Answer Key

Chapter 1

Section 1.1

15. Recalling the definitions for population and sample, we know that the population is a particular group of interest, and a sample is a subset of the population from which data are collected. In this case, we are looking at the number of billable man-hours logged per week by employees at Deloitte. This is the group of interest, and therefore the statement describes a population.

17. Recalling the definitions for population and sample, we know that the population is a particular group of interest, and a sample is a subset of the population from which data are collected. In this case, we are looking at the final rankings of 5 candidates who applied for the open CFO position in your organization. This is a subset of the 22 candidates who applied, and therefore the statement describes a sample.

19. Recalling the definitions for population and sample, we know that the population is a particular group of interest, and a sample is a subset of the population from which data are collected. In this case, the particular group of interest is all the employees who work at a company in Silicon Valley. Therefore, the sample is the 35 employees who work at a company in Silicon Valley.

21. Recalling the definitions for population parameter and sample statistic, we know that a parameter is a numerical description of a particular population characteristic, and a statistic is the actual numerical description of a particular sample characteristic. Therefore, first consider whether the statement refers to a sample or a population. To be a population parameter, it must describe all members being studied, not just a portion of them.

In this case, the average number of hours is a population parameter because it is based on all of the students in your statistics class.

Section 1.2

5. a. Answers will vary.

b. Answers will vary.

c. Answers will vary.

Section 1.3

3. Recalling the definitions of descriptive and inferential statistics, we know that a descriptive statistic gathers, sorts, summarizes, and displays data while an inferential statistic involves using descriptive statistics to estimate population parameters.

In this case, the average price of a car at the new car dealership in town is a descriptive statistic because it describes all of the cars at the new dealership.

Section 1.4

3. a. Answers will vary.

b. Answers will vary.

c. Answers will vary.

d. Answers will vary.

Chapter 1 Additional Exercises

1. a. Internet users

b. The amount of time users view photos of products on websites

c. Inferential

3. a. Elderly citizens

b. Self-esteem after reading a news story

c. 276 elderly citizens

d. Answers will vary.

e. Inferential

5. a. Mobile phone users

b. The percentage of people who use their mobile phones to access the Internet for different reasons

c. 500 American adults 18 years of age and older

d. Answers will vary.

e. Inferential

7. a. Married couples

b. The percentage of married couples that met online

c. 7000 adults married in the past 5 years

d. Answers will vary.

9. a. U.S. states with coastlines

b. Coastline length

Chapter 2

Section 2.1

29. a. Well-defined

b. Well-defined

c. Not well-defined

d. Well-defined

e. Not well-defined

31. There is no well-defined scale to measure cleanliness or aesthetics. Answers will vary.

33. Answers will vary.

35. 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 self-selection bias. We would not necessarily know the cause of the weight reduction.

37. 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.

39. Jacob's knee could feel better simply because he took a week break from playing basketball. Answers will vary.

41. 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 going to church. Answers will vary.

43. Generally people that have more money will seek the help of a financial advisor since advisors are paid for their services. Answers will vary.

Section 2.2

9. Volume – the satellites would collect a huge volume of data since they are continuously monitoring one position.

Velocity – the data are being collected every second.

Veracity – the data will be of very good quality and can be trusted since it is automatically collected by a government source.

Variety – images and numbers are part of the data.

11. Both predictive and prescriptive analytics can be done on these data. Predictive analytics can be used to develop a model to help determine any trends or patterns in the BMI by country and prescriptive analytics can be used to implement or prescribe nutritional programs.

13. Since most people tend to have their mobile phones with them at all times, the marketing company can use the location services in addition to what the user searched for to provide coupons or discounts for products at nearby retail stores. Based on purchasing history and other factors, the marketing company can offer very specific ads to individual users.

15. LBS can be used to prevent credit card fraud by matching the user location from the smartphone to a credit card transaction. Tying the smartphone's location to a credit card allows the credit card company to flag transactions made across several geographic locations over a short time or determine if the card is being used outside of its home zone.

Section 2.3

13. a. Discrete
b. Continuous
c. Continuous
d. Continuous
e. Discrete

15. 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. A lot of information is gathered about the cadets; to be able to derive some conclusions from the information, a data summary answering the relevant questions is required.

17. a. Quantitative
b. Ratio
b. Nominal
c. Ordinal

19. a. Ratio
d. Ratio

Section 2.4

9. a. Yes
b. Nonstationary

11. The table contains time series data. Answers will vary.

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.

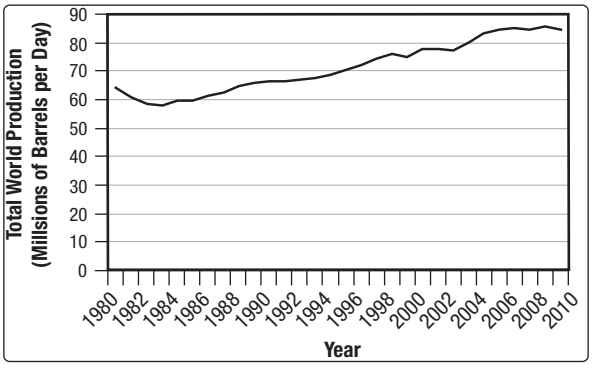
3. a. Types of investments available, rate of return, amount, etc.
b. Rates of return on various types of investments: Returns for diversified portfolio of stocks: e.g. S&P 500, Dow Jones Index Returns for diversified portfolio of bonds.

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.

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?

9. a. Interval
b. Interval
c. Ordinal
d. Nominal
e. Ratio

11. a. The level of measurement for millions of barrels of oil per day is ratio.
b. Time series, Nonstationary



13. Answers will vary.

15. a. Nominal.
b. Time Series.
c. If respondents were owners or renters, answers will vary.

Chapter 3

Section 3.1

7. a. Nominal

b. Qualitative

c.

Air Conditioner	16
Lawn Mower	10
Fan	7
Washing Machine	6
Miscellaneous	9

9. a. Answers will vary.

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

b. Answers will 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. Results may vary depending on who prepared the data.

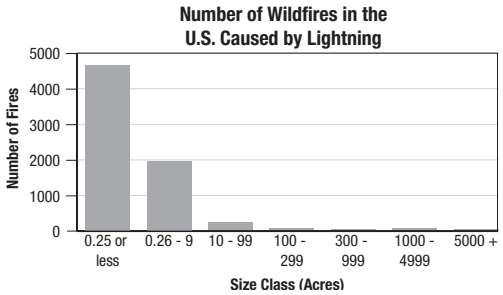
d. Yes. Given that you assign a complaint to only one category, the categories are mutually exclusive. Given that you assign each complaint to a category, the categories are exhaustive.

Section 3.2

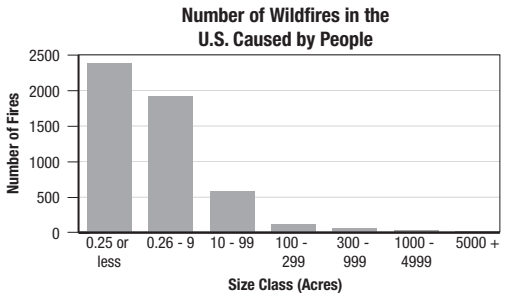
11. a. Bars are not proportional. The bar for the Dodge Intrepid (27) should be shorter than the bar for the Ford Taurus (28). The graph also lacks labels on the horizontal axis. Answers will vary.

b. It is hard to tell if bars are proportional because there are no horizontal axis labels. The names of the vehicles would be easier to read if they were listed outside the graph area. Horizontal and vertical axis labels are needed. Answers will vary.

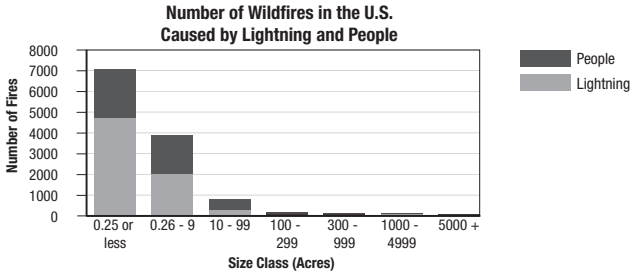
13. a.



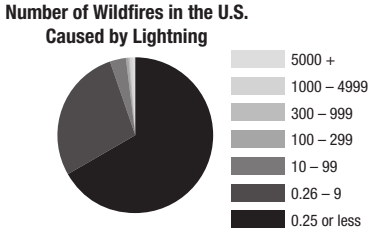
b.



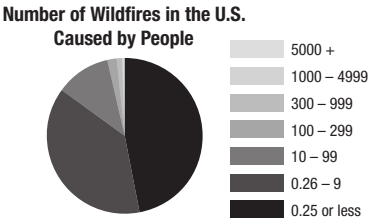
c.



d.

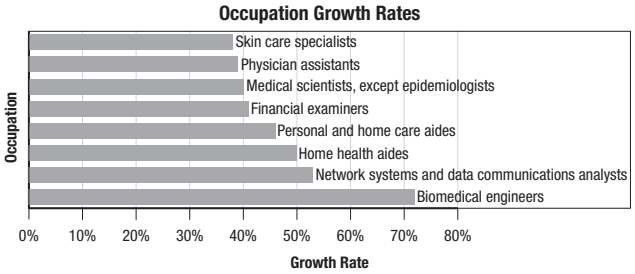


e.



f. A majority of fires both lightning-caused and people-caused occur in the 0.25 or less acre size class. For the 0.25 or less acre size class, about twice as many fires are caused by lightning than by people. However, in the 0.26-9 acre size class, the number of wildfires caused by lightning and by people are almost the same. The number of wildfires caused by both lightning and people is much smaller for the size classes above 0.26-9 acres.

15. a.



b. Biomedical engineers have the highest projected growth rate. The rest of the growth rates appear to be between 35% and 55%, whereas the rate for biomedical engineers is above 70%.

Section 3.3

Solutions given here are only examples of frequency distributions. Students' answers may be different if they choose different classes.

7.

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

9.

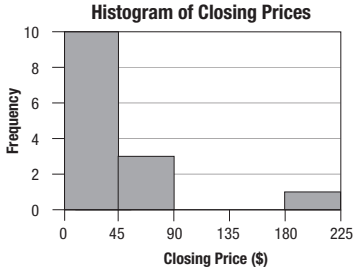
Average Temp (°F)	Frequency	Relative Frequency	Cumulative Frequency
40 – 49	3	0.20	3
50 – 59	7	0.467	10
60 – 69	4	0.267	14
70 – 79	1	0.067	15

Section 3.4

17. a.

Closing Price	Frequency
$0 – $44.00	10
$45.00 – $89.00	3
$90.00 – $134.00	0
$135.00 – $179.00	0
$180.00 – $224.00	1

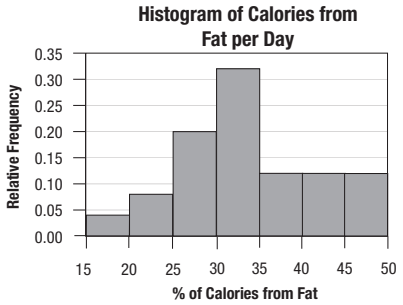
b.



19. a., b.

% of Calories From Fat	Frequency	Relative Frequency
15% – 19%	1	0.04
20% – 24%	2	0.08
25% – 29%	5	0.20
30% – 34%	8	0.32
35% – 39%	3	0.12
40% – 44%	3	0.12
45% – 49%	3	0.12

c.



d. A majority of the subjects in the sample consumed between 30% and 34.9% of calories from fat per day. The sample percentages appear to have a bell-shaped distribution. Relatively few of the subjects consumed less than 20% of calories per day from fat. Answers will vary.

21. a. Ratio

b.

Stem	Leaf
12	3 5 5
13	5 5 5
14	7 7 8 8 8 9
15	6 6 6 6 8 8
16	9
17	8 8 9 9
18	9 9 9
19	8 9 9
20	
21	4 5 5
22	
23	5 9
24	8 8 9
25	6 7 8 8 8 9 9 9
26	5 8 9 9
27	
28	8
Key: 12	3 = $123

c. There appear to be two different clusters of daily rates for semi-private rooms. One of the clusters seems to center around $150 per day. The other cluster seems to center around $250 per day. One explanation of this could be the location of the surveyed hospitals, it could be that some of the hospitals surveyed were in large metropolitan areas and some were in smaller cities or suburban areas. Answers will vary.

23. a.

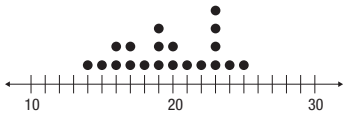
Stem	Leaf
2	3 6 7 8 8 8 9 9
3	1 3 4 6 6 7 7 8
4	0 2 6
Key: 2	3 = 23 (Miles per Gallon)

b. The majority of the miles per gallon are between 25 and 38. Miles per gallon ratings above 40 are uncommon. Answers will vary.

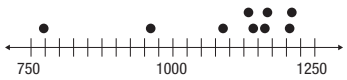
25. a. –4, 0, 10, 10, 16, 19, 19, 20, 20, 22, 23, 24, 24, 27, 33, 37

b. Only 1 quarter showed a loss, and most quarters showed growth between 20% and 30%. Answers will vary.

27.



29.

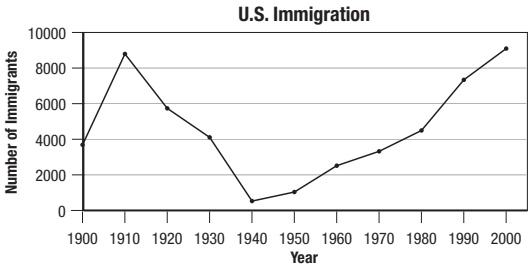


31. a. All of the rates seem to change in the same way. The longer the period of time, the higher the interest rate charged. Answers will vary.

b. Stationary. Answers will vary.

33. a. Year – Ordinal; Number – Ratio; Rate – Ratio

b.



c. 146.61%

d. –39.62%. Though the number of immigrants has increased drastically, the general population of the U.S. has increased also, so the rate of annual immigration (per 1000 people) has actually decreased from 1900 to 1990. Answers will vary.

Section 3.5

5. Answers will 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 45%

b. No. The area of the graphic for November 2017 has an area more than double that of the November 2010 graphic, even though the value is only 45% more.

c. Answers will 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 will vary. Graph A is better because the vertical scale starts at 0.

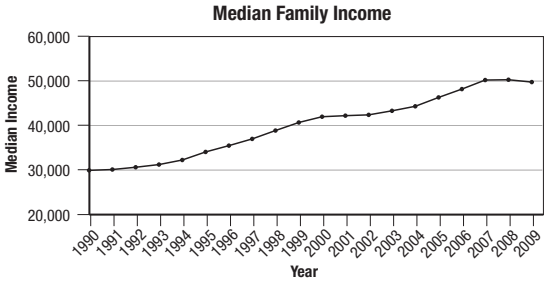
b. Answers will vary. Graph B causes more concern as the increase in robberies appears to be much more dramatic.

c. The 2016 bar in Graph B is approximately 8 times taller than the 2013 bar. There were actually only about 1.3 times as many robberies in 2016 than in 2013.

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.

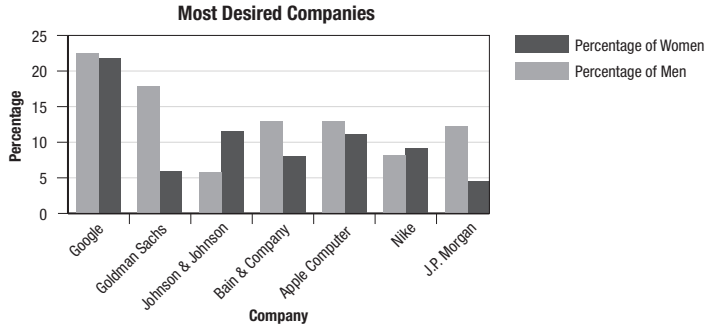
b. Answers will vary.



c. Median household income has an upward trend, but this could likely be due to inflation. 2009 is the only year in which median household income decreased from the previous year. Answers will vary.

3. a. A side-by-side bar chart showing the percentage of both men and women who ranked each company in their top 5. Answers will vary.

b.

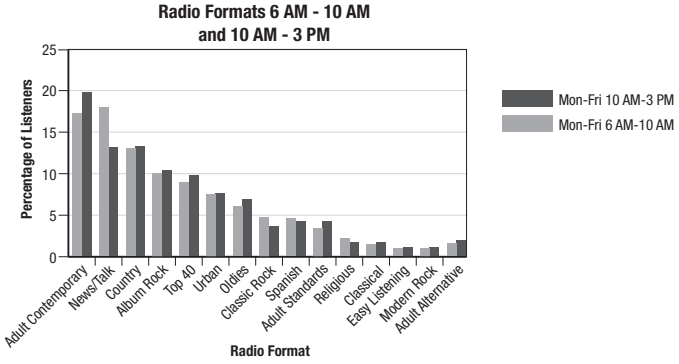


c. Google appears to be the most desirable company for both men and women. Goldman Sachs appears to be desirable for men, but not so much for women. Johnson & Johnson appears to be popular among women, but not so much among men. Answers will vary.

5. a. Bar charts and pie charts would both be appropriate for displaying the data. Answers will vary.

b. Answers will vary.

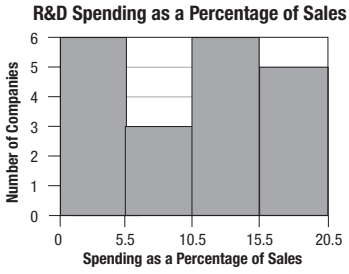
c. Answers will vary.



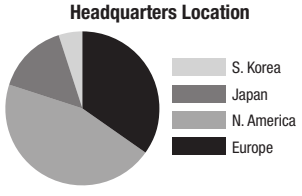
7. a. Spending as a percentage of sales is a more useful measure of research and development expenditures for comparative purposes because it standardizes research and development expenditures.

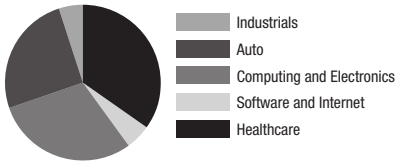
b. A bar chart for R&D expenditures, pie charts representing industries, and headquarters locations of the top 20 R&D spenders, a histogram to represent spending as a percentage of sales. Answers will vary.

c.



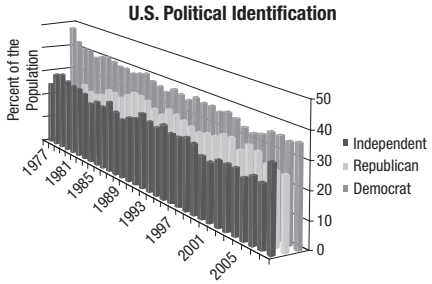
d.

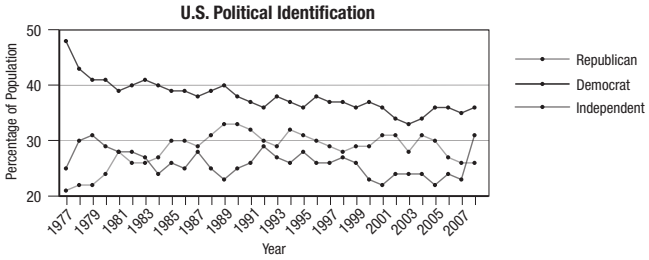




9. a. A line graph displaying the percentage of voters in each category and a pie chart showing political identification in a particular year might be helpful in visualizing the data. Answers will vary.

b.





c. The time series appears to be stationary. Recently it appears that more people identify themselves as Democrats than Republicans or Independents. Answers will vary.

Chapter 4

Section 4.1

13. mean = 15, median = 15, mode = 11, 20% trimmed mean = 15

15. a. mode

b. median (some very high incomes may skew data)

c. mean

d. median or mode

17. a. 92.9667

b. 92

c. 88

d. 92.75

e. Mean, because there are no extreme values. Answers may vary.

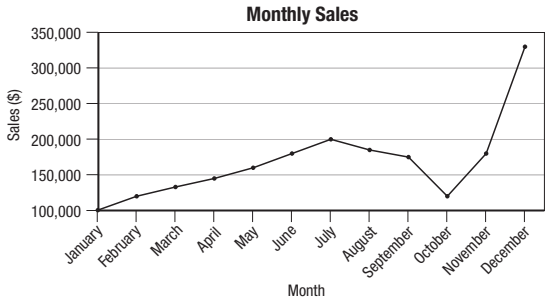
19. a. Ratio

b. mean = 99.9

10 % trimmed mean = 107.5625, 20 % trimmed mean = 107.25

c. Answers will vary.

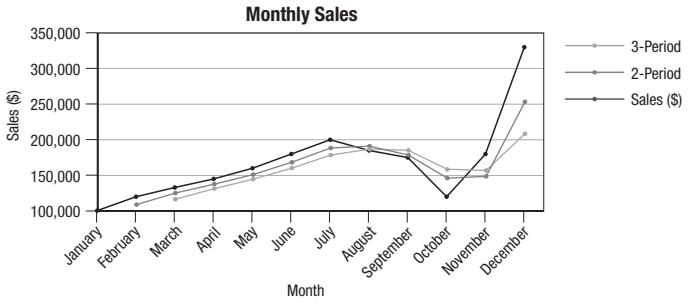
21. a.



b., c.

Month	Sales	2-period	3-period
Jan	$100,500		
Feb	$120,000	$110,250	
Mar	$133,000	$126,500	$117,833.33
Apr	$145,000	$139,000	$132,666.66
May	$160,000	$152,500	$146,000
June	$180,000	$170,000	$161,666.66
July	$200,000	$190,000	$180,000
Aug	$185,000	$192,500	$188,333.33
Sept	$175,000	$180,000	$186,666.66
Oct	$120,000	$147,500	$160,000
Nov	$180,000	$150,000	$158,333.33
Dec	$330,000	$255,000	$210,000

d.



e. Answers will vary.

Section 4.2

13. 180

15. a. 14.2143

b. 3.7702 feet

c. 11

d. Athletic ability, sex, height, weight, etc. Answers will vary.

17. a. both averages = 76.8571

b. male variance = 350.8095

female variance = 205.8095

c. male std. dev. = 18.7299

female std. dev. = 14.3461

d. Average scores are the same for males and females but the standard deviation of scores of females is lower, implying more consistent scores.

e. Answers will vary.

19. a. mean of original data = 89.4167

std. dev. of original data = 7.1916

mean of adjusted data = 109.4167, std. dev. of adjusted data = 7.1916

b. mean of adjusted data = mean of original data + 20; the standard deviations are the same.

c. These results hold in general. If you add a constant to each data point, the mean of the adjusted data is equal to the mean of the original data + the constant. The standard deviation of the adjusted data is the same as the standard deviation of the original data (i.e., the way in which the data vary does not change if you simply add a constant to each data point).

21. 5 to 47

23. a. 68

b. 95

c. In order to use the empirical rule, we must assume that the distribution of the amounts of chowder eaten is approximately bell-shaped.

25. a. Job grade 25: \$19,000 to \$25,000

Job grade 33: \$31,000 to \$39,000

Job grade 40: \$35,000 to \$55,000

b. In order to use the empirical rule, we must assume that the distribution of the salaries is approximately bell-shaped.

27. a. 12

b. 10

c. Since the coefficient of variation is smaller for Machine Y than Machine X, the standard deviation in the diameter of the bolts produced is smaller relative to the average diameter for Machine Y than for Machine X. Thus, Machine Y more consistently produces bolts of the correct diameter.

Section 4.3

11. a. 1

b. 6

c. 25th percentile: approximately 25% of the salespeople sold one or fewer copiers in a day.

90th percentile: approximately 90% of salespeople sold 6 or fewer copiers in a day.

d. 83rd percentile

e. 33rd percentile

13. a. .185

b. .255

c. .285

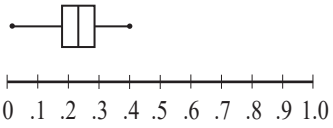
d. 1st quartile: approximately 25% of the batting averages are at or below .185,

2nd quartile: approximately 50% of the batting averages are at or below .255,

3rd quartile: approximately 75% of the batting averages are at or below .285.

e. 0.1

f. 0.02 is an outlier. This likely represents the pitcher's batting average.



g. –2.61

h. 1.14

i. The player with the .020 batting average has a batting average 2.61 standard deviations below the mean. The player with a .330 batting average has a batting average 1.14 standard deviations above the mean.

j. 70th percentile

k. 15th percentile

15. First exam: $z = 0.10$, Second exam: $z = -0.43$. On the first exam, the student's score was 0.1 standard deviations above the mean; but on the second exam, the student's

score was 0.43 standard deviations below the mean. Thus, although the student achieved a higher absolute score on the second exam, the student performed relatively better on the first exam than on the second.

Section 4.4

3. a. Nominal: Beer ID, Beer Name, Beer Style, Brewery ID, Brewery Name, City, and State; Ratio: ABV, IBU, and Ounces

b. Style, Ounces, Brewery Name/ID, and State.

c. 8

d. Elevation Triple India Pale Ale, Renegade Brewing Company.

e. 2, American IPA and American Double/Imperial IPA

f. mean=0.059, st. deviation = 0.013

g. CVRenegade=28.16%. CVWynkoop=21.30%; The Wynkoop Brewery has more consistent ABV values since it has a smaller coefficient of variation.

5. a. The mean is 18,065.79, the mode is 3,060, the median is 7,026.

b. The variance is 381,922,239 the standard deviation is 19,542.83, and the range is 74,876.

c. The 1st quartile is 2,898.75, the 3rd quartile is 37,712.75

d. Age, age group, gender, ethnicity

e.

Age Group	Average Expenditure
0 to 5	1415.28
6 to 12	2226.86
13 to 17	3922.61
18 to 21	9888.54
22 to 50	40209.28
51+	53521.90

f. The highest average Expenditure occurs in the 51+ age group.

It seems that as an individual ages, more is spent.

Answers will vary, the differences could be because as the person ages, they may no longer have family to care for them. Therefore, public institutions may have to assume the cost of their care. Also, as a person ages, perhaps there are more health issues to deal with that will increase the amount spent on that age group.

g. The standard deviation and coefficient of variation for each age group are listed below.

Age Group	Standard Deviation	CV
0 to 5	612.6	43.3%
6 to 12	830.9	37.3%
13 to 17	1012.7	25.8%
18 to 21	2940.6	29.7%
22 to 50	6287.3	15.6%
51+	6283.8	11.7%

The highest level of dispersion is in the lowest age group.

Answers will vary, the differences could be the result of greater levels of care needed at younger ages when trying to determine the appropriate level of service needed.

h.

Ethnicity	Expenditures
American Indian	36438.25
Asian	18392.37
Black	20884.59
Hispanic	11065.57
Multi Race	4456.73
Native Hawaiian	42782.33
Other	3316.50
White not Hispanic	24697.55

i. Divide the average Expenditure for each Ethnicity by the total Expenditure to get the proportion for each group as shown below.

Ethnicity	Proportion
American Indian	22.5%
Asian	11.4%
Black	12.9%
Hispanic	6.8%
Multi Race	2.8%
Native Hawaiian	26.4%
Other	2.0%
White not Hispanic	15.2%

j. Answers will vary. There seems to be some inequity by age and by ethnicity. The reasons for the inequity is unknown. Also, there is much more variability in the age group 6-12 in terms of spending.

7. a. The mean is 138,528.40, the mode is 93,896.18, the median is 127,850.10.

b. The variance is 2,292,419,896, the standard deviation is 47,879.22, and the range is 447,679.40.

c. The 1st quartile is 102,031.16, the 3rd quartile is 167,463.53

d. Base pay, total pay, whether there was any overtime, and benefits. The data could also be arranged by job title, although there are several job titles listed.

e. The relative frequency per category should be determined. The table could look like this:

Class	Frequency	Relative Frequency
0-25000	0	0.00%
25001-50000	775	3.47%
50001-75000	8054	36.06%
75001-100000	6023	26.97%
100001-125000	4244	19.00%
125001-150000	2073	9.28%
150001-175000	505	2.26%
175001-200000	508	2.27%
200001-225000	105	0.47%
225001-250000	24	0.11%
250001-275000	14	0.06%
275001-300000	4	0.02%
300001-350000	5	0.02%

f. The highest group is the 75,000. Between 75,000 up to 125,000 makes up about 82% of the salaries for the data set. Only a small percentage is below 75,000 and above 175,000. The differences could be because of experience levels or the specific type of job. When the data were sorted by salaries, public service aides were among the lowest paid in the group. The higher paying jobs were managers and department heads.

g. Approximately 60% of the jobs have OT pay listed. (No OT = 9010 rows)

h.

	Mean	Standard Deviation	CV
No OT	100711.41	39013.68	39%
OT	105395.36	41734.74	40%

Both have very similar dispersion about the mean. In addition, the averages are quite close.

i. Answers will vary. For example, the pay one could expect seems to be between about 75,000 and 125,000 based on the data set. Seventy-five percent (75%) of people are getting paid more than approximately 102,000 in total pay. However, this of course is contingent upon the position and other individual factors of the applicant such as experience and degree. Also, it seems that most of the positions do have OT associated with the job (60%). But that OT doesn't provide a huge difference in total pay for the individual. Finally, it must be noted that this data is from 2014 and is quite out of date at this point. So, this analysis can provide some perspective, but it is limited in its applicability to the job market today.

Section 4.5

3. Mean = 17.5 days, Variance = 55.3043 days^2

5. a. 0.061

b. 0.0002

c. 0.014

Section 4.6

5. a. 0.5

b. 0.3415

c. 0.4661

d. 0.6341

e. It appears that the supplier executives are optimistic, while the original equipment managers appear to be skeptical about the economic recovery. Answers will vary.

7. 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 will vary.

9. a. 835 **b.** 2884

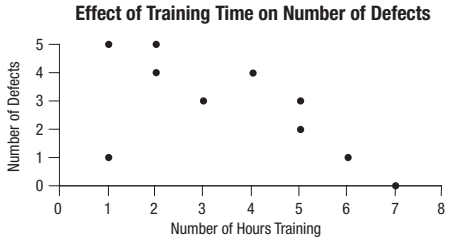
Section 4.7

11. a. Yes, the pattern roughly follows a straight line. The pattern is upward sloping; as x increases, y generally increases. The data values are widely dispersed. The fanning of the data could be considered a significant deviation from the pattern.

b. Yes, the pattern roughly follows a straight line. The pattern is downward sloping; as x increases, y decreases. The data values are tightly clustered; in fact, they exactly fall in a straight line. There are no significant deviations from the pattern.

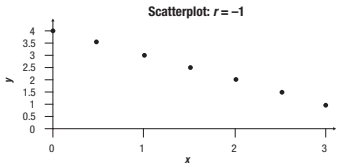
13. a. Yes, the data collected seem to be appropriate to study the relationship between training and countertop defects. An example of a bias could be that some employees could have more experience than others aside from the training. The data are collected using an observational study. Answers will vary.

b.



c. Yes, the pattern roughly follows a straight line. The pattern is downward sloping; as the number of hours of training increases the number of defects tends to decrease. The data values are tightly clustered. There is one significant deviation from the pattern: the employee with only one hour of training who had only one defect.

15. Answers may vary, but points should be plotted in a downward sloping, perfectly straight line.



17. a. Tightly clustered in a positive linear fashion.

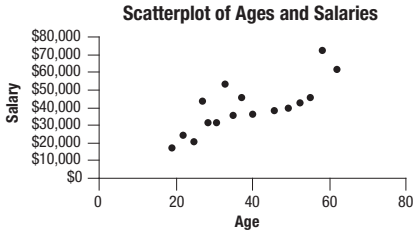
b. Loosely clustered in a positive linear fashion.

c. Tightly clustered in a negative linear fashion.

d. Loosely clustered in a negative linear fashion.

e. Loosely clustered in a positive linear fashion.

19. a.



b. $r = 0.7707$

c. The correlation coefficient indicates a moderate positive linear relationship. This seems consistent with the scatterplot.

21. a. Summer

b. Valentine's Day, Christmas

c. Winter

Chapter 4 Additional Exercises

1. No. Because you did not want any variation in the length of the boards. Answers will vary.

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.

7. a. 92.0929

b. 9.1383

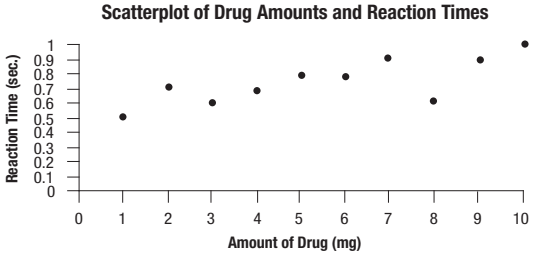
c. 9

d. 12

e. The literacy rates are normally distributed.

9. a. Yes, the variables measured seem appropriate to study the relationship between reaction time and the amount of drug in the bloodstream. Biases could include differences in reaction time without the drug. The data are ratio data. Answers will vary.

b.



c. Yes, the pattern roughly follows a straight line. The pattern is upward sloping; as the amount of drug increases, the reaction time tends to increase. The data values are tightly clustered. There is one significant deviation from the pattern: the person who was administered 8 mg of the drug had a reaction time of only 0.6 sec. Answers will vary.

11. a. Strong positive linear relationship

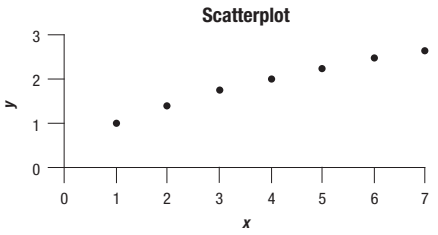
b. Weak positive linear relationship

c. Strong negative linear relationship.

d. Weak negative linear relationship.

e. Weak positive linear relationship.

13. a.



b. $r = 0.9916$

c. There is a strong positive relationship between x and y. However, the scatterplot does not appear to be linear. It appears that y is the square root of x. Answers will vary.

Chapter 5

Section 5.1

21. a. $S = \{\text{Very Attentive, Somewhat Attentive, Not Attentive}\}$

b. $A = \{\text{Somewhat Attentive, Not Attentive}\}$

23. $P(\text{Yellow}) = 0.3$, $P(\text{Red}) = 0.5$, $P(\text{Blue}) = 0.2$

25. 0.6667

27. 0.375

d. relative frequency

e. classical

29. a. classical

b. subjective

c. subjective

d. classical

e. subjective

f. relative frequency

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

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

Section 5.2

7. a. Yes

b. No, probabilities cannot be greater than 1.

c. Yes

d. No, probabilities cannot be less than 0.

e. Yes

9. a. The event cannot occur.

b. The event is certain to occur.

c. Relative frequency interpretation: If an experiment is performed 100 times, the event will occur, on average, 45 times.

d. Relative frequency interpretation: If an experiment is performed 100 times, the event will occur, on average, 65 times.

e. Not a valid probability because it is negative.

11. a. 0.9696

b. 0.0523

13. 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 \$150,000 in insurance and the husband could have between \$50,000 and \$100,000 of insurance. Answers may vary.

Section 5.3

3. a. 0.3194

b. 0.3472

c. 0.6327

d. 0.6032

5. **a.** 0.8571
b. 0.1429

7. **a.** 0.005
b. 0.0417

Section 5.4

7. No, the events are dependent. If A = husband has more than \$150,000 insurance and B = wife has more than \$50,000 insurance, $P(A|B) \neq P(A)$.

9. 0.0001

11. **a.** 0.9980
b. 0.0020
c. 0.000001

13. **a.** 0.0004
b. 0.0004

Section 5.5

5. $P(\text{Dem}|\text{Favor}) = 0.7404$

7. $P(\text{Def}|\text{Insp}) = 0.2069$

9. $P(\text{Woman's name}|\text{Man chosen}) = 0.7347$

Section 5.6

5. 120

7. **a.** 1
b. 6
c. 120
d. 5040

9. 665,280

11. 50,400

13. 55

Chapter 5 Additional Exercises

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

Chapter 6

Section 6.1

5. **a.** Discrete
b. Continuous
c. Discrete
d. Continuous
e. Discrete

7. **a.** Discrete
b. Continuous
c. Continuous
d. Discrete
e. Continuous

Section 6.2

13. Yes

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

17. No. Probabilities cannot be negative.

19. Yes

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

21.

x	$p(x)$	$xp(x)$	$(x-\mu)^2 p(x)$
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

$E(X) = 474$

$\sigma^2 = 724$

$\sigma = 26.9072$

23.

x	$p(x)$	$xp(x)$	$(x-\mu)^2 p(x)$
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. $\sigma = 1.2490$

d. $P(X = 5) = 0.2$

e. $P(X \geq 2) = 0.9$

f. $P(X \leq 3) = 0.6$

g. $P(X < 2) = 0.1$

25. a.

x	$p(x)$	$xp(x)$	$(x-\mu)^2 p(x)$
\$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

27. a. Cereal A = \$200,000; Cereal B = \$276,000.

b. Cereal A = \$188,414.40; Cereal B = \$309,619.10

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 will vary.

Section 6.3

5. 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	$\frac{3}{24}$
3	$\frac{4}{24}$
4	$\frac{4}{24}$
5	$\frac{4}{24}$
6	$\frac{4}{24}$
7	$\frac{3}{24}$
8	$\frac{1}{24}$

d. $E(X) = 4.5$

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

Section 6.4

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. $\sigma = 0.9487$

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 strike.
$P(X = 4) = 0.0026$. There is a 0.26% chance that exactly 4 of the plants will strike.
$P(X = 7) = 0$.
There is a negligible chance that all 7 plants will strike.

c. $E(X) = 0.7$

d. $\sigma = 0.794$. The standard deviation is larger than the expected value. The standard deviation is expressed as the number of plants that strike. Answers may vary.

15. a. $P(X = 2) = 0.375$

b. $P(X = 4) = 0.0625$

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. $E(X) = 2.2222$,
$\sigma^2 = 1.7284$

Section 6.5

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. $P(X = 6) = 0.0771$

Section 6.6

5. a. X has a hypergeometric distribution with $N = 50$, $k = 3$ and $n = 10$.

b. $E(X) = 0.6$

c. $\sigma = 0.6785$

d. $P(X \geq 1) = 0.4959$

e. $P(X \leq 2) = 0.9939$

f. $P(X > 3) = 0$

7. a. $E(X) = 5$

b. $\sigma = 1.5076$

c. $P(X = 10) = 0.0006$

d. $P(X = 0) = 0.0006$

Chapter 6 Additional Exercises

1. a. $E(X) = 2.3$

b. $\sigma^2 = 1.41$

c. $\sigma = 1.1874$

d. $P(X = 4.0) = 0.15$

e. $P(X \geq 2.0) = 0.75$

f. $P(X \leq 1.0) = 0.25$

g. $P(X > 3.0) = 0.15$

3. a. $P(4 \leq X \leq 6) = 0.6563$

b. $P(X \geq 8) = 0.0547$

c. $P(X = 1) = 0.0098$

5. a. $E(X) = 1.6667$, $\sigma^2 = 1.3889$

b. $E(X) = 3.25$, $\sigma^2 = 2.4375$

c. $E(X) = 8.8$, $\sigma^2 = 1.056$

d. $E(X) = 0.8333$, $\sigma^2 = 0.4419$

e. $E(X) = 3.5$, $\sigma^2 = 2.9167$

7. $P(X \geq 1) = 0.9615$

9. \$0.25

11. $P(X \geq 1) = 0.9933$

13. $P(X \geq 3) = 0.2962$

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. $P(X = 3) = 0.25$

Chapter 7

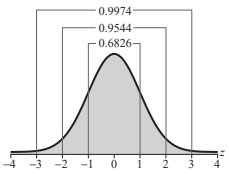
Section 7.1

7. a. $\mu = 60$
b. $\sigma = 1.7321$
c. 0.3333
d. 0.5
e. 0.1667
f. 0

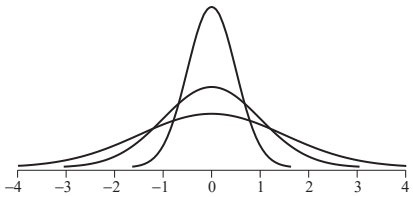
9. a. $\mu = 8{:}15$ am
b. $\sigma = 0.1443$
c. 0.3333
d. 0.1667
e. 0.5
f. 0

Section 7.2

9.

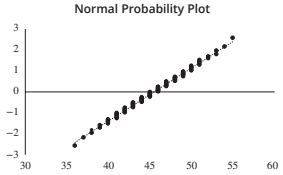


11.

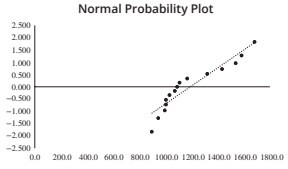


Section 7.3

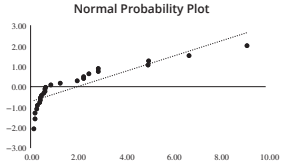
5. The data seem to fit a line very closely. We would assume from the normal probability plot that the population is normally distributed.



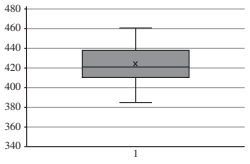
7. By examining the normal probability plot, we notice a substantial deviation from a linear pattern. The data do not appear to be normally distributed.

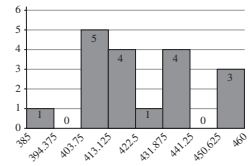


9. By examining the normal probability plot, we notice a substantial deviation from a linear pattern. The data do not appear to be normally distributed.

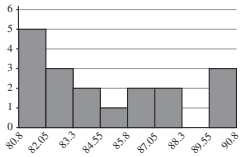


11. The data in this problem are not from a normal distribution. The histogram allows one to make a decision more easily.





13. When the data set is small the box plot and the histogram are not definitive in assessing normality. Care needs to be taken when dealing with statistics not to jump to conclusions. An interesting read is *How to Lie With Statistics* by Darrell Huff.



Section 7.4

5. a. 0.2486

b. 0.4500

c. 0.4750

d. 0.4950

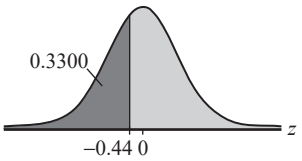
7. a. 0.6680

b. 0.6710

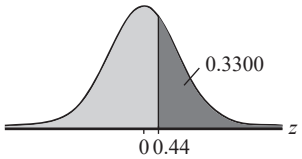
c. 0.9631

d. 0.9422

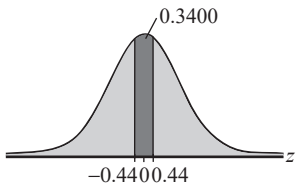
9. a. $P(z \le -0.44) = 0.3300$



b. $P(z \ge 0.44) = 0.3300$

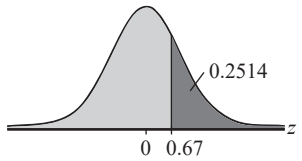


c. $P(-0.44 \le z \le 0.44) = 0.3400$

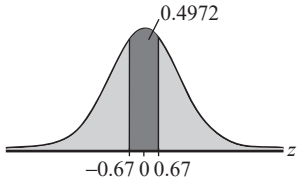


d. $P(z \le -0.67) = 0.2514$

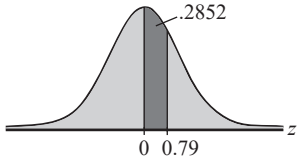
e. $P(z \ge 0.67) = 0.2514$



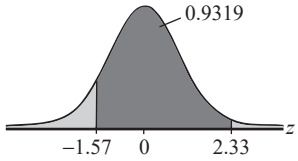
f. $P(-0.67 \le z \le 0.67) = 0.4972$



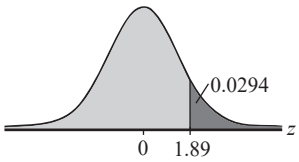
11. a. $P(0 \le z \le 0.79) = 0.2852$



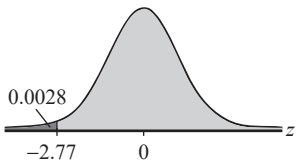
b. $P(-1.57 \le z \le 2.33) = 0.9319$



c. $P(z \ge 1.89) = 0.0294$



d. $P(z \le -2.77) = 0.0028$



13. 1.645

15. 1.28

17. −2.33

19. 1.14

21. 1.645

23. a. 0.7333

b. 0.0548

c. 0.0228

25. 0.4101

27. a. $631

b. 0.2206

c. 0.1190

d. 0.4235

e. Answers will vary.

29. a. 0.0668

b. 0.0668

c. 0.6826

31. 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.

33. a. At least 131

b. 0.13%

c. 6.5%

d. At least 135

Section 7.5

5. a. 45

b. 2.1213

c. 0

d. 0.9830

e. 0.0011

7. a. 120

b. 6.9282

c. 0

d. 0.5264

e. 0.0853

9. a. 2.2361

b. 0.0318

c. 0.0222

d. 0.0318 is the more accurate probability. When using the normal approximation, the probability is underestimated. Answers will vary.

11. a. 30

b. 5.4772

c. 0.9977

d. 0.5359

e. 0.8413

Chapter 7 Additional Exercises

1. a. 20

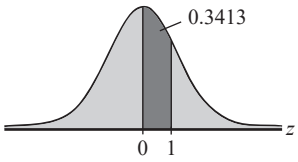
b. 8.6603

c. 0.1667

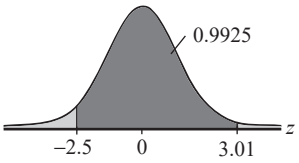
d. 57.73%

e. According to the empirical rule 68% of the results will fall within one standard deviation of the mean. The discrepancy is due to the fact that the empirical rule applies to distributions that are bell-shaped. The uniform distribution is not bell-shaped.

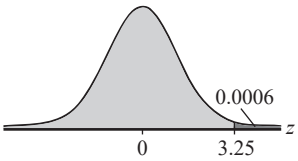
3. a. $P(0 \le z \le 1.00) = 0.3413$



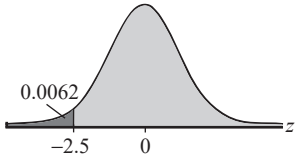
b. $P(-2.50 \le z \le 3.01) = 0.9925$



c. $P(z \ge 3.25) = 0.0006$



d. $P(z \le -2.50) = 0.0062$



5. 1.28

7. a. 5

b. 2.2349

c. Using the normal approximation: 0.9778; Using tables: 0.9933

d. Using the normal approximation: 0.9418; Using tables: 0.9596

e. Yes, $np = 5$, which is almost too small for the approximation to the binomial to produce accurate results. Answers may vary.

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

11. 535 days

13. $\mu \approx 5.71$ ml

15. The bulbs should be replaced after approximately 397 hours.

17. $\mu = \$61{,}000$, $\sigma = \$12{,}000$

Chapter 8

Section 8.1

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

b. People with strong opinions would be more likely to reply and the categories are not specific so people could have different ideas of what "Frequently" means, for example. Answers will 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.

15. a. Only students with very strong opinions may have responded, and the response categories are not well-defined. Answers will vary.

b. No. Only 150 surveys were mailed, and it is possible that only students with strong opinions responded. Answers will vary.

c. No. Nothing about business majors is mentioned so we don't know how many, if any, business majors responded to this survey. Answers will vary.

d. Instead of using a voluntary survey they could have surveyed all students in a particular class, for example. Answers will vary.

e. A question that asked whether or not the participant would enroll in the program if it were created, for example. Answers will vary.

17. a. Residents of the community

b. Post office, DMV, or the IRS, for example. Answers will vary.

c. Not everyone has a car so some would be excluded from the DMV frame. People move often so the frame obtained from the post office could be inaccurate. The IRS may be able to provide a list of tax payers for that community. This would probably be the best option since the politician is concerned about opinions on a tax increase. Answers will vary.

Section 8.2

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

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

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

13. 0.0174

15. a. 1 **b.** 1 **c.** 1 **d.** 0.0793

17. a. 0.2912 **b.** 0.0006

19. a. 0.0016 **b.** 0.0016 **c.** 0.9232

21. a. 0.2981 **b.** 0.0170 **c.** 0.6261

Section 8.3

7. a. 0.3

b. 0.7

9. a. $\mu_{\hat{p}} = 0.35,\ \sigma_{\hat{p}} = 0.0774$

b. $\mu_{\hat{p}} = 0.35,\ \sigma_{\hat{p}} = 0.0661$

c. $\mu_{\hat{p}} = 0.35,\ \sigma_{\hat{p}} = 0.0551$

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

11. 0.1151

13. a. 0.8461 **b.** 0.0011 **c.** 0.4793 **d.** 0.4989

15. a. 0.0023 **b.** 0.0384

17. a. 0.9767 **b.** 0.6536

Section 8.4

9. A systematic sample is not a random sample because it has potential for bias if there is some inherent data pattern.

11. a. Families in the state of Florida

b. The number of children per family

c. Ratio

d. Simple random sampling: develop sampling frame, choose participants randomly from sampling frame; Cluster sampling: create clusters, randomly select clusters, survey all members of the chosen clusters; Stratified sampling: create strata, randomly select participants from each stratum such that the population characteristics are adequately represented. Answers will vary.

e. Cluster sampling, because travel costs would likely be minimized. Answers will vary.

13. a. Convenience sampling

b. All of the people surveyed may not be residents of Orlando, Florida because there are a large number of tourists visiting the area.

c. No, the sample was not representative of the population of interest. Answers will vary.

Chapter 8 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 will 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 will vary.

c. No, because the sample is biased.

7. a. $\mu_{\bar{x}} = 15\%$, $\sigma_{\bar{x}} = 6.957\%$

b. 8.043% to 21.957%

9. a. 0.0031

b. Not necessarily. The samples may not have been representative of the population and employees may have not been honest about the amount of time they spend texting at work. Answers will vary.

c. Survey respondents might not be completely truthful when answering the survey since it is a representation of job performance. Answers will vary.

11. a. 0.0132

b. 0

c. No. Noise in excess of 103 decibels only occurs 1.32% of the time.

13. a. $\mu_{\hat{p}} = 0.90$, $\sigma_{\hat{p}} = 0.0134$

b. $\mu_{\hat{p}} = 0.85$, $\sigma_{\hat{p}} = 0.0179$

c. 0.8638

d. 0.9750

15. a. 0.9951

b. 0.8788

c. 0.1212

17. a. $P(\bar{x} > 35) \approx 1$, indicating that the researchers are likely correct in claiming that the U.S. average is greater than 35 hours (i.e., 50.4 hours).

b. $\sigma \approx 16.5$

Chapter 9

Section 9.1

25. a. 1.96

b. 2.575

c. 1.645

27. a. 2.33

b. 1.88

c. 1.75

29. (237, 263)

31. (4.7, 5.3)

33. (4964, 5036).

35. (29,232, 30,768)

37. (8.5, 10.7)

Section 9.2

15. 2.518

17. a. 2.201

b. 2.898

c. 1.721

19. (77.3579, 83.3887)

21. 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.

23. a. (101.36, 128.64)

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 \$101.36 and \$128.64. We are assuming the prices are normally distributed.

25. $n = 98$

27. $n = 31$

Section 9.3

9. a. 65%

b. (62.52, 67.48)

11. a. (0.2176, 0.3324)

b. No, 0.40 falls outside the confidence interval.

13. a. (0.3856, 0.5644)

b. Yes, 0.665 falls above the interval.

15. a. (0.0625, 0.1475)

b. Yes, 0.05 falls below the interval.

17. $n = 637$

19. $n = 208$

Section 9.4

5. a. (0.0557, 0.0941) We are 95% confident that the standard deviation of the bolt diameters is between 0.0557 inch and 0.0941 inch.

b. The diameters of the bolts have an approximately normal distribution.

7. a. (0.2841, 0.4608) We are 90% confident that the standard deviation of the share prices of the bond fund is between \$0.28 and \$0.46.

b. The share prices of the bond fund have an approximately normal distribution.

9. a. (5.1561, 7.2954) We are 80% confident that the standard deviation of the life of the touch screens is between 5.16 months and 7.30 months.

b. The life of the touch screens as measured by the consumer advocacy group has an approximately normal distribution.

Chapter 9 Additional Exercises

1. $n = 27$

3. a. $n = 68$

b. (0.2598, 0.3402) We are 95% confident that the true proportion of Fontana residents who think safety is a significant factor in their decision about whether or not to ride a bus is between 0.2598 and 0.3402.

5. (9.56, 9.94)

7. $n = 1038$

9. (29,630, 32,570)

11. (621, 679)

13. (0.3965, 0.4835)

Chapter 10

Section 10.1

15. $H_0: p = 0.47$, $H_a: p > 0.47$

17. $H_0: p = 0.29$, $H_a: p \neq 0.29$

19. $H_0: \mu = 10.9$, $H_a: \mu \neq 10.9$

21. $H_0: \mu = 56.8$, $H_a: \mu < 56.8$

23. a. $H_0: \mu = 30{,}000$, $H_a: \mu > 30{,}000$

Type I error: They will risk needing to replace tires.
Type II error: The company will research ways to make their tires last longer, even though that may be unnecessary.

b. $H_0: \mu = 240$, $H_a: \mu > 240$

Type I error: Mrs. Russell doesn't research ways to improve the bar hooks even though she may need to.
Type II error: Mrs. Russell will do research to improve the bar hooks, even though that may be unnecessary.

Section 10.2

5. a. $z = -2.33$

b. $z = 1.28$

c. $z = 1.96$ and -1.96

7. $H_0: \mu = 55$, $H_a: \mu \neq 55$, Critical values = -1.96, 1.96, $z = -8.95$, Reject H_0.

9. a. μ = the average number of days which the checks are late (or early).

b. H_0: $\mu = 0$, H_a: $\mu > 0$,

c. $z = 8.57$

d. Reject the null hypothesis if the calculated value of z is greater than or equal to 1.645.

e. Yes

f. There is sufficient evidence that the veterans organization's complaints are warranted and the checks arrive later than the 10th of the month, on average.

11. a. All computer systems sold by the retail computer store.

b. Service costs of the systems in the second year of operation.

c. Ratio

d. H_0: $\mu = 50$, H_a: $\mu \neq 50$

Critical values = –1.645, 1.645, $z = -10.39$, Reject H_0.

e. Answers will vary.

13. a. Yes. H_0: $\mu = 5$, H_a: $\mu < 5$, Critical value = –2.33, $z = -36.51$, Reject H_0.

b. 4.99 lb

15. No. H_0: $\mu = 5$, H_a: $\mu > 5$

Critical value = 1.28, $z = 0.47$, Fail to reject H_0.

17. a. IRS customers

b. Yes. H_0: $\mu = 45$, H_a: $\mu > 45$

Critical value = 1.28, $z = 13.33$, Reject H_0.

19. a. Reject H_0

b. Reject H_0

c. Fail to reject H_0

d. Reject H_0

21. a. P-value = 0.0228, Fail to reject H_0.

b. P-value = 0.0071, Reject H_0.

c. P-value = 0.0070, Reject H_0.

Section 10.3

7. a. $t = -2.624$

b. $t = 1.328$

c. $t = 2.365$ and -2.365

9. H_0: $\mu = 10$, H_a: $\mu \neq 10$

Critical values = –2.571, 2.571, $t = 1.337$, Fail to reject H_0.

11. H_0: $\mu = 0.5$, H_a: $\mu < 0.5$

Critical values = –1.440, $t = -1.542$, $0.05 < P\text{-value} < 0.10$ (tables), P-value = 0.0863 (exact), Reject H_0.

13. a. The patients of the Sisters of Mercy Hospital.

b. Yes. H_0: $\mu = \$1240$, H_a: $\mu > \$1240$

Critical values = 1.328, $t = 3.282$, Reject H_0.

c. The daily charges for patients of Sisters of Mercy Hospital have an approximately normal distribution.

15. a. No. H_0: $\mu = 5$, H_a: $\mu < 5$

Critical values = –2.583, $t = -0.825$, Fail to reject H_0.

b. The times customers spend watching the in-store video have an approximately normal distribution.

17. a. No. H_0: $\mu = 8$, H_a: $\mu > 8$

Critical values = 1.833, $t = 1.581$, Fail to reject H_0.

b. The times required to install 130 square feet of bathroom tile have an approximately normal distribution.

19. a. No. H_0: $\mu = 9$, H_a: $\mu < 9$

Critical values = –1.318, $t = -0.75$, Fail to reject H_0.

b. The diameters of the pellet patterns have an approximately normal distribution.

21. a. P-value = 0.0086, Reject H_0.

b. P-value = 0.0233, Fail to reject H_0.

c. P-value = 0.0652, Fail to reject H_0.

23. a. 0.0314

b. Yes, there is sufficient evidence to support the claim that the boots can remain immersed for more than 12 hours without leaking.

25. The test is statistically significant since the null hypothesis that the average daily growth of the shrub is equal to 1 cm per day was rejected. However, it is unlikely that a difference of 0.10 cm growth will be noticeable to the untrained eye. Answers will vary.

27. The test is not statistically significant because the conclusion was to fail to reject the null hypothesis that the average time customers watch the video is equal to 5 minutes The test is practically significant because it lets the store know that customers are spending time watching the new in-store video. Answers will vary.

Section 10.4

3. a. (40.09, 44.91), Reject H_0

b. Heights of children are approximately normally distributed.

5. a. (1120.82, 1213.86)

b. No

7. a. Since the hypothesized value of 18.27 falls outside the given confidence interval (20.36, 26.54), we reject the null hypothesis at a 5% level of significance. There is sufficient evidence to conclude that the population average completion time is different from 18.27 minutes.

b. Since the hypothesized value of 24.96 falls within the given confidence interval (20.36, 26.54), we fail to reject the null hypothesis at a 5% level of significance. There is not sufficient evidence to conclude that the population average completion time is different from 24.96 minutes.

c. Since the hypothesized value of 29.53 falls outside the given confidence interval (20.36, 26.54), we reject the null hypothesis at a 5% level of significance. There is not sufficient evidence to conclude that the population average completion time is different from 29.53 minutes.

9. Since the hypothesized value of 5020 falls within the confidence interval (4964.6, 5036), we fail to reject the null hypothesis at a 1% level of significance. There is not sufficient evidence to conclude that the true mean breaking strength of the metal link chain is different from 5020 pounds.

11. a. (5.6, 7.2)

b. Since the hypothesized value of 5.4 falls outside the confidence interval (5.6, 7.2), we reject the null hypothesis at a 5% level of significance. There is sufficient evidence to conclude that the true mean length of stay for patients having abdominal surgery is different from 5.4 days.

Section 10.5

5. a. $z = -1.645$

b. $z = 2.33$

c. $z = 1.645$ and -1.645

7. Yes. $H_0: p = 0.38$, $H_a: p > 0.38$

Critical values = 2.33, $z = 4.04$, Reject H_0.

9. a. Yes. $H_0: p = 0.013$, $H_a: p \neq 0.013$

Critical values = -1.96, 1.96, $z = 2.75$, Reject H_0.

b. Teenagers may not be honest when answering a survey like this. Answers will vary.

11. a. No. $H_0: p = 0.20$, $H_a: p < 0.20$

Critical values = -2.33, $z = -1.77$, Fail to reject H_0.

b. No, $np_0 \geq 5$ and $n(1 - p_0) \geq 5$.

13. a. No. $H_0: p = 0.05$, $H_a: p > 0.05$

Critical value = 1.28, $z = -0.46$, Fail to reject H_0.

b. Yes, $np_0 \geq 5$ and $n(1 - p_0) \geq 5$, but $np_0 = 5$, which could be a concern. Answers may vary.

15. No. $H_0: p = 0.15$, $H_a: p > 0.15$

Critical value = 2.33, $z = 0.31$, Fail to reject H_0.

17. a. No. $H_0: p = 0.002$, $H_a: p > 0.002$

Critical value = 1.645, $z = 0.12$, Fail to reject H0

b. 0.4522

c. No

19. a. $H_0: p = 0.40$, $H_a: p < 0.40$

$z = -0.73$, P-value = 0.2327

b. No

21. Yes. $H_0: p = 0.32$, $H_a: p < 0.32$

Critical value = -1.645, $z = -2.12$, Reject H_0.

23. a. Yes. $H_0: p = 0.49$, $H_a: p > 0.49$

Critical value = 1.645, $z = 3.82$, Reject H_0.

b. 110 people

Section 10.6

3. a. $df = 19$, $\chi^2 = 36.191$

b. $df = 23$, $\chi^2 = 35.172$

c. $df = 4$, $\chi^2 = 14.860$

5. a. Yes. $H_0: \sigma^2 = 0.0025$, $H_a: \sigma^2 > 0.0025$

Critical value = 42.557, $\chi^2 = 56.84$,
P-value = 0.0015, Reject H_0.

b. The diameters of the bolts have an approximately normal distribution.

7. a. Yes. $H_0: \sigma^2 = 0.0625$, $H_a: \sigma^2 > 0.0625$

Critical value = 42.980, $\chi^2 = 47.04$,
P-value = 0.0033, Reject H_0.

b. The share prices of the bond fund have an approximately normal distribution.

Chapter 10 Additional Exercises

1. a. $H_0: \mu = 18, H_a: \mu \neq 18$

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 $= -1.28$, $z = -4.34$, Reject H_0.

5. Yes. $H_0: \mu = 13.20, H_a: \mu < 13.20$

$z = -2.53$, P-value $= 0.0057$, Reject H_0.

7. No. $H_0: \mu = 895, H_a: \mu > 895$

Critical value $= 1.28$, $z = 1.19$, 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 $= 1.645$, $z = 0.90$, Fail to reject H_0.

13. a. P-value $= 0.0062$, Reject H_0.

b. P-value $= 0.0256$, Fail to reject H_0.

c. P-value $= 0.0002$, Reject H_0.

b. Delivery times are approximately normally distributed.

15. a. P-value $= 0.0464$, Reject H_0.

b. P-value $= 0.0050$, Reject H_0.

c. P-value $= 0.0510$, Fail to reject H_0.

17. Yes. $H_0: \sigma^2 = 0.00156, H_a: \sigma^2 > 0.00156$

Critical value $= 118.498$, $\chi^2 = 134.615$, 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. Yes. $H_0: \sigma^2 = 0.01, H_a: \sigma^2 < 0.01$

Critical value $= 3.325$, $\chi^2 = 1.44$, Reject H_0.

Chapter 11

Section 11.1

7. a. $z = -1.75$

b. $z = 1.41$

c. $z = 2.33$ and -2.33

9. a. $(-7.50, -2.50)$ We are 95% confident that Mr. Ellis' expenses are between \$2.50 and \$7.50 less than Mr. Ford's.

b. Yes. $H_0: \mu_1 - \mu_2 = 0, H_a: \mu_1 - \mu_2 \neq 0,$

Critical values $= -1.96, 1.96$, $z = -3.93$, Reject H_0.

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

11. a. Yes. $H_0: \mu_1 - \mu_2 = 0, H_a: \mu_1 - \mu_2 < 0,$

Critical value $= -1.28$, $z = -1.283$, Reject H_0.

b. P-value $= 0.0997$

c. Yes, we would fail to reject H_0 at $\alpha = 0.05$.

Section 11.2

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

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

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

7. a. $(-8.19, -2.21)$ We are 95% confident that the Dodge Grand Caravan ES takes between 8.19 and 2.21 fewer seconds to accelerate from 0 to 60 mph.

b. Yes. $H_0: \mu_1 - \mu_2 = 0, H_a: \mu_1 - \mu_2 \neq 0,$

Critical values $= -2.048, 2.048$, $t = -3.560$,

Reject H_0.

c. An independent experimental design is used, both populations are approximately normal, population variances are equal.

9. a. $(-3.19, 1.19)$ We are 99% confident that the hourly wage in City A is between \$3.19 lower and \$1.19 higher than in City B.

b. No. $H_0: \mu_1 - \mu_2 = 0, H_a: \mu_1 - \mu_2 \neq 0,$

Critical values $= -2.024, 2.024$, $t = -1.240$,

Fail to reject H_0.

c. P-value $= 0.2226$

d. An independent experimental design is used, both populations are approximately normal, population variances are equal.

11. a. An independent experimental design is used, both populations are approximately normal, population variances are equal.

b. Yes. H_0: $\mu_1 - \mu_2 = 0$, H_a: $\mu_1 - \mu_2 > 0$,
Critical value = 1.725, $t = 1.868$, Reject H_0.

13. a. No. H_0: $\mu_1 - \mu_2 = 0$, H_a: $\mu_1 - \mu_2 > 0$,
Critical value = 2.764, $t = 0.585$, Fail to reject H_0.

b. An independent experimental design is used, both populations are approximately normal, population variances are equal.

c. H_0: $\mu_1 - \mu_2 = 0$, H_a: $\mu_1 - \mu_2 > 0$,
Critical value = 2.764, $t = 0.585$, Fail to reject H_0.

d. The results of the hypothesis test did not change. Answers will vary.

Section 11.3

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 register and the new register so the samples can be paired. Answers will vary.

b. The differences have an approximately normal distribution.

c. Answers will vary.

d. (–4.15, 0.43) With 95% confidence, cashiers using the old cash register process between 4.15 fewer and 0.43 more items than using the new cash register.

e. Yes. H_0: $\mu_d = 0$, H_a: $\mu_d < 0$,
Critical value = –1.943, $t = -1.983$, Reject H_0.

Section 11.4

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, $z = -0.91$, Fail to reject H_0.
There is not sufficient evidence to support the fund-raiser's theory.

b. P-value = 0.1814. This is the probability of making a Type I error. Answers will vary.

c. (−0.1098, 0.0403) We are 95% confident that the proportion of men who answered "Yes" when asked to donate to a worthy cause is between 0.1098 less than and 0.0403 greater than the proportion of women who answered "Yes."

11. a. Yes, the sample sizes are sufficiently large.(−0.261, −0.149) We are 99% confident that the proportion of people age 30 or younger that believe alcoholic beverage commercials are targeted at teenagers is between 0.149 and 0.261 lower than the proportion of people in the older than 30 age group.

b. Yes. $H_0: p_1 - p_2 = 0, H_a: p_1 - p_2 < 0$,
Critical value = –2.33, $z = -9.22$, Reject H_0.

Section 11.5

7. a. 5.6864

b. 3.8807

c. 2.5289

d. 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. $F_{0.950} = 0.3744$; reject H_0 if $F \le 0.3744$; reject H_0.

b. $F_{0.010} = 3.0558$; reject H_0 if $F \ge 3.0558$; fail to reject H_0.

c. $F_{0.975} = 0.4148$, $F_{0.025} = 2.2505$; reject H_0 if $F \le 0.4148$ or $F \ge 2.2505$; reject H_0.

13. $H_0: \sigma_1^2 = \sigma_2^2; H_a: \sigma_1^2 > \sigma_2^2$

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

c. $F \approx 1.4067$

d. $F_{0.100} = 1.9532$; reject H_0 if $F \ge 1.9532$.
P-value ≈ 0.2543; fail to reject H_0. At the 0.10 level of significance, there is not sufficient evidence 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.

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

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

c. $F \approx 0.9571$

d. $F_{0.995} = 0.1910$, $F_{0.005} = 4.9884$; reject H_0 if $F \le 0.1910$ or $F \ge 4.9884$. P-value ≈ 0.9490; fail to reject H_0. At the 0.01 level of significance, there is not sufficient evidence to support the coach's claim that the variance in heights of adult male basketball players is different than that of the general population of men.

Chapter 11 Additional Exercises

1. Yes. H_0: $\mu_1 - \mu_2 = 0$, H_a: $\mu_1 - \mu_2 < 0$,
Critical value = –1.645, z = –2.59, 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 will vary.

b. The differences have an approximately normal distribution.

c. Answers will vary.

d. No. H_0: $\mu_d = 0$, H_a: $\mu_d > 0$,
Critical value = 3.143, t = 1.769, Fail to reject H_0.

5. H_0: $\mu_1 - \mu_2 = 0$, H_a: $\mu_1 - \mu_2 \ne 0$,
Critical values = –2.878, 2.878, t = –1.388, Fail to reject H_0.

7. a. No. H_0: $\mu_1 - \mu_2 = 0$, H_a: $\mu_1 - \mu_2 > 0$,
Critical value = 1.645, z = 2.00, Reject H_0.

b. (0.08, 0.52)

9. H_0: $\mu_1 - \mu_2 = 10$, H_a: $\mu_1 - \mu_2 > 10$,
Critical value = 2.014, t = 0.380, Fail to reject H_0. There is not sufficient evidence that Arrangement A has a lower mean net annual income by at least $10.

11. H_0: $\mu_1 - \mu_2 = 0$, H_a: $\mu_1 - \mu_2 \ne 0$,
Critical values = –2.101, 2.101, t = 8.771, Reject H_0. There is sufficient evidence of a difference in the average weights.

13. a. H_0: $p = 0.10$, H_a: $p < 0.10$, Critical value = –1.645, 2009: z = –16.29, Reject H_0, 2010: z = –5.84, Reject H_0, 2011: z = 0.53, Fail to reject H_0.

b. (–0.0316, –0.0133) We are 95% confident that the proportion of travelers using tablets or e-readers in 2011 is between 1.33% and 3.16% greater than the proportion of travelers using tablets or e-readers in 2010.

c. Yes. H_0: $p_1 - p_2 = 0$, H_a: $p_1 - p_2 < 0$,
Critical value = –1.645, z = –4.77, Reject H_0. There is sufficient evidence that the proportion of travelers using tablets or e-readers has increased between 2010 and 2011.

15. a. Yes. H_0: $p_1 - p_2 = 0$, H_a: $p_1 - p_2 \ne 0$,
Critical values = –1.96, 1.96, z = 4.78, Reject H_0.

b. (0.0450, 0.0750) We are 95% confident that the percentage of people with maxed out credit cards with IQs of 90 is between 4.55% and 7.5% greater than the percentage of people with maxed out credit cards with IQs greater than 125.

c. An independent experimental design was used.

17. P-value = $P(F \le 0.5888) = 0.2211$. Fail to reject H_0

Chapter 12

Section 12.1

15. a. Probably not, the boxes overlap quite a bit. Variation appears to be large in comparison to the difference in the middle values. Answers will vary.

b. Probably, the variation in scores for juniors is larger than for freshmen but the middle value is greater for juniors than for freshmen. Answers will vary.

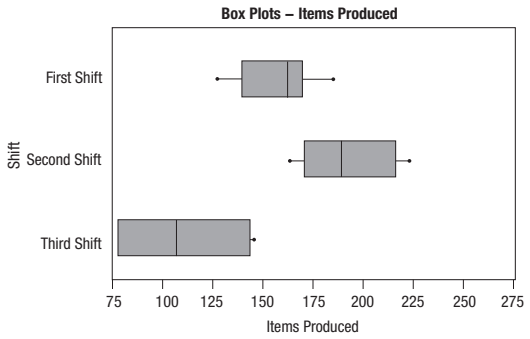
c. Probably not, the boxes overlap quite a bit. Variation appears to be large in comparison to the difference in the middle values. Answers will vary.

17. a. The experimental units are the items produced and the treatment is the shift.

b. First shift: mean = 158, median = 161; Second shift: mean = 191.8571, median = 182; Third shift: mean = 111.5714, median = 111

c. First shift: Min = 127, Max = 181, Q_1 = 140, Q_3 = 173,
Second shift: Min = 162, Max = 224, Q_1 = 168, Q_3 = 219,
Third shift: Min = 77, Max = 147, Q_1 = 77, Q_3 = 145,

d.



e. Probably, answers will vary.

f. Yes, answers will vary.

g. Probably, answers will vary.

h. Second shift, answers will vary.

19. a. 1.8464 is the variance of the sample means.

b. 2

c. 3.6928

d. 15

e. 14.1345

Section 12.2

9. 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.

11. 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.

13. Shapiro-Wilk Test: The *P*-values for each state in the test output are greater than 0.05 (Iowa: 0.1417; Hawaii: 0.5628; California: 0.6889).

Anderson-Darling Test: The *P*-values for each state in the test output are greater than 0.05 (Iowa: 0.1445; Hawaii: 0.5122; California: 0.4609).

Conclusion: The foot length measurements of all adult males come from normally distributed populations for each of the three states: Iowa, Hawaii, and California. The assumption of normality is satisfied.

15. For this survey the three samples of consumers are surveyed from a shopping center close to the surveyor's residence. Thus, the samples are selected through a convenience sample as the participants of the sample were "conveniently" selected from shoppers at a nearby shopping center. The three samples cannot be considered random samples. One of the assumptions of the ANOVA test is that the samples are randomly selected.

Therefore, the potential error of this ANOVA test is that the samples considered are not random samples. Thus, the results are not reliable.

17. It is possible that a student could be in more than one of these groups. In this case, the samples are not independent.

Section 12.3

11. a. Yes. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, H_a: at least one μ_i is different, $F = 16.8582$, $F_\alpha = 3.2389$, Reject H_0.

b. Hourly wages for employees are approximately normally distributed with equal variances. Observations were collected in an independent and random fashion. Answers will vary.

13. a. No. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, H_a: at least one μ_i is different, $F = 3.4667$, $F_\alpha = 4.0662$, 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 will vary.

15. a. Yes. H_0: $\mu_1 = \mu_2 = \mu_3$; H_a: at least one m_i is different, $F = 5.1012$, $F_\alpha = 2.5893$, 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 will 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 will vary.

Section 12.4

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

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

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. $df = n_T - k = 30 - 4 = 26$

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

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. MSE from the ANOVA output is 102.2.

$df = n_T - k = 30 - 3 = 27$

$$\frac{\alpha}{2} = \frac{0.1}{2} = 0.05$$

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

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

13. a. $df = n_T - k = 15 - 3 = 12$, $a = 0.05$

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.

Section 12.5

13. a. 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.

b. Yes. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, H_a: at least one m_i is different, $F = 696.8608$, $F_\alpha = 3.2874$, Reject H_0.

c. Yes. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$, H_a: at least one μ_i is different, $F = 101.7798$, $F_\alpha = 2.9013$, Reject H_0.

15. a. 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.

b. Yes. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, H_a: at least one μ_i is different, $F = 58.8261$, $F_\alpha = 6.9919$, Reject H_0.

c. Yes. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, H_a: at least one μ_i is different, $F = 15.2609$, $F_\alpha = 6.9919$, Reject H_0.

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

b. Yes. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, H_a: at least one m_i is different, $F = 9.9883$, $F_\alpha = 3.2874$, Reject H_0.

c. Yes. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$, H_a: at least one m_i is different, $F = 135.8303$, $F_\alpha = 2.9013$, Reject H_0.

Section 12.6

13. a. Yes, there appears to be interaction between airport location and major rental car company for all three cities.

b. Yes. H_0: There is no interaction, H_a: There is interaction, $F = 13.8127$, $F_\alpha = 2.9277$, Reject H_0.

c. We cannot test for effect of company on average daily rental rates because there is interaction.

15. a. There appears to be slight interaction between operator and machine. If there was no interaction, the lines would be parallel. Answers may vary.

b. 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.

c. Yes. H_0: $\mu_1 = \mu_2 = \mu_3$, H_a: at least one μ_i is different, $F = 4.5$, $F_\alpha = 2.6240$, Reject H_0.

d. Yes. H_0: $\mu_1 = \mu_2 = \mu_3$, H_a: at least one μ_i is different, $F = 14.2895$, $F_\alpha = 2.6240$, Reject H_0.

17. a.

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		

b. No. H_0: There is no interaction, H_a: There is interaction, $F = 0.1159$, $F_\alpha = 2.8131$, Fail to reject H_0.

c. Yes. H_0: $\mu_1 = \mu_2$, H_a: at least one μ_i is different, $F = 14.7246$, $F_\alpha = 4.0427$, Reject H_0.

d. No. H_0: $\mu_1 = \mu_2$, H_a: at least one μ_i is different, $F = 2.8986$, $F_\alpha = 4.0427$, Fail to reject H_0.

Chapter 12 Additional Exercises

1. a. Dividing the students of each class into blocks categorized as Below Average, Average, and Above Average. Answers will vary.

b. Dividing participants into blocks, categorized as <100 lb overweight, 50-100 lb overweight, and 0-50 lb overweight. Answers will vary.

c. Dividing the persons into blocks categorized as Low IQ, Average IQ, and High IQ. Answers will vary.

3. a. Yes. H_0: $\mu_1 = \mu_2 = \mu_3$, H_a: at least one μ_i is different, $F = 15.6429$, $F_\alpha = 6.3589$, 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 will vary.

c. The data are obtained from servings. There are no variables associated with the servings that we can block with. Answers will vary.

5. a. $\bar{\bar{x}} = 3.3064$, SST = 20.1246

b. $F = 35.3327$

c. Yes. H_0: $\mu_1 = \mu_2 = \mu_3$, H_a: at least one μ_i is different, $F = 35.3327$, $F_\alpha = 4.7623$, 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. H_0: The average median starting salary for all four types of majors is the same, H_a: At least one of the median starting salaries is different.

b. Yes, there is sufficient evidence that at least one of the starting salaries is different.

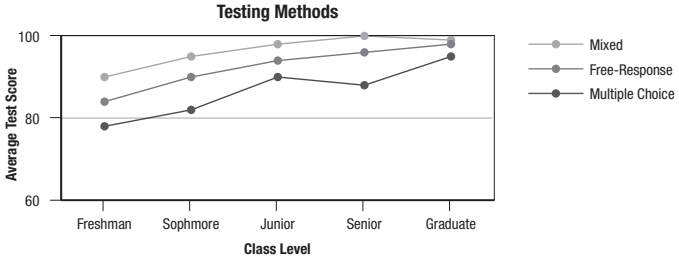
H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, H_a: at least one μ_i is different, $F = 17.1205$, $F_\alpha = 4.4156$, Reject H_0.

c. H_0: The average median mid-career salary for all four types of majors is the same, H_a: At least one of the median mid-career salaries is different.

d. Yes, there is sufficient evidence that at least one of the mid-career salaries is different.

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, H_a: at least one μ_i is different, $F = 10.9840$, $F_\alpha = 4.4156$, Reject H_0.

9. a. 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 will vary.



b. Yes. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$, H_a: at least one μ_i is different, $F = 22.0786$, $F_\alpha = 3.8379$, Reject H_0.

c. Yes, because the blocking effects were significant. Answers will vary.

11. 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 μ_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.

13. a. H_0: $\mu_1 = \mu_2 = \mu_3$, H_a: at least one μ_i is different.

b.

Source	SS	df	MS	F
Block	1103	3	367.6667	31.2908
Treatment	158.1667	2	79.0833	6.7305
Error	70.5	6	11.75	
Total	1331.6667	11		

c. Yes, at least one of the revenues is significantly different. H_0: $\mu_1 = \mu_2 = \mu_3$, H_a: at least one μ_i is different, $F = 6.7305$, $F_\alpha = 5.1433$, Reject H_0.

d. Yes, H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, H_a: at least one μ_i is different, $F = 31.2908$, $F_\alpha = 4.7571$, Reject H_0.

Chapter 13

Section 13.1

33. $\hat{y}_i = b_0 + b_1 x_i$

35. a. Sales volume, since it is the value we want to predict.

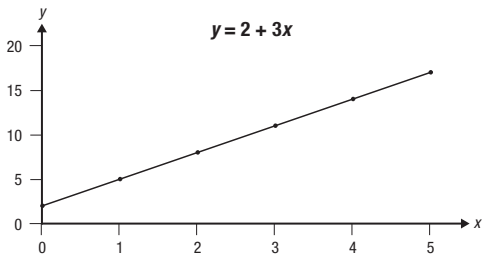
b. Advertising expenditures, since we use it to predict sales volume.

c. \$69,650

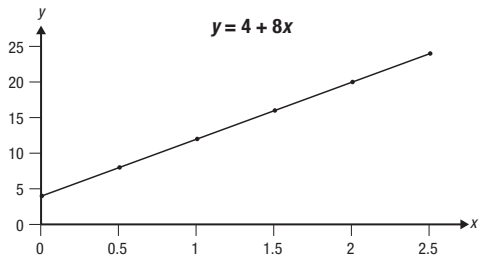
d. \$99,510

e. Random error. This may cause the company to under or over-estimate sales volume, causing budgeting problems. Answers will vary.

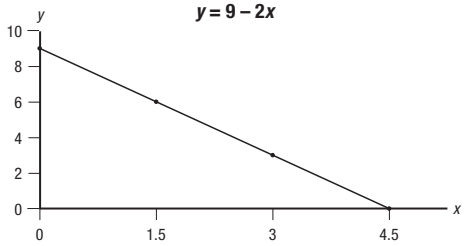
37. a.



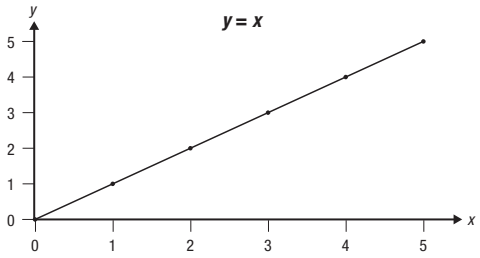
b.



c.



d.



39. a.

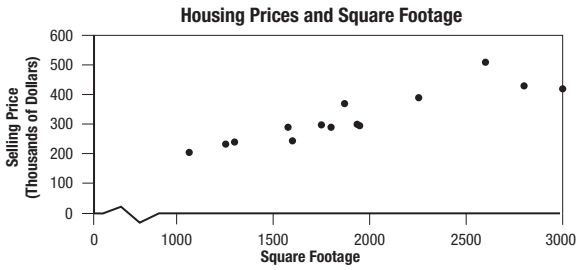
$\hat{y}$
15
45
65
85
95

b. Positive

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

41. a. Estimated Selling Price $= b_0 + b_1$(Square Footage)

b.

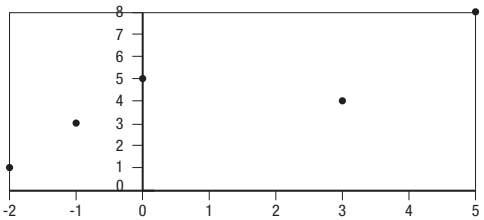


c.

Predicted Selling Price (Thousands of Dollars)	Error	Squared Error
201.45	–1.55	2.4025
227.91	0.09	0.0081
234.35	0.65	0.4225
273.13	11.87	140.8969
276.35	–37.35	1395.0225
297.35	–4.35	18.9225
304.35	–19.35	374.4225
314.15	50.85	2585.7225
323.25	–28.25	798.0625
325.07	–35.07	1229.9049
367.91	17.09	292.0681
416.35	88.65	7858.8225
444.35	–19.35	374.4225
472.35	–57.35	3289.0225

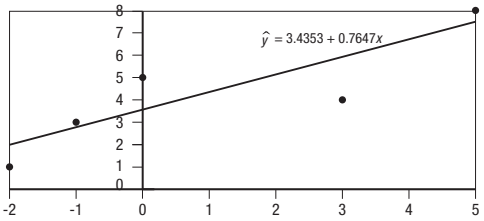
d. SSE = 18,360.123

43. a.



b. $\hat{y} = 3.4353 + 0.7647x$

c.



d. –0.9059, 0.3294, 1.5647, –1.7294, 0.7412

45. a. Current value of home

b. Annual salary

c. $112,811

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 $12,331.

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 will vary.

47. 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 will vary.

49. a.

Predicted *y*	Error	Squared Error
124.8372	–14.8372	220.1425
131.1462	3.8538	14.8518
145.8672	4.1328	17.0800
143.7642	5.2358	27.4136
147.9702	10.0298	100.5969
177.4122	–8.4122	70.7651

b. 450.8499

c. 112.7125

d. 10.6166

e. The standard error of the model is relatively small in comparison with the predicted *y*-values, so yes. Answers will vary.

51. a. $1.79 + 1.95 \cdot 5 + 1.57 \cdot 2 = 14.68$ minutes

b. 16–14.68=1.32

c. The residual of 1.32 means that it took 1.32 minutes more than the predicted time to deliver 5 pizzas at a distance of 2 miles.

Section 13.2

5. The given scatterplot of *y* on *x* shows that the linearity assumption is violated.

7. The plot of residuals of the regression model against the experience shows that the constant variance assumption is violated.

9. No, this model is not appropriate for predicting the price of the car using the age of the car because the residuals of the model show a nonlinear relationship.

11. The histogram of the regression residuals is closely normal, so it can be argued that residuals are normally distributed, and the assumption of the regression model is satisfied. However, one should interpret the model with caution.

13. The normal probability plot of the regression residuals shows that the points form an approximately straight line, and hence, the assumption of the normality of the regression residuals is satisfied.

15. a. The scatterplot of the data, along with the least squares line, shows there is a linear trend that as the number of kilometers run in the 4 weeks prior increases, the marathon time decreases.

b. The scatterplot of Marathon Time vs. Km Run in 4 Weeks Prior in a. shows that the data follow a linear pattern, validating the linearity assumption.

The residual plot shows a random scatter about $y = 0$, indicating that the linear fit is appropriate. Also, since it doesn't show a pattern, it upholds the assumption that the errors are independent of each other. As x increases the spread of the residuals remains fairly constant; therefore the assumption of equal variance is not violated.

By looking at the normal probability plot of the residuals, the residuals reasonably follow the diagonal line, indicating the errors are normally distributed.

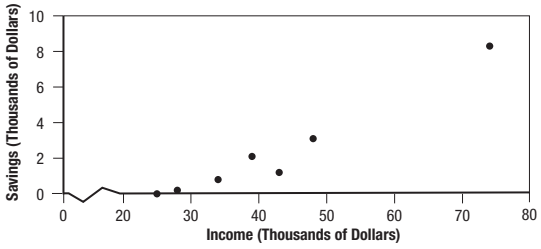
Section 13.3

13. a. Ratio

b. Savings

c. Income

d.



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

e. $\hat{y} = -4.8658 + 0.171x$

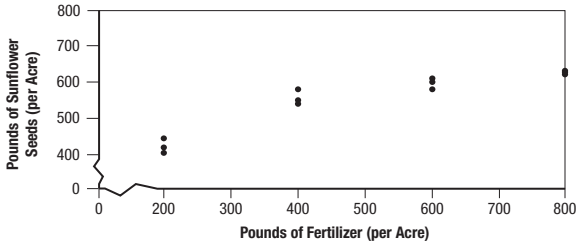
f. \$3684.20

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 variation annual income.

15. a. Controlled experiment

b.



c. Pounds of sunflower seeds because the researchers are interested in the effect of fertilizer on the yield of sunflower seeds.

d. $\hat{y} = 389 + 0.323x$

e. As the amount of fertilizer per acre increases by one pound, the yield of sunflower seeds is predicted to increase by 0.323 of a pound.

f. $R^2 = 0.8456$. Approximately 84.56% of the variation in the sunflower yield can be explained by the variation in the amount of fertilizer used.

g. 550.5 seeds

17. Coefficient of determination $= r^2 = 0.64^2 = 0.4096$.

Section 13.4

7. a. Independent variable: Month; Dependent variable: Sales

b. $\hat{y} = 403.7576 + 40.9091x$

c. MSE $= 2622.248$, $s_e = 51.2079$

d. \$935,575.90

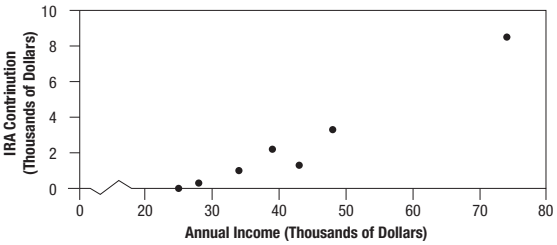
e. Approximately 90.12% of the variation in sales is explained by the time trend model, therefore the model seems to fit the data accurately.

Section 13.5

11. a. (33.1498, 48.6684)

b. We are 90% confident that the true increase in sales for each additional month is between \$33,149.80 and \$48,668.40.

13. a. The variables appear to have a positive correlation.



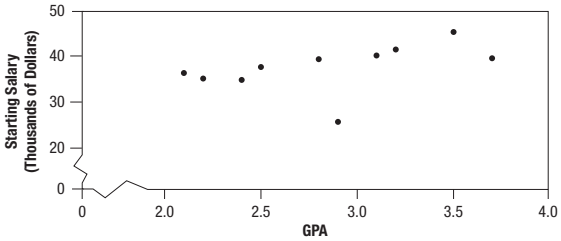
b. $\hat{y} = -4.8603 + 0.1740x$

c. (0.1293, 0.2186) We are 95% confident that the true increase in IRA contribution for each additional thousand dollars of income is between $129.30 and $218.60.

d. The error term is a normally distributed random variable, the expected value of the error term is zero, the variance of the error term is constant, and the errors are independent of each other.

e. We found the 95% confidence interval for the true slope (β_1) to be between 0.1293 and 0.2186. Our hypotheses in this case are H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 > 0$. Since the confidence interval is in entirely positive, we reject the null hypothesis that the true slope is zero. Thus, we conclude at 5% level of significance that the IRA contribution increases with the increase in income of the subject.

15. a. The two variables appear to have a positive correlation.



b. $\hat{y} = 24.9542 + 4.4175x$

c. No, H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$, $t = 1.471$, P-value = 0.1794, Fail to reject H_0.

d. $35,997.95

e. For each additional one point increase in GPA, starting salary is expected to increase by approximately $4417.50.

f. Approximately 21.3%

g. The error term is a normally distributed random variable, the expected value of the error term is zero, the variance of the error term is constant, and the errors are independent.

17. a. $\hat{y} = 42.5154 + 0.3914x$

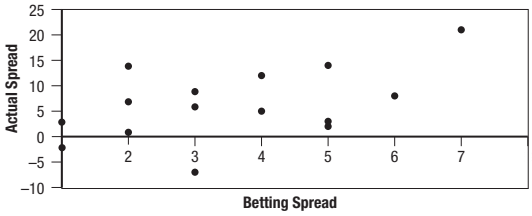
b. Approximately 20.76%

c. Yes, H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$, $t = 2.231$, P-value = 0.0379, Reject H_0.

d. 72

Section 13.6

5. a. There appears to be a weak linear relationship between actual spreads and betting spreads.



b. $\hat{y} = -0.3834 + 1.9198x$

c. No, H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$, $t = 2.043$, P-value = 0.0619, Fail to reject H_0.

d. Approximately 24.3%

e. The estimated increase in the actual spread for a one point increase in the betting spread is approximately 1.9198 points.

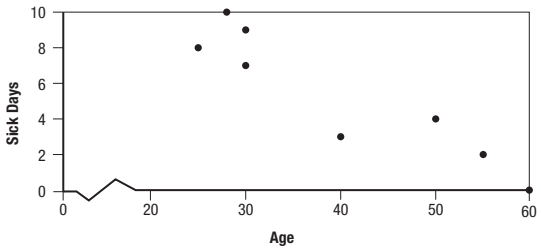
f. (–0.1104, 3.9500) We are 95% confident that the change in actual spread for a one point increase in betting spread is between –0.1104 and 3.9500 points.

g. 9.2156

h. (–5.27, 23.70) We are 95% confident that the actual spread when the betting spread is 5 is between –5.27 and 23.70.

i. (4.59, 13.85)

7. a. There appears to be a moderate negative linear relationship between age and sick days.



b. $s_e = 1.4765$. This is the estimated standard deviation of the errors associated with the model.

c. With each additional year of age, the number of sick says is expected to decrease by approximately 0.247.

d. Approximately 85.8%

e. Yes, H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$, $t = -6.01$, P-value = 0.001, Reject H_0.

f. (–0.3473, –0.1464) We are 95% confident that the true decrease in sick days for a one year increase in age is between 0.3473 and 0.1464 days.

g. Approximately 6.5 days

h. (5.184, 7.911) We are 95% confident that the average number of sick days for a 35-year-old employee is between 5.184 and 7.911.

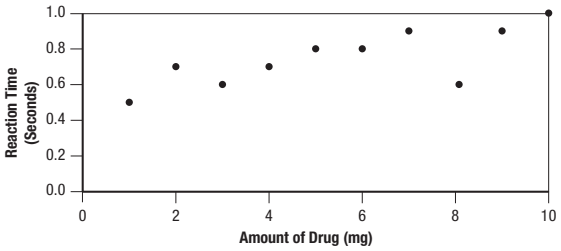
i. (2.686, 10.409) We are 95% confident that a new 35-year-old employee will take between 2.686 and 10.409 sick days.

j. 4 days

9. We have found the 99% confidence interval for the slope to be (–1809.1701, –973.0039). Our hypotheses in this case are H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 \neq 0$). Since the confidence interval does not include 0, we reject the null hypothesis that the true slope is zero. Thus, we conclude at 1% level of significance that there is a significant relation between the age of the car and its price.

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 milligram of the drug.

d. Approximately 56.9%. Other factors could include age, weight, etc. Answers will vary.

e. H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$, $t = 3.25$, P-value = 0.012, Reject H_0 at the 0.05 level, Fail to reject H_0 at the 0.01 level.

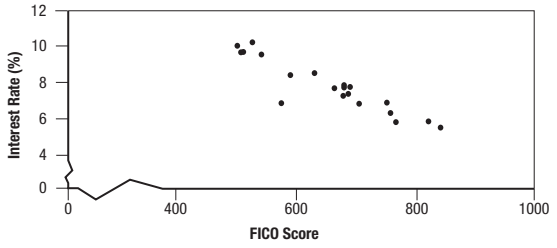
f. (0.011, 0.067) We are 95% confident that the true change in reaction time for each additional milligram of the drug is between 0.011 and 0.067 seconds.

g. 0.6909 seconds

h. (0.600, 0.782) We are 95% confident that the average reaction time for an individual with 4 mg 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-value = 0.0000, 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.

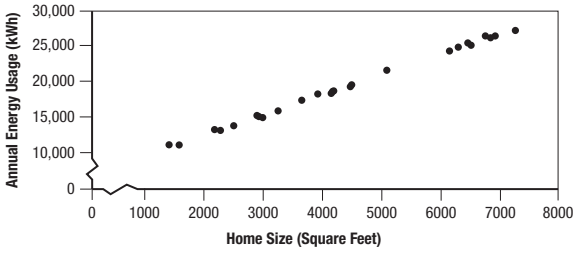
h. $r = -0.9244$. There is a strong negative linear relationship between FICO score and interest rate.

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-value = 0.0000, 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

Section 14.1

11. a. $b_0 = 6.1342$, $b_1 = 0.0108$, $b_2 = -0.0100$

b. $\hat{y} = 6.1342 + 0.0108x_1 - 0.0100x_2$

c. Yes, the coefficient of the number of pages is positive, indicating that more pages increases printing cost. The coefficient for the number of copies is negative, indicating that if you buy in bulk, the cost is less per book. The magnitudes of $0.01 per page and $0.01 per copy also seem reasonable. Answers may vary.

d. Type of paper, black & white vs. color printing, type of binding, etc. Answers will vary.

13. a. # of employees, average salary, advertising expenditures, research and development expenditures, charitable gifts to the community, etc. Answers will vary.

b. Revenue $= \beta_0 + \beta_1(\text{R\&D}) + \beta_2(\text{Advertising}) + \beta_3(\text{Salary Paid}) + \varepsilon_i$

c. Answers will vary. The coefficient could be positive because R&D expenditures may result in more and better products, meaning more sales, which would increase revenue. Alternatively, spending on R&D may reduce revenue if the resulting sales do not overcome the amount spent on development.

d. R&D expenditures and advertising expenditures are both costs that should increase revenue in the long run. However, in the short run, spending could decrease revenues, so the model may not be reliable. Answers will 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. His 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.

Section 14.2

11. a. $R_a^2 = 0.6607$

b. The adjusted R^2 for this model is larger than the adjusted R^2 from Exercise 10. The adjusted R^2 value indicates that approximately 66.07% of the variation in price is explained by the variation in the independent variables, compared with only 29.19% in the previous model. Answers will vary.

c. Yes, because the adjusted R^2 value is significantly larger for this model. Answers will vary.

13. a. Revenue $= \beta_0 + \beta_1(\text{Television}) + \beta_2(\text{Newspaper}) + \beta_3(\text{Mail}) + \varepsilon_i$

b. $\hat{y} = 73.9320 + 2.3830x_1 + 1.4544x_2 + 1.8160x_3$

c. For each additional $1000 spent on TV advertising, revenue is expected to increase by approximately $2383, assuming newspaper and mail advertising expenditures remain constant.

d. $R_a^2 = 0.8865$. Approximately 88.65% of the variation in revenue is explained by the variation in the three independent variables.

e. The adjusted R^2 value in this model is larger than the R^2 value in the previous model, so this model appears to be more useful. Answers may vary.

f. $R^2 = 0.9352$. The adjusted R^2 value should be used to compare this model to the simple model because additional independent variables have been added.

Section 14.3

17. a. H_0: $\beta_1 = \beta_2 = 0$, H_a: At least one $\beta_i \neq 0$.

b. $F = 10.8947$

c. Yes. P–value = 0.0001, Reject H_0.

d. (1466.6644, 4243.1181) We are 95% confident that the true increase in annual salary for each additional year of education is between \$1466.66 and \$4243.12.

e. H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$

f. Yes. $t = 4.140$, P–value = 0.0001, Reject H_0.

19. a. Rent = $\beta_0 + \beta_1$(Population) + β_2(Income) + ε_i

b. We would expect both coefficients to be positive since larger cities tend to have more expensive rental rates and more expensive rental rates would be expected in areas with greater incomes. Answers will vary.

c. $b_0 = 138.5023$; $b_1 = 0.1199$; $b_2 = 16.8207$; $\hat{y} = 138.5023 + 0.1199x_1 + 16.8207x_2$. A city with 0 population and 0 income would have an expected monthly rent of about \$138.50. For each additional 1000 people in the city, monthly rent is expected to increase by about \$0.12. For each additional \$1000 in average median income, monthly rent is expected to increase by about \$16.82.

d. Yes; H_0: $\beta_1 = \beta_2 = 0$; H_a: At least one $\beta_i \neq 0$; $F = 30.6224$; P–value = 0.0000; Reject H_0.

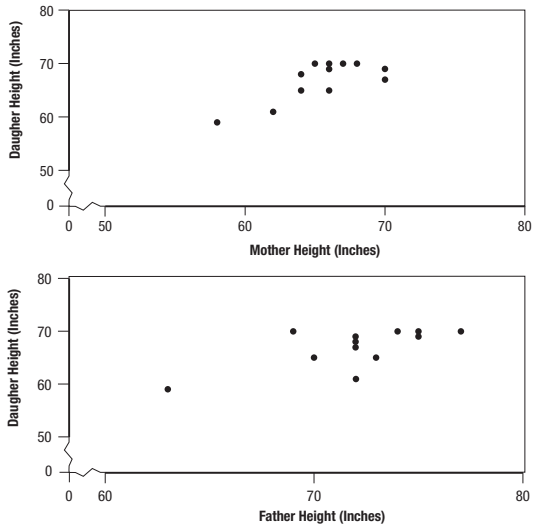
e. (11.4624, 22.1789); We are 95% confident that the true increase in monthly rent for each additional \$1000 increase in median income is between \$11.46 and \$22.18.

f. Population: No; H_0: $\beta_1 = 0$; H_a: $\beta_1 \neq 0$; $t = 1.423$; P–value = 0.1803; Fail to reject H_0. Income: Yes; $\beta_2 = 0$; H_a: $\beta_2 \neq 0$; $t = 6.840$; P–value = 0.0000; Reject H_0.

g. Yes, the population variable should be removed, as it is not a significant predictor of monthly rent. Answers may vary.

Section 14.4

7. a.



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. See plot on previous page.

b. $\hat{y} = -4.6456 + 0.5939x_1 + 0.4523x_2$

c. Yes; H_0: $\beta_1 = \beta_2 = 0$; H_a: At least one $\beta_i \neq 0$; $F = 11.2521$; P–value = 0.0028; Reject H_0.

d. No; H_0: $\beta_2 = 0$; H_a: $\beta_2 \neq 0$; $t = 2.176$; P–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–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. $\hat{y} = -15.2395 + 0.1319x_1 + 0.0869x_2$

b. Yes; H_0: $\beta_1 = \beta_2 = 0$; H_a: At least one $\beta_i \neq 0$; $F = 21.6521$; P–value = 0.0000; Reject H_0.

c. Approximately 59.89%. This is slightly lower than the percentage for the model with the 3 independent variables (59.99%). 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. Though the R^2 value is lower, the adjusted R^2 value is larger, indicating that the

difference is likely due to the addition of the first downs variable, which is not useful in predicting points scored. Answers will vary.

e. Approximately 4 (3.6890)

f. (–2.96, 11.16); We are 95% confident that the average points scored is between 0 and 11 points (since points scored cannot be negative in football) when the offense has 102 rushing yards and 63 passing yards.

g. (–12.70, 20.91); We are 95% confident that in this particular game against Miami, Buffalo will score between 0 and 21 points (since points scored cannot be negative in football).

Section 14.5

9. a. Yes, the overall model is significant. H_0: $\beta_1 = \beta_2 = \beta_3 = 0$, H_a: At least one $\beta_i \neq 0$, $F = 235.1310$, P–value = 0.0000, Reject H_0.

b. Approximately 51.06%

c. Yes. H_0: $\beta_3 = 0$, H_a: $\beta_3 \neq 0$, $t = -3.051$, P–value = 0.0024, Reject H_0.

d. Yes, all of the independent variables appear to be significant in predicting GPA. However, the R^2 value is only about 51%. This indicates that there may be other factors influencing GPA that have not been accounted for. Answers will vary.

e. Answers will vary. Quantitative: hours of study time, number of credit hours, SAT score, etc. Qualitative: major, gender, extracurricular activities, etc.

11. a.

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

b. $\hat{y} = 105.9195 + 2.5203x_1 - 0.0033x_2 - 65.3808x_3$

c. Yes. H_0: $\beta_1 = \beta_2 = \beta_3 = 0$, H_a: At least one $\beta_i \neq 0$, $F = 35.6335$, P–value = 0.0000, Reject H_0.

d. The coefficient for police is positive, indicating that as the number of police increases, so does the number of crimes. This is not what would be expected. The coefficient for enrollment is negative, indicating that as enrollment increases, the number of crimes decreases. This is also surprising, as one would think that a larger student body would result in increased crimes. The coefficient for private is negative, indicating that private schools tend to have less crimes than public schools. This is somewhat expected since private schools are more expensive and generally smaller in size. Answers will vary.

e. H_0: $\beta_3 = 0$, H_a: $\beta_3 \neq 0$, $t = -2.624$, P–value = 0.0184, Reject H_0 at the 0.05 level, so it supports the officials' belief, Fail to reject H_0 at the 0.01 level, so yes, the decision would change.

Section 14.6

11. a. Age and experience are likely correlated since the older you are, the more experience you tend to have. Answers will vary.

b. By using statistical software to find the correlation coefficient for each pair of variables. The closer the correlation coefficients are to 1 or –1, the more likely it is that collinearity exists. Answers will vary.

c. $58,921.91

d. Yes, because the overall model is not significant. None of the independent variables appear to be useful in predicting salary. Also, the age and experience values for which we are predicting salary are outside the range of the observed data, so extrapolation could be an issue. Answers will vary.

13. There appears to be a strong positive correlation between x_1 and x_3. There also appears to be a moderate positive correlation between x_1 and x_4 and x_3 and x_4. Answers will vary.

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

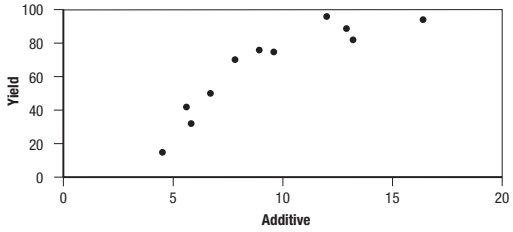
b. $\hat{y} = 10419.8636 + 1501.2807x_1 + 13047.4495x_2$

c. \$1501.28

d. Yes. H_0: $\beta_2 = 0$, H_a: $\beta_2 \neq 0$, $t = 3.845$, P–value = 0.0000, 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 \$13,047.45, which is greater than \$10,000.

e. There are many other factors that influence annual salary. Answers will vary.

3. a. The relationship appears to be positive, but it does not look linear. Answers will vary.



b. $\hat{y} = 5.9860 + 6.3361x_1$, $R^2 = 0.7971$, $s_e^2 = 163.5177$

c. Linear: 107.3638, Polynomial: 91.09. The observed value when the additive was 16.4 was a yield of 94, which is far closer to the polynomial result. Answers will vary.

d. The polynomial model has a higher R^2 value and a lower s_e^2 value, indicating that the polynomial model fits the data better than the linear model. Answers may vary.

e. The polynomial model because it has a higher R^2 value and a lower s_e^2 value. Answers will vary.

5. a. Answers may vary. GPA = $\beta_0 + \beta_1$(Xoom) + β_2(Galaxy) + β_3(iPad) + ε_i

b. Answers may vary. $\hat{y} = 3.4867 - 0.2761$(Xoom) $- 0.2503$(Galaxy) $- 0.4182$(iPad)

c. No. H_0: $\beta_1 = \beta_2 = \beta_3 = 0$, H_a: At least one $\beta_i \neq 0$, $F = 0.4740$, P–value = 0.7031, Fail to reject H_0. Answers may vary.

d. Answers may vary. $\hat{y} = 3.5473 + 0.1115$(\$30,000) $- 0.6059$(\$30,000–\$49,999) $- 0.3448$(\$50,000–\$74,999). This model is also not significant at the 0.05 level. There is not sufficient evidence that type of tablet or income is a significant predictor of GPA.

e. No, none of the tablet types are significant in predicting GPA. Answers will vary.

7. a. Benefit = $\beta_0 + \beta_1$(Family Size) + β_2(Income) + ε_i

b. Yes. $\hat{y} = 40.7903 + 3.6594x_1 + 0.1461x_2$, H_0: $\beta_1 = \beta_2 = 0$, H_a: At least one $\beta_i \neq 0$, $F = 397.9462$, P-value = 0.0000, Reject H_0.

c. Family size is significant at the 0.10 level ($t = 1.702$, P-value = 0.0972), and monthly income is significant at the 0.01 level ($t = 27.955$, P-value = 0.0000). Monthly income appears to be a much better predictor for benefits than family size. Answers may vary.

d. (410.70, 430.77), We are 95% confident that the average monthly benefit for a 4-person family with a monthly income of \$2500 is between \$410.70 and \$420.77.

e. (338.93, 502.53), We are 99% confident that the monthly benefit for a particular 4-person family with a monthly income of \$2500 is between \$338.93 and \$502.53.

f. The prediction interval is wider than the confidence interval by 143.53. This is due to the increased confidence level (99% vs. 95%) and the prediction interval accounts for individual variation. Answers may vary.

Chapter 15

Section 15.1

9. Yearly: Linear trend, big drop after 2013.

Monthly: More stationary. February is always the lowest point and August the highest.

11. There is a positive trend; however it is broken during 2008 (during the economy collapse) and looks like the sales numbers are reaching back to that level again.

13. The rate has been decreasing since 1981 and it seems to be stabilizing around 4%.

Section 15.2

7. 192,686.67
Passenger: 582,790.80

9. Truck: 128,487.40,
5-month WMA: 188,083.05

11. 4-year SMA: 4.11%, 4-year WMA: 4.23%

Section 15.3

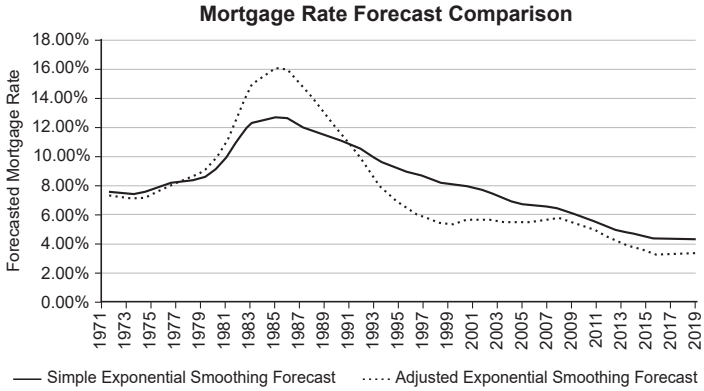
7.

Year	Truck Crossings	Forecast
2011	1,474,775.00	1,474,775.00
2012	1,541,150.00	1,474,775.00
2013	1,533,049.00	1,494,687.50
2014	1,554,152.00	1,506,195.95
2015	1,544,702.00	1,520,582.77
2016	1,598,017.00	1,527,818.54
2017	1,574,771.00	1,548,878.07
2018	1,581,443.00	1,556,645.95
2019		**1,564,085.07**

9. a. Simple exponential smoothing forecast for 2020: 4.35%

Adjusted exponential smoothing forecast for 2020: 3.45%

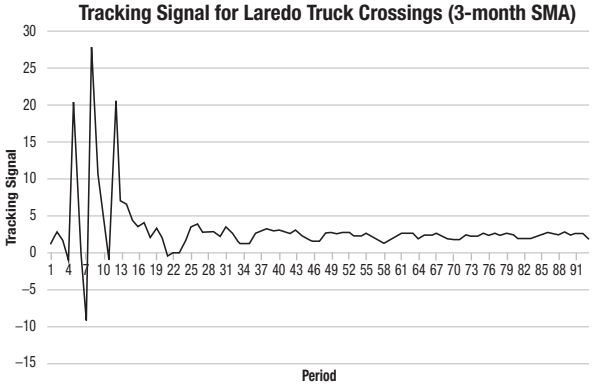
b. Until 1993, the adjusted exponential smoothing forecast is above the simple exponential smoothing forecast. After that, it's always below.



Section 15.4

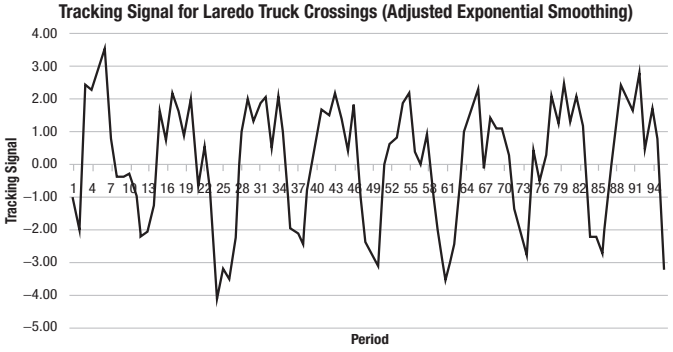
11. MAPD = 4.62%

From the tracking signal, we see it looks biased and out of control. Note in many places it is above 4. The forecast looks good from an MAPD point of view, but it is biased.



13. MAPD = 0.63%

No bias, all values of the *TS* are between 4 and –4. The forecast looks very accurate based on the MAPD since it's less than 1%. However, seasonality still needs to be addressed.



15. MAPD = 8.37%

17.

Error Metrics
MSE
MAPD (best), MAPE and MAD
TS and E
MAPD

Section 15.5

7.

Month	Seasonality Factor
Jan	0.0788
Feb	0.0781
Mar	0.0863
Apr	0.0837

May	0.0854
Jun	0.0851
Jul	0.0833
Aug	0.0864
Sep	0.0828
Oct	0.0893
Nov	0.0832
Dec	0.0775

9. Forecast = 380010.875 + 5335.250 Q_1 + 21527.000 Q_2 + 3351.625 Q_3

11. The summer months, particularly May and June are the highest.

Month	Seasonality Factor
Jan	0.0767
Feb	0.0781
Mar	0.0818
Apr	0.0849
May	0.0875
Jun	0.0878
Jul	0.0870
Aug	0.0866
Sep	0.0868
Oct	0.0839
Nov	0.0806
Dec	0.0782

Chapter 15 Additional Exercises

1. MAD = 0.8333
E(WMA) = –0.8

3. E(SMA) = –1.6667;

5. $SF_1 = 16.28$; $SF_2 = 11.63$; $SF_3 = 8.73$; $SF_4 = 21.53$

Chapter 16

Section 16.1

9. a. 41.401
b. 28.845
c. 5.024
d. 107.565
e. 24.769

11. a. 66.766
b. 27.488
c. 7.378
d. 33.196
e. 63.167

13. a. 0.042982
b. 30.0874
c. 16.013

Section 16.2

9. a. No.

$H_0: p_1 = p_2 = \cdots = p_5 = \frac{1}{5}$

H_a: Any possible difference.

$\chi^2_\alpha = 9.488, \chi^2 = 7.647$

P-value ≈ 0.1054

Fail to reject H_0

b. The distribution of the number of service calls has a multinomial probability distribution.

11. a. Yes.

$H_0: p_1 = p_2 = \cdots = p_{12} = \frac{1}{12}$

H_a: Any possible difference.

$\chi^2_\alpha = 24.725, \chi^2 = 26.376$

P-value ≈ 0.0057

Reject H_0

b. The distribution of fatal accidents has a multinomial probability distribution.

13. a. Yes.

$H_0: p_1 = 0.15, p_2 = p_3 = 0.30, p_4 = 0.25$

H_a: Any possible difference.

$\chi^2_\alpha = 11.345, \chi^2 = 28.000$

P-value ≈ 0.0000036

Reject H_0

b. The distribution of the survey responses has a multinomial probability distribution.

Section 16.3

13. a. Yes.
H_0: Age and political affiliation are independent.
H_a: Age and political affiliation are dependent.
$\chi_\alpha^2 = 9.488, \chi^2 = 31.881$
P-value ≈ 0.00000202
Reject H_0

b. Each variable satisfies the properties of a multinomial distribution.

15. a. No.
H_0: Type of abuse and gender are independent.
H_a: Type of abuse and gender are dependent.
$\chi_\alpha^2 = 7.815, \chi^2 = 3.575$
P-value ≈ 0.311164
Fail to reject H_0

b. Each variable satisfies the properties of a multinomial distribution.

Chapter 16 Additional Exercises

1. a. Yes.
$H_0: p_1 = 0.50, p_2 = p_3 = 0.25$
H_a: Any possible difference.
$\chi_\alpha^2 = 5.991, \chi^2 = 68.400$
P-value ≈ 0.0000
Reject H_0

b. The underlying distribution is a multinomial probability distribution.

3. Yes.
$H_0: p_1 = 0.51, p_2 = 0.33, p_3 = 0.16$
H_a: Any possible difference.
$\chi_\alpha^2 = 5.991, \chi^2 = 10.861$
P-value ≈ 0.00438
Reject H_0

5. No.
$H_0: p_1 = 0.226, p_2 = 0.611, p_3 = 0.075, p_4 = 0.088$
H_a: Any possible difference.
$\chi_\alpha^2 = 11.345, \chi^2 = 8.012$
P-value ≈ 0.045764
Fail to reject H_0

b. Each variable satisfies the properties of a multinomial distribution.

7. Yes.
H_0: Opinion and region are independent.
H_a: Opinion and region are dependent.
$\chi_\alpha^2 = 16.812, \chi^2 = 96.416$
P-value ≈ 0.0000
Reject H_0

Chapter 17

Section 17.1

15. H_0: Median = \$25,000, H_a: Median > \$25,000, Critical value = −1.28, $z = -3.25$, Reject H_0

17. a. H_0: Median = 14,400 hrs, H_a: Median < 14,400 hrs, Critical value = 2, $X = 6$, Fail to Reject H_0

b. The data are randomly selected.

19. H_0: # of Positive Signs = # of Negative Signs
H_a: # of Positive Signs > # of Negative Signs
Critical value = 6
$X = 9$
Fail to Reject H_0

21. H_0: # of Negative Signs = # of Positive Signs
H_a: # of Negative Signs ≠ # of Positive Signs
Critical value = 0
$X = 0$
Reject H_0

Section 17.2

11.

Mutual Fund	Price	Rank
American Funds	24.4	7.5
Columbia Management	9.41	2.5
Morgan Stanley	23.74	6
Fidelity Investments	24.40	7.5
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	9
The Vanguard Group	34.72	10

13. a. H_0: DHEA-S is the same
H_a: DHEA-S is increased
Critical value = 0
$X = 2$
Fail to Reject H_0

b. The data are randomly selected.

c. H_0: DHEA-S is the same
H_a: DHEA-S is increased
Critical value = 4
$T_+ = 5$
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 will vary.

15. a. The paired differences have an approximately normal distribution.

b. H_0: Model A = Model B
H_a: Model A ≠ 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.

Section 17.3

11. a. H_0: Mr. Ellis = Mr. Ford
H_a: Mr. Ellis ≠ Mr. Ford
Critical value = 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: Dramas = Comedies
H_a: Dramas > Comedies
Critical value = 36
$T_x = 32$
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.

Section 17.4

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 will 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

Section 17.5

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

Section 17.6

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 will 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 will 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

Chapter 18

Section 18.1

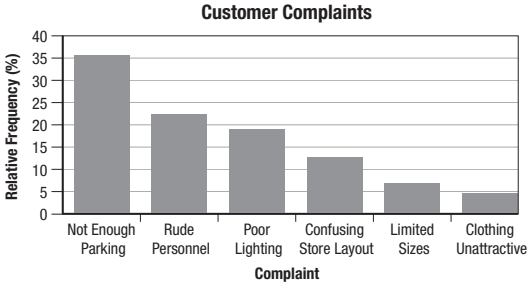
11. a. In order to assist customers in the best possible manner, the receptionist should understand how to handle various customer calls. Answers will vary.

b. The flowchart appears to be incomplete, since it does not tell you what to do in the case of something other than shipping questions, billing questions, problems with software function, and product information. The arrow following the "other" box goes nowhere. Answers will vary.

13. a.

Complaint	Relative Frequency (%)
Not Enough Parking	35.6
Rude Personnel	22.2
Poor Lighting	18.7
Confusing Store Layout	12.4
Limited Sizes	6.7
Clothing Unattractive	4.4

b.

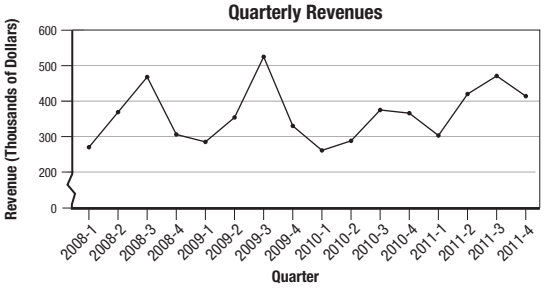


c. Not enough parking, rude personnel, poor lighting

d. No, 80% of the problems are explained by 50% of the causes. Answers will vary.

e. The store should implement more parking, better training for its employees, and better lighting. The parking issue should be addressed first because it is generating the largest percentage of complaints. Answers will vary.

15. a.



b. The revenues appear to be cyclical. Answers will vary.

c. Based on the run chart, yes. Answers will vary.

d. Families probably tend to eat out more when the weather is nicer. Answers will vary.

e. There is a cyclical pattern, so the process could be considered unstable. The restaurant could offer coupons or incentives in months where revenues tend to be lower. Answers will vary.

Section 18.2

9. a. UCL = 14.2, LCL = 5.8, Centerline = 10.0

b. The average value of the measurements is 10. Measurements between 5.8 and 14.2 are considered in control, and can be attributed to normal process variation.

c. B, F, H, K, N

11. a. UCL = 36, LCL = 16, Centerline = 26

b. The average value of the measurements is 26. Measurements between 16 and 36 are considered in control, and can be attributed to normal process variation.

c. G, K, M

Section 18.3

11. UCL = 31.125, LCL = 28.875

13. a. UCL = 4.3354, LCL = 3.6646, Centerline = 4

b. UCL = 4.6244, LCL = 0, Centerline = 2.1875

c. 1, 3, 4, 5, 8, 11, 13, 14, and 15

d. 13

15. a. UCL = 25.134, LCL = 23.046, samples 1, 4, 9, and 10 are out of control.

b. UCL = 9.1988, LCL = 2.4012

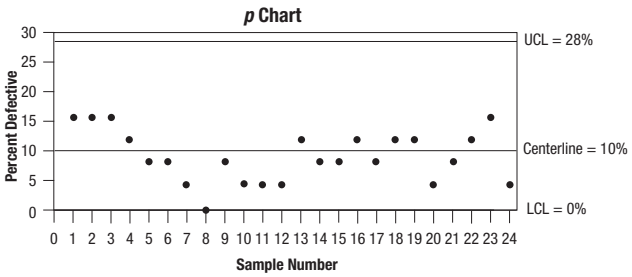
17. UCL = 0.235, LCL = 0.070, samples 5 and 9 are out of control.

Section 18.4

9. UCL = 3.98%, LCL = 0%

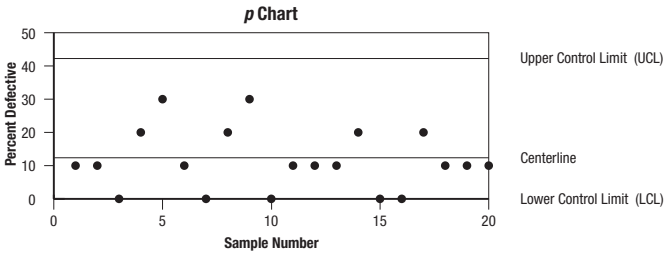
11. a. UCL = 28%, LCL = 0%, Centerline = 10%

b.



c. No, no samples are out of control.

13. UCL = 41.77%, LCL = 0%, Centerline = 11.5%, no samples are out of control.

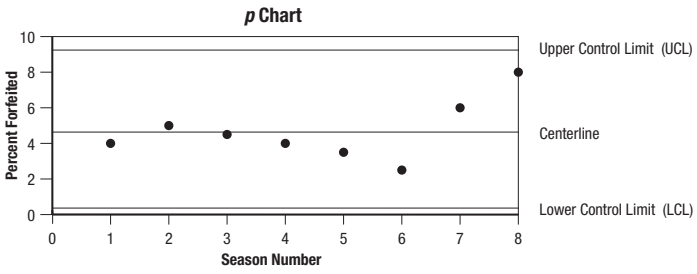


Chapter 18 Additional Exercises

1. a. UCL = 221.232, LCL = 218.768, Centerline = 220

b. UCL = 7.108, LCL = 0.892, Centerline = 4

3. UCL = 9.17%, LCL = 0.21%, there are no seasons out of control.



5. UCL = 54.03%, LCL = 0%, Centerline = 21.96%, there are no samples out of control, so the VP should not be concerned about the number of mistakes being made on financial statements.

7. $\bar{x}$ chart: UCL = 42.4908, LCL = 39.1800,

Centerline = 40.8354, no samples are out of control.

R chart: UCL = 6.0651, LCL = 0, Centerline = 2.869, no samples are out of control.