

# Answer Key

## Chapter 1

---

### 1.1 Exercises

7. The bricks were the results from the analysis of the data. The clay was the data.
9. Yes, you could reasonably hypothesize that the marbles in each bag are all red. It would be more likely that this is true for the bag with 100 marbles as we already know that 50% of them are red. For the bag with 500 marbles, we know for sure that 10% of them are red.
11. Induction. The repeated observation is that geese have been coming to the pond every May for 10 years. This leads to the general conclusion that the trend will repeat, and they will come in May every year.
13. Deduction. You would use the general principle of the quadratic formula to solve this equation using the axioms or rules of algebra.
15. Pigeons have two legs.
17. You are allergic to bees.
11.
  - a. Students in all colleges and universities in the United States.
  - b. The proportion that binge drink.
  - c. A total of 17,592 students selected from 140 US 4-year colleges.
  - d. The data depends on what the student admits rather than an experimental measurement. The sample covers only students of 4-year colleges within a two week period.
13.
  - a. All employees of the company.
  - b. Average reading comprehension.
  - c. 20 employees.

### Chapter 1 Additional Exercises

1. Not necessarily. It could be that people who wear seat belts are also more cautious drivers.
3.
  - a. All automobile accident claims of the company.
  - b. The number of auto accident claims by region.
  - c. 50 claims.
  - d. Descriptive statistics.

### 1.2 Exercises

9.
  - a. Adult females.
  - b. Average blood pressure.
  - c. 20 females who have high blood pressure.

## Chapter 2

---

### 2.1 Exercises

13. A yottabyte ( $10^{24}$  bytes) is larger than an exabyte ( $10^{18}$  bytes) by a factor of 1,000,000. In other words, a million exabytes equal a yottabyte.
15. The world's data storage capacity grew most rapidly from 2019 to 2020.

### 2.2 Exercises

17.
  - a. interval
  - b. ratio
  - c. ordinal
  - d. nominal

- e. ratio
- f. nominal

19. Ratio, because browser usage share percentages have a meaningful zero and their ratios are meaningful.

21. a. ordinal  
b. ratio  
c. nominal  
d. ratio

## 2.3 Exercises

17. a. By randomly assigning women to two groups and using one of the groups as a “control” group, the experiment should produce data that will reveal the impact of the different diets.  
b. Difference in diet (The first group received 1200 calorie per day diet for the entire period whereas the second group received 420 calorie per day diet for 16 weeks and then were shifted to 1200 calorie per day diet for the rest of the experimental period.)  
c. Weight loss.  
d. Yes, the women receiving the 1200 calorie per day diet represent a control group.  
e. Observational studies are subject to selfselection bias. We would not necessarily know the cause of the weight reduction.
19. a. Phase 1: Gather information about the phenomenon being studied.  
b. Controlled experiment.  
c. Number of major attacks of Multiple Sclerosis.  
d. Bovine myelin.  
e. Fifteen individuals in the early stages of MS fed bovine myelin.  
f. Fifteen individuals in the early stages of MS given a placebo.
21. Jacob’s knee could feel better simply because he took a week break from playing basketball. Answers may vary.

23. There are many factors that affect whether or not someone is happy. Additionally, since both questions require a yes or no reply, there is no way to quantify happiness, or the type and frequency of yoga practiced from a yes or no response. Answers may vary.

## 2.4 Exercises

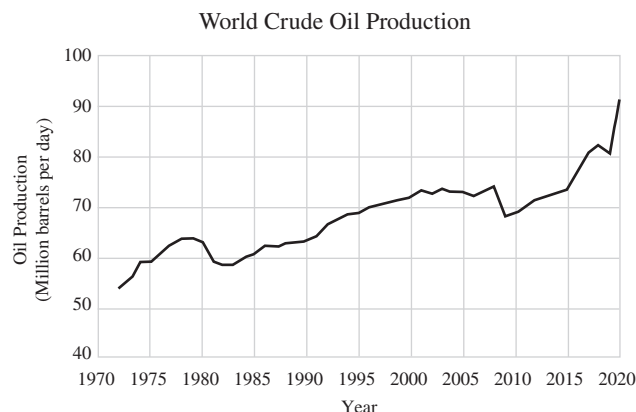
7. a. Cadet height, weight, state from which the cadet was appointed, father’s occupation, parents’ income, type of home residence  
b. State from which the cadet was appointed, father’s occupation, type of home residence are qualitative variables. Cadet height and weight, parents’ income, are quantitative variables.  
c. State from which the cadet was appointed – Nominal;  
Father’s occupation – Nominal;  
Type of home residence –Nominal;  
Cadet Weight – Ratio;  
Parent income – Ratio;  
Cadet Height – Ratio  
d. Answers may vary. For example, health test results such as x-rays, photo ID, recruiter interview notes, etc.
9. a. Discrete  
b. Continuous  
c. Continuous  
d. Continuous  
e. Discrete
11. a. Quantitative  
b. Discrete  
c. Qualitative  
d. Neither

## 2.5 Exercises

9. a. Time series  
b. Nonstationary
11. Stationary; ratio

## Chapter 2 Additional Exercises

1.
  - a. Standardized aptitude test, GPA, etc.
  - b. Number of student suspensions, number of student detentions, etc.
  - c. Survey students to rate teacher preparedness: Always prepared, Sometimes prepared, Never prepared; or measure the number of times teachers are unprepared on a spot check by the principal.
  - d. Number of days missed, number of hours missed (for both teachers and students).
  - e. Survey students, survey teachers: e.g., rate cafeteria food as Excellent, Very Good, Good, Poor, Very Poor.
3.
  - a. Rate of return of certain types of investments, volatility of the investments, etc.
  - b. You could consider cross-sectional data to understand rates of return comparatively across investments and time series data to learn of any trends in rates of return or to evaluate the volatility of the investments.
5. One example: Make a list of standard grocery items for several randomly selected families. Have families buy groceries from one store and then buy exactly the same groceries from another store and then compare prices. Answers may vary.
7. Survey of customer satisfaction: How would you rate the service you received when purchasing the car: Poor, Average, Good? How would you rate the service after you purchased the car: Poor, Average, Good? How would you rate the car's performance: Poor, Average, Good? How satisfied are you with your purchase: Not Satisfied, Satisfied, Very Satisfied? Answers may vary.
9.
  - a. Interval
  - b. Interval
  - c. Ordinal
  - d. Nominal
  - e. Ratio
11.
  - a. Ratio
  - b. Time series, nonstationary, since the graph possesses time-varying behavior, Answers may vary.



13. Answers may vary.
15.
  - a. Nominal.
  - b. Time series.
  - c. If respondents were owners or renters, Answers may vary.
17. Variables you might want to measure are the number of your company's policyholders in Florida with hurricane or flood insurance and the average claim for a sample of covered policyholders in Florida. One approach of collecting the claim data would be to find a reasonable sample of covered policyholders and send an adjuster to determine the average claim for each policyholder in the sample. Answers may vary.
19. Answers may vary. We would certainly want to look at the study with an open mind because it might be insightful. What kind of data is used in the study, observational or from a designed experiment? If the data were created by a designed experiment, then the effects of potential confounding variables could potentially be removed and perhaps causality could be eventually established with further research. However, this was not the case. Anytime anyone claims to have discovered a causal relationship using observational data we need to be very skeptical! In fact, making such a claim instead of just presenting the finding should increase one's skepticism even further. Establishing causality using an observational study is more of a matter of salesmanship than the science.

# Chapter 3

## 3.1 Exercises

13. a. Answers may vary.

Type of Complaint	March	July
Comfort	17	28
Price	11	15
Service	18	14
Schedule	29	33

b. Answers may vary.

Type of Complaint	March	July
Plane	22	34
Personnel	8	3
Building/ Equipment	17	16
Other	28	37

c. No, another person would not necessarily have assigned the various complaints to the same categories. One implication may be that action taken by the airline to rectify these complaints may be focused on the wrong category when another category should have been the focus. Answers may vary.

d. Yes. Given that you assign a complaint to only one category, the categories are mutually exclusive. Given that you assign each complaint to a category, the categories are exhaustive.

15. a. Nominal

b. Qualitative

c.

House Style	Frequency
Charleston Single	17
Colonial	7
Condo Regime	9
Condominium	7
Contemporary	4
Cottage	9
Craftsman	10
Patio	1
Ranch	10
Townhouse	18
Traditional	153

17. Table below is for parts a. and b.

Days Traveling	Frequency	Relative Frequency	Cumulative Frequency
0 – 6	15	0.20	15
7 – 13	21	0.28	36
14 – 20	27	0.26	63
21 – 27	9	0.12	72
28 – 34	2	0.03	74
35 and above	1	0.01	75

19. a. Interval

b. 32–40

Table below accompanies parts c., and d.

Average January Temperature (°F)	Number of Cities	Relative Frequency
14 – 22	1	0.025 = 2.5%
23 – 31	4	0.1 = 10%
32 – 40	11	0.275 = 27.5%
41 – 49	10	0.25 = 25%
50 – 58	8	0.2 = 20%
59 – 67	4	0.1 = 10%
68 – 76	2	.05 = 5%

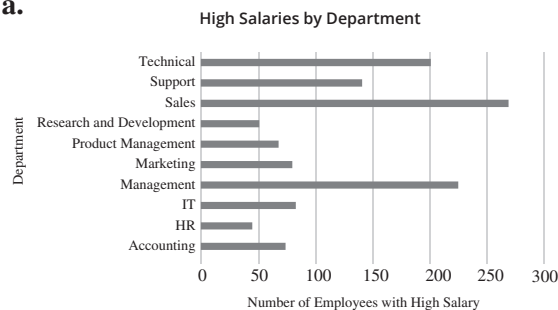
c. 12.5%

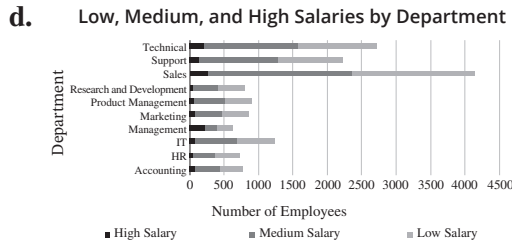
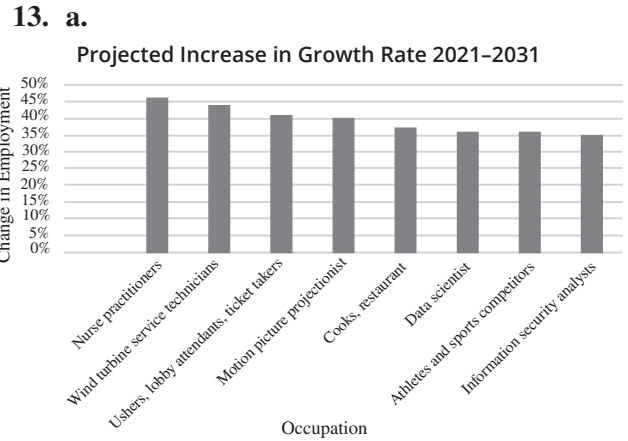
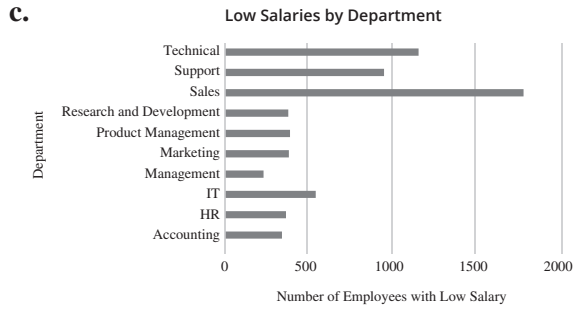
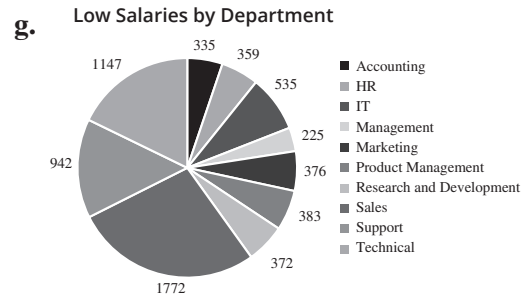
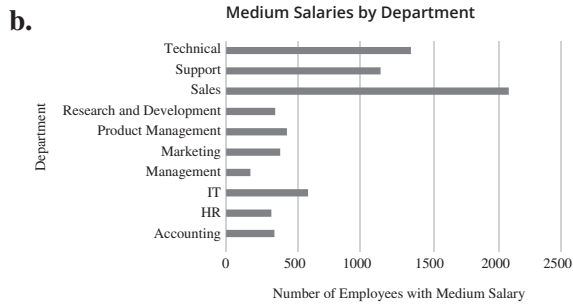
d. 35%

e. It helps determine the percentage of values below or above a certain threshold value by summing the relative frequencies for the classes that meet the criteria. Answers may vary.

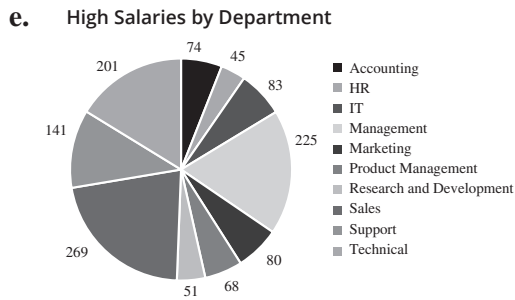
## 3.2 Exercises

11. a.

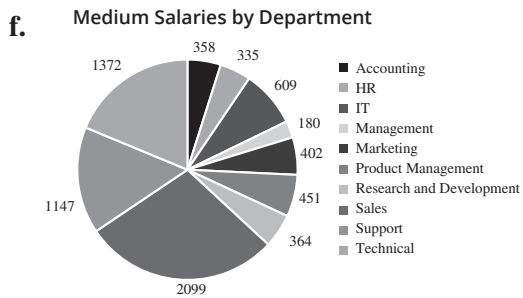




**b.** Answers may vary. For example, the occupation with the highest projected percent change in employment for 2021 through 2031 is nurse practitioners.



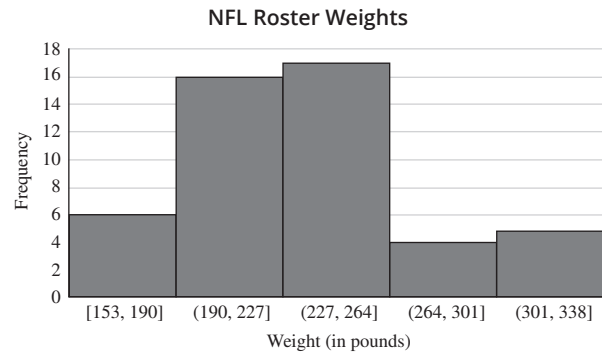
### 3.3 Exercises



**17. a.** Answers may vary.

NFL Roster Weights	
Weight range	Frequency
153 – 190	6
191 – 227	16
228 – 264	17
265 – 301	4
302 – 338	5

**b.** Answers may vary.

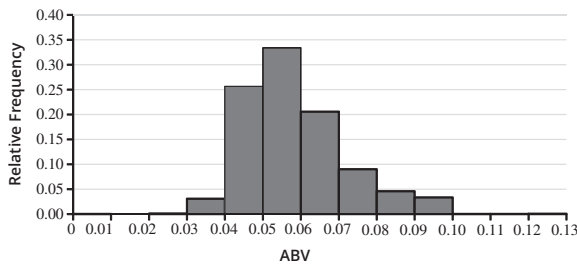


Note that results shown are using Excel.

19. Table below is for parts a. and b.

ABV	Frequency	Relative Frequency
0.001 – 0.010	1	0.0004
0.011 – 0.020	0	0.0000
0.021 – 0.030	3	0.0013
0.031 – 0.040	74	0.0309
0.041 – 0.050	615	0.2567
0.051 – 0.060	800	0.3339
0.061 – 0.070	493	0.2058
0.071 – 0.080	216	0.0902
0.081 – 0.090	110	0.0459
0.091 – 0.100	80	0.0334
0.101 – 0.110	1	0.0004
0.111 – 0.120	1	0.0004
0.121 – 0.130	2	0.0008

c. Relative Frequency Distribution ABV



d. The majority of American brewed beers have between 4% and 7% ABV. Beers with more than 10% ABV are very uncommon. Answers may vary.

21. a. Ratio

b.

Stem	Leaf
18	8 8 8
19	4 8 8 8 8 8 8 8 8 9
20	5 5 9 9
21	0 7 7
22	1
23	6 6
24	6

Key : 18|8 = \$188

c. The distribution is skewed to the right.

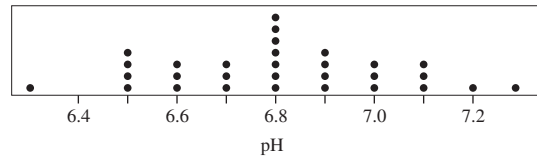
23. a. Interval

b. Answers may vary. For example, both distributions appear to be bell-shaped but the average temperatures in June for Charleston are about ten degrees higher than Milwaukee, Wisconsin.

c. 64.5%, 80%

d. 82 degrees

25. a. pH of Drinking Water



Note the dot plot was created in Minitab.

b. 6.8

c. No, one of the data values is below 6.5.

d. The water tends to be acidic as the majority of the pH values are below 7.

27. a. Total contributions have increased from 2000 to 2023. The largest spike happens between 2019 and 2021.

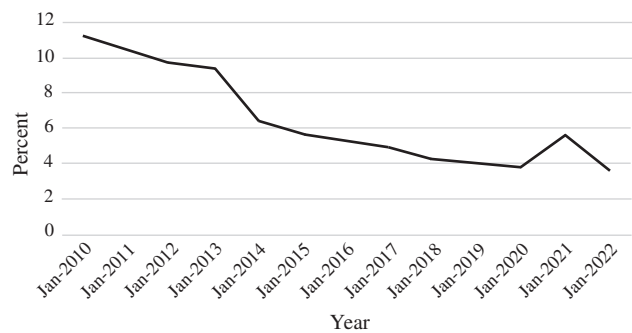
b. Yes

c. Nonstationary

29. a. Ratio

b.

North Carolina Unemployment Rate

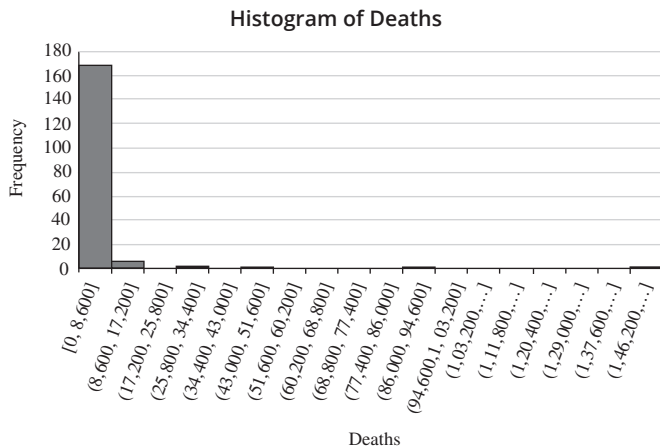


c. Answers may vary. The unemployment rate in North Carolina is steadily decreasing from 2010 to 2020 with a slight upswing in 2021 and then decreasing again in 2022 which indicates a nonstationary time series.

### 3.4 Exercises

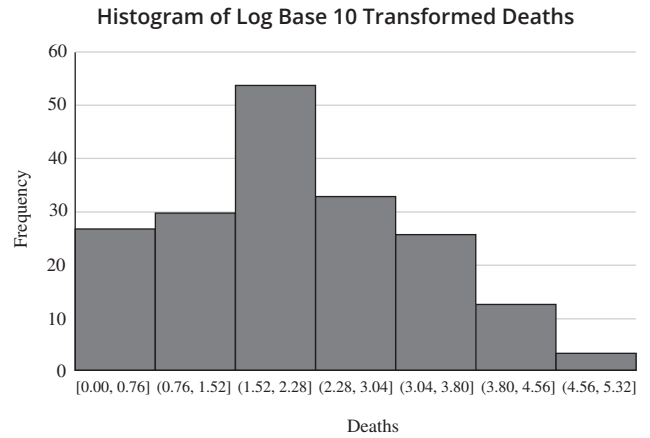
5. Answers may vary. A possible answer is that the scale is not a good choice since a dramatic change is shown even though the change is only a few cents.
7. a. Approximately 30%  
 b. No. The graphic for June 2023 has an area more than double that of the November 2017 graphic, even though the value is only about 30% more.  
 c. Answers may vary. Ensure that the width of the two graphics is consistent and only alter the height, thus ensuring that the area increase matches the percentage increase.
9. a. Answers may vary. Graph A is better because the vertical scale starts at 0.  
 b. One could investigate the robbery rates per capita which may normalize the data.  
 c. The 2020 bar in Graph B is approximately 6 times taller than the 2012 bar though there were only about 1.5 times as many robberies in 2020 than in 2012.

11. a.



No. It is just a count of the number of deaths for each country.

b.

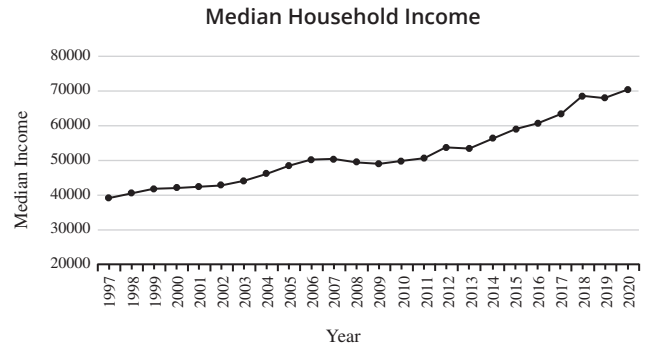


c. It is generally bell-shaped with a slight right skew.

### Chapter 3 Additional Exercises

1. a. Time sequence plot or line graph of both the median family income and percent change in median family income.

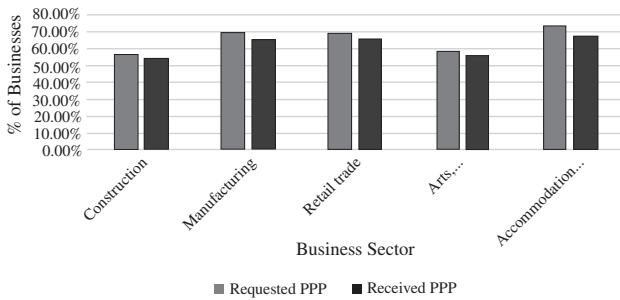
b.



c. Median household income has an upward trend, but this could likely be due to inflation. 2008–2009 and 2019 are the only years in which median household income decreased from the previous year. Answers may vary.

3. a. A side-by-side bar chart showing the percentage of both Requested PPP and Received PPP gives a comparison of the denial percentage of loans. Also, a bar chart of just the Requested PPP would show the business sector with the highest and lowest percentage of requested loans. Answers may vary.

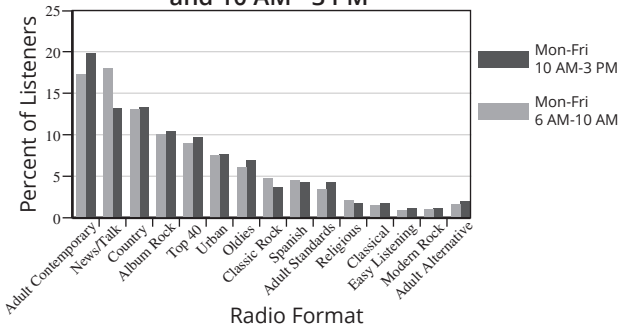
**b. COVID-19 Financial Assistance**



**c.** The percentage difference in those businesses requesting PPP loans and those receiving them is consistent across the business sectors except for the accommodation and food services industry, which received almost 7% less than requested compared to about 3% for the other sectors.

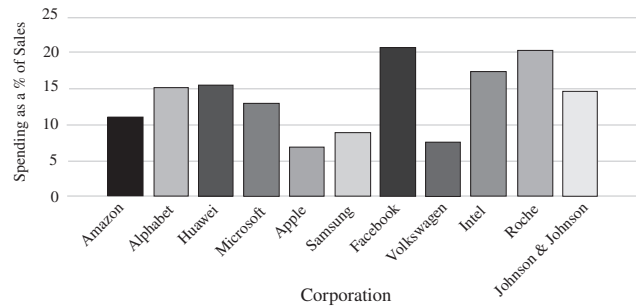
- 5. a.** Bar charts and pie charts would both be appropriate for displaying the data. Answers may vary.
- b.** Answers may vary.
- c.** Answers may vary.

**Radio Formats 6 AM - 10 AM and 10 AM - 3 PM**



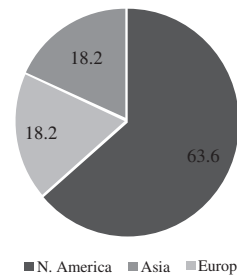
**7. a.** Spending as a % of sales is a more useful measure of research and development expenditures for comparative purposes because it standardizes research and development expenditures.

**b. Top Global Corporations Research & Development Spending**

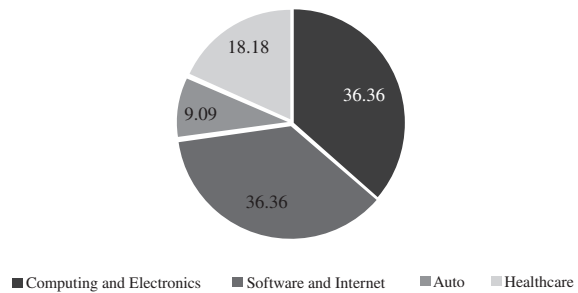


A bar chart of R&D expenditures as a percent of sales by company. The top global company spending the most on research with regard to its percentage of sales is Facebook. The company spending the least is Apple. Answers may vary.

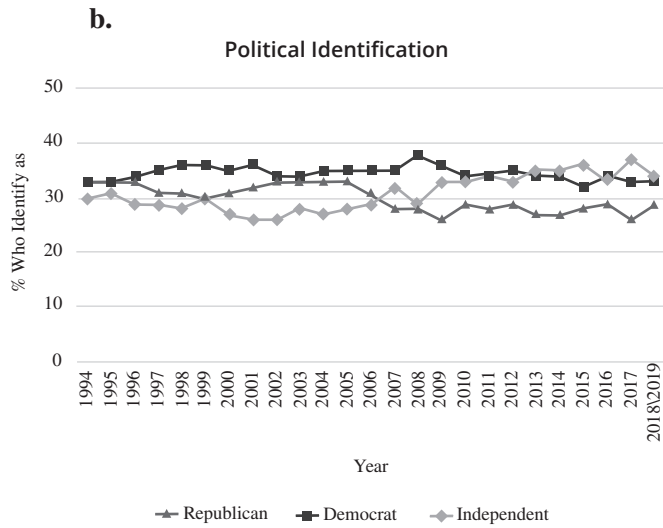
**c. Headquarters Location**



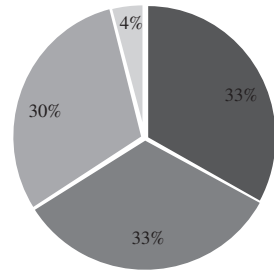
**Industry**



**9. a.** A line graph displaying the percentage of voters in each category and a pie chart showing political identification in a particular year might be helpful in visualizing the data. Answers may vary.



Political Identification in 1994



■ Republican ■ Democrat ■ Independent ■ Other

**c.** In the year 1994 the political identification percentages for the three groups were very close, around 30%. The percentage of people who identify as Republican has decreased since 2005. The percentage of people who identify as Independent has grown since 2004. Answers may vary.

# Chapter 4

## 4.1 Exercises

**13.** mean,  $\bar{x} = 378.88$ , median = 358.6, 20% trimmed mean = 351.19, mode = 293.2

**15.** mean,  $\bar{x} = 12$ , median = 12, mode = 12, 10% trimmed mean = 12.

**17. a.** mean/median

**b.** mode

**c.** median or mode

**d.** median (some very high home prices may skew data)

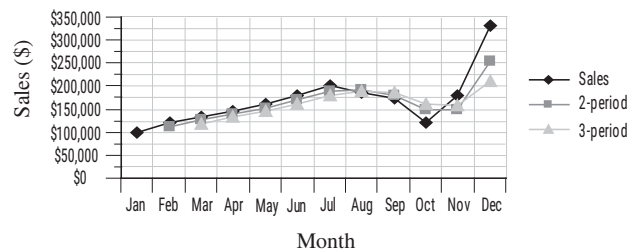
**19. a.** ratio

**b.** mean,  $\bar{x} = 109.2$ ,  
10% trimmed mean = 108.25,  
20% trimmed mean = 107.92

**c.** 10% trimmed mean. Looking at a plot of the blood pressure readings, the top and bottom 10% seem to be outliers.

**d.** empiricism

**21.** Answers for parts **a.**, **b.**, **c.**, and **d.** shown in the graph and table below.



Month	Sales	2-period	3-period
Jan	\$100,500		
Feb	\$120,000	\$110,250	
Mar	\$133,000	\$126,500	\$117,833
Apr	\$145,000	\$139,000	\$132,667
May	\$160,000	\$152,500	\$146,000
June	\$180,000	\$170,000	\$161,667
July	\$200,000	\$190,000	\$180,000
Aug	\$185,000	\$192,500	\$188,333
Sept	\$175,000	\$180,000	\$186,667
Oct	\$120,000	\$147,500	\$160,000
Nov	\$180,000	\$150,000	\$158,333
Dec	\$330,000	\$255,000	\$210,000

**e.** Answers may vary.

23. 84.86%

## 4.2 Exercises

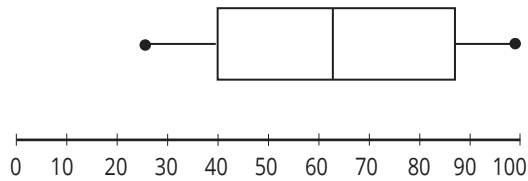
11. a. sample variance = 90.033  
 b. sample standard deviation = 9.489  
 c. range = 44  
 d. Pass attempts, opponent defensive ability, playbook, etc. Answers may vary.
13. a. 0.2917  
 b. 0.54 or 54 percentage points  
 c. 1.5%  
 d. Answers may vary.
15. a. Ratio  
 b. Answers may vary.  
 c. Portfolio A:  
 Average market value = \$150,000,  
 Standard deviation = \$7,906,  
 Maximum Value = \$160,000,  
 Minimum Value = \$140,000,  
 Portfolio B:  
 Average market value = \$150,000,  
 Standard deviation = \$37,249,  
 Maximum Value = \$195,000,  
 Minimum Value = \$100,000  
 d. Portfolio B is more risky because it has the larger standard deviation.
17. a. mean,  $\bar{x} = 10.4$ , standard deviation = 5.80  
 b. mean,  $\bar{x} = 104$ , standard deviation = 57.97.  
 The mean and standard deviation of the adjusted data are each ten times that of the original data. Answers may vary.  
 c. mean,  $\bar{x} = 5.2$ , standard deviation = 2.90.  
 The mean and standard deviation of the adjusted data are each half that of the original data. Answers may vary.  
 d. Multiplying or dividing each data point by the same constant will result in a mean and standard deviation that are each multiplied or divided by that constant. Answers may vary.
19. 718 to 1226

21. a. 68%  
 b. 2.5%  
 c. 16%
23. a. 2  
 b. 2.8  
 c. Since the coefficient of variation is smaller for Brand A than Brand B, the standard deviation of the amounts of Vitamin C in the tablets is smaller relative to the average size of the tablets for Brand A than for Brand B. Thus, Brand A more consistently produces tablets as advertised. Answers may vary.

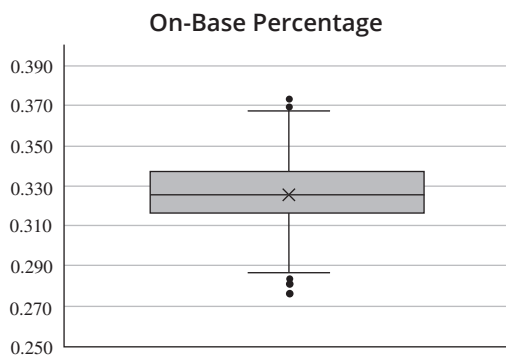
## 4.3 Exercises

11. a. ratio  
 b. 20<sup>th</sup> percentile, or score in 8<sup>th</sup> position = 66  
 c. 95<sup>th</sup> percentile, or score in 37<sup>th</sup> position = 91  
 d. 20% of the safety scores are less than or equal to 66; 95% of the safety scores are less than or equal to 91.
13. a. ratio  
 b. 0.161  
 c. 0.323  
 d. 20<sup>th</sup> percentile: 20% of the counties in the USA have less than or equal to 16.1% of their adult population that smoke. 95<sup>th</sup> percentile: 95% of the counties in the USA have less than or equal to 32.3% of their adult population that smoke.
15. a. 238 lbs.  
 b. 76  
 c. about 75%  
 d. about 14
17. a. Ratio  
 b. 40  
 c. 61  
 d. 84  
 e. 25<sup>th</sup> percentile: 25% of scores were at or below 40. 50<sup>th</sup> percentile: 50% of scores

were at or below 61. 75<sup>th</sup> percentile:  
75% of scores were at or below 84.



- f. 44
- g. No outliers
- h. 0.826
- i. 0.006.
- j. Score of 81 is 0.826 standard deviations above the mean. Score of 62 is 0.006 standard deviation above the mean.
19. a. Ratio
- b. 0.312
- c. 0.318
- d. 0.327
- e. 25<sup>th</sup> percentile: 25% of the on-base percentages are at or below 0.312, 50<sup>th</sup> percentile: 50% of the on-base percentages are at or below 0.318, 75<sup>th</sup> percentile 75% of the on-base percentages are at or below 0.327.
- f. 0.015
- g. Yes, there are a couple outliers both above  $Q_3$  and below  $Q_1$ .



- h.  $-3.086$ , 0<sup>th</sup> percentile
- i. 1.910, 96<sup>th</sup> percentile
- j. The player with the 0.280 on-base percentage has a batting average 3.086 standard deviations below the average on-base percentage of 0.326, the player with a .355 on-base percentage has a batting

average 1.910 standard deviations above the average on-base percentage of 0.326.

21. a. 0.762, a score of 80 is 0.762 standard deviations above the mean.
- b. 0, a score of 64 is 0 standard deviations above the mean.
- c.  $-1.143$ , a score of 40 is 1.143 standard deviations below the mean.
23. a. 19.656 in., 49.93 cm.
- b. 1.077 in., 2.735 cm.
- c. The values of the  $z$ -scores should be the same regardless of whether the baby's length was measured in inches or centimeters because the  $z$ -score standardizes the measurement. Answers may vary.
- d.  $-0.377$
- e.  $-0.377$
- f. Answers may vary.

## 4.4 Exercises

5. a. Nominal: ID, Duplex, Subdivision, New, House Style, Misc. Exterior, Pool, Dock, Fenced Yard, Screened Porch, Golf Course, and Fireplace; Ratio: List Price, No. of bedrooms, No. of bathrooms, Stories, Square Footage, Acreage, Covered Parking Spot, No. of Fireplaces; Interval: Year Built
- b. All but Misc. Exterior and Amenities.
- c. Create intervals that break the variable into a few (2-5) categories such as 'Less than or equal to \$500,000' and 'Greater than \$500,00' in List Price or 'Acreage less than or equal to 0.5 acres' and 'Acreage greater than 0.5 acres'.
- d. 11; Charleston Single, Colonial, Condo Regime, Condominium, Contemporary, Cottage, Craftsman, Patio, Ranch, Townhouse, and Traditional
- e. 103; Carolina Park with 42 new homes
- f. \$552,882
- g. Minimum is in Carolina Park with a List Price of \$369,900; Maximum is in Park West with a List Price of \$1,800,000.

- h. The minimum priced home is \$205.84 per square foot and the maximum priced home is \$333.27 per square foot.
- i. Answers may vary. Possible reasons include: 5 beds, 5.5 baths, 5401 sq. ft., has a pool, and has a dock.

## 4.5 Exercises

- 3. Mean = 17.5; Variance = 53
- 5. a. 18.29%
- b. 0.0007
- c. 0.0257 (calculated directly) or 0.0265 (square root of variance)

## 4.6 Exercises

- 5. 0.565
- 7. 0.0833
- 9. a. 28%
- b. 22%
- c. Americans are not in the habit of saving money in case of a financial emergency. Around a quarter would not be able to cope at all. Answers may vary.
- 11. a. lavish dinners
- b. 0.264
- c. 0.24
- d. Many of the clients did not receive a perk and still joined the Wall Street firm. Given that the perks are so expensive, firms might want to reevaluate whether or not offering the perks is worth it. Answers may vary.

## Chapter 4 Additional Exercises

- 1. a. mean,  $\bar{x} = 56.05$ , median = 55, mode = 51
- b. min = 42,  $Q_1 = 51$ , median = 55,  $Q_3 = 61$ , max = 78
- c.  $IQR = 10$
- d. Yes, 78 is just above the value of 76 that delineates outliers.

- 3. a. 0.3438
- b. Mean = 2.4531, Standard Deviation = 2.5754
- c. 0 to 5.0285
- d. 87.5%
- e. The empirical rule predicts that 68.26% of the data falls within one standard deviation of the mean. The percent of the data in this problem falling within one standard deviation is 87.5%, which is not very close to the empirical rule. Answers may vary.
- 5. a. Machine A = 3.1429, Machine B = 3.1429
- b. Machine A = 5.8095, Machine B = 0.8095
- c. Machine A = 2.4103, Machine B = 0.8997
- d. Machine B is probably a better machine because the average number of defects produced by the 2 machines is the same but Machine B is much more consistent in the number of defective circuit boards it produces. Answers may vary.
- e. Answers may vary.
- 7. a. 93
- b.  $s = 5.06$
- c. 68%, or about 13 countries
- d. 15 countries have literacy rates between 87.94 and 98.06.
- e. We assumed that the distribution of literacy rates is symmetric and bell-shaped.
- f. Overall, the Asian countries have a slightly lower mean literacy rate of 91.1 but a larger standard deviation of 13.25.
- 9. a. Mean = 38.12
- b. Median = 21.4
- c. Bimodal, Mode 1 = 14.50, Mode 2 = 8.50
- d. 25.76
- e. The skewness of the distribution will be towards the left.
- 11. a.  $\sigma = \frac{73.24 - 67.8}{3.4} = 1.6$  km/s/Mpc
- b. Answers may vary.

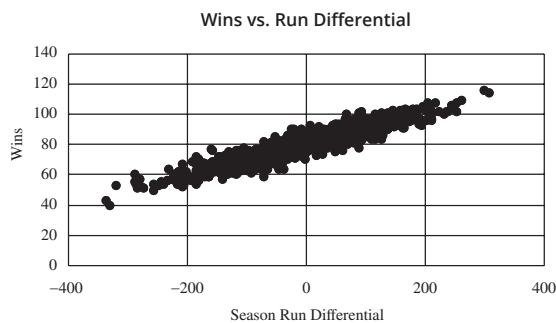
# Chapter 5

## 5.1 Exercises

11. a. The pattern in Scatterplot A does not follow a linear pattern, but the pattern in Scatterplot B does.
- b. There is no discernible pattern in Scatterplot A. The pattern in Scatterplot B is downward sloping.
- c. The data values are widely dispersed in Scatterplot A, but they are tightly clustered around a line in Scatterplot B.
- d. Scatterplot A does not have a pattern from which to deviate, but there are no significant deviations in Scatterplot B.

13. a.
- What is the run differential necessary to win 95 games? Answers may vary.
  - Yes
  - Possible errors in recording of runs scored and runs allowed. Answers may vary.
  - Run Differential: Ratio; Wins: Ratio
  - Observation

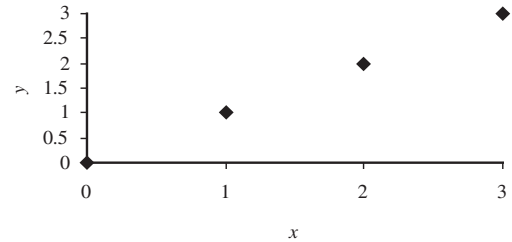
b.



c.

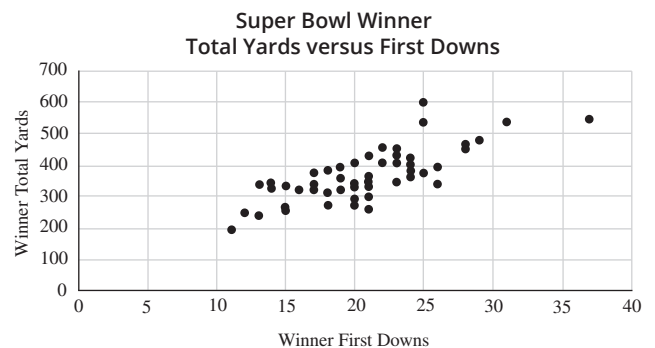
- Yes, the data follows an obvious linear pattern.
- Upward sloping
- Tightly clustered
- Yes, there are a few observations at each end of the plot that are worthy of further analysis.

15.



17. a. Tightly clustered in a positive linear fashion.
- b. Loosely clustered in a positive linear fashion.
- c. Tightly clustered in a negative linear fashion.
- d. Loosely clustered in a negative linear fashion.
- e. No linear relationship.
- f. The scatterplots look linear.
19. a. Strong positive linear relationship.
- b. Moderate positive linear relationship.
- c. Strong negative linear relationship.
- d. Moderate negative linear relationship.
- e. No linear relationship.
- f. The scatterplots look linear.

21. a.



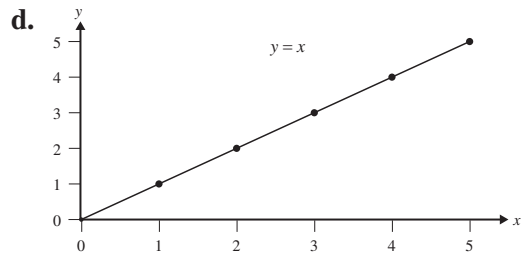
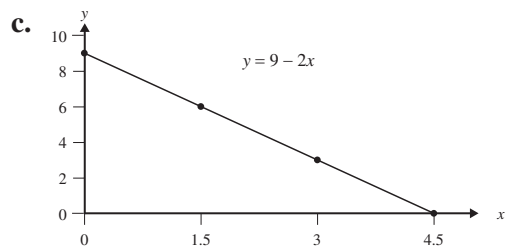
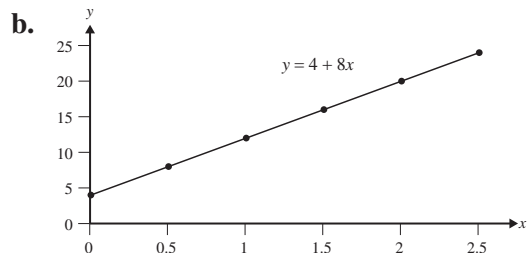
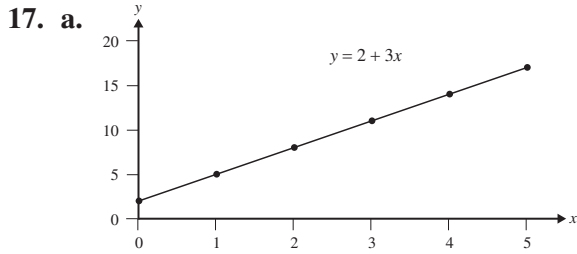
- b. There is a positive relationship between the variables.
- c.  $r = 0.764$

23. a. Social media usage

- b. Hours of sleep
- c. Number of hours worked

## 5.2 Exercises

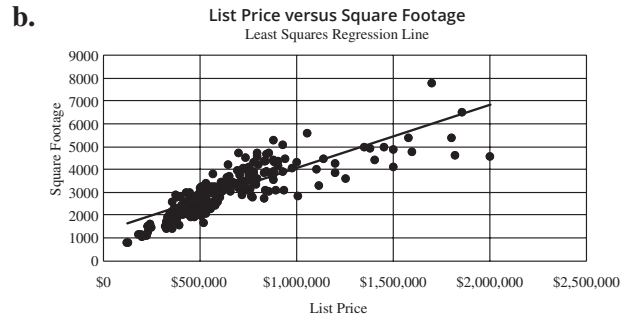
15. a. Sales volume, since it is the value we want to predict.  
 b. Advertising expenditures, since we use it to predict sales volume.  
 c. \$69,650  
 d. \$99,510  
 e. Random error. This may cause the company to under or over-estimate sales volume, causing budgeting problems. Answers may vary.



19. a. 15, 45, 65, 85, 95  
 b. Positive

- c. We would expect  $r$  to be positive, since the variables have a positive relationship.

21. a.  $List\ Price = b_0 + b_1 Square\ Footage$



c.

Predicted Selling Price	Error	Squared Error
316,513	58,487	3,420,680,730
403,168	20,432	417,455,775
341,638	106,677	11,379,941,045
477,004	38,246	1,462,729,599
573,658	-17,258	297,833,041
636,470	-35,970	1,293,825,685
654,416	-70,796	5,012,082,041
678,772	4253	18,090,450
661,595	-26,345	694,036,258
555,712	59,588	3,550,781,109
712,101	19,309	372,856,713
758,504	102,246	10,454,156,319
941,813	-106,813	11,408,968,006
840,545	-25,045	627,228,348

- d.  $SSE = 50,410,665,118$   
 e. No, *Square Footage* is the best predictor for *List Price* in the data set. Any other variable will have a wide dispersion of data values.

23. a. The negative correlation coefficient indicates that as latitude increases the average July temperature decreases.

b.  $Average\ July\ Temperature = 115.4885 - 0.8152\ Latitude$

c. 84 °F

25. a. *Current Value of Home*

b. *Annual Salary*

c. \$235,611

- d. \$15,700
- e. For each additional dollar earned in annual salary, the current value of home is predicted to increase by \$3.14.
- f. If someone is not earning any annual income, the predicted value of his or her home would be \$72,331.
- g. It is possible because if you have a greater annual salary you have more money to spend on a home. However, we cannot conclude that there is a causality from the estimated regression equation. Answers may vary.

27. a.  $\hat{y} = 5.7333 + 0.6667x$   
 b. 16.4  
 c. No.  $\hat{y} = 10.9732 + 0.3065x$   
 d. 15.8772  
 e. There is not likely a causal relationship between these two variables. Answers may vary.

### 5.3 Exercises

- 21. a. 0.0641  
 b. 0.2531
- 23. a. Estimated  $RS = -1079.02442 + 5486.29728 OBP$   
 b. Estimated  $RS = -315.51172 + 2610.88272 SLG$   
 c. Estimated  $RA = -865.94701 + 4921.06617 OOBP$   
 d. Estimated  $RA = -448.44236 + 2908.69139 OSLG$

e.

Predicted RS	Error	Squared Error
671	16	256
841	56	3136
677	47	2209
863	60	3600
693	-51	2601

- f. SSE = 11,802.000; Standard Error = 108.637
- g.

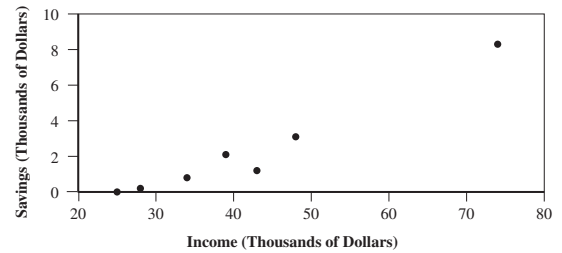
Predicted RA	Error	Squared Error
709	4	16
831	-25	625
651	-24	576
988	-20	400
823	-57	3249

- h. SSE = 4866.000; Standard Error = 69.757
25. a. Ratio

b. *Savings*

c. *Income*

d. **Income and Savings**



Yes, the data points in the scatterplot appear approximately linear with a positive slope trend.

- e.  $\hat{y} = -4.8658 + 0.171x$
- f. \$3684.20
- g. Savings is estimated to increase by approximately \$171 for each additional \$1000 earned in annual income.
- h.  $R^2 = 0.9495$  Approximately 94.95% of the variation in savings is explained by the variable annual income.

### 5.4 Exercises

- 5. a. Upward
- b. Independent variable: *Year*; Dependent variable: *CO<sub>2</sub> Emissions per Capita*
- c.  $CO_2 \text{ Emissions per Capita} = \beta_0 + \beta_1 \text{ Year} + \epsilon_i$
- d.  $CO_2 \text{ Emissions per Capita} = -179.22 + 0.0912 \text{ Year}$
- e.  $CO_2 \text{ Emissions per Capita} = -179.22 + 0.0912(2020) = 5.004$
- f. Yes, we can subtract our estimate from the actual observed CO<sub>2</sub> emissions per capita for Chile for the year 2020 to determine the error. Answers may vary.

7. a. Independent variable: *Year*; Dependent variable: *Beer Production*
- b.  $\hat{y} = 3325.0213 - 1.5563x$
- c. Yes,  $R^2 = 0.8901$
- d. *Production*  
 $= 3325.0213 - 1.5563(2022) = 178.1827$
- e. Seasonal events such as the NFL football season e.g., the Super Bowl), economic growth leading to more disposable income, increase in number of local breweries, increased advertising, government tax incentives, less expensive prices on beer ingredients, supply chain stability, etc. Answers may vary.
- f. Increased taxes on alcoholic beverages, increased competition from alternatives such as wine and liquor, more stringent marketing restrictions, shortages of key beer ingredients, an economic downturn leading to less disposable income, increasing health and wellness trends, supply chain disruptions, public health campaigns about drunk driving or alcohol-related diseases, etc. Answers may vary.

f.

x	y	pred y	error	error <sup>2</sup>
1	1	-4	5	25
2	4	4	0	0
3	9	12	-3	9
4	16	20	-4	16
5	25	28	-3	9
6	36	36	0	0
7	49	44	5	25
Total			0	84

g. Average value of the model's errors  $= \frac{0}{7} = 0$

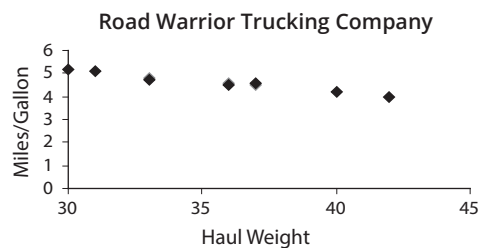
h. Variance of the model's errors  
 $= \frac{84}{(7-2)} = \frac{84}{5} = 16.8$

3. a. ratio

b. *Miles per gallon.*

c. *Haul Weight.*

d. Data appears to be approximately linear. Thus a linear model is appropriate.



e. *Miles per gallon*  $= b_0 + b_1$  *Haul Weight.*

f. *Miles per gallon*  
 $= 7.94 - 0.0935$  *Haul Weight.*

g. The slope is  $-0.0935$ . This means that for each additional thousand pounds of haul weight, the miles per gallon decrease by 0.0935 miles per gallon.

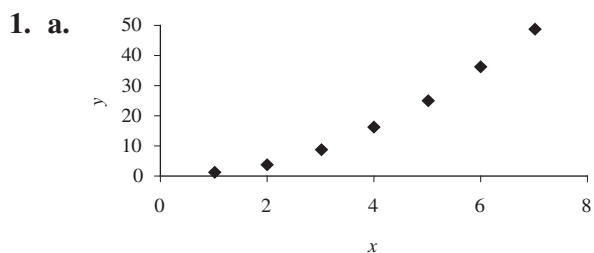
h. *Miles per gallon*  
 $= 7.94 - 0.0935(38) = 4.387$ .

i. It seems reasonable to think that a larger haul weight would contribute to the truck having a lower gas mileage. But, regression analysis does not prove the causality.

j.  $r = -0.977$ , indicating a strong negative linear relationship between miles/gallon and haul weight.

k.  $R^2 = 95.4\%$ , indicating that 95.4% of the variation in the observed miles/gallon is explained by the differences in haul weight.

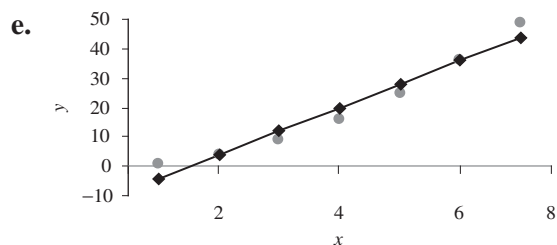
## Chapter 5 Additional Exercises

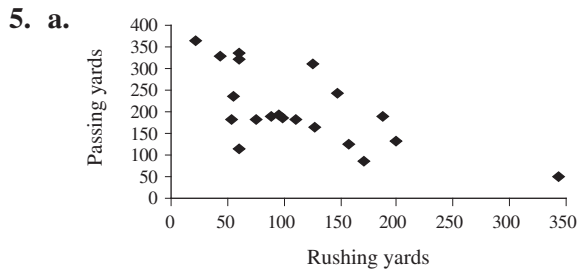


b.  $r = 0.977$

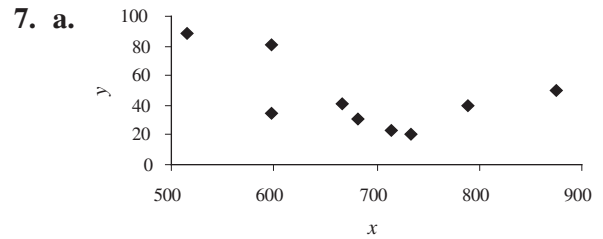
c. The correlation coefficient indicates a strong positive linear relationship-but notice that the scatterplot is not linear.

d.  $b_1 = 8$   $b_0 = -12$   
 $y = b_0 + b_1x = -12 + 8x$





- b.  $-0.6633$ ; yes, the negative sign makes sense as the number of rushing yards increases, the number of passing yards tends to decrease as teams score usually via one method or the other.
- c.  $0.4399$ ; it is the percentage of variation in passing yards explained by the model using rushing yards as the explanatory variable.
- d. The coefficient of determination is the square of the correlation coefficient.



b. Negative relationship

c.  $-0.51883$

d.

$X$	$Y$	$V$	$W$
876	50.1	176	0.1
516	88.2	-184	38.2
598	80.7	-102	30.7
789	39.6	89	-10.4
734	20.5	34	-29.5
667	40.9	-33	-9.1
682	30.6	-18	-19.4
714	22.9	14	-27.1
598	34.8	-102	-15.2

e.  $-0.51883$

f. There is no effect on the correlation coefficient by subtracting a constant.

## Chapter 6

### 6.1 Exercises

13. a.  $P(1) \approx 0.4242$ ,  
 $P(2) \approx 0.2121$ ,  
 $P(3) \approx 0.1515$ ,  
 $P(4) \approx 0.0606$ ,  
 $P(5) \approx 0.0909$ ,  
 $P(6) \approx 0.0606$

b. 1

c. Empirically

15. 0.88

17. 0.9

19. a. Subjective

b. Relative frequency

c. Subjective

d. Relative frequency

e. Classical

21. a.  $\frac{1}{5} = 0.2$

b.  $\frac{4}{5} = 0.8$

23. a.  $\frac{261}{682} \approx 0.3827$

b.  $\frac{391}{682} \approx 0.5733$

c. Relative frequency

### 6.2 Exercises

9. a. Yes

b. Yes

c. No, probabilities cannot be greater than 1.

d. Yes

e. No, probabilities cannot be less than 0.

11. a. 1

b. 0

c.  $\frac{3}{6} = \frac{1}{2} = 0.5$

d. 0

13. a. 0.9696

b. 0.0523

15. a. 0.27

b. 0.08

c. 0.165

d. 0.4

e. 0.3

f. 0.51

g. Relative frequency

h. No, the wife could have more than \$1,000,000 in insurance while the husband could have between \$250,000 and \$499,999 of insurance. Answers may vary.

## 6.3 Exercises

9. a. 0.3194

b. 0.3472

c. 0.6327

d. 0.6032

11. a. 0.8571

b. 0.1429

13. a. 0.005

b. 0.0417

15. No, the events are dependent. If  $A$  = husband has more than \$1,000,000 insurance and  $B$  = wife has more than \$249,999 insurance,  $P(A|B) \neq P(A)$ .

17. 0.0001

19. a. 0.9980

b. 0.0020

c. 0.000001

21. a. 0.0004

b. 0.0004

23. a.  $\frac{4}{42} \cdot \frac{13}{41} = \frac{26}{861} \approx 0.0302$

b.  $\frac{(9+4+4+2)}{42} \cdot \frac{(9+4+4+2-1)}{41}$   
 $= \frac{57}{287} \approx 0.1986$

## 6.4 Exercises

5. 120

7. a. 1

b. 6

c. 120

d. 5040

9. 19,958,400

11. 2520

13. 2,598,960

15.  ${}_{15}C_{10} = 3003$

17.  $\frac{8!}{4!2!2!} = 420$

19. The only word without any repeated letters is the first word, TEARS. Therefore, it will have the most five-letter arrangements possible.

## 6.5 Exercises

5.  $P(\text{Defective} | \text{Inspected}) = 0.3947$

7.  $P(\text{Adult} | \text{Cried}) = 0.2553$

9.  $P(\text{Online} | \text{Fail}) = 0.1429$

## Chapter 6 Additional Exercises

1. a.  $S = \{\text{MMM}, \text{MMF}, \text{MFM}, \text{MFF}, \text{FMM}, \text{FMF}, \text{FFM}, \text{FFF}\}$

b.  $\frac{1}{8} = 0.125$

c.  $\frac{7}{8} = 0.875$

3. a.  $\frac{18}{38} = \frac{9}{19} \approx 0.4737$

- b.  $\frac{12}{38} = \frac{6}{19} \approx 0.3158$   
 c.  $\frac{2}{38} = \frac{1}{19} \approx 0.0526$   
 d.  $\frac{1}{38} \approx 0.0263$   
 e.  $\frac{35}{38} \approx 0.9211$
5. 0.9989
7. a. 1 to 5  
b. 1 to 1
- c. 7 to 1  
d.  $\frac{8}{11} \approx 0.7273$
9.  $\frac{11}{14} \approx 0.7857$
11. a. 8%  
b. 16%  
c. 14%  
d. 84%
13.  ${}_8P_3 = 336$
15.  $\frac{{}_{28}C_2({}_{41}C_2){}_{35}C_0({}_{18}C_0)}{{}_{122}C_4} = \frac{4428}{125,477} \approx 0.0353$

## Chapter 7

### 7.1 Exercises

17. a. Continuous  
b. Discrete  
c. Discrete  
d. Discrete  
e. Continuous
19. a. Discrete  
b. Discrete  
c. Continuous  
d. Discrete  
e. Continuous

21. Yes

23. No. The sum of the probabilities is less than 1.

25. No. Probabilities cannot be negative.

27. Yes

$x$	$P(X = x)$
1	$\frac{1}{30}$
2	$\frac{4}{30}$
3	$\frac{9}{30}$
4	$\frac{16}{30}$

### 7.2 Exercises

9.

$x$	$p(x)$	$xp(x)$	$\frac{(x - \mu)^2}{p(x)}$
400	0.0	0	0.0
420	0.1	42	291.6
440	0.1	44	115.6
460	0.2	92	39.2
480	0.2	96	7.2
500	0.4	200	270.4
Total	1.0	474	724.0

$E(X) = 474; \sigma^2 = 724; \sigma \approx 26.9072$

11.

$x$	$p(x)$	$xp(x)$	$\frac{(x - \mu)^2}{p(x)}$
1	0.1	0.1	0.484
2	0.2	0.4	0.288
3	0.3	0.9	0.012
4	0.2	0.8	0.128
5	0.2	1.0	0.648
Total	1.0	3.2	1.560

- a.  $E(X) = 3.2$   
 b.  $\sigma^2 = 1.56$   
 c.  $\sigma \approx 1.2490$   
 d.  $P(X = 5) = 0.2$   
 e.  $P(X \geq 2) = 0.9$   
 f.  $P(X \leq 3) = 0.6$   
 g.  $P(X < 2) = 0.1$

13. a.

$x$	$p(x)$	$xp(x)$	$\frac{(x - \mu)^2}{p(x)}$
\$50,000	0.4	20,000	518,400,000
-\$10,000	0.6	-6000	345,600,000
Total	1.0	14,000	864,000,000

- b. \$14,000  
 c. \$29,393.88

15. a. Cereal A = \$260,000; Cereal B = \$284,000.  
 b. Cereal A = \$88,881.94;  
 Cereal B = \$302,297.87  
 c. Cereal B has a greater value for expected sales, but also a much greater standard deviation. The difference in the expected sales is much smaller than the difference in the standard deviation, so Cereal A is probably the best choice. Answers may vary.

## 7.3 Exercises

7.  $\frac{2}{10} = 0.2$

9. a.  $\{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}$

b.  $\frac{1}{12} \approx 0.0833$

c.  $\frac{3}{12} = 0.25$

## 7.4 Exercises

7. a. 5

b. 45

c. 15

d. 1

9. a.  $E(X) = 0.9$

b.  $\sigma = 0.9$

c.  $P(X = 2) = 0.1722$

d.  $P(X \leq 3) = 0.9917$

e.  $P(X \geq 2) = 0.2252$

f.  $P(X < 5) = 0.9991$

11. a. Binomial distribution with  $n = 10$  and  $p = 0.10$

b.  $E(X) = 1$

c.  $\sigma \approx 0.9487$

d.  $P(X = 1) = 0.3874$

e.  $P(X = 5) = 0.0015$

f.  $P(X \geq 3) = 0.0702$

13. a. Binomial distribution with  $n = 7$  and  $p = 0.1$

- b.  $P(X = 0) = 0.4783$ ; there is a 47.83% chance that none of the plants will have a power outage in the next month.

$P(X = 4) = 0.0026$ ; there is a 0.26% chance that exactly four of the plants will have a power outage in the next month.

$P(X = 7) \approx 0$ ; there is a negligible chance that all seven of the plants will have a power outage in the next month.

c.  $E(X) = 0.7$

- d.  $\sigma \approx 0.794$ ; the standard deviation is larger than the expected value. The standard deviation is expressed as the number of plants that have a power outage in the next month. Answers may vary.

15. a.  $P(X = 2) = 0.2646$

b.  $P(X \geq 2) = 0.9163$

17. a.  $P(X \leq 1) = 0.8290$

b.  $E(X) = 0.75$

19. a.  $\frac{2}{9}$  or 0.2222

- b.  $P(X = 5) = 0.0389$   
 c.  $P(X = 0) = 0.0810$   
 d.  $E(X) = 2.2222,$   
 $\sigma^2 = 1.7284$

## 7.5 Exercises

7.  $P(X = 2) = 0.0842$
9. a.  $P(X = 0) = 0.1353$   
 b.  $P(X = 0) = 0.0003$   
 c.  $\mu = \lambda = 2$   
 d.  $\mu = \lambda = 8$   
 e.  $\sigma = 2.8284$   
 f.  $P(X \geq 4) = 0.9576$
11. a.  $\lambda = 20$  (If 5 people arrive on average in 15 minutes, then 20 will arrive on average in 60 minutes.)  
 b.  $P(X = 0) = 0$   
 c.  $P(X > 6) = 0.2378$
13.  $P(X = 6) = 0.0771$

## 7.6 Exercises

5. a.  $X$  has a hypergeometric distribution with  $N = 50, A = 3$  and  $n = 10$ .  
 b.  $E(X) = 0.6$   
 c.  $\sigma = 0.6785$   
 d.  $P(X > 1) = 0.4959$   
 e.  $P(X < 2) = 0.9939$   
 f.  $P(X > 3) = 0$
7. a.  $E(X) = 5; \sigma = 1.5076$   
 b.  $P(X = 10) = 0.0006$   
 c.  $P(X = 0) = 0.0006$

## Chapter 7 Additional Exercises

1. a.  $E(X) = 1.25$   
 b.  $\sigma^2 = 1.1875$   
 c.  $\sigma \approx 1.0897$   
 d.  $P(X = 1) = 0.20$

- e.  $P(X \geq 1) = 0.65$   
 f.  $P(X \leq 1) = 0.55$   
 g. The most likely outcome of the World Cup Final according to the empirical data is a goal differential of 0 as this has the highest probability.

3. a.  $P(X \leq 1) = 0.0691$   
 b.  $P(3 \leq X \leq 5) = 0.5982$   
 c.  $P(X = 10) = 0.0020$
5. a.  $P(X = 0) = 0.0067$   
 b.  $P(X = 3) = 0.1404$   
 c.  $P(X \geq 1) = 0.9933$
7. a.  $P(X \geq 1) = 0.1934$   
 b.  $P(X = 0) = 0.8069$
9. a.  $E(X) = 12$   
 b.  $P(X = 15) = 0.0352$   
 c.  $P(5 < X < 9) = 0.0179$
11.  $P(X \geq 1) = 0.9648$
13.  $P(X \geq 3) \approx 1$
15. a. Binomial distribution with  $n =$  the number of buildings inspected and  $p = 0.5$ . The binomial is used rather than the hypergeometric because it is not known how many buildings in the population of new buildings have violations. Answers may vary.  
 b.  $P(X = 3) = 0.25$
17. a. HH1, HH2, HH3, HH4, HH5, HH6, HT1, HT2, HT3, HT4, HT5, HT6, TH1, TH2, TH3, TH4, TH5, TH6, TT1, TT2, TT3, TT4, TT5, TT6  
 b.  $X =$  Sum of the number of heads on the two coins and number of dots on the die.  
 $X = \{1, 2, 3, 4, 5, 6, 7, 8\}$

c.

$x$	$P(X = x)$
1	$\frac{1}{24}$
2	$\frac{3}{24}$

$x$	$P(X=x)$
3	$\frac{4}{24}$
4	$\frac{4}{24}$
5	$\frac{4}{24}$
6	$\frac{4}{24}$

$x$	$P(X=x)$
7	$\frac{3}{24}$
8	$\frac{1}{24}$

d. 4.5

19. a. 0.3115

b. 0.1094

c. 0.3115

d. 0.1094

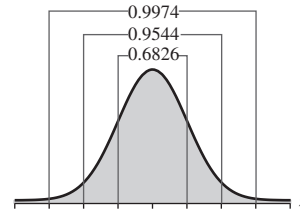
## Chapter 8

### 8.1 Exercises

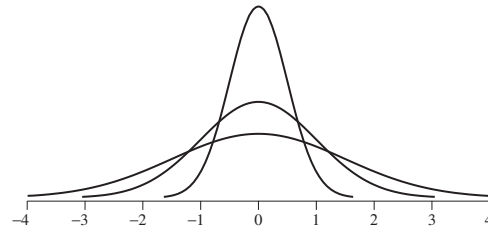
5. a.  $\mu = 40$   
 b.  $\sigma \approx 17.3205$   
 c. 0.4167  
 d. 0.3333  
 e. 0.4167  
 f. 0
7. a.  $\mu = 60$   
 b.  $\sigma \approx 1.7321$   
 c. 0.3333  
 d. 0.5  
 e. 0.1667  
 f. 0
9. a.  $\mu = 8:15$  AM  
 b.  $\sigma \approx 0.1443$   
 c. 0.3333  
 d. 0.1667  
 e. 0.5  
 f. 0

### 8.2 Exercises

9.



11.



13.

Stem	Leaf
10	7 8
11	5 9
12	2 9
13	1 3 4 7
14	0 0 0 8 8 8 9 9
15	1 4 5 6 8
16	1 1 2 4 6
17	0 3 3 7 9
18	4 7
19	0 4 5 7 7

Key :  $10|7 = 10.7$  feet

Yes, the data appear to be normally distributed because the histogram appears to resemble a bell shape. Answers may vary.

## 8.3 Exercises

7. a. 0.2486

b. 0.4500

c. 0.4750

d. 0.4950

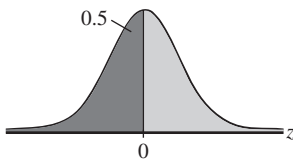
9. a. 0.6046

b. 0.4176

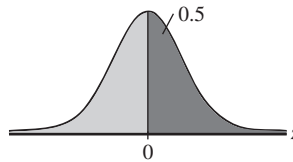
c. 0.9167

d. 0.9856

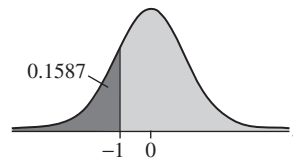
11. a.  $P(z \leq 0) = 0.5$



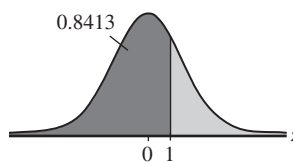
b.  $P(z \geq 0) = 0.5$



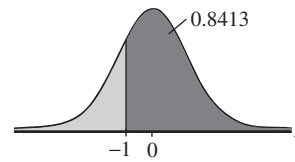
c.  $P(z \leq -1) = 0.1587$



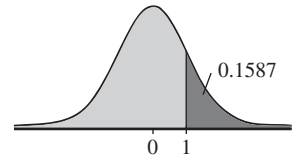
d.  $P(z \leq 1) = 0.8413$



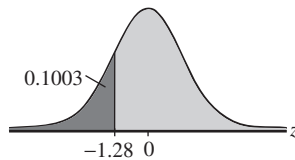
e.  $P(z \geq -1) = 0.8413$



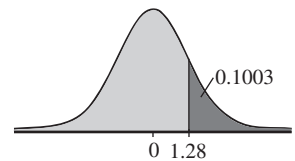
f.  $P(z \geq 1) = 0.1587$



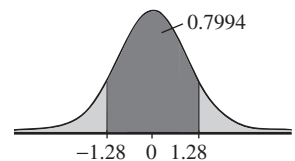
13. a.  $P(z \leq -1.28) = 0.1003$



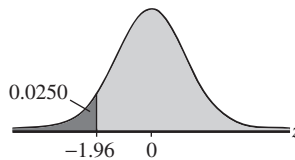
b.  $P(z \geq 1.28) = 0.1003$



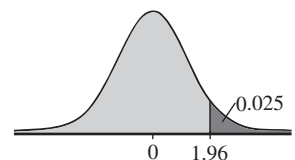
c.  $P(-1.28 \leq z \leq 1.28) = 0.7994$



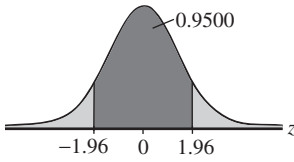
d.  $P(z \leq -1.96) = 0.0250$



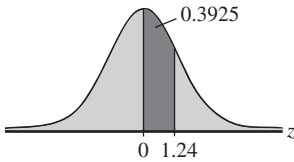
e.  $P(z \geq 1.96) = 0.0250$



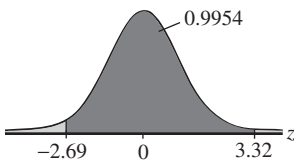
f.  $P(-1.96 \leq z \leq 1.96) = 0.9500$



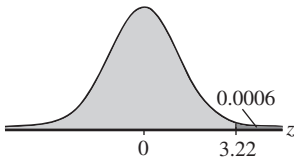
15. a.  $P(0 \leq z \leq 1.24) = 0.3925$



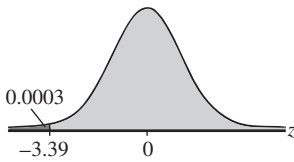
b.  $P(-2.64 \leq z \leq 3.32) = 0.9954$



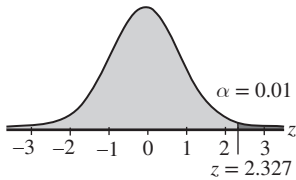
c.  $P(z \geq 3.22) = 0.0006$



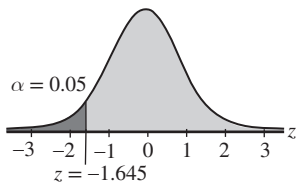
d.  $P(z \leq -3.39) = 0.0003$



17.



19.

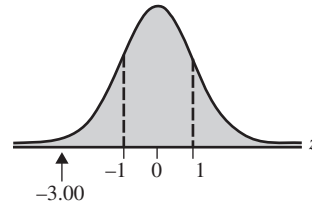


21. -1.28

23. 1.96

## 8.4 Exercises

3.



5. a. 0.0779

b. 0.0808

c. 0.6554

7. 0.4101

9. a. \$938

b. 0.2676

c. 0.0838

d. 0.4640

e. Answers may vary.

11. a. 0.253

b. 0.091

c. 0.409

13. a. 92.24

b. No. The score must be at least a 92.24 to be in the top 10% of scores.

c. 75.28

d. The student who scored a 65 would receive an F because the score is less than 71.76 and thus is in the lowest 10% of the scores.

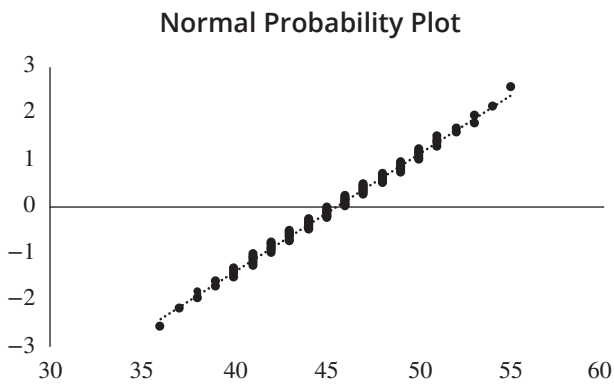
15. a. 13.29

b. Yes, because a weight of more than 13.29 lb is in the top 5%.

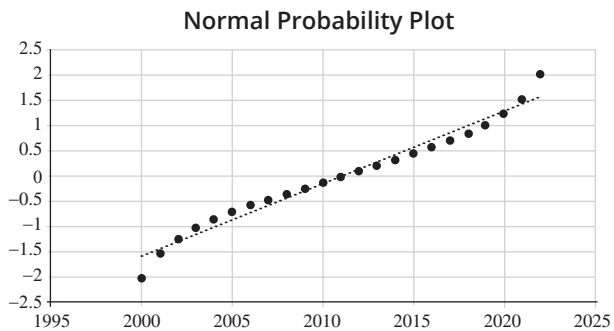
c. No, because a weight of less than 13.29 lb is not in the top 5%.

## 8.5 Exercises

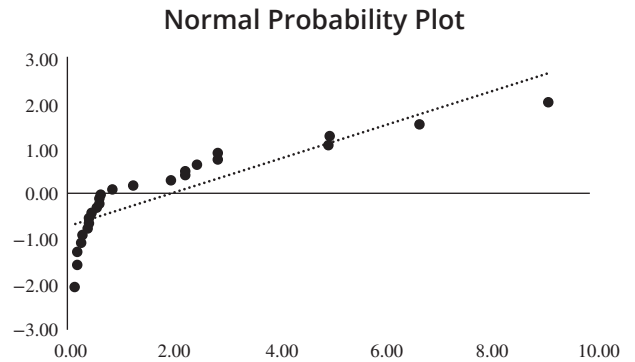
5. The data seem to fit a line very closely. We would also assume from the normal probability plot that the population is normally distributed. Answers may vary.



7. By examining the normal probability plot, we notice a systematic pattern in the data that is not linear. The data do not appear to be normally distributed. Answers may vary.



9. By examining the normal probability plot, we notice a substantial deviation from a linear pattern. The data do not appear to be normally distributed. Answers may vary.

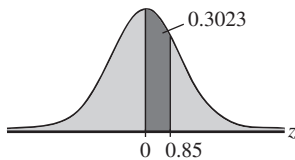


## 8.6 Exercises

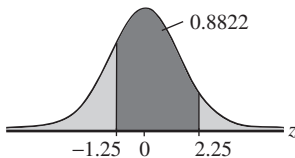
5. No; currently,  $np = 4$ .
7. a. 9  
b. 2.7659  
c. 0.5714  
d. 0  
e. 0.9390
9. a. 60  
b. 7.6681  
c. 0  
d. 0.9924  
e. 0.1075

## Chapter 8 Additional Exercises

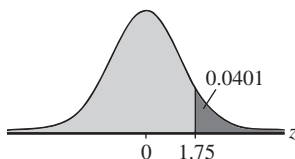
1. a.  $P(0 \leq z \leq 0.85) = 0.3023$



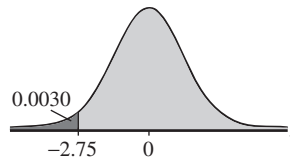
b.  $P(-1.25 \leq z \leq 2.25) = 0.8822$



c.  $P(z \geq 1.75) = 0.0401$



d.  $P(z \leq -2.75) = 0.0030$



3. 2.575

5. a. 0.2266

b. 0.1056

c. 0.6677

d. 5.97 lbs.; Yes, because the lowest 10% of weights are below 5.97 lbs. and 5 is less than 5.97.

7.  $\mu \approx 507, \sigma \approx 114$

9. 535 days

11. The time limit should be 100 minutes.

13. 0.6915

15.  $\mu \approx 825, \sigma \approx 133.3333$

## Chapter 9

### 9.1 Exercises

17. Answers may vary.

19. a. Yes, readers voluntarily sent in their responses.

b. People with strong opinions would be more likely to reply and “Anger” is not well-defined and reader response may be influenced by whether or not they are angry at the moment, for example. Answers may vary.

c. No. All Americans do not read the magazine, and even if they did, it is likely that only those readers with strong opinions would have responded to the survey. Answers may vary.

21. a. voluntary sampling

b. Only people with Twitter accounts were able to participate. Only people

with strong opinions may have felt compelled to participate in the poll.

c. It is only reasonable to believe that the results of the survey reflect the attitudes of the Twitter users who participated in the poll—not necessarily all Twitter users.

### 9.2 Exercises

11. a.  $\mu_{\bar{x}} = 50, \sigma_{\bar{x}} = 1.5811$

b.  $\mu_{\bar{x}} = 50, \sigma_{\bar{x}} = 1.3484$

c.  $\mu_{\bar{x}} = 50, \sigma_{\bar{x}} = 1$

d. The standard error decreases, reflecting the additional information provided by a larger sample size.

13. a. 0.4448

b. 0.9994

15. a. 0.0329  
b. 0.0094  
c. 1

17. 0

### 9.3 Exercises

7. a.  $\mu_{\hat{p}} = 0.45$ ,  
 $\sigma_{\hat{p}} = 0.0908$   
b.  $\mu_{\hat{p}} = 0.45$ ,  
 $\sigma_{\hat{p}} = 0.0742$   
c.  $\mu_{\hat{p}} = 0.45$ ,  
 $\sigma_{\hat{p}} = 0.0617$   
d. The standard deviation decreases, reflecting the additional information provided by a larger sample size.
9. 0.1867
11. a. 0.0105  
b. 0.0829  
c. 0.7518  
d. 0.2338
13. a. 0.0516  
b. 0.5
15. a. 0.1894  
b. 0.0384  
c. 0.6212

### 9.4 Exercises

9. Stratified sampling because the population is split into sub-populations (hourly vs. salaried) and then participants are randomly selected. If the randomly chosen employees are not required to respond, this could also be a convenience sample or a voluntary sample. Possible deficiencies in the procedure are that there are 3 times as many hourly employees than salaried employees, but they represent only half the sample. Further, the number of employees that have children may affect the perceived desirability of certain benefits like childcare. Answers may vary.

11. a. All stocks traded on the New York Stock Exchange  
b. i. Simple random sampling: develop sampling frame, choose participants randomly from sampling frame;  
ii. Cluster sampling: create clusters, randomly select clusters, survey all members of the chosen clusters;  
iii. Stratified sampling: create strata, randomly select participants from each stratum such that the population characteristics are adequately represented. Answers may vary.

## Chapter 9 Additional Exercises

1. a. Voluntary sampling  
b. All Americans do not watch the news program. It is likely that only those with strong opinions responded.  
c. No. Answers may vary.
3. 0
5. a. Convenience sample  
b. Pre-med majors may be over-represented since there are more pre-med majors in a biology class than in most other classes. Answers may vary.  
c. No, because the sample is biased.
7. The sample mean is approximately normally distributed with  $\mu_{\bar{x}} = 18$  and  $\sigma_{\bar{x}} = 0.6532$ .
9. a. 0.0062  
b. The population is infinite.
11. a.  $\mu_{\hat{p}} = 0.60$ ,  
 $\sigma_{\hat{p}} = 0.0490$   
b. 0.9586
13. 0.0045
15. 0.0571
17. a. 0.0618  
b. 0.7324  
c. 0.2676

# Chapter 10

## 10.1 Exercises

13. The sample mean is the best point estimate:

$$\bar{x} = 8.8 \text{ credit hours}$$

15. The sample proportion is the best point estimate:

$$\hat{p} = \frac{32}{250} = 0.128$$

17. The sample variance is the best point estimate:

$$s^2 = 2.4 \text{ centimeters}^2$$

- b. We are 90% confident that the true average price of a regular room with a king-size bed in the resort community is between \$146.36 and \$173.64. We are assuming the prices are normally distributed.

19.  $n = 35$

## 10.2 Exercises

11. a. 2.33

b. 1.88

c. 1.75

d. 2.05

e. 1.555

f. 1.44

13. (157, 179)

15. (2.973, 3.027)

## 10.3 Exercises

9. 2.160

11. a. 2.145

b. 2.861

c. 1.895

d. 2.201

e. 2.898

f. 1.721

13. (77.3579, 83.3887)

15 a. (5.6, 7.2)

- b. We are 95% confident that the true average length of stay for the hospital's abdominal surgery patients is between 5.6 days and 7.2 days. We are assuming that the lengths of stay are approximately normally distributed.

17. a. (146.36, 173.64)

## 10.4 Exercises

5. a. 65%

b. (62.52, 67.48)

7. a. (0.2176, 0.3324)

- b. No, 0.40 falls outside the confidence interval.

9. a. (0.4796, 0.6954)

- b. No, 0.64 falls inside the confidence interval.

11.  $n = 944$

13.  $n = 41$

## 10.5 Exercises

7. a.  $df = 17$ ,  $\chi^2 = 30.191$

b.  $df = 40$ ,  $\chi^2 = 55.759$

9. a. (0.0079, 0.0364); We are 99% confident that the variance of the amount of drug in each tablet is between 0.0079 mg and 0.0364 mg.

- b. The amounts of drug in the tablets have an approximately normal distribution.

11. a. (4.8519, 12.1330); We are 98% confident that the true population standard deviation of the life of the automobile batteries is between 4.85 and 12.13 months.

- b. The life of batteries produced by the automobile manufacturer has an approximately normal distribution.

## Chapter 10 Additional Exercises

1. a.  $\bar{x} = 56.1333$ ,  $s = 2.5317$

- b. (54.7311, 57.5355)  
 c. The heights are normally distributed.
3. (0.8453, 0.8947)
5. (0.4982, 0.6018)
7. a. (25,697, 38,303); We are 99% confident that the true mean valuation of the customer accounts is between \$25,697 and \$38,303.  
 b. The distribution of the account valuations is approximately normal.
9.  $n = 610$
11. (0.3965, 0.4835)

## Chapter 11

### 11.1 Exercises

17.  $H_0: \mu = 450$   
 $H_a: \mu > 450$

19.  $H_0: \mu = 7.9$   
 $H_a: \mu < 7.9$

21.  $H_0: \mu = 7.6$   
 $H_a: \mu > 7.6$

23.  $H_0: p = 0.05$   
 $H_a: p > 0.05$

25.  $H_0: \mu = 2784$   
 $H_a: \mu < 2784$

27. a.  $H_0: \mu = 0.5$   
 $H_a: \mu \neq 0.5$

Consequences of a Type I error: the company will work on fixing the machines when they are working correctly.

Consequences of a Type II error: the company will not fix the machines, even though they are not working correctly.

b.  $H_0: \mu = 3$   
 $H_a: \mu > 3$

Consequences of a Type I error: the company will not do research to make the flares last longer, even though they last less than three hours.

Consequences of a Type II error: the company will do research to make the flares last longer, even though it is not necessary.

c.  $H_0: \mu = \text{advertised mileage on one battery charge}$   
 $H_a: \mu < \text{advertised mileage on one battery charge}$

Consequences of a Type I error: the car manufacturer may invest time and money that isn't necessary to improve the mileage the EV gets on a single battery charge.

Consequences of a Type II error: the car manufacturer may end up with upset customers and a damaged reputation as it becomes apparent that the vehicles do not get the mileage on a single battery charge. It could put customers in danger of being stranded as well.

### 11.2 Exercises

7. a. Fail to reject  $H_0$   
 b. Reject  $H_0$   
 c. Reject  $H_0$   
 d. Fail to reject  $H_0$

9. a.  $P$ -value  $\approx 0.0228$ , Fail to reject  $H_0$ .  
 b.  $P$ -value  $\approx 0.0071$ , Reject  $H_0$ .  
 c.  $P$ -value  $\approx 0.0070$ , Reject  $H_0$ .

11. a. Insured homes in southeast Florida that were damaged by hurricane Ian.  
 b.  $\mu$ , the average size of a claim  
 c.  $H_0: \mu = 24,000$   
 $H_a: \mu > 24,000$

d. Critical Value: 2.33

Test Statistic:  $z = 7.64$

$P$ -value  $\approx 0$ , Reject  $H_0$ .

e. Yes. We should reject the null hypothesis; there is sufficient evidence at the 0.01 level of significance that the average claim size for homes damaged by hurricane Ian is greater than the historical average.

13. Yes.

$H_0: \mu = 0.2$

$H_a: \mu < 0.2$

Critical Value:  $-1.645$

Test Statistic:  $z = -11.18$

$P$ -value  $\approx 0$

Reject  $H_0$ . There is sufficient evidence to conclude that the growth rate of the shrub is less than 0.2 cm per day.

15. a.  $H_0: \mu = 147$

$H_a: \mu < 147$

Critical Value:  $-2.33$

Test Statistic:  $z = -1.46$

$P$ -value  $\approx 0.0721$  (tables)

$P$ -value  $\approx 0.0722$  (technology)

Fail to reject  $H_0$ .

b. The times that Jennifer spends on social media each day have an approximately normal distribution.

9.  $H_0: \mu = 0.0647$ ,  $H_a: \mu > 0.0647$ ,

Critical value = 1.680,

$t = 1.451$ ,  $0.05 < P$ -value  $< 0.10$  (table),

$P$ -value  $\approx 0.0769$  (technology), Fail to reject  $H_0$ .

11. No.  $H_0: \mu = 5$ ,  $H_a: \mu > 5$ , Critical value = 1.29,

$t = 0.471$ ,  $P$ -value  $\approx 0.3189$ , Fail to reject  $H_0$ .

13. a. NarStor hard drives

b. Time until failure

c. Ratio

d. No.  $H_0: \mu = 14400$ ,  $H_a: \mu < 14400$ , Critical value =  $-2.602$ ,  $t = -1.727$ ,  $P$ -value  $\approx 0.0524$ , Fail to reject  $H_0$ .

e. We must assume that the time until failure of the NarStor hard drives has an approximately normal distribution.

15. a. No.  $H_0: \mu = 600$ ,  $H_a: \mu \neq 600$ ,

Critical value = 1.753,  $t = 1.6$ ,

$P$ -value  $\approx 0.1304$ , Fail to reject  $H_0$ .

b. We assumed that the maximum horsepower for the engines have an approximately normal distribution.

## 11.4 Exercises

5. (20.047, 26.228); Reject  $H_0$

7. a. (40.35, 44.65); Reject  $H_0$

b. Heights of children are approximately normally distributed.

9. a. (1120.82, 1213.86)

b. No

11.  $H_0: \mu = 200$

$H_a: \mu > 200$

Critical Value: 1.645

Test Statistic:  $z = 2.11$

$P$ -value  $\approx 0.0174$  (table)

$P$ -value  $\approx 0.0175$  (technology), Reject  $H_0$ .

The test is statistically significant since the  $P$ -value is less than 0.05, yet a 4 unit increase in cholesterol does not seem to be a practical difference from 200 mg/dl. Answers may vary.

## 11.3 Exercises

5. a. 6 year old vegetarian children

b. Yes.  $H_0: \mu = 45.75$ ,  $H_a: \mu \neq 45.75$ .

Critical values =  $-1.96$ ,  $1.96$ ,  $z = -2.963$ ,

$P$ -value  $\approx 0.0030$ , Reject  $H_0$ .

c. We assumed that the heights of 6 year old vegetarian children have an approximately normal distribution.

7. a.  $P$ -value  $\approx 0.0750$ , Fail to reject  $H_0$ .

b.  $P$ -value  $\approx 0.0068$ , Reject  $H_0$ .

c.  $P$ -value  $\approx 0.0744$ , Fail to reject  $H_0$ .

13.  $H_0: \mu = 45$

$H_a: \mu > 45$

Critical Value: 1.282

Test Statistic:  $t = 13.333$  $P$ -value  $< 0.005$  (table) $P$ -value  $\approx 0$  (technology), Reject  $H_0$ .

There is sufficient evidence at the 0.10 level of significance that the waiting time to see an IRS representative is greater than 45 minutes. The test is also practically significant since we are dealing with customer complaints and based on the sample, customers are having to wait 10 more minutes longer to see an IRS representative. Answers may vary.

Even though Jennifer's average time spent on social media seems less than the average, the large standard deviation results in the test not being significant. Fifteen minutes less time on social media when the standard deviation is 35 minutes is not practically significant. Answers may vary.

## 11.5 Exercises

5. a.  $z = -1.645$

b.  $z = 2.33$

c.  $z = 1.645$  and  $-1.645$

7. Yes.

$H_0: p = 0.68$

$H_a: p > 0.68$

Critical value = 2.33,  $z = 4.08$ , $P$ -value  $\approx 0$ , Reject  $H_0$ .

9. a.  $H_0: p = 0.2$

$H_a: p < 0.2$

Critical Value:  $-1.645$ ,Test Statistic:  $z = -2.511$ ,  $P$ -value  $\approx 0.0060$ 

Reject  $H_0$ . There is sufficient evidence to conclude that the survey return rate is less than 20% at a 0.05 level of significance

b. The problem doesn't state how long survey recipients had to fill out the survey. Answers may vary.

11. a. No.

$H_0: p = 0.20$

$H_a: p < 0.20$

Critical value =  $-2.33$ ,  $z = -1.77$ , $P$ -value  $\approx 0.0384$ , Fail to reject  $H_0$ .

b. Yes,  $np_0 \geq 10$  and  $n(1 - p_0) \geq 10$ , but  $np_0 = 10$ , which could be a concern. Answers may vary.

c. Yes; perhaps a larger sample size should be taken before a new hypothesis is constructed. Answers may vary.

13. No,

$H_0: p = 0.70$

$H_a: p > 0.70$

Critical value = 1.645,  $z = -0.62$ , $P$ -value  $\approx 0.7315$ , Fail to reject  $H_0$ .

15. No.

$H_0: p = 0.50$

$H_a: p > 0.50$

Critical value = 2.33,  $z = 0.20$ , $P$ -value  $\approx 0.4205$ , Fail to reject  $H_0$ .

17. a.  $H_0: p = 0.80$

$H_a: p > 0.80$

 $z = 0.62$ ,  $P$ -value  $\approx 0.2676$ 

b. No

19. a.  $H_0: p = 0.50$

$H_a: p > 0.50$

 $z = 2.26$ ,  $P$ -value  $\approx 0.0119$ 

b. Yes

21. a.  $H_0: p = 0.44$

$H_a: p < 0.44$

Critical Value:  $-1.645$ Test Statistic:  $z = -1.425$  $P$ -value  $\approx 0.0771$ 

Reject  $H_0$ . There is not enough evidence to conclude that the proportion of households with cable television is lower in 2024 than in 2021 at a 0.05 level of significance.

b.  $x = 76$

## 11.6 Exercises

7. a.  $df=19, \chi^2 = 36.191$   
 b.  $df = 23, \chi^2 = 35.172$   
 c.  $df = 4, \chi^2 = 14.860$
9. a. Yes.  
 $H_0: \sigma^2 = 0.0025$   
 $H_a: \sigma^2 > 0.0025$   
 Critical value = 42.557,  $\chi^2 = 56.84$ ,  
 $P$ -value  $\approx 0.0015$ , Reject  $H_0$ .
- b. The diameters of the bolts have an approximately normal distribution.
11.  $H_0: \sigma = 3.8$   
 $H_a: \sigma > 3.8$   
 Critical Value: 49.588  
 Test Statistic:  $\chi^2 = 38.881$   
 $P$ -value  $\approx 0.104$   
 Fail to reject  $H_0$ . There is not enough evidence to conclude that the standard deviation of cooler temperatures is greater than 3.8 degrees Fahrenheit at a 0.01 level of significance.

## Chapter 11 Additional Exercises

1. a.  $H_0: \mu = 62$   
 $H_a: \mu \neq 62$   
 b. The company believes that the average time to replace a set of 4 tires has changed when in fact the average time is unchanged.  
 c. The company believes that the average time to replace a set of 4 tires remains unchanged when in fact the average time has changed.
3. Yes.  
 $H_0: \mu = 3.5$   
 $H_a: \mu < 3.5$   
 Critical value =  $-1.299$ ,  $t = -4.34$ ,  
 $P$ -value  $\approx 0$ , Reject  $H_0$ .
5. Yes.
- $H_0: \mu = 13.20$   
 $H_a: \mu < 13.20$   
 Critical value =  $-2.33$ ,  $z = -2.53$ ,  
 $P$ -value  $\approx 0.0057$ , Reject  $H_0$ .
7. No.  
 $H_0: \mu = 895$   
 $H_a: \mu > 895$   
 Critical value = 1.28,  $z = 1.19$ ,  
 $P$ -value  $\approx 0.1770$ , Fail to reject  $H_0$ .
9. a. No.  
 $H_0: p = 0.003$   
 $H_a: p < 0.003$   
 Critical value =  $-1.645$ ,  $z = -1.42$ ,  
 Fail to reject  $H_0$ .
- b. 0.0778  
 c. Yes,  $H_0$  would be rejected.
11. No.  
 $H_0: p = 0.2632$   
 $H_a: p > 0.2632$   
 Critical value = 1.645,  $z = 0.90$ ,  
 $P$ -value  $\approx 0.1841$ , Fail to reject  $H_0$ .
13. a.  $P$ -value  $\approx 0.0062$ , Reject  $H_0$ .  
 b.  $P$ -value  $\approx 0.0256$ , Fail to reject  $H_0$ .  
 c.  $P$ -value  $\approx 0.0002$ , Reject  $H_0$ .
15. a.  $P$ -value  $\approx 0.0464$ , Reject  $H_0$ .  
 b.  $P$ -value  $\approx 0.0050$ , Reject  $H_0$ .  
 c.  $P$ -value  $\approx 0.0510$ , Fail to reject  $H_0$ .
17. Yes.  
 $H_0: \sigma^2 = 0.00156$   
 $H_a: \sigma^2 > 0.00156$   
 Critical value = 118.498,  $\chi^2 = 134.615$ ,  
 $P$ -value  $\approx 0.0120$ , Reject  $H_0$ .
19. a. No.  
 $H_0: p = 0.90$   
 $H_a: p > 0.90$   
 Critical value = 1.28,  $z = 1.25$ , Fail to reject  $H_0$ .
- b. 0.1056

21. No.

$$H_0: \sigma^2 = 0.01$$

$$H_a: \sigma^2 < 0.01$$

Critical value = 3.325,  $\chi^2 = 1.44$ ,  
 $P$ -value  $\approx 0.0024$ , Reject  $H_0$ .

## Chapter 12

### 12.1 Exercises

7.  $(-1.2461, 2.5139)$
9. a.  $(-7.0741, -3.3259)$  (table),  
 $(-7.074, -3.326)$  (technology); we are 95% confident that the true mean difference in acceleration times between the Dodge and Volkswagen minivans is between 3.326 and 7.074 seconds, where the Volkswagen takes this many seconds longer to reach 60 mph than the Dodge. Answers may vary.
- b. Yes; since the confidence interval does not include zero, the data suggest that the difference in acceleration times is statistically significant at the 0.05 level.
- c. The samples are independent random samples, both populations are approximately normally distributed, and the population standard deviations are known.

11.  $H_0: \mu_1 - \mu_2 = 0$

$$H_a: \mu_1 - \mu_2 < 0$$

Critical Value:  $-1.28$ , Test Statistic:  $z = -2.37$  $P$ -value  $\approx 0.0089$  (table) $P$ -value  $\approx 0.0088$  (technology)

Reject  $H_0$ . There is enough evidence at the 0.10 level of significance to support the claim that the braking distance of tires made with compound A is less than that of tires made with compound B.

13.  $H_0: \mu_1 - \mu_2 = 0$

$$H_a: \mu_1 - \mu_2 > 0$$

Critical Value:  $1.28$ , Test Statistic:  $z = 1.34$  $P$ -value  $\approx 0.0901$  (table) $P$ -value  $\approx 0.0908$  (technology)

Reject  $H_0$ . There is enough evidence at the 0.10 level of significance to support the claim that the test preparation course improved students' SAT Math scores.

### 12.2 Exercises

5. a.  $t = -1.774$ ,  $df = 78$

b.  $t = 1.28$ ,  $df = 78$

c.  $t = 2.375$  and  $-2.375$ ,  $df = 78$

7. a.  $(-7.566, -2.436)$  We are 95% confident that Mrs. Ellis' expenses are between \$2.44 and \$7.57 less than Mr. Ford's.

b. Yes.  $H_0: \mu_1 - \mu_2 = 0$ ,  $H_a: \mu_1 - \mu_2 \neq 0$ ,  
 Critical values =  $-2.015, 2.015$ ,  $t = -3.926$ ,  
 $P$ -value  $\approx 0.0002$ , Reject  $H_0$ .

c. The confidence interval only contains negative values indicating that with 95% confidence the expenses for Mrs. Ellis will always be less than those of Mr. Ford.

9. a. No.  $H_0: \mu_1 - \mu_2 = 0$ ,  $H_a: \mu_1 - \mu_2 < 0$ ,  
 Critical value =  $-1.294$ ,  $t = -1.2834$ ,  
 Fail to reject  $H_0$ .b.  $P$ -value  $\approx 0.1019$ c. No, we would fail to reject  $H_0$  at  $\alpha = 0.05$ .

11. a.  $df = 23$ ,  $t = -1.714$

b.  $df = 18$ ,  $t = 1.330$

c.  $df = 10$ ,  $t = 3.169$  and  $-3.169$

13. a.  $(-1.956, -0.045)$ , We are 90% confident that the fat content in Fiber Oat Flakes is between 1.956 and 0.045 grams less than the fat content in Bran Flakes Plus.

b. Yes.

$$H_0: \mu_1 - \mu_2 = 0,$$

$$H_a: \mu_1 - \mu_2 < 0,$$

Critical value =  $-1.311$ ,  $t = -1.778$ , $P$ -value  $\approx 0.0429$ , Reject  $H_0$ .

c. The samples are independent random samples, both populations are approximately

normally distributed, and the variances are unknown but assumed equal.

15. a.  $(-3.467, 63.467)$ , We are 95% confident that dramas generate between \$3.5 million less and \$63.5 million more in box office revenues than comedies. However, since the confidence interval contains zero we cannot say that the revenues from comedies and dramas are significantly different.
- b. No.  
 $H_0: \mu_1 - \mu_2 = 0$ ,  
 $H_a: \mu_1 - \mu_2 > 0$ ,  
 Critical value = 2.681,  $t = 1.953$ ,  
 Fail to reject  $H_0$ .
- c.  $P$ -value = 0.0373
- d. The samples are independent random samples, both populations are approximately normally distributed, and the variances are unknown and assumed not equal.

17. a. The samples are independent random samples, both populations are approximately normally distributed, and the variances are unknown but assumed equal.
- b. Yes.  
 $H_0: \mu_1 - \mu_2 = 0$ ,  
 $H_a: \mu_1 - \mu_2 > 0$ ,  
 Critical value = 2.500,  $t = 2.703$ ,  
 $P$ -value  $\approx 0.0063$ , Reject  $H_0$ .
- c.  $H_0: \mu_1 - \mu_2 = 0$ ,  
 $H_a: \mu_1 - \mu_2 > 0$ ,  
 Critical value = 2.821,  $t = 2.778$ ,  
 $P$ -value  $\approx 0.0107$ , Fail to reject  $H_0$ .
- d.  $H_0$  is rejected in part b. but not in part c. if the variance assumption changes. Answers may vary.

## 12.3 Exercises

7. a.  $df = 11$ ,  $t = -3.106$   
 b.  $df = 4$ ,  $t = 2.776$   
 c.  $df = 24$ ,  $t = 1.711$  and  $-1.711$
9. a. Yes. The same cashier is using the old scanner and the new scanner so the samples can be paired. Answers may vary.

- b. The differences have an approximately normal distribution.
- c. Answers may vary.
- d.  $(-4.15, 0.43)$ ; With 95% confidence, cashiers using the old scanner process between 4.15 fewer and 0.43 more items than using the new scanner. However, since the confidence interval contains zero we cannot say that the number of grocery items processed between the old scanner and the new scanner are significantly different.
- e. Yes.  
 $H_0: \mu_d = 0$ ,  
 $H_a: \mu_d < 0$ ,  
 Critical value =  $-1.943$ ,  $t = -1.983$ , Reject  $H_0$ .  $P$ -value  $\approx 0.0473$ , which is less than  $\alpha = 0.05$ , Reject  $H_0$ .

11.  $(-0.1002, 0.4752)$

## 12.4 Exercises

7. a.  $z = -2.33$   
 b.  $z = 1.645$   
 c.  $z = 1.645$  and  $-1.645$
9. a. Yes, the sample sizes are sufficiently large.  
 $H_0: p_1 - p_2 = 0$ ,  
 $H_a: p_1 - p_2 < 0$ ,  
 Critical value =  $-1.28$ ,  $z = -1.03$ , Fail to reject  $H_0$ . There is not sufficient evidence to support the fund-raiser's theory.
- b.  $P$ -value  $\approx 0.1519$ ; this is the probability of getting the results obtained assuming the null hypothesis is true and the assumptions of the test were met.
- c.  $(-0.0888, 0.0278)$ ; We are 95% confident that true the proportion of those individuals between 25–40 years of age who answered “Yes” to donating to a worthy cause is between about 8.8% less than and 2.8% higher than the proportion of those over 40 who answered “Yes.” However, since the confidence interval contains zero, we cannot say that the proportions who answered “Yes” to donating to a worthy cause are significantly different.

11. Yes, the sample sizes are sufficiently large.

$$H_0: p_1 - p_2 = 0,$$

$$H_a: p_1 - p_2 > 0,$$

$$\text{Critical value} = 1.645, z = 0.18,$$

Fail to reject  $H_0$ .  $P$ -value  $\approx 0.4275$ , which is greater than  $\alpha = 0.05$ , Fail to reject  $H_0$ . The manufacturer will choose Supplier A.

d.  $F_{0.10,14,16} = 1.9532$

$$P\text{-value} \approx 0.2543 \text{ (technology)}$$

Fail to reject  $H_0$ ; there is not enough evidence at the 0.10 level of significance to support the inspector's claim that the variance in the diameters of soda cans is greater for soda cans produced by Machine A than for soda cans produced by Machine B.

## 12.5 Exercises

5. a.  $F_{0.100,7,22} = 2.0084$

b.  $F_{0.050,7,22} = 2.4638$

c.  $F_{0.025,7,22} = 2.9338$

d.  $F_{0.010,7,22} = 3.5867$

7. a.  $F_{0.025,1,25} = 5.6864$

b.  $F_{0.025,6,11} = 3.8807$

c.  $F_{0.025,8,40} = 2.5289$

d.  $F_{0.025,3,18} = 3.9539$

9. a.  $H_0: \sigma_1^2 = \sigma_2^2$

$$H_a: \sigma_1^2 < \sigma_2^2$$

b.  $H_0: \sigma_1^2 = \sigma_2^2$

$$H_a: \sigma_1^2 > \sigma_2^2$$

11. a. Critical Value:  $F_{0.95,13,10} = 0.3744$

$$\text{Test Statistic: } F = 0.3389$$

Conclusion:  $F < F_{0.95,13,10}$ , so we reject the null hypothesis that the variances are equal.

b. Critical Value:  $F_{0.01,11,25} = 3.0558$

$$\text{Test Statistic: } F = 1.7758$$

Conclusion:  $F < F_{0.01,11,25}$ , so we fail to reject the null hypothesis that the variances are equal.

c. Critical Values:  $F_{0.975,19,28} = 0.4148$ ,

$$F_{0.025,19,28} = 2.2505$$

$$\text{Test Statistic: } F = 2.301$$

Conclusion:  $F > F_{0.025,19,28}$ , so we reject the null hypothesis that the variances are equal.

13. a.  $H_0: \sigma_1^2 = \sigma_2^2$

$$H_a: \sigma_1^2 > \sigma_2^2$$

b.  $F$ -distribution,  $\alpha = 0.10$

c.  $F = 1.4067$

15. a.  $H_0: \sigma_1^2 = \sigma_2^2$

$$H_a: \sigma_1^2 \neq \sigma_2^2$$

b.  $F$ -distribution,  $\alpha = 0.01$

c.  $F = 0.9571$

d.  $F_{0.005,11,12} = 4.9884$ ,  $F_{0.995,11,12} = 0.1910$

$$P\text{-value} \approx 0.9490 \text{ (technology)}$$

Fail to reject  $H_0$ ; there is not sufficient evidence at the 0.01 level of significance to support the coach's belief that the variance in heights of adult male basketball players is different than that of the general population of men.

## Chapter 12 Additional Exercises

1. Yes.

$$H_0: \mu_1 - \mu_2 = 0,$$

$$H_a: \mu_1 - \mu_2 < 0,$$

$$\text{Critical value} = -1.669, t = -2.543,$$

$$P\text{-value} \approx 0.0067, \text{ Reject } H_0.$$

3. a. Yes. The cholesterol levels are measured for the same person before and after the diet so the samples can be paired. Answers may vary.

b. The differences have an approximately normal distribution.

c. Answers may vary.

- d. No.

$$H_0: \mu_d = 0,$$

$$H_a: \mu_d > 0,$$

$$\text{Critical value} = 3.143, t = 1.769,$$

$$P\text{-value} \approx 0.0637, \text{ Fail to reject } H_0.$$

5. a.  $H_0: \mu_1 - \mu_2 = 0$

$$H_a: \mu_1 - \mu_2 \neq 0$$

Critical Values:  $\pm 1.980$  (using 120  $df$  from table)

Critical Values:  $\pm 1.978$  (using 135  $df$  from technology)

Test Statistic:  $t = 1.819$

$P$ -value  $\approx 0.0711$  (technology)

Fail to reject  $H_0$ . There is not enough evidence at the 0.05 level of significance to support a difference in speed between wide receivers and cornerbacks.

- b.  $(-0.0026, 0.0626)$  formula,  $(-0.0026, 0.0626)$  technology; we expect the speeds of wide receivers to be between 0.002 seconds slower to 0.06 seconds faster than cornerbacks. However, since the confidence interval contains zero we cannot say that the speeds are significantly different.

7. a. No.

$$H_0: \mu_1 - \mu_2 = 0,$$

$$H_a: \mu_1 - \mu_2 \neq 0,$$

Critical values =  $-1.96, 1.96, z = -0.48,$

Fail to reject  $H_0$ .

- b.  $P$ -value = 0.6328

9.  $H_0: \mu_1 - \mu_2 = 0,$

$$H_a: \mu_1 - \mu_2 \neq 0,$$

Critical values =  $-2.101, 2.101, t = 8.771,$

$P$ -value is approximately 0, Reject  $H_0$ . There is sufficient evidence of a difference in the average weights.

11. a. Yes. Answers may vary.

b. We must assume that the differences have an approximately normal distribution.

c. Answers may vary.

$$d. H_0: \mu_d = 0$$

$$H_a: \mu_d < 0$$

Critical value =  $-3.747; t = -2.75,$

$P$ -value  $\approx 0.0258$

Reject  $H_0$ . At the 0.05 level of significance, there is sufficient evidence to conclude that the Vest method is more effective in increasing the diameter of the blood vessels, therefore the hospital will recommend the method to its patients.

13.  $H_0: \sigma_1^2 = \sigma_2^2$

$$H_a: \sigma_1^2 < \sigma_2^2$$

$$F = 0.5888$$

$$F_{0.95,9,9} = 0.3146$$

$P$ -value  $\approx 0.2211$  (technology)

Fail to reject  $H_0$ ; there is not sufficient evidence at the 0.05 level of significance to support Shirley's belief that the variance in expenditures when her family eats out is less when she uses cash than when she uses a credit card.

## Chapter 13

### 13.1 Exercises

5. a. Independent: *Number of Absences*

Dependent: *Grade*

$$y_i = \beta_0 + \beta_1 x_i$$

- b.  $\text{Estimated Grade} = 5.427885 - 0.44135$

*Number of Absences*

- c.  $\text{MSE} \approx 0.108635$

7. a. *Estimated Birth Rate*

$$= 61.99724 - 0.50156 \cdot \text{Female Literacy Rate}$$

- b.  $R^2 \approx 0.612148$ . The estimated model explains about 61% of the variation in birth rates.

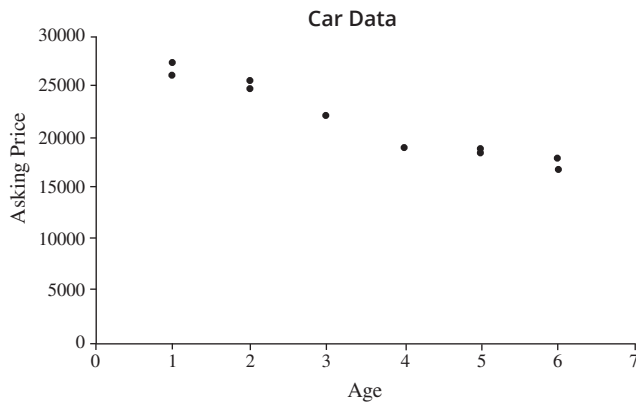
- c. 26.89 births per 1,000

- d. 16.86 births per 1,000

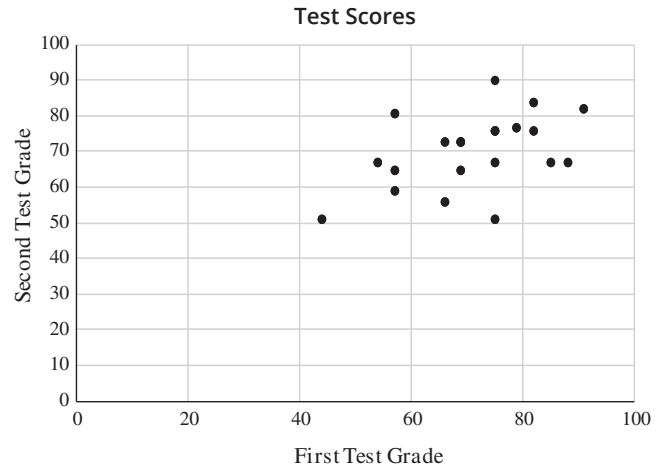
- e. As female literacy rates increase, birth rates decrease. Yes, this makes sense as a higher literacy rate implies a larger chance of a female working or pursuing further education, which may result in less time to have children or less desire to have a large family. Answers may vary.

### 13.2 Exercises

11. a. There appears to be a strong negative linear relationship between age and asking price.



- b.  $\hat{y} = 28525.504 - 1948.087x$   
 c. 152.6682  
 d. (-2460.3480, -1435.8260)  
 e. We are 99% confident that the true decrease in asking price for a car for each additional year is between \$1435.83 and \$2460.35.
13. a.  $\hat{y} = -2.1324 + 1.5643x$   
 b. Yes,  $H_0: \beta_1 = 0$ ,  $H_a: \beta_1 \neq 0$ ,  $t = 2.635$ ,  $P\text{-value} \approx 0.0388$ , Reject  $H_0$ .  
 c. Yes,  $P\text{-value} \approx 0.0388$  is greater than  $\alpha$  at the 0.01 level, so there is not sufficient evidence of a linear relationship between age and salary.  
 d. Approximately 0.5365 or 53.65%
15. a. There appears to be a weak positive linear relationship between the two test grades.



- b.  $\hat{y} = 42.5154 + 0.3914x$   
 c. Approximately 0.2076 or 20.76%  
 d. Yes,  $H_0: \beta_1 = 0$ ,  $H_a: \beta_1 \neq 0$ ,  $t = 2.231$ ,  $P\text{-value} \approx 0.0379$ , Reject  $H_0$ .  
 e. 72
17. a.  $\hat{y} = -0.00936 + 0.38128x$   
 b. For each additional 1% increase in *Adult.obesity.percent*, *Diabetes.percent* increases by approximately 0.38128%.  
 c. (0.36834, 0.39423), We are 95% confident that the increase in *Diabetes.percent* for each additional *Adult.obesity.percent* is between 0.36834% and 0.39423%.  
 d. Yes,  $H_0: \beta_1 = 0$ ,  $H_a: \beta_1 \neq 0$ ,  $t = 57.741$ ,  $P\text{-value} \approx 0$ , Reject  $H_0$ .  
 e. Approximately 0.515 or 51.5%

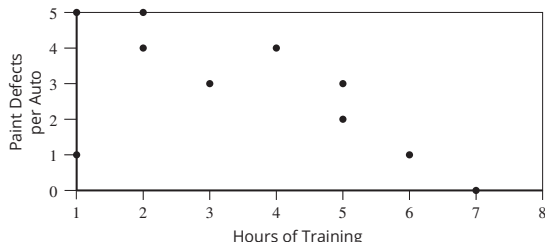
### 13.3 Exercises

5. a.  $s_e = 372.478$ . This is the estimated standard deviation of the errors associated with the model.  
 b. For each additional one million dollar increase in net sales, net income is expected to increase by approximately 18.4 million dollars.  
 c. Approximately 0.407 or 40.7%  
 d. Yes,  $H_0: \beta_1 = 0$ ,  $H_a: \beta_1 \neq 0$ ,  $t = 4.15$ ,  $P\text{-value} \approx 0$ , Reject  $H_0$ .  
 e. (9.2752, 27.5928); We are 95% confident that the true increase in net income for a

one million dollar increase in net sales is between 9.2752 and 27.5928 million dollars.

- f. \$1005.3 million
- g. (702.4, 1308.2); We are 95% confident that the average net income when net sales is equal to \$50 million is between \$702.4 million and \$1308.2 million.
- h. (180.5, 1830.0); We are 95% confident that when net sales is equal to \$50 million, net income is between \$180.5 million and \$1308.2 million.
- i. \$1189.64 million

7. a. There appears to be a weak negative linear relationship between hours of training and the number of paint defects per auto produced.

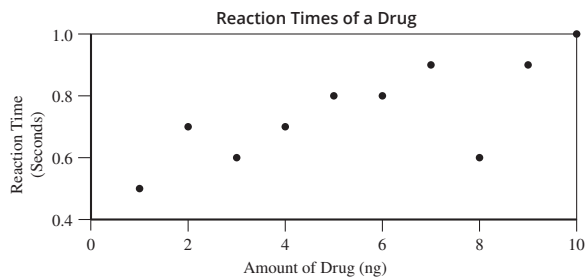


- b.  $s_e = 1.4531$ . This is the estimated standard deviation of the errors associated with the model.
- c. With each additional hour of training, the number of paint defects per auto is expected to decrease by approximately 0.515.
- d. Approximately 0.388 or 38.8%. Other factors include years of experience, size or model of auto, etc. Answers may vary.
- e.  $H_0: \beta_1 = 0, H_a: \beta_1 \neq 0, t = -2.25, P\text{-value} = 0.054$ , Fail to reject  $H_0$  at the 0.05 level, Reject  $H_0$  at the 0.10 level.
- f. (-1.0420, 0.0123); We are 95% confident that the true change in the number of paint defects per auto for each additional hour of training is between -1.0420 and 0.0123.
- g. 2.594 paint defects per auto
- h. (1.514, 3.674); We are 95% confident that the average number of paint defects per auto for an employee with 4 hours of training is between 1.514 and 3.674 defects.

- i. (-0.927, 6.115); We are 95% confident that a new employee with 4 hours of training will have between -0.927 and 6.115 defects. Since the number of defects cannot be negative, we can say that we are 95% confident that the number of defects for a new employee with 4 hours of training will be between 0 and 6.
- j. 1 paint defect per auto

## Chapter 13 Additional Exercises

1. a. There appears to be a moderate positive linear relationship between amount of drug and reaction time.



- b.  $s_e = 0.1101$ . This is the estimated standard deviation of the errors associated with the model.
- c. Reaction time increases by approximately 0.0394 seconds for each additional nanogram of the drug.
- d. Approximately 0.569 or 56.9%. Other factors could include age, weight, etc. Answers may vary.
- e.  $H_0: \beta_1 = 0, H_a: \beta_1 \neq 0, t = 3.25, P\text{-value} = 0.012$ , Reject  $H_0$  at the 0.05 level, Fail to reject  $H_0$  at the 0.01 level.
- f. (0.011, 0.067); We are 95% confident that the true change in reaction time for each additional nanogram of the drug is between 0.011 and 0.067 seconds.
- g. 0.6909 seconds
- h. (0.600, 0.782); We are 95% confident that the average reaction time for an individual with 4 ng of the drug in the bloodstream is between 0.600 and 0.782 seconds.
- i. (0.421, 0.961)

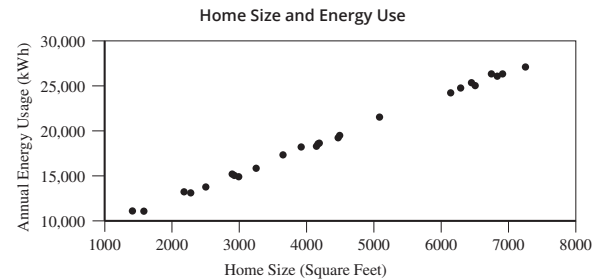
3. a. There appears to be a strong negative relationship between FICO score and interest rate.



- b.  $\hat{y} = 16.2146 - 0.0129x$
- c. 0.3202
- d. Yes.  $H_0: \beta_1 = 0$ ,  $H_a: \beta_1 \neq 0$ ,  $t = -10.281$ ,  $P\text{-value} \approx 0$ , Reject  $H_0$ .
- e. Interest rate decreases by approximately 0.0129 percent for each additional one point increase in FICO score.
- f.  $(-0.0155, -0.0102)$ ; We are 95% confident that the true change in interest rate for each additional one point increase in FICO score is between  $-0.0155$  and  $-0.0102$  percentage points.
- g.  $R^2 = 0.8545$ . Approximately 85.45% of the variation in interest rate is explained by the variation in FICO scores.
- h.  $r = -0.9244$ . There is a strong negative linear relationship between FICO score and interest rate.
- i. 6.8621%
- j.  $(6.613, 7.148)$ ; We are 95% confident that the average interest rate for a FICO score of 725 is between 6.613 and 7.148 percent.
- k.  $(5.864, 7.898)$ ; We are 95% confident that a person with a FICO score

of 725 will receive an interest rate between 5.864 and 7.898 percent.

5. a. There appears to be a strong positive linear relationship between home size and annual energy usage.



- b.  $\hat{y} = 6774.4571 + 2.8406x$
- c.  $R^2 = 0.9982$ . Approximately 99.82% of the variation in annual energy usage is explained by the variation in home size.
- d. Yes.  $H_0: \beta_1 = 0$ ,  $H_a: \beta_1 \neq 0$ ,  $t = 112.114$ ,  $P\text{-value} \approx 0$ , Reject  $H_0$ .
- e. For each additional square foot of home size, annual energy usage is expected to increase by approximately 2.8406 kWh.
- f.  $(2.7695, 2.9117)$ ; We are 99% confident that the true increase in annual energy usage for each additional square foot of home size is between 2.7695 and 2.9117 kWh.
- g.  $r = 0.9991$ . There is a strong positive linear relationship between home size and annual energy usage.
- h. 15,864.3771 kWh
- i.  $(15,572.2, 15,976.5)$ ; We are 95% confident that the James family will use on average between 15,572.2 and 15,976.5 kWh in their first year in the home.

## Chapter 14

### 14.1 Exercises

9. a.  $b_0 = 11,214.1992$ ,  $b_1 = 2854.8913$ ,  $b_2 = 839.6360$
- b.  $\hat{y} = 11,214.1992 + 2854.8913x_1 + 839.6360x_2$
- c. Age, median income for your location, industry, etc. Answers may vary.
- d. \$47,152.17

11. a.  $Weight = \beta_0 + \beta_1 Height + \beta_2 Age + \beta_3 Calories + \beta_4 Exercise + \varepsilon_i$   
 b. *Height, Age, Calories, Exercise*  
 c. Positive: *Height, Age, Calories*;  
 Negative: *Exercise*  
 d. Medical history (diabetes, thyroid issues, etc.), Smoker/non-smoker, etc. Answers may vary.
13. The coefficient became negative. This doesn't make sense as you would expect additional bedrooms to raise the selling price of the house. This variable is likely highly correlated with square footage. Answers may vary.
15. a. Yes, the coefficient is positive, meaning that as years of education increases, so does estimated salary. The magnitude of \$2854.89 for each additional year of education seems reasonable.  
 b. Yes, the coefficient is positive, meaning that as years of experience increases, so does estimated salary. The magnitude of \$839.64 for each additional year of experience seems reasonable.  
 c. For each additional year of experience, annual salary is expected to increase by \$839.64, assuming years of education remains constant.  
 d. \$38,251.52  
 e. Her annual salary would be expected to increase by \$839.64, assuming years of education remains constant.  
 f. The employee with the master's degree is expected to earn approximately \$5709.78 more than the employee with the bachelor's degree.

## 14.2 Exercises

7. a.  $\hat{y} = -69,280.13 + 142,935.73 Bedrooms + 369,879.29 Acreage$   
 b.  $SSR = 1.1312E+13$ ;  $SSE = 1.0668E+13$ ;  
 $TSS = 2.3785E+13$   
 c.  $R^2 = 0.5515$ . Approximately 55.15% of the variation in price is explained by the variation in the number of bedrooms and the lot size.

- d.  $R_a^2 = 0.5478$ . Approximately 54.78% of the variation in price is explained by the variation in the number of bedrooms and the lot size, adjusted for the number of independent variables in the model.  
 e. The value of  $R^2$  is larger than the adjusted  $R^2$  by 0.0037. The adjusted  $R^2$  statistic should be used when comparing with other models. Answers may vary.

9. a.  $\hat{y} = 74.6989 + 1.8548x$   
 b.  $R^2 = 0.7365$ . Approximately 73.65% of the variation in revenues is explained by the variation in advertising expenditures.  
 c.  $R_a^2 = 0.6926$ . No, because there is only one independent variable in the model.  
 d. Advertising expenditures appear to be significant in explaining weekly revenues. The coefficient for advertising expenditures is statistically significant ( $P$ -value = 0.0064) and the  $R^2$  value is large. However, there are many additional factors that could also affect revenue. Answers may vary.

## 14.3 Exercises

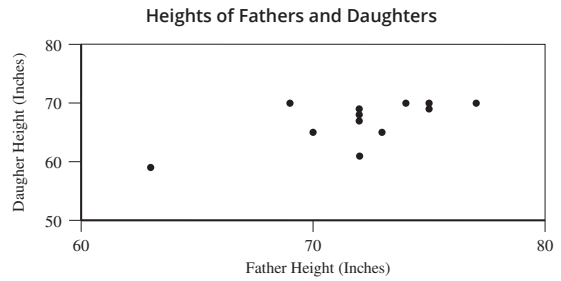
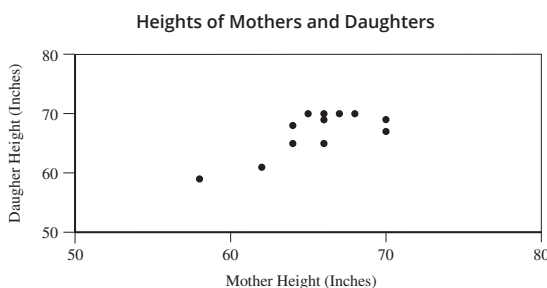
15. a.  $\hat{y} = -811.662 + 2830.705 OBP + 1517.578 SLG$   
 b. Yes.  $H_0: \beta_1 = \beta_2 = 0$ ,  $H_a$ : At least one  $\beta_i \neq 0$ ,  $F = 7556.757$ ,  $P$ -value = 0, Reject  $H_0$ .  
 c. Approximately 92.48%  
 d. OBP: Yes;  $H_0: \beta_1 = 0$ ;  $H_a$ :  $\beta_1 \neq 0$ ;  $t = 36.319$ ;  $P$ -value = 0; Reject  $H_0$ . SLG: Yes;  $\beta_2 = 0$ ;  $H_a$ :  $\beta_2 \neq 0$ ;  $t = 43.146$ ;  $P$ -value  $\approx 0$ ; Reject  $H_0$ .  
 e. No, both variables are significant predictors of runs scored (RS).
17. a. Approximately 97.54%  
 b. Yes.  $H_0: \beta_1 = \beta_2 = 0$ ,  $H_a$ : At least one  $\beta_i \neq 0$ ,  $F = 336.5465$ ,  $P$ -value  $\approx 0$ , Reject  $H_0$ .  
 c. (-0.0012, 0.0228). We are 99% confident that the true change in printing cost for a one-page increase in the number of pages is between -\$0.0012 and \$0.0228.  
 d.  $H_0: \beta_1 = 0$ ,  $H_a$ :  $\beta_1 \neq 0$ ,  $t = 2.604$ ,  $P$ -value = 0.0185, Reject  $H_0$  at the 5%

level (so yes, it is useful), Fail to reject  $H_0$  at the 1% level (so no, it is not useful).

- e.  $(-0.0211, 0.0012)$ . We are 95% confident that for each additional copy, the change in printing cost per book will be between  $-\$0.0211$  and  $\$0.0012$ .
  - f. No.  $H_0: \beta_2 = 0, H_a: \beta_2 \neq 0, t = -1.888, P\text{-value} = 0.0762$ , Fail to reject  $H_0$ . Yes, the publisher should consider removing the variable from the model because there is not sufficient evidence of a linear relationship between the number of copies printed and printing cost.
- 19.**
- a.  $\hat{y} = 284.3965 + 0.0152x_1$
  - b. Yes;  $H_0: \beta_1 = 0; H_a: \beta_1 \neq 0; F = 70.2604; P\text{-value} = 0.0000013$ ; Reject  $H_0$ .
  - c. Yes;  $H_0: \beta_1 = 0; H_a: \beta_1 \neq 0; t = 8.382; P\text{-value} = 0.0000013$ ; Reject  $H_0$ .
  - d. Approximately 84.39%. This is slightly lower than the percentage explained by income and population, 84.62%.
  - e. Though the model including both the income and population variables has a higher  $R^2$  value, the simple linear regression model is probably better because the population variable is not significant in predicting rental rates. The coefficient of determination increases as more independent variables are added to the model, but this does not necessarily mean that the model is better or more useful. Answers may vary.

## 14.4 Exercises

- 7. a.** There appears to be a positive relationship between each parent's height and the child height. The mother-daughter plot appears more linear than the father-daughter plot.



- b.  $\hat{y} = -4.6456 + 0.5939x_1 + 0.4523x_2$
  - c. Yes;  $H_0: \beta_1 = \beta_2 = 0; H_a: \text{At least one } \beta_i \neq 0; F = 11.2521; P\text{-value} = 0.0028$ ; Reject  $H_0$ .
  - d. No;  $H_0: \beta_2 = 0; H_a: \beta_2 \neq 0; t = 2.176; P\text{-value} = 0.0546$ ; Fail to reject  $H_0$ .
  - e. No;  $H_0: \beta_1 = 0; H_a: \beta_1 \neq 0; t = 2.628; P\text{-value} = 0.0253$ ; Fail to reject  $H_0$ .
  - f. For each additional inch in the mother's height, the daughter's height is expected to increase by approximately 0.5939 inch. For each additional inch in the father's height, the daughter's height is expected to increase by approximately 0.4523 inch.
  - g. Mother: (0.0903, 1.0974); We are 95% confident that for each additional inch in the mother's height, the daughter will be between 0.0903 and 1.0974 inches taller. Father: (-0.0108, 0.9153); We are 95% confident that for each additional inch in the father's height, the daughter will be between 0.0108 inch shorter and 0.9153 inch taller.
  - h. 66.8297 inches, or 5 feet 6.8297 inches
  - i. (61.533, 72.126); We are 95% confident that a particular daughter whose mother is 5 foot 4 and father is 6 foot 2 will be between 61.533 and 72.126 inches tall.
  - j. (64.785, 68.875)
- 9. a.**  $\hat{y} = -13.2883 + 0.1373x_1 + 0.0861x_2$
- b. Yes;  $H_0: \beta_1 = \beta_2 = 0; H_a: \text{At least one } \beta_i \neq 0; F = 14.2452; P\text{-value} = 0.00007$ ; Reject  $H_0$ .
  - c. Approximately 53.26%. This is slightly lower than the percentage for the model with the three independent variables (56.94%). However, the difference is likely due to the additional independent variable in the previous model.
  - d. The model without the first downs variable is likely the better model. It is a simpler model

that only includes variables that are useful in predicting points scored. Answers may vary.

- e. Approximately 23 points (23.4)
- f. (20.3945, 26.3410); We are 95% confident that the average number of points scored will be between 20 and 26 points when the Buffalo offense has 102 rushing yards and 263 passing yards.
- g. (8.2322, 38.5033); We are 95% confident that in this particular game against Miami, Buffalo will score between 8 and 39 points when the Buffalo offense has 102 rushing yards and 263 passing yards.

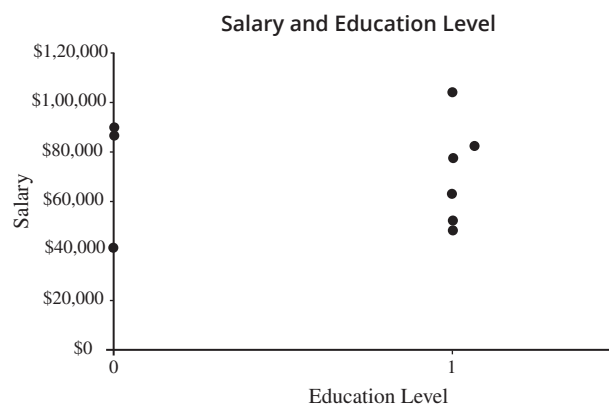
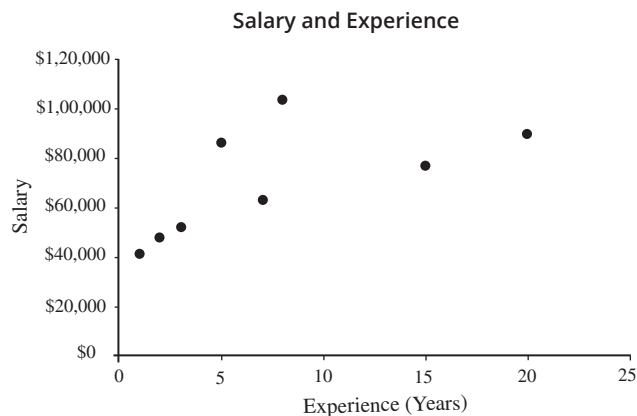
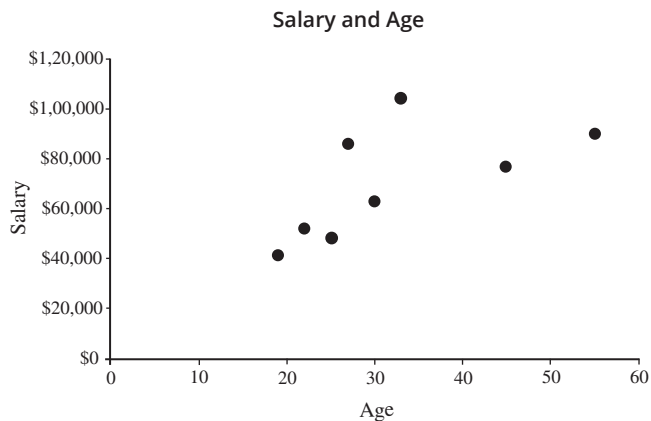
## 14.5 Exercises

- 7. a. 4 indicator variables
- b. One way to set up the model is as follows:
  - $x_1 = 1$  if very small, 0 otherwise
  - $x_2 = 1$  if small, 0 otherwise
  - $x_3 = 1$  if medium, 0 otherwise
  - $x_4 = 1$  if large, 0 otherwise
- c. For the setup in part b., the constant term represents the contribution of the state 'very large' in the model. In general, the constant term will represent the contribution of the state not defined with an indicator variable.

- 9. a. The coefficient for attendance is positive, indicating that as the number of classes attended increases, Cumulative GPA increases, which seems reasonable. The coefficient for exam score is positive, indicating that as exam score increases, Cumulative GPA increases, which also makes sense. The coefficient for freshman is negative, indicating that freshmen tend to have lower Cumulative GPAs than other classes. This may not be expected, but it is not unreasonable. Answers may vary.
- b. For each additional class attended, one would expect Cumulative GPA to increase by 0.0672 of a point.
- c. For each additional point earned on the final exam, one would expect Cumulative GPA to increase by 0.0678 of a point.

- d. Freshmen tend to have Cumulative GPAs that are, on average, 0.1436 lower than students in other classes.
- e. 0.1436

11. a. No;



- b.  $\hat{y} = 45,839.1608 + 457.3427x_1 + 1311.1888x_2 - 75.1049x_3$
- c. No;  $H_0: \beta_1 = \beta_2 = \beta_3 = 0, H_a: \text{At least one } \beta_i \neq 0, F = 0.9068, P\text{-value} = 0.5125, \text{Fail to reject } H_0.$
- d. No;  $H_0: \beta_1 = 0, H_a: \beta_1 \neq 0, t = 0.077, P\text{-value} = 0.9424, \text{Fail to reject } H_0.$

- e. No;  $H_0: \beta_2 = 0$ ,  $H_a: \beta_2 \neq 0$ ,  $t = 0.120$ ,  $P$ -value = 0.9102, Fail to reject  $H_0$ .
- f. No;  $H_0: \beta_3 = 0$ ,  $H_a: \beta_3 \neq 0$ ,  $t = -0.004$ ,  $P$ -value = 0.9967, Fail to reject  $H_0$ .
- g. For each additional year of age, salary is expected to increase by \$457.34. For each additional year of experience, salary is expected to increase by \$1311.19. With a college degree or above, salary is expected to decrease by \$75.10. The coefficient interpretation may not make sense when the model is not significant.
- h. Approximately \$74,883
- i. (131,537, 149,634); We are 95% confident that the annual salary for a particular employee who is 35 years old with a Master's degree and 10 years of experience is between \$132 and \$149,634. This interval is not particularly useful because the model is not statistically significant, and the range of salaries contained in the interval is very large.
- j. (35,912.7, 113,853); We are 95% confident that the average annual salary for an employee who is 35 years old with a PhD and 10 years of experience is between \$35,913 and \$113,853. This interval is not particularly useful because the model is not statistically significant, and the range of salaries contained in the interval is very large.

13. a.  $x_1 = 1$  if Toyota, 0 otherwise  
 $x_2 = 1$  if Honda, 0 otherwise
- b.  $\hat{y} = -3744.5301 - 345.6692 \text{ Toyota} + 338.4950 \text{ Honda} - 176.7995 \text{ Number of Doors} + 4.5545 \text{ Curb Weight} + 40.9486 \text{ Engine Size} + 42.3026 \text{ City MPG} - 79.5580 \text{ Highway MPG}$
- c. Only curb weight and engine size are significant at a 0.05 level. All of the other variables are not significant in predicting the price of a used six-year-old car.
- d. Approximately 85% of the variation in car prices is explained by the model.
- e. Approximately 83.8% of the variation in car prices is explained by the model

with only curb weight and engine size as the independent variables.

## Chapter 14 Additional Exercises

1. a.

Observation	Degree
1	0
2	1
3	0
4	0
5	1
6	1
7	1
8	1
9	0
10	1
11	0
12	1
13	0
14	0
15	1
16	0
17	1
18	1
19	0
20	0

- b.  $\hat{y} = 20,839.7273 + 3002.5613x_1 + 26,094.8990x_2$
- c. \$3002.56
- d. Having a Master's degree results in a salary that is \$26,094.90 higher than the salaries of people who do not have a Master's degree.
- e. Yes.  $H_0: \beta_2 = 0$ ,  $H_a: \beta_2 \neq 0$ ,  $t = 3.845$ ,  $P$ -value = 0.0013, Reject  $H_0$ . This indicates that a master's degree is significant in predicting salary, and the estimated increase in annual salary for people with master's degrees is \$26,094.90, which is greater than \$20,000.
- f. There are many other factors that influence annual salary such as area of study, industry employed in, etc. Answers may vary.

3. a. Yes.  $\hat{y} = 3.6409 - 0.0144 \text{ Share Price} + 0.3333 \text{ Dividend per Share}$ . The model is significant at the 0.01 level ( $F = 5.4731$ ,  $P\text{-value} = 0.0073$ ), and both estimated coefficients are statistically significant in predicting dividend yield at the 0.01 level. Answers may vary.

- b. Dividends per share is more significant in predicting dividend yield. Answers may vary.

## Chapter 15

### 15.1 Exercises

19. a. 153.8095  
 b. 7  
 c. 3  
 d. 21  
 e. Total  $df = 20$ , SST  $df = 2$ , SSE  $df = 18$ ,  $20 = 2 + 18$ , so the relationship holds.
21. Sample 1: No, the histogram represents data drawn from a population that has a negative exponential distribution. Middle histogram Sample 2: No, the histogram represents data drawn from a population that has an unknown distribution. Sample 3: Yes, the histogram represents data drawn from a population that has a normal distribution.
23. No, because the boxes are not basically the same width. The box plot at the bottom appears to be less than half the size of the box plot in the middle, indicating that the variation for this population is considerably smaller than that of the other two populations. Answers may vary.
25. a. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 16.8582$ ,  $F_\alpha = 3.2389$ , Reject  $H_0$ .  $P\text{-value} \approx 0.00003$ , which is less than  $\alpha = 0.05$ , Reject  $H_0$ .  
 b. Hourly wages for employees are approximately normally distributed with equal variances. Observations were collected in an independent and random

fashion. The data does appear to satisfy the necessary assumptions. Answers may vary.

27. a. No.  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 3.4667$ ,  $F_\alpha = 4.0662$ , Fail to reject  $H_0$ .  $P\text{-value} \approx 0.0709$ , which is greater than  $\alpha = 0.05$ , Fail to reject  $H_0$ .  
 b. Maximum heart rates for each workout are approximately normally distributed with equal variances. Observations were selected in an independent and random fashion. Answers may vary.
29. a. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 5.1012$ ,  $F_\alpha = 2.5893$ , Reject  $H_0$ .  $P\text{-value} \approx 0.0176$ , which is less than  $\alpha = 0.10$ , Reject  $H_0$ .  
 b. Dividends per share for each industry are approximately normally distributed with equal variances. Observations were selected in an independent and random fashion. Answers may vary.  
 c. It appears that the transportation industry pays lower dividends per share than the banking and energy industries, but the ANOVA test does not tell us which population mean(s) differ significantly. Answers may vary.

### 15.2 Exercises

5. There will be 6 pairwise comparisons that need to be tested individually.

$$\begin{aligned}
 P(\text{at least 1 type I error}) &= 1 - P(\text{no type I error}) \\
 &= 1 - (1 - \alpha)^6 = 1 - (1 - 0.01)^6 \\
 &= 0.05852 \approx 0.06
 \end{aligned}$$

Thus, it can be seen that the probability of a Type I error increases from 0.01 to approximately 0.06 as the number of pairs for comparing means increases.

$$7. \quad df = n_T - k = 30 - 4 = 26$$

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

Therefore, the critical value of  $t$  corresponding to 26 degrees of freedom and a 0.025 level of significance is equal to 2.0555.

$$9. \quad \text{MSE from the ANOVA output is } 102.2.$$

$$df = n_T - k = 30 - 3 = 27; \frac{\alpha}{2} = \frac{0.1}{2} = 0.05$$

$$\text{Fisher's LSD} = 7.700725$$

$$|\bar{x}_A - \bar{x}_B| < 7.700725$$

$$|\bar{x}_A - \bar{x}_C| > 7.700725$$

$$|\bar{x}_B - \bar{x}_C| > 7.700725$$

There is a significant difference in the mean fasting blood glucose levels between treatments A and C and between treatments B and C.

$$11. \quad df = (n_T - k) = (64 - 4) = 60$$

The studentized range value corresponding to 60 degrees of freedom, four treatments, and  $\alpha = 0.05$  is 3.737.

$$13. \quad \text{a. } df = n_T - k = 15 - 3 = 12, \alpha = 0.05$$

The studentized range value corresponding to 12 degrees of freedom, three groups and 0.05 level of significance is 3.77.

The confidence interval is (2.50, 18.70).

Since the interval does not include the value of 0, the null hypothesis is rejected. Thus, there is a significant difference between the mean test scores of students who took curriculum A versus those who took curriculum B.

b. The confidence interval is (4.10, 20.3). Since the interval does not include the value of 0, the null hypothesis is rejected. Thus, there

is a significant difference between the mean test scores of students who took curriculum B versus those who took curriculum C.

## 15.3 Exercises

9. a. Dependent variable: *Gas Mileage*

Treatment variable: *Car Models*

Blocking variable: *Drivers*

b. Yes, because the dealer believes that the average gas mileage of a particular car will vary depending on the person who is driving the car due to different driving styles. Blocking will reduce the variation in gas mileage which is not due to the type of car.

c. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 696.8608$ ,  $F_\alpha = 3.2874$ , Reject  $H_0$ .  $P$ -value is approximately 0, which is less than  $\alpha = 0.05$ , Reject  $H_0$ .

d. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 101.7798$ ,  $F_\alpha = 2.9013$ , Reject  $H_0$ .  $P$ -value is approximately 0, which is less than  $\alpha = 0.05$ , Reject  $H_0$ .

11. a. Dependent variable: *On-Time Flights*

Treatment variable: *Airlines*

Blocking variable: *Airports*

b. Yes, because the FAA believes that the number of on-time arrivals varies by airport. Blocking will reduce the variation in on-time arrivals which is not due to airline.

c. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 58.8261$ ,  $F_\alpha = 6.9919$ , Reject  $H_0$ .  $P$ -value  $\approx 0.000003$ , which is less than  $\alpha = 0.01$ , Reject  $H_0$ .

d. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 15.2609$ ,  $F_\alpha = 6.9919$ , Reject  $H_0$ .  $P$ -value  $\approx 0.0007$ , which is less than  $\alpha = 0.01$ , Reject  $H_0$ .

13. a. Dependent variable: *Blood Pressure*

Treatment variable: *Measurement Method/Device*

Blocking variable: *Patients*

b. So that any variation not due to the type of device used to measure systolic blood pressure can be reduced.

- c. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 9.9883$ ,  $F_\alpha = 3.2874$ , Reject  $H_0$ .  $P$ -value  $\approx 0.00072$ , which is less than  $\alpha = 0.05$ , Reject  $H_0$ .
- d. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 135.8303$ ,  $F_\alpha = 2.9013$ , Reject  $H_0$ .  $P$ -value is approximately 0, which is less than  $\alpha = 0.05$ , Reject  $H_0$ .

- Reject  $H_0$ .  $P$ -value  $\approx 0.0260$ , which is less than  $\alpha = 0.10$ , Reject  $H_0$ .
- e. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 14.2895$ ,  $F_\alpha = 2.6240$ , Reject  $H_0$ .  $P$ -value  $\approx 0.0002$ , which is less than  $\alpha = 0.10$ , Reject  $H_0$ .

17. a. Dependent variable: *Emotional Reaction*  
 Factors: *Power* and *Knowledge*  
 Factor A: *Power*  
 Factor B: *Knowledge*

## 15.4 Exercises

13. a. Dependent variable: *Rental Rates*  
 Factors: *Rental Car Company* and *Airport Location*  
 Factor A: *Rental Car Company* (Sample)  
 Factor B: *Airport Location* (Columns)
- b. Yes, there appears to be interaction between airport location and major rental car company for all three cities.
  - c. Yes.  $H_0$ : There is no interaction,  $H_a$ : There is interaction,  $F = 13.8127$ ,  $F_\alpha = 2.9277$ , Reject  $H_0$ .  $P$ -value  $\approx 0.000025$ , which is less than  $\alpha = 0.05$ , Reject  $H_0$ .
  - d. We cannot test for effect of company on average daily rental rates because there is interaction.

b.

Source	SS	df	MS	F
Power	1.270	1	1.270	14.7246
Knowledge	0.250	1	0.250	2.8986
Interaction	0.010	1	0.010	0.1159
Error	4.140	48	0.0863	
Total	5.670	51		

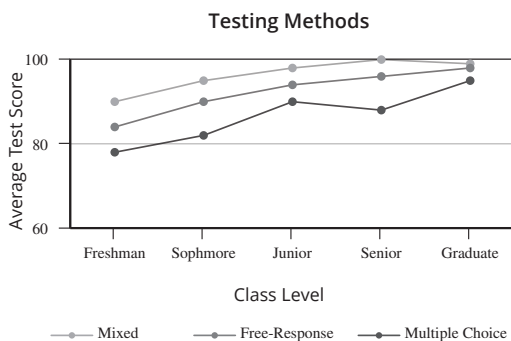
15. a. Dependent variable: *Number of Defects Produced per Day*  
 Factors: *Machine* and *Operator*  
 Factor A: *Machine* (Sample)  
 Factor B: *Operator* (Columns)
- b. There appears to be slight interaction between operator and machine. If there was no interaction, the lines would be parallel. Answers may vary.
  - c. No, there is not significant interaction. This agrees with part a., we only thought the interaction was slight.  $H_0$ : There is no interaction,  $H_a$ : There is interaction,  $F = 0.0789$ ,  $F_\alpha = 2.2858$ , Fail to reject  $H_0$ .  $P$ -value  $\approx 0.9878$ , which is greater than  $\alpha = 0.10$ , Fail to reject  $H_0$ .
  - d. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 4.5$ ,  $F_\alpha = 2.6240$ ,

- c. No.  $H_0$ : There is no interaction,  $H_a$ : There is interaction,  $F = 0.1159$ ,  $F_\alpha = 2.8131$ , Fail to reject  $H_0$ .  $P$ -value  $\approx 0.7350$ , which is greater than  $\alpha = 0.05$ , Fail to reject  $H_0$ .
- d. Yes.  $H_0: \mu_1 = \mu_2$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 14.7246$ ,  $F_\alpha = 4.0427$ , Reject  $H_0$ .  $P$ -value  $\approx 0.0004$ , which is less than  $\alpha = 0.05$ , Reject  $H_0$ .
- e. No.  $H_0: \mu_1 = \mu_2$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 2.8986$ ,  $F_\alpha = 4.0427$ , Fail to reject  $H_0$ .  $P$ -value  $\approx 0.0951$ , which is greater than  $\alpha = 0.05$ , Fail to reject  $H_0$ .

## Chapter 15 Additional Exercises

- 1. a. Dividing the students of each class into blocks categorized as Below Average, Average, and Above Average. Answers may vary.
- b. Dividing participants into blocks, categorized as <100 lb overweight, 50-100 lb overweight, and 0-50 lb overweight. Answers may vary.
- c. Dividing the persons into blocks categorized as Low IQ, Average IQ, and High IQ. Answers may vary.

3. a. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 15.6429$ ,  $F_\alpha = 6.3589$ , Reject  $H_0$ .  $P$ -value  $\approx 0.0002$ , which is less than  $\alpha = 0.01$ , Reject  $H_0$ .
- b. Fat contents for each brand of margarine are approximately normally distributed with equal variances. Observations were collected in an independent and random fashion. Answers may vary.
- c. The data are obtained from servings. There are no variables associated with the servings that we can block with. Answers may vary.
5. a.  $\bar{x} = 3.3064$ ,  
SST = 20.1246
- b.  $F = 35.3327$
- c. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 35.3327$ ,  $F_\alpha = 4.7623$ , Reject  $H_0$ .  $P$ -value is approximately 0, which is less than  $\alpha = 0.01$ , Reject  $H_0$ .
- d. The distributions of all of the populations of interest are approximately normal with equal variances, each of the  $k$  samples must be selected independently from each other and in a random fashion. They cannot be checked in this instance because the raw data are not available.
7. a. Dependent variable: *Test Scores*  
Treatment variable: *Testing Method*  
Blocking variable: *Class Level*
- b. The mixed question tests appear to result in the highest average test scores. It also appears that graduate students typically have higher test scores, on average. Answers may vary.



- c. Yes.  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 22.0786$ ,  $F_\alpha = 3.8379$ , Reject  $H_0$ .  $P$ -value  $\approx 0.0006$ , which is less than  $\alpha = 0.05$ , Reject  $H_0$ .
- d. Yes, because the blocking effects were significant. Answers may vary.

9. a.

Source	SS	df	MS	F
Block	2154.1333	9	239.3481	96.8876
Treatment	30.2	2	15.1	6.1124
Error	44.4667	18	2.4704	
Total	2228.8	29		

- b. Yes,  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 6.1124$ ,  $F_\alpha = 3.5546$ , Reject  $H_0$ .
- c.  $P = 0.0094$ . The probability of a Type I error for this hypothesis test is 0.0094.

11. a. Dependent variable: TSP Revenues  
Treatment variable: Marketing Strategy  
Blocking variable: Quarter

- b.  $H_0: \mu_1 = \mu_2 = \mu_3$ ,  
 $H_a$ : at least one  $\mu_i$  is different.

c.

Source	SS	df	MS	F
Block	1103.0	3	367.6667	31.2908
Treatment	158.1667	2	79.0833	6.7305
Error	70.5	6	11.75	
Total	1331.6667	11		

- d. Yes, at least one of the revenues is significantly different.  $H_0: \mu_1 = \mu_2 = \mu_3$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 6.7305$ ,  $F_\alpha = 5.1433$ , Reject  $H_0$ .  $P$ -value  $\approx 0.0293$ , which is less than  $\alpha = 0.05$ , Reject  $H_0$ .
- e. Yes,  $H_0: \mu_1 = \mu_2 = \mu_3$ ,  $H_a$ : at least one  $\mu_i$  is different,  $F = 31.2908$ ,  $F_\alpha = 4.7571$ , Reject  $H_0$ .  $P$ -value  $\approx 0.0005$ , which is less than  $\alpha = 0.05$ , Reject  $H_0$ .

# Chapter 16

## 16.1 Exercises

9. a. 66.766  
 b. 27.488  
 c. 7.378  
 d. 33.196  
 e. 63.167
11. a. 0.140346  
 b. 98.2422  
 c. 16.01
13. a. 0.236643  
 b. 12.60  
 c. 9

## 16.2 Exercises

9. a. No.  
 $H_0: p_1 = p_2 = \dots = p_5 = \frac{1}{5}$   
 $H_a$ : Any possible difference.  
 $\chi^2_{\alpha} = 9.488$ ;  $\chi^2 = 7.647$ ;  
 $P$ -value  $\approx 0.1054$ ; Fail to reject  $H_0$
- b. The distribution of the number of service calls has a multinomial probability distribution.
11. a. Yes.  
 $H_0: p_1 = p_2 = \dots = p_{12} = \frac{1}{12}$   
 $H_a$ : Any possible difference.  
 $\chi^2_{\alpha} = 24.725$ ;  $\chi^2 = 26.376$ ;  $P$ -value  $\approx 0.0057$ ; Reject  $H_0$
- b. The distribution of fatal accidents has a multinomial probability distribution.
13. a.  $H_0: p_1 = 0.20, p_2 = 0.30, p_3 = 0.50$   
 $H_a$ : Any possible difference.
- b.  $E(n_{\text{Whole Life}}) = 48$
- c.  $\chi^2 = 4.2$
- d.  $\chi^2_{\alpha} = 5.99$ ;  $P$ -value = 0.122; Fail to reject  $H_0$

## 16.3 Exercises

11. a. No.  
 $H_0$ : Education and crime are independent.  
 $H_a$ : Education and crime are dependent.  
 $\chi^2_{\alpha} = 6.251$ ;  $\chi^2 = 6.077$ ;  $P$ -value  $\approx 0.1080$ ;  
 Fail to reject  $H_0$
- b. The experiment satisfies the properties of a multinomial experiment, the null hypothesis is true and the expected counts in each cell are all at least 5.
13. a. No.  
 $H_0$ : Causes of fires and region are independent.  
 $H_a$ : Causes of fires and region are dependent.  
 $\chi^2_{\alpha} = 30.58$ ;  $\chi^2 = 21.36$ ;  $P$ -value = 0.1257;  
 Fail to reject  $H_0$
- b. The experiment satisfies the properties of a multinomial experiment, the null hypothesis is true and the expected counts in each cell are all at least 5.
15. a.  $E(n_{\text{On campus, Above Average}}) \approx 41.767$
- b.  $E(n_{\text{Off campus, Above Average}}) \approx 15.233$
- c.  $\chi^2 = 0.143$
- d. No;  $P$ -value  $\approx 0.9000$ ; Fail to reject  $H_0$ .  
 There is not enough evidence at the  $\alpha = 0.01$  level of significance to support the claim that academic performance dependent on living location.

## Chapter 16 Additional Exercises

1. a. Yes.  
 $H_0: p_1 = 0.50, p_2 = p_3 = 0.25$   
 $H_a$ : Any possible difference.  
 $\chi^2_{\alpha} = 5.991$ ;  $\chi^2 = 68.400$   
 $P$ -value  $\approx 0$ ; Reject  $H_0$

- b. The underlying distribution is a multinomial probability distribution.
3. a.  $P(\text{Lane 1}) = 0.4, P(\text{Lane 2}) = 0.2,$   
 $P(\text{Lane 3}) = 0.2, P(\text{Lane 4}) = 0.2$
- b. Yes.  
 $H_0: p_1 = 0.4, p_2 = p_3 = p_4 = 0.2$   
 $H_a: \text{Any possible difference.}$   
 $\chi^2_\alpha = 7.815; \chi^2 = 20.638$   
 $P\text{-value} \approx 0.0001; \text{Reject } H_0$
5. a. Yes  
 $H_0: \text{Rating and brand are independent.}$   
 $H_a: \text{Rating and brand are dependent.}$   
 $\chi^2_\alpha = 26.296; \chi^2 = 62.895; P\text{-value} \approx 0$   
 Reject  $H_0$
- b. The experiment satisfies the properties of a multinomial experiment, the null hypothesis is true and the expected counts in each cell are all at least 5.

## Chapter 17

### 17.1 Exercises

15.  $H_0: \text{Median} = \$25,000$   
 $H_a: \text{Median} > \$25,000$   
 Critical value =  $-1.28$   
 $z = -3.25; \text{Reject } H_0$
17. a.  $H_0: \text{Median} = 14,400 \text{ hrs}$   
 $H_a: \text{Median} < 14,400 \text{ hrs}$   
 Critical value =  $4$   
 $X = 6; \text{Fail to reject } H_0$
- b. The data are randomly selected.
19.  $H_0: \# \text{ of Positive Signs} = \# \text{ of Negative Signs}$   
 $H_a: \# \text{ of Positive Signs} > \# \text{ of Negative Signs}$   
 Critical value =  $6$   
 $X = 9; \text{Fail to Reject } H_0$
21.  $H_0: \# \text{ of Negative Signs} = \# \text{ of Positive Signs}$   
 $H_a: \# \text{ of Negative Signs} \neq \# \text{ of Positive Signs}$   
 Critical value =  $0$   
 $X = 0; \text{Reject } H_0$

### 17.2 Exercises

11.

Mutual Fund	Price (\$)	Rank
American Funds	49.30	9
Columbia Management	9.41	2.5
Morgan Stanley	88.50	10
Fidelity Investments	24.40	6
John Hancock	9.41	2.5
DWS Investments	15.57	5
UBS	12.15	4
Prudential Investments	9.23	1
Value Line Funds	32.82	7
The Vanguard Group	34.72	8

13. a.  $H_0: \text{DHEA-S is not increased}$   
 $H_a: \text{DHEA-S is increased}$   
 Critical value =  $0$   
 $X = 2; \text{Fail to Reject } H_0$
- b. The data are randomly selected.
- c.  $H_0: \text{DHEA-S is not increased}$   
 $H_a: \text{DHEA-S is increased}$   
 Critical value =  $4$   
 $T_+ = 5; \text{Fail to Reject } H_0$
- d. Pairs of data have been randomly selected and are such that the absolute values of their differences can be ranked.

- e. The signed-rank test because the magnitudes of the differences are not ignored. Answers may vary.
15. a. The paired differences have an approximately normal distribution.
- b.  $H_0$ : Model A = Model B  
 $H_a$ : Model A  $\neq$  Model B  
 Critical value = 1  
 $T_+ = 0$ ; Reject  $H_0$
- c. Pairs of data have been randomly selected and are such that the absolute values of their differences can be ranked.
- d. In all three tests the null hypothesis is rejected in favor of the alternative.

## 17.3 Exercises

11. a.  $H_0$ : Mrs. Ellis = Mr. Ford  
 $H_a$ : Mrs. Ellis  $\neq$  Mr. Ford  
 Critical values = 37, 68  
 $T_x = 42, 63$ ; Fail to Reject  $H_0$
- b. The data are such that they can be ranked. The two samples are selected in an independent and random fashion.
13. a.  $H_0$ : New Battery = Old Battery  
 $H_a$ : New Battery < Old Battery  
 Critical value = 28  
 $T_x = 35$ ; Fail to Reject  $H_0$
- b. The data are such that they can be ranked. The two samples are selected in an independent and random fashion.
15. a.  $H_0$ : Drama = Comedy  
 $H_a$ : Drama > Comedy  
 Critical value = 36  
 $T_x = 18$ ; Fail to Reject  $H_0$
- b. The data are such that they can be ranked. The two samples are selected in an independent and random fashion.
17. a.  $H_0$ : Service City = Sunshine City  
 $H_a$ : Service City < Sunshine City  
 Critical value = 52  
 $T_x = 47$ ; Fail to Reject  $H_0$
- b. The data are such that they can be ranked. The two samples are selected in an independent and random fashion.

## 17.4 Exercises

11. a.  $H_0$ :  $\rho_s = 0$   
 $H_a$ :  $\rho_s \neq 0$   
 $r_s = -0.8042$   
 Critical values =  $-0.591, 0.591$   
 Reject  $H_0$
- b. Yes, but the assumption must be made that the relationship between the variables is linear. Answers may vary.
13.  $H_0$ :  $\rho_s = 0$   
 $H_a$ :  $\rho_s \neq 0$   
 $r_s = 0.0182$   
 Critical values =  $-0.648, 0.648$   
 Fail to Reject  $H_0$

## 17.5 Exercises

9.  $H_0$ : The sequence is random.  
 $H_a$ : The sequence is not random.  
 $N = 26$ ,  $m = 10$ ,  $n = 16$ ,  $R = 14$   
 Critical values = 8, 19  
 Fail to Reject  $H_0$
11.  $H_0$ : The sequence is random.  
 $H_a$ : The sequence is not random.  
 $N = 31$ ,  $m = 15$ ,  $n = 16$ ,  $R = 19$   
 Critical values = 10, 23  
 Fail to Reject  $H_0$

## 17.6 Exercises

9.  $H_0$ : The milk production for all schedules is the same.  
 $H_a$ : The milk production for at least one of the schedules is different.  
 $H = 7.0234$   
 Critical value = 6.2514  
 Reject  $H_0$

## Chapter 17 Additional Exercises

1. a. The paired differences have an approximately normal distribution.
- b.  $H_0$ : Vest Treatment = Traditional Treatment  
 $H_a$ : Vest Treatment > Traditional Treatment  
 $X = 0$ ; Fail to Reject  $H_0$
- c. The data are randomly selected.

- d.  $H_0$ : Vest Treatment = Traditional Treatment  
 $H_a$ : Vest Treatment > Traditional Treatment  
 $T_+ = 0$ ; Fail to Reject  $H_0$
- e. Pairs of data have been randomly selected and are such that the absolute values of their differences can be ranked.
- f. The signed-rank test because the magnitudes of the differences are not ignored. Answers may vary.
- g. There is not sufficient evidence that the diameter of blood vessels in the lungs is significantly larger after using the vest treatment when performing a paired difference test. There is sufficient evidence using the nonparametric methods. Answers may vary.  
 $H_0$ : Vest Treatment = Traditional Treatment  
 $H_a$ : Vest Treatment > Traditional Treatment  
 $t = -2.75$   
Critical value =  $-3.747$   
Fail to Reject  $H_0$
3. a. The Wilcoxon rank-sum test because we are not dealing with paired data.
- b.  $H_0$ : Diet B = Diet A  
 $H_a$ : Diet B > Diet A
- c.  $H_0$ : Diet B = Diet A  
 $H_a$ : Diet B > Diet A  
Critical value = 131  
 $T_x = 113$ ; Fail to Reject  $H_0$
- d. The data are such that they can be ranked. The two samples are selected in an independent and random fashion.
5. a.  $r_s = 0.2727$  these two variables have a weak positive relationship.
- b.  $H_0: \rho_s = 0$   
 $H_a: \rho_s \neq 0$   
 $r_s = 0.2727$   
Critical values =  $-0.648, 0.648$   
Fail to Reject  $H_0$
7.  $H_0$ : The sequence is random.  
 $H_a$ : The sequence is not random.  
 $N = 9, m = 5, n = 4, R = 4$   
Critical values = 2, 9  
Fail to Reject  $H_0$
9.  $H_0$ : The rankings are the same for the three conferences.  
 $H_a$ : The ranking of at least one conference is different.  
 $H = 13.5755$   
Critical value = 4.6052  
Reject  $H_0$