidence that the way the hospital deals with twin pregnancies has changed.

15. a) Based on these data, the average annual rainfall in LA is between 11.65 and 17.39 inches, with 90% confidence.
b) About 46 years
c) No. The regression equation is $\widehat{Rain} = -51.684 + 0.033 \times Year$. $R^2 = 0.1\%$. For the slope, $t = 0.12$ with P-value $= 0.9029$.

17. a) Linear regression is meaningless—the data are categorical.
b) This is a two-way table that is appropriate. H_0: *Eye* and *Hair* color are independent. H_A: *Eye* and *Hair* color are not independent. However, four cells have expected counts less than 5, so the χ^2 analysis is not valid unless cells are merged. However, with a χ^2 value of 223.6 with 16 df and a P-value < 0.0001, the results are not likely to change if we merge appropriate eye colors.

19. a) H_0: $p_Y - p_O = 0$; H_A: $p_Y - p_O \neq 0$. $z = 3.56$; P-value $= 0.0004$. With such a small P-value, we reject H_0. We conclude there is evidence of a difference in effectiveness; it appears the methods are not as good for older women.
b) $\chi^2 = 12.70$; P-value $= 0.0004$. Same conclusion.
c) The P-values are the same; $z^2 = (3.563944)^2 = 12.70 = \chi^2$.

21. a) Positive direction, generally linear trend; moderate scatter.
b) H_0: There is no linear relationship between *Interval* and *Duration*. $\beta_1 = 0$. H_A: There is a linear relationship, $\beta_1 \neq 0$.
c) Yes; histogram is unimodal and roughly symmetric; residuals plot shows random scatter.
d) $t = 27.1$; P-value ≤ 0.001. With such a small P-value, we reject H_0. There is evidence of a positive linear relationship between duration and time to next eruption of Old Faithful.
e) The average time to next eruption after a 2-minute eruption is between 53.24 and 56.12 minutes, with 95% confidence.
f) Based on this regression, we will have to wait between 63.23 and 87.57 minutes after a 4-minute eruption, with 95% confidence.

23. a) $t = 1.42$, df $= 459.3$, P-value $= 0.1574$. Since $P > 0.05$, we do not reject H_0. There's no evidence the two groups differed in ability at the start of the study.
b) $t = 15.11$; P-value < 0.0001. The group taught using the accelerated Math program showed a significant improvement.
c) $t = 9.24$; P-value < 0.0001. The control group showed a significant improvement in test scores.
d) $t = 5.78$; P-value < 0.0001. The Accelerated Math group had significantly higher gains than the control group.

25. a) The regression—he wanted to know about association.
b) There is a moderate relationship between cottage cheese and ice cream sales; for every million pounds of cottage cheese, 1.19 million pounds of ice cream are sold, on average.
c) Testing if the mean difference is 0 (matched *t*-test). Regression won't answer this question.
d) The company sells more cottage cheese than ice cream, on average.
e) part (a)—linear relationship; residuals have a Normal distribution; residuals are independent with equal variation about the line. (c)—Observations are independent; differences are approximately Normal; less than 10% of all possible months' data.
f) About 71.32 million pounds. g) (0.09, 2.29)
h) From this regression, every million pounds of cottage cheese sold is associated with an increase in ice cream sales of between 0.09 and 2.29 million pounds.

27. Based on these data, the average weight loss for the clinic is between 8.24 and 10.06 pounds, with 95% confidence. The clinic's claim is plausible.

29. $\chi^2 = 8.23$; P-value $= 0.0414$. There is evidence of an association between *cracker type* and *bloating*. Standardized residuals for the gum cracker are -1.32 and 1.58. Prospects for marketing this cracker are not good.