Chapter 1 Solutions

2. a. The 10 elements are the 10 tablet computers

b. Five variables: Cost ($), Operating System, Display Size (inches), Battery Life (hours), and CPU Manufacturer

c Categorical variables: Operating System and CPU Manufacturer

Quantitative variables: Cost ($), Display Size (inches), and Battery Life (hours)

d.

Variable	Measurement Scale
Cost ($)	Ratio
Operating system	Nominal
Display size (inches)	Ratio
Battery life (hours)	Ratio
CPU manufacturer	Nominal

4. a. There are eight elements in this data set; each element corresponds to one of the eight models of cordless telephones.

b. Categorical variables: Voice Quality and Handset on Base

Quantitative variables: Price, Overall Score, and Talk Time

c. Price: ratio measurement

Overall score: interval measurement

Voice quality: ordinal measurement

Handset on base: nominal measurement

Talk time: ratio measurement

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

6. a. Categorical

b. Quantitative

c. Categorical

d. Quantitative

e. Quantitative

8. a. 762

b. Categorical

c. Percentages

d. .67(762) = 510.54
510 or 511 respondents said they want the amendment to pass.

10. a. Categorical

b. Percentages

c. 44 of 1080 respondents, approximately 4%, strongly agree with allowing drivers of motor vehicles to talk on a hand-held cell phone while driving.

d. 165 of the 1,080 respondents or 15% of said they somewhat disagree and 741 or 69% said they strongly disagree. Thus, there does not appear to be general support for allowing drivers of motor vehicles to talk on a hand-held cell phone while driving.

12. a. The population is all visitors coming to the state of Hawaii.

b. Because airline flights carry the vast majority of visitors to the state, the use of questionnaires for passengers during incoming flights is a good way to reach this population. The questionnaire actually appears on the back of a mandatory plants and animals declaration form that passengers must complete during the incoming flight. A large percentage of passengers complete the visitor information questionnaire.

c. Questions 1 and 4 provide quantitative data indicating the number of visits and the

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

number of days in Hawaii. Questions 2 and 3 provide categorical data indicating the categories of reason for the trip and where the visitor plans to stay.

14. a. The graph of the time series follows:

b. In 2007 and 2008, Hertz was the clear market share leader. In 2009 and 2010, Hertz and Avis have approximately the same market share. The market share for Dollar appears to be declining.

c. The bar chart for 2010 follows, based on cross-sectional data.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

16. a. Time series

b.

c. Sales appear to be increasing in a linear fashion.

18. a. 684/1,021; or approximately 67%

b. 612

c. Categorical

20. a. 43% of managers were bullish or very bullish; 21% of managers expected health care to be the leading industry over the next 12 months.

b. We estimate the average 12-month return estimate for the population of investment managers to be 11.2%.

c. We estimate the average over the population of investment managers to be 2.5 years.

22. a. The population consists of all clients who currently have a home listed for sale with the agency or who have hired the agency to help them locate a new home.

b. Some of the ways that could be used to collect the data are as follows:

- A survey could be mailed to each of the agency's clients.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

- Each client could be sent an e-mail with a survey attached.
- The next time one of the firms agents meets with a client they could conduct a personal interview to obtain the data.

24. a. This is a statistically correct descriptive statistic for the sample.

    b. An incorrect generalization because the data were not collected for the entire population.

    c. An acceptable statistical inference based on the use of the word *estimate*.

    d. Although this statement is true for the sample, it is not a justifiable conclusion for the entire population.

    e. This statement is not statistically supportable. Although it is true for the particular sample observed, it is entirely possible and even highly likely that at least some students will be outside the 65 to 90 range of grades.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 2 Solutions

2. a. 1 – (.22 + .18 + .40) = .20

   b. .20(200) = 40

   c/d.

Class	Frequency	Percent Frequency
A	.22(200) = 44	22
B	.18(200) = 36	18
C	.40(200) = 80	40
D	.20(200) = 40	20
Total	200	100

4. a. These data are categorical.

   b.

Website	Frequency	% Frequency
FB	8	16
GOOG	14	28
WIKI	9	18
YAH	13	26
YT	6	12
Total	50	100

   c. The most frequently visited website is google.com (GOOG); the second is yahoo.com (YAH).

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

6. a.

Network	Relative Frequency	% Frequency
ABC	6	24
CBS	9	36
FOX	1	4
NBC	9	36
Total:	25	100

b. For these data, NBC and CBS tie for the number of top-rated shows. Each has nine (36%)

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

of the top 25. ABC is third with six (24%) and the much younger FOX network has 1(4%).

8. a.

Position	Frequency	Relative Frequency
Pitcher	17	0.309
Catcher	4	0.073
1st base	5	0.091
2nd base	4	0.073
3rd base	2	0.036
Shortstop	5	0.091
Left field	6	0.109
Center field	5	0.091
Right field	7	0.127
	55	1.000

b. Pitchers (almost 31%)

c. 3rd base (3%–4%)

d. Right field (almost 13%)

e. Infielders (16 or 29.1%) to outfielders (18 or 32.7%)

10. a.

Rating	Frequency

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Excellent	187
Very good	252
Average	107
Poor	62
Terrible	41
Total	649

b.

Rating	Percent Frequency
Excellent	29
Very good	39
Average	16
Poor	10
Terrible	6
Total	100

c.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. At the Lakeview Lodge, 29% + 39% = 68% of the guests rated the hotel as excellent or very good, but 10% + 6% = 16% of the guests rated the hotel as poor or terrible.

e. The percent frequency distribution for the Timber Hotel follows:

Rating	Percent Frequency
Excellent	48
Very good	31
Average	12
Poor	6
Terrible	3
Total	100

At the Lakeview Lodge, 48% + 31% = 79% of the guests rated the hotel as excellent or very good, and 6% + 3% = 9% of the guests rated the hotel as poor or terrible.

Compared to ratings of other hotels in the same region, both of these hotels received very favorable ratings. But in comparing the two hotels, guests at the Timber Hotel provided somewhat better ratings than guests at the Lakeview Lodge.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

12.

Class	Cumulative Frequency	Cumulative Relative Frequency
Less than or equal to 19	10	.20
Less than or equal to 29	24	.48
Less than or equal to 39	41	.82
Less than or equal to 49	48	.96
Less than or equal to 59	50	1.00

14. a.

b/c.

Class	Frequency	Percent Frequency
6.0–7.9	4	20
8.0–9.9	2	10
10.0–11.9	8	40
12.0–13.9	3	15
14.0–15.9	3	15
	20	100

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

16. Leaf unit = 10

1	6
1	
1	0 2
2	
1	0 6 7
3	
1	2 2 7
4	
1	5
5	
1	0 2 8
6	
1	0 2 3
7	

18. a.

PPG	Frequency
10–12	1
12–14	3
14–16	7
16–18	19
18–20	9

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

20–22	4
22–24	2
24–26	0
26–28	3
28–30	2
Total	50

b.

PPG	Relative Frequency
10–12	0.02
12–14	0.06
14–16	0.14
16–18	0.38
18–20	0.18
20–22	0.08
22–24	0.04
24–26	0.00
26–28	0.06
28–30	0.04
Total	1.00

c.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

PPG	Cumulative Percent Frequency
Less than 12	2
Less than 14	8
Less than 16	22
Less than 18	60
Less than 20	78
Less than 22	86
Less than 24	90
Less than 26	90
Less than 28	96
Less than 30	100

d.

e. There is skewness to the right.

f. (11/50)(100) = 22%

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

20. a. Lowest = 12, Highest = 23

b.

Hours in Meetings per Week	Frequency	Percent Frequency (%)
11–12	1	4
13–14	2	8
15–16	6	24
17–18	3	12
19–20	5	20
21–22	4	16
23–24	4	16
	25	100

c.

The distribution is slightly skewed to the left.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

22. a.

No. U.S. Locations	Frequency	Percent Frequency
0–4,999	10	50
5,000–9,999	3	15
10,000–14,999	2	10
15,000–19,999	1	5
20,000–24,999	0	0
25,000–29,999	1	5
30,000–34,999	2	10
35,000–39,999	1	5
Total:	20	100

b.

c. The distribution is skewed to the right. The majority of the franchises in this list have fewer than 20,000 locations (50% + 15% + 15% = 80%). McDonald's, Subway, and

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

7-Eleven have the highest number of locations.

24.

Starting Median Salary

4	6 8
5	1 2 3 3 5 6 8 8
6	0 1 1 1 2 2
7	1 2 5

Mid-Career Median Salary

8	0 0 4
9	3 3 5 6 7
10	5 6 6
11	0 1 4 4 4
12	2 3 6

There is a wider spread in the mid-career median salaries than in the starting median salaries. Also, as expected, the mid-career median salaries are higher that the starting median salaries. The mid-career median salaries were mostly in the $93,000 to $114,000 range while the starting median salaries were mostly in the $51,000 to $62,000 range.

26. a.

2	1 4

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

2	6	7		
3	0	1	1	1
3	5	6	7	7
4	0	0	3	3
4	6	6	7	9
5	0	0	0	2
5	5	6	7	9
6	1	4		
6	6			
7	2			

b. Most frequent age group: 40-44 with 9 runners

c. 43 was the most frequent age with 5 runners

28. a.

		y				
		20–39	40–59	60–79	80–100	Grand Total
	10–29			1	4	5
x	30–49	2		4		6
	50–69	1	3	1		5
	70–90	4				4
	Grand Total	7	3	6	4	20

b.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

		y				
		20–39	40–59	60–79	80–100	Grand Total
	10–29			20.0	80.0	100
x	30–49	33.3		66.7		100
	50–69	20.0	60.0	20.0		100
	70–90	100.0				100

c.

		y			
		20–39	40–59	60–79	80–100
	10–29	0.0	0.0	16.7	100.0
x	30–49	28.6	0.0	66.7	0.0
	50–69	14.3	100.0	16.7	0.0
	70–90	57.1	0.0	0.0	0.0
	Grand Total	100	100	100	100

d. Higher values of *x* are associated with lower values of *y* and vice versa.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

30. a.

	Year					
Average Speed	1988–1992	1993–1997	1998–2002	2003–2007	2008–2012	Total
130–139.9	16.7	0.0	0.0	33.3	50.0	100
140–149.9	25.0	25.0	12.5	25.0	12.5	100
150–159.9	0.0	50.0	16.7	16.7	16.7	100
160–169.9	50.0	0.0	50.0	0.0	0.0	100
170–179.9	0.0	0.0	100.0	0.0	0.0	100

b. It appears that most of the faster average winning times occur before 2003. This could be the result of new regulations that take into account driver safety, fan safety, the environmental impact, and fuel consumption during races.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

32. a. Row percentages follow.

Region	Under $15,000	$15,000 to $24,999	$25,000 to $34,999	$35,000 to $49,999	$50,000 to $74,999	$75,000 to $99,999	$100,000 and Higher	Total
Northeast	12.72	10.45	10.54	13.07	17.22	11.57	24.42	100.00
Midwest	12.40	12.60	11.58	14.27	19.11	12.06	17.97	100.00
South	14.30	12.97	11.55	14.85	17.73	11.04	17.57	100.00
West	11.84	10.73	10.15	13.65	18.44	11.77	23.43	100.00

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The percent frequency distributions for each region now appear in each row of the table. For example, the percent frequency distribution of the West region is as follows:

Income Level	Percent Frequency
Under $15,000	11.84
$15,000 to $24,999	10.73
$25,000 to $34,999	10.15
$35,000 to $49,999	13.65
$50,000 to $74,999	18.44
$75,000 to $99,999	11.77
$100,000 and over	23.43
Total	100.00

b. West: 18.44 + 11.77 + 23.43 = 53.64%

South: 17.73 + 11.04 + 17.57 = 46.34%

c.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The largest difference appears to be a higher percentage of household incomes of $100,000 and higher for the Northeast and West regions.

d. Column percentages follow.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Region	Under $15,000	$15,000 to $24,999	$25,000 to $34,999	$35,000 to $49,999	$50,000 to $74,999	$75,000 to $99,999	$100,000 and Higher
Northeast	17.83	16.00	17.41	16.90	17.38	18.35	22.09
Midwest	21.35	23.72	23.50	22.68	23.71	23.49	19.96
South	40.68	40.34	38.75	39.00	36.33	35.53	32.25
West	20.13	19.94	20.34	21.42	22.58	22.63	25.70
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Each column is a percent frequency distribution of the region variable for one of the household income categories. For example, for an income level of $35,000 to $49,999 the percent frequency distribution for the region variable is as follows:

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Region	Percent Frequency
Northeast	16.90
Midwest	22.68
South	39.00
West	21.42
Total	100.00

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

34. a.

	Brand Revenue ($ billions)						
Industry	0–25	25–50	50–75	75–100	100–125	125–150	Total
Automotive and luxury	10	1	1		1	2	15
Consumer packaged goods	12						12
Financial services	2	4	2	2	2	2	14
Other	13	5	3	2	2	1	26
Technology	4	4	4	1	2		15
Total	41	14	10	5	7	5	82

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b.

Brand Revenue ($ Billion)	Frequency
0–25	41
25–50	14
50–75	10
75–100	5
100–125	7
125–150	5
Total	82

c. Consumer packaged goods have the lowest brand revenues; each of the 12 consumer packaged goods brands in the sample data had a brand revenue of less than $25 billion. Approximately 57% of the financial services brands (8 out of 14) had a brand revenue of $50 billion or greater, and 47% of the technology brands (7 out of 15) had a brand revenue of at least $50 billion.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d.

	One-Year Value Change (%)						
Industry	–60–41	-40–21	–20–1	0–19	20–39	40–60	Total
Automotive and luxury				11	4		15
Consumer packaged goods			2	10			12
Financial services		1	6	7			14
Other			2	20	4		26
Technology	1	3	4	4	2	1	15
Total	1	4	14	52	10	1	82

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

e.

One-Year Value Change (%)	Frequency
–60–41	1
–40–21	4
–20–1	14
0–19	52
20–39	10
40–60	1
Total	82

f. The automotive & luxury brands all had a positive one-year value change (%). The technology brands had the greatest variability.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

36. a.

b. There is a negative relationship between *x* and *y; y* decreases as *x* increases.

38. a.

b.

40. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. Colder average low temperature seems to lead to higher amounts of snowfall.

c. Two cities have an average snowfall of nearly 100 inches of snowfall: Buffalo, New York, and Rochester, New York. Both are located near large lakes in the state.

42. a.

b. After increasing in ages 25–34, smartphone ownership decreases with increasing age. The percentage of people with no cell phone increases with age. There is less variation across age groups in the percentage who own other cell phones.

c. Unless a newer device replaces the smartphone, we would expect smartphone ownership would become less sensitive to age. This would be true because current

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

users will become older and because the device will become to be seen more as necessity than luxury.

44. a.

Class	Frequency
800–999	1
1000–1199	3
1200–1399	6
1400–1599	10
1600–1799	7
1800–1999	2
2000–2199	2
Total	30

b. The distribution if nearly symmetrical. It could be approximated by a bell-shaped

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

curve.

c. Ten of 30, or 33%, of the scores are between 1400 and 1599. The average SAT score looks to be slightly more than 1500. Scores below 800 or above 2200 are unusual.

46. a.

Population in Millions	Frequency	% Frequency
0.0–2.4	15	30.0
2.5–4.9	13	26.0
5.0–7.4	10	20.0
7.5–9.9	5	10.0
10.0–12.4	1	2.0
12.5–14.9	2	4.0
15.0–17.4	0	0.0
17.5–19.9	2	4.0
20.0–22.4	0	0.0
22.5–24.9	0	0.0
25.0–27.4	1	2.0
27.5–29.9	0	0.0
30.0–32.4	0	0.0
32.5–34.9	0	0.0
35.0–37.4	1	2.0

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

37.5–39.9	0	0.0
More	0	0.0

b. The distribution is skewed to the right.

c. Fifteen states (30%) have a population less than 2.5 million. More than half of the states have populations of less than 5 million (28 states, or 56%). Only seven states have a population greater than 10 million (California, Florida, Illinois, New York, Ohio, Pennsylvania, and Texas). The largest state is California (37.3 million). and the smallest states are Vermont and Wyoming (600.000).

48. a.

Industry	Frequency	% Frequency
Bank	26	13%
Cable	44	22%
Car	42	21%

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Cell	60	30%
Collection	28	14%
Total	200	100%

b.

c. The cellular phone providers had the highest number of complaints.

d. The percentage frequency distribution shows that the two financial industries (banks and collection agencies) had about the same number of complaints. Also, new car dealers and cable and satellite television companies also had about the same number of complaints.

50. a.

Level of Education	Percent Frequency
High school graduate	32,773/65,644(100) = 49.93
Bachelor's degree	22,131/65,644(100) = 33.71
Master's degree	9003/65,644(100) = 13.71
Doctoral degree	1737/65,644(100) = 2.65

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Total	100.00

13.71 + 2.65 = 16.36% of heads of households have a master's or doctoral degree.

b.

Household Income	Percent Frequency
Less than $25,000	13,128/65,644(100) = 20.00
$25,000 to $49,999	15,499/65,644(100) = 23.61
$50,000 to $99,999	20,548/65,644(100) = 31.30
$100,000 and higher	16,469/65,644(100) = 25.09
Total	100.00

31.30 + 25.09 = 56.39% of households have an income of $50,000 or more.

c.

	Household Income			
Level of Education	Under $25,000	$25,000 to $49,999	$50,000 to $99,999	$100,000 and Higher
High School graduate	75.26	64.33	45.95	21.14
Bachelor's degree	18.92	26.87	37.31	47.46
Master's degree	5.22	7.77	14.69	24.86
Doctoral degree	0.60	1.03	2.05	6.53
Total	100.00	100.00	100.00	100.00

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

There is a large difference between the level of education for households with an income of less than $25,000 and households with an income of $100,000 or more. For instance, 75.26% of households with an income of less than $25,000 are households in which the head of the household is a high school graduate, but only 21.14% of households with an income level of $100,000 or more are households in which the head of the household is a high school graduate. It is interesting to note, however, that 45.95% of households with an income of $50,000 to $99,999 are households in which the head of the household his a high school graduate.

52 a.

	Size of Company			
Job Growth (%)	Small	Midsized	Large	Total
–10–0	4	6	2	12
0–10	18	13	29	60
10–20	7	2	4	13
20–30	3	3	2	8
30–40	0	3	1	4
60–70	0	1	0	1
Total	32	28	38	98

b. Frequency distribution for growth rate.

Job Growth (%)	Total
–10–0	12

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

0–10	60
10–20	13
20–30	8
30–40	4
60-70	1
Total	98

Frequency distribution for size of company.

Size	Total
Small	32
Medium	28
Large	38
Total	98

c. Cross-tabulation showing column percentages.

	Size of Company		
Job Growth (%)	Small	Midsized	Large
–10–0	13	21	5
0–10	56	46	76
10–20	22	7	11
20–30	9	11	5

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

30–40	0	11	3
60–70	0	4	0
Total	100	100	100

d. Cross-tabulation showing row percentages.

	Size of Company			
Job Growth (%)	Small	Midsized	Large	Total
–10–0	33	50	17	100
0–10	30	22	48	100
10–20	54	15	31	100
20–30	38	38	25	100
30–40	0	75	25	100
60–70	0	4	0	100

e. Twelve companies had negative job growth: 13% were small companies, 21% were midsized companies, and 5% were large companies. So in terms of avoiding negative job growth, large companies were better off than small and midsized companies. But even though 95% of the large companies had a positive job growth, the growth rate was below 10% for 76% of these companies. In terms of better job growth rates, midsized companies performed better than either small or large companies. For instance, 26% of the midsized companies had a job growth of at least 20% as compared to 9% for small companies and 8% for large companies.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

54. a.

	Percent Graduating													
Year Founded	35–40	40–45	45–50	50–55	55–60	60–65	65–70	70–75	75–80	80–85	85–90	90–95	95–100	Grand Total
1600–1649													1	1
1700–1749													3	3
1750–1799												1	3	4
1800–1849						1	2	4	2	3	4	3	2	21
1850–1899			1	2	4	3	11	5	9	6	3	4	1	49
1900–1949	1	1	1		1	3		3	2	4	1	1		18
1950–2000	1		1	3			2							7
Grand Total	2	1	3	5	5	7	15	12	13	13	8	9	10	103

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b.

	% Graduate													
Year Founded	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100	Grand Total
1600-1649													100.00	100
1700-1749													100.00	100
1750-1799												25.00	75.00	100
1800-1849						4.76	9.52	19.05	9.52	14.29	19.05	14.29	9.52	100
1850-1899			2.04	4.08	8.16	6.12	22.45	10.20	18.37	12.24	6.12	8.16	2.04	100
1900-1949	5.56	5.56	5.56		5.56	16.67		16.67	11.11	22.22	5.56	5.56		100
1950-2000	14.29		14.29	42.86			28.57							100

c. Older colleges and universities tend to have higher graduation rates.

56. a.

b. There appears to be a strong positive relationship between Tuition and Fees and Percent Graduating.

58. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Zoo attendance appears to be dropping over time.

b.

c. General attendance is increasing, but not enough to offset the decrease in member attendance. School membership appears fairly stable.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 3 Solutions

2. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{96}{6} = 16$

10, 12, 16, 17, 20, 21

$$\text{Median} = \frac{16+17}{2} = 16.5$$

4.

Period	Rate of Return (%)
1	–6.0
2	–8.0
3	–4.0
4	2.0
5	5.4

The mean growth factor over the five periods is:

$$\bar{x}_g = \sqrt[n]{(x_1)(x_2)\cdots(x_5)} = \sqrt[5]{(0.940)(0.920)(0.960)(1.020)(1.054)} = \sqrt[5]{0.8925} = 0.9775$$

So the mean growth rate (0.9775 – 1)100% = –2.25%.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

6. $\text{Mean} = \frac{\Sigma x_i}{n} = \frac{657}{11} = 59.73$

Median = 57 sixth item

Mode = 53 It appears three times.

8. a. Median = 80 or $80,000. The median salary for the sample of 15 middle-level managers working at firms in Atlanta is slightly lower than the median salary reported by the *Wall Street Journal*.

b. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{1260}{15} = 84$

Mean salary is $84,000. The sample mean salary for the sample of 15 middle-level managers is greater than the median salary. This indicates that the distribution of salaries for middle-level managers working at firms in Atlanta is positively skewed.

c. The sorted data are as follow:

53 55 63 67 73 75 77 80 83 85 93 106 108 118 124

$$i = \frac{p}{100}(n+1) = \frac{25}{100}(16) = 4$$

First quartile or 25th percentile is the value in position 4 or 67.

$$i = \frac{p}{100}(n+1) = \frac{75}{100}(16) = 12$$

Third quartile or 75th percentile is the value in position 12 or 106.

10. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{1318}{20} = 65.9$

Order the data from the lowest rating (42) to the highest rating (83)

Position	Rating	Position	Rating

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

1	42	11	67
2	53	12	67
3	54	13	68
4	61	14	69
5	61	15	71
6	61	16	71
7	62	17	76
8	63	18	78
9	64	19	81
10	66	20	83

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(20+1) = 10.5$$

Median or 50th percentile = 66 + . 5(67 − 66) = 66.5

Mode is 61.

b. $L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(20+1) = 5.25$

First quartile or 25th percentile = 61

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(20+1) = 15.75$$

Third quartile or 75th percentile = 71

c. $L_{90} = \frac{p}{100}(n+1) = \frac{90}{100}(20+1) = 18.9$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

90th percentile = 78 + .9(81 − 78) = 80.7

Ninety percent of the ratings are 80.7 or less; 10% of the ratings are 80.7 or greater.

12. a. The mean for the previous year is 34,182; the median for the previous year is 34,000; the mode for the previous year is 34,000.

b. The mean for the current year is 35,900; the median for the current year is 37,000; the mode for the current year is 37,000.

c. The data are first arranged in ascending order.

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(11+1) = 3$$

25th percentile = 33 (or 33,000)

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(11+1) = 9$$

75th percentile = 35 (or 35,000)

d. The data are first arranged in ascending order.

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(10+1) = 2.75$$

25th percentile = 33 + .75(35 – 33) = 34.5 (or 34,500).

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(10+1) = 8.25$$

75th percentile = 37 + .25(37 − 37) = 37 (or 37,000).

e. The mean, median, mode, Q1, and Q3 values are all larger for the current year than for the previous year. This indicates that there have been more consistently more downloads in the current year compared to the previous year.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

14. For the previous year:

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(50+1) = 12.75$$

First quartile or 25th percentile = 6.8 + .75(6.8 – 6.8) = 6.8

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(50+1) = 25.5$$

Second quartile or median = 8 + .5(8 −8) = 8

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(50+1) = 38.25$$

Third quartile or 75th percentile = 9.4 + . 25(9.6 – 9.4) = 9.45

For the current year:

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(50+1) = 12.75$$

First quartile or 25th percentile = 6.2 + .75(6.2 – 6.2) = 6.2

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(50+1) = 25.5$$

Second quartile or median = 7.3 + .5(7.4 − 7.3) = 7.35

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(50+1) = 38.25$$

Third quartile or 75th percentile = 8.6 + . 25(8.6 −8.6) = 8.6

It may be easier to compare these results if we place them in a table.

	Previous Year	Current Year

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

First Quartile	6.80	6.20
Median	8.00	7.35
Third Quartile	9.45	8.60

The results show that in the current year approximately 25% of the states had an unemployment rate of 6.2% or less, lower than in the previous year. And the median of 7.35% and the third quartile of 8.6% in the current year are both less than the corresponding values in the previous year, indicating that unemployment rates across the states are decreasing.

16. a.

Grade x_i	Weight W_i
4 (A)	9
3 (B)	15
2 (C)	33
1 (D)	3
0 (F)	0
	60 Credit Hours

$$\bar{x} = \frac{\Sigma w_i x_i}{\Sigma w_i} = \frac{9(4)+15(3)+33(2)+3(1)}{9+15+33+3} = \frac{150}{60} = 2.50$$

b. Yes; satisfies the 2.5 grade point average requirement.

18.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Assessment	Deans	f_iM_i	Recruiters	f_iM_i
5	44	220	31	155
4	66	264	34	136
3	60	180	43	129
2	10	20	12	24
1	0	0	0	0
Total	180	684	120	444

Deans: $\bar{x} = \frac{\Sigma_i fM_i}{n} = \frac{1745}{500} = 3.49$

Recruiters: $\bar{x} = \frac{\Sigma_i fM_i}{n} = \frac{1745}{500} = 3.49$

20.

	Stivers		Trippi	
Year	End of Year Value ($)	Growth Factor	End of Year Value ($)	Growth Factor
1	11,000	1.100	5,600	1.120
2	12,000	1.091	6,300	1.125
3	13,000	1.083	6,900	1.095
4	14,000	1.077	7,600	1.101
5	15,000	1.071	8,500	1.118

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

6	16,000	1.067	9,200	1.082
7	17,000	1.063	9,900	1.076
8	18,000	1.059	10,600	1.071

For the Stivers mutual fund, we have:

$$18000=10000\left[(x_1)(x_2)\cdots(x_8)\right], \text{ so } \left[(x_1)(x_2)\cdots(x_8)\right]=1.8$$

and

$$\overline{x}_g=\sqrt[n]{(x_1)(x_2)\cdots(x_8)}=\sqrt[8]{1.80}=1.07624$$

So the mean annual return for the Stivers mutual fund is (1.07624 – 1)100 = 7.624%.

For the Trippi mutual fund, we have:

$$10600=5000\left[(x_1)(x_2)\cdots(x_8)\right], \text{ so } \left[(x_1)(x_2)\cdots(x_8)\right]=2.12$$

and

$$\overline{x}_g=\sqrt[n]{(x_1)(x_2)\cdots(x_8)}=\sqrt[8]{2.12}=1.09848$$

So the mean annual return for the Trippi mutual fund is (1.09848 – 1)100 = 9.848%.

Although the Stivers mutual fund has generated a nice annual return of 7.6%, the annual return of 9.8% earned by the Trippi mutual fund is far superior.

22. 25,000,000 = 10,000,000$\left[(x_1)(x_2)\cdots(x_6)\right]$, so $\left[(x_1)(x_2)\cdots(x_6)\right]$=2.50,

and so

$$\overline{x}_g=\sqrt[n]{(x_1)(x_2)\cdots(x_6)}=\sqrt[6]{2.50}=1.165$$

So the mean annual growth rate is (1.165 – 1)100 = 16.5%

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

24. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{75}{5} = 15$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{64}{4} = 16$$

$$s = \sqrt{16} = 4$$

26. a. $\bar{x} = \frac{\sum_i x_i}{n} = \frac{74.4}{20} = 3.72$

b. $s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1.6516}{20-1}} = \sqrt{.0869} = .2948$

c. The average price for a gallon of unleaded gasoline in San Francisco is much higher than the national average. This indicates that the cost of living in San Francisco is higher than it would be for cities that have an average price close to the national average.

28. a. The mean annual sales amount is $315,643, the variance is 13,449,632, and the standard deviation is $115,973.

b. Although the mean sales amount has increased from the previous to most recent fiscal year by more than $15,000, this amount is extremely small compared to the standard deviation in either fiscal year. Therefore, there is a strong likelihood that this change is the result of simple randomness rather than a true change in demand for these products.

30. Dawson Supply: Range = 11 – 9 = 2

$$s = \sqrt{\frac{4.1}{9}} = 0.67$$

J. C. Clark: Range = 15 – 7 = 8

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$s = \sqrt{\frac{60.1}{9}} = 2.58$$

32. a. Automotive:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{39201}{20} = 1960.05$$

Department store:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{13857}{20} = 692.85$$

b. Automotive:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{4,407,720.95}{19}} = 481.65$$

Department store:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{456804.55}{19}} = 155.06$$

c. Automotive: 2901 – 598 = 2303

Department store:

1011 – 448 = 563

d. Order the data for each variable from the lowest to highest.

	Automotive	Department Store
1	598	448
2	1,512	472

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

3	1,573	474
4	1,642	573
5	1,714	589
6	1,720	597
7	1,781	598
8	1,798	622
9	1,813	629
10	2,008	669
11	2,014	706
12	2,024	714
13	2,058	746
14	2,166	760
15	2,202	782
16	2,254	824
17	2,366	840
18	2,526	856
19	2,531	947
20	2,901	1011

$$i = \frac{p}{100}(n+1) = \frac{25}{100}(21) = 5.25$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Automotive: First quartile or 25th percentile =

$$1{,}714 + .25(1{,}720 - 1{,}714) = 1{,}715.5$$

Department store: First quartile or 25th percentile =

$$589 + .25(597 - 589) = 591$$

$$i = \frac{p}{100}(n+1) = \frac{75}{100}(21) = 15.75$$

Automotive: Third quartile or 75th percentile =

$$2{,}202 + .75(2{,}254 - 2{,}202) = 2{,}241$$

Department store: First quartile or 75th percentile =

$$782 + .75(824 - 782) = 813.5$$

e. Automotive spends more on average, has a larger standard deviation, larger max and min, and larger range than department store. Autos have all new model years and may spend more heavily on advertising.

34. Quarter milers

$$s = 0.0564$$

Coefficient of variation =

$$(s/\bar{x})100\% = (0.0564/0.966)100\% = 5.8\%$$

Milers

$$s = 0.1295$$

Coefficient of variation =

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$(s/\bar{x})100\% = (0.1295/4.534)100\% = 2.9\%$$

Yes; the coefficient of variation shows that as a percentage of the mean the quarter milers' times show more variability.

36.

$$z = \frac{520-500}{100} = +.20$$

$$z = \frac{650-500}{100} = +1.50$$

$$z = \frac{500-500}{100} = 0.00$$

$$z = \frac{450-500}{100} = -.50$$

$$z = \frac{280-500}{100} = -2.20$$

38. a. Approximately 95%

    b. Almost all

    c. Approximately 68%

40. a. \$3.33 is one standard deviation below the mean, and \$3.53 is one standard deviation above the mean. The empirical rule says that approximately 68% of gasoline sales are in this price range.

    b. Part (a) shows that approximately 68% of the gasoline sales are between \$3.33 and \$3.53. Because the bell-shaped distribution is symmetric, approximately half of 68%, or 34%, of the gasoline sales should be between \$3.33, and the mean price of \$3.43. \$3.63 is two standard deviations above the mean price of \$3.43. The empirical rule

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

says that approximately 95% of the gasoline sales should be within two standard deviations of the mean. Thus, approximately half of 95%, or 47.5%, of the gasoline sales should be between the mean price of \$3.43 and \$3.63. The percentage of gasoline sales between \$3.33 and \$3.63 should be approximately 34% + 47.5% = 81.5%.

c. \$3.63 is two standard deviations above the mean, and the empirical rule says that approximately 95% of the gasoline sales should be within two standard deviations of the mean. Thus, 1 – 95% = 5% of the gasoline sales should be more than two standard deviations from the mean. Because the bell-shaped distribution is symmetric, we expected half of 5%, or 2.5%, would be more than \$3.63.

42. a. $z = \frac{x-\mu}{\sigma} = \frac{2300-3100}{1200} = -.67$

b. $z = \frac{x-\mu}{\sigma} = \frac{4900-3100}{1200} = 1.50$

c. \$2300 is .67 standard deviations below the mean. \$4900 is 1.50 standard deviations above the mean. Neither is an outlier.

d. $z = \frac{x-\mu}{\sigma} = \frac{13000-3100}{1200} = 8.25$

\$13,000 is 8.25 standard deviations above the mean. This cost is an outlier.

44. a. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{765}{10} = 76.5$

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{442.5}{10-1}} = 7$$

b. $z = \frac{x-\bar{x}}{s} = \frac{84-76.5}{7} = 1.07$

This is approximately one standard deviation above the mean. Approximately 68% of

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

the scores are within one standard deviation. Thus, half of (100–68), or 16%, of the games should have a winning score of 84 or more points.

$$z = \frac{x - \bar{x}}{s} = \frac{90 - 76.5}{7} = 1.93$$

This is approximately two standard deviations above the mean. Approximately 95% of the scores are within two standard deviations. Thus, half of (100–95), or 2.5%, of the games should have a winning score of more than 90 points.

c. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{122}{10} = 12.2$

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{559.6}{10-1}} = 7.89$$

Largest margin 24: $z = \frac{x - \bar{x}}{s} = \frac{24 - 12.2}{7.89} = 1.50$. No outliers.

46. 15, 20, 25, 25, 27, 28, 30, 34

Smallest = 15

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(8+1) = 2.25$$

First quartile or 25th percentile = 20 + .25(25 − 20) = 21.25

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(8+1) = 4.5$$

Second quartile or median = 25 + .5(27 − 25) = 26

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(8+1) = 6.75$$

Third quartile or 75th percentile = 28 + . 75(30 − 28) = 29.5

Largest = 34

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

48. 5, 6, 8, 10, 10, 12, 15, 16, 18

Smallest = 5

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(9+1) = 2.5$$

First quartile or 25th percentile = 6 + .5(8 − 6) = 7

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(9+1) = 5.0$$

Second quartile or median = 10

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(9+1) = 7.5$$

Third quartile or 75th percentile = 15 + . 5(16 − 15) = 15.5

Largest = 18

A box plot follows:

50. a. The first place runner in the men's group finished $109.03 - 65.30 = 43.73$ minutes ahead of the first place runner in the women's group. Lauren Wald would have finished in 11th place for the combined groups.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. Men: $i = \left(\frac{50}{100}\right)22 = .50(22) = 11$

Use the 11th and 12th place finishes.

$$\text{Median} = \frac{109.05 + 110.23}{2} = 109.64$$

Women: $i = \left(\frac{50}{100}\right)31 = .50(31) = 15.5$

Use the 16th place finish.

$$\text{Median} = 131.67$$

Using the median finish times, the men's group finished $131.67 - 109.64 = 22.03$ minutes ahead of the women's group.

Also note that the fastest time for a woman runner, 109.03 minutes, is approximately equal to the median time of 109.64 minutes for the men's group.

c. Men: Lowest time = 65.30; highest time = 148.70

Q_1: $i = \left(\frac{25}{100}\right)(n+1) = .25(23) = 5.75$ Q_1 = 70.87 + .75(87.18-70.87) = 83.1025

Q_3: $i = \left(\frac{75}{100}\right)(n+1) = .75(23) = 17.25$ Q_3 = 128.4 + .25(130.90—128.40) = 129.025

Five-number summary for men: 65.30, 83.1025, 109.64, 129.025, 148.70

Women: Lowest time = 109.03; highest time = 189.28

Q_1: $i = \left(\frac{25}{100}\right)(n+1) = .25(32) = 8.0$

Use 8th position.

Q_1 = 122.08

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$Q_3: \quad i = \left(\frac{75}{100}\right)(n+1) = .75(32) = 24.0$$

Use 24th position.

$$Q_3 = 147.18$$

Five-number summary for women: 109.03, 122.08, 131.67, 147.18, 189.28

d. Men: IQR = 129.025 – 83.1025 = 45.9225

Lower limit = $Q_1 - 1.5(\text{IQR})$ = 83.1025 – 1.5(45.9225) = 14.22

Upper limit = $Q_3 + 1.5(\text{IQR})$ = 129.025 + 1.5(45.9225) = 197.91

There are no outliers in the men's group.

Women: IQR = $147.18 - 122.08 = 25.10$

Lower limit = $Q_1 - 1.5(\text{IQR}) = \; = 122.08 - 1.5(25.10) = 84.43$

Upper limit = $Q_3 + 1.5(\text{IQR}) = 147.18 + 1.5(25.10) = 184.83$

The two slowest women runners with times of 189.27 and 189.28 minutes are outliers in the women's group.

e.

The box plots show the men runners with the faster or lower finish times. However, the

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

box plots show the women runners with the lower variation in finish times. The interquartile ranges of 41.22 minutes for men and 25.10 minutes for women support this conclusion.

52. a. Median $n = 20$; 10th and 11th positions

$$\text{Median} = \frac{73+74}{2} = 73.5$$

b. Five-number summary: 68, 71.25, 73.5, 74.75, 77

c. IQR = Q3 – Q_1 = 74.75 – 71.25 = 3.5

Lower limit = Q_1 – 1.5(IQR) = 71.25 – 1.5(3.5) = 66

Upper limit = Q_3 + 1.5(IQR) = 74.75 + 1.5(3.5) = 80

All ratings are between 66 and 80. There are no outliers for the T-Mobile service.

d. Using the solution procedures shown in parts a, b, and c, the five number summaries and outlier limits for the other three cell phone services are as follows.

AT&T	66, 68, 71, 73, 75	Limits: 60.5 and 80.5
Sprint	63, 65, 66, 67.75, 69	Limits: 60.875 and 71.875
Verizon	75, 77, 78.5, 79.75, 81	Limits: 72.875 and 83.875

There are no outliers for any of the cell phone services.

e.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The box plots show that Verizon is the best cell phone service provider in terms of overall customer satisfaction. Verizon's lowest rating is better than the highest AT&T and Sprint ratings and is better than 75% of the T-Mobile ratings. Sprint shows the lowest customer satisfaction ratings among the four services.

54.

a. Mean = 173.24 and median (second quartile) = 89.5

b. First quartile = 38.5, and the third quartile = 232

c. 21, 38.5, 89.5, 232, 995

d. A box plot follows.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The box plot shows the distribution of number of personal vehicle crossings is skewed to the right (positive). Three ports of entry are considered outliers:

NY: Buffalo-Niagara Falls 707

TX: El Paso 807

CA: San Ysidro 995

56. a.

b. Positive relationship

c/d. $\Sigma x_i = 80 \quad \bar{x} = \frac{80}{5} = 16 \quad \Sigma y_i = 50 \quad \bar{y} = \frac{50}{5} = 10$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 106 \quad \Sigma(x_i - \bar{x})^2 = 272 \quad \Sigma(y_i - \bar{y})^2 = 86$$

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{106}{5-1} = 26.5$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{272}{5-1}} = 8.2462$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{86}{5-1}} = 4.6368$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{26.5}{(8.2462)(4.6368)} = .693$$

A positive linear relationship

58. Let x = miles per hour and y = miles per gallon.

$$\Sigma x_i = 420 \quad \bar{x} = \frac{420}{10} = 42 \quad \Sigma y_i = 270 \quad \bar{y} = \frac{270}{10} = 27$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -475 \quad \Sigma(x_i - \bar{x})^2 = 1660 \quad \Sigma(y_i - \bar{y})^2 = 164$$

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-475}{10-1} = -52.7778$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1660}{10-1}} = 13.5810$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{164}{10-1}} = 4.2687$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-52.7778}{(13.5810)(4.2687)} = -.91$$

A strong negative linear relationship exists. For driving speeds between 25 and 60 miles per hour, higher speeds are associated with lower miles per gallon.

60. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. DJIA : $\bar{x} = \frac{\sum x_i}{n} = \frac{227.57}{25} = 9.10 \quad s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{5672.61}{24}} = 15.37$

Russell 1000: $\bar{x} = \frac{\sum x_i}{n} = \frac{227.29}{25} = 9.09 \quad s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{7679.81}{24}} = 17.89$

c. $r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{263.611}{(15.37)(17.89)} = .959$

d. Based on this sample, the two indexes are strikingly similar. They have a strong positive correlation. The variance of the Russell 1000 is slightly larger than that of the DJIA.

62. The data in ascending order follow.

Position	Value	Position	Value
1	0	11	3
2	0	12	3
3	1	13	3
4	1	14	4

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

5	1	15	4
6	1	16	5
7	1	17	5
8	2	18	6
9	3	19	6
10	3	20	7

a. The mean is 2.95, and the median is 3.

b. $L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(20+1) = 5.25$

First quartile or 25th percentile = 1 + . 25(1 −1) = 1

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(20+1) = 15.75$$

Third quartile or 75th percentile = 4 + . 75(5 − 4) = 4.75

c. The range is 7 and the interquartile range is 4.75 – 1 = 3.75.

d. The variance is 4.37, and standard deviation is 2.09.

e. Because most people dine out a relatively few times per week and a few families dine out frequently, we would expect the data to be positively skewed. The skewness measure of 0.34 indicates the data are somewhat skewed to the right.

f. The lower limit is –4.625, and the upper limit is 10.375. No values in the data are less than the lower limit or greater than the upper limit, so there are no outliers.

64. a. The mean and median patient wait times for offices with a wait-tracking system are 17.2 and 13.5, respectively. The mean and median patient wait times for offices

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

without a wait-tracking system are 29.1 and 23.5, respectively.

b. The variance and standard deviation of patient wait times for offices with a wait-tracking system are 86.2 and 9.3, respectively. The variance and standard deviation of patient wait times for offices without a wait-tracking system are 275.7 and 16.6, respectively.

c. Offices with a wait-tracking system have substantially shorter patient wait times than offices without a wait-tracking system.

d. $z = \frac{37 - 29.1}{16.6} = 0.48$

e. $z = \frac{37 - 17.2}{9.3} = 2.13$

As indicated by the positive *z*–scores, both patients had wait times that exceeded the means of their respective samples. Even though the patients had the same wait time, the *z*–score for the sixth patient in the sample who visited an office with a wait-tracking system is much larger because that patient is part of a sample with a smaller mean and a smaller standard deviation.

f. The *z*–scores for all patients follow.

Without Wait-Tracking System	With Wait-Tracking System
–0.31	1.49
2.28	–0.67
–0.73	–0.34
–0.55	0.09
0.11	–0.56
0.90	2.13

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

–1.03	–0.88
–0.37	–0.45
–0.79	–0.56
0.48	–0.24

The z–scores do not indicate the existence of any outliers in either sample.

66. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{20665}{50} = 413.3$

This is slightly higher than the mean for the study.

b. $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{69424.5}{49}} = 37.64$

c. $i = \frac{p}{100}(n) = \frac{25}{100}(50) = 12.5$ (13th position) $Q_1 = 384$

$i = \frac{p}{100}(n) = \frac{75}{100}(50) = 37.5$ (38th position) $Q_3 = 445$

IQR = 445 – 384 = 61

LL = Q_1 – 1.5 IQR = 384 – 1.5(61) = 292.5

UL = Q_3 + 1.5 IQR = 445+ 1.5(61) = 536.5

There are no outliers.

68. a. The data in ascending order follow:

46.5 48.7 49.4 51.2 51.3 51.6 52.1 52.1 52.2 52.4 52.5 52.9 53.4 64.5

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(14+1) = 7.5$$

Median or 50th percentile = 52.1 + . 5(52.1 − 52.1) = 52.1

b. Percentage change = $\left(\frac{52.1 - 55.5}{55.5}\right)100 = -6.1\%$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. $L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(14+1) = 3.75$

25th percentile = 49.4 + .75(51.2 − 49.4) = 50.75

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(14+1) = 11.25$$

75th percentile = 52.5 + .25(52.9 − 52.5) = 52.6

d. 46.5 50.75 52.1 52.6 64.5

e. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{730.8}{14} = 52.2$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{208.12}{13} = 16.0092$$

$$s = \sqrt{16.0092} = 4.0012$$

The z-scores = $\frac{x_i - \bar{x}}{s}$ are shown below:

–1.42 –0.87 –0.70 –0.25 –0.22 –0.15 –0.02 –0.02 0.00 0.05 0.07 0.17 0.30 3.07

The last household income (64.5) has a z-score > 3 and is an outlier.

Lower limit = $Q_1 - 1.5(\text{IQR}) = 50.75 - 1.5(52.6 - 50.75) = 47.98$

Upper limit = $Q_3 + 1.5(\text{IQR}) = 52.6 + 1.5(52.6 - 50.75) = 55.38$

Using this approach, the first observation (46.5) and the last observation (64.5) would be consider outliers.

The two approaches will not always provide the same results.

70. a. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{4368}{12} = 364$ rooms

b. $\bar{y} = \frac{\Sigma y_i}{n} = \frac{5484}{12} = \$457$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c.

It is difficult to see much of a relationship. When the number of rooms becomes larger, there is no indication that the cost per night increases. The cost per night may even decrease slightly.

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
	499	–144	42	20.736	1,764	–6,048
727	340	363	–117	131,769	13,689	–42,471
285	585	–79	128	6,241	16,384	–10,112
273	495	–91	38	8,281	1,444	–3,458
145	495	–219	38	47,961	1,444	–8,322
213	279	–151	–178	22,801	31,684	26,878
398	279	34	–178	1,156	31,684	–6,052

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

343	455	–21	–2	441	4	42
250	595	–114	138	12,996	19,044	–15,732
414	367	50	–90	2,500	8,100	–4,500
400	675	36	218	1,296	47,524	7,848
700	420	336	–37	112,896	1,369	–12,432
					174,13	
			Total	369,074	4	–74,359

d.

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-74{,}359}{11} = -6{,}759.91$$

$$s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{369{,}074}{11}} = 183.17$$

$$s_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{174{,}134}{11}} = 125.82$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-6759.91}{(183.17)(125.82)} = -.293$$

There is evidence of a slightly negative linear association between the number of rooms and the cost per night for a double room. Augh this is not a strong relationship, it suggests that the higher room rates tend to bessociated with the smaller hotels.

This tends to make sense when we think about the economies of scale for larger hotels. Many of the amenities—pools, equipment, spas, restaurants, and so on—exist for all hotels in the *Travel + Leisure* top 50 hotels in the world. The smaller

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

hotels tend to charge more for their rooms. The larger hotels can spread their fixed costs over many rooms and may actually be able to charge less per night and still achieve nice profits. The larger hotels may also charge slightly less in an effort to obtain a higher occupancy rate. In any case, there is apparently a slightly negative linear association between the number of rooms and the cost per night for a double room at the top hotels.

72. a.

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
.407	.422	–.1458	–.0881	.0213	.0078	.0128
.429	.586	–.1238	.0759	.0153	.0058	–.0094
.417	.546	–.1358	.0359	.0184	.0013	–.0049
.569	.500	.0162	–.0101	.0003	.0001	–.0002
.569	.457	.0162	–.0531	.0003	.0028	–.0009
.533	.463	–.0198	–.0471	.0004	.0022	.0009
.724	.617	.1712	.1069	.0293	.0114	.0183
.500	.540	–.0528	.0299	.0028	.0009	–.0016
.577	.549	.0242	.0389	.0006	.0015	.0009
.692	.466	.1392	–.0441	.0194	.0019	–.0061
.500	.377	–.0528	–.1331	.0028	.0177	.0070
.731	.599	.1782	.0889	.0318	.0079	.0158

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

.643	.488	.0902	–.0221	.0081	.0005	–.0020
.448	.531	–.1048	.0209	.0110	.0004	–.0022
			Total	.1617	.0623	.0287

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{.0287}{14-1} = .0022$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{.1617}{14-1}} = .1115$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{.0623}{14-1}} = .0692$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{.0022}{.1115(.0692)} = +.286$$

b. There is a low positive correlation between a Major League Baseball team's winning percentage during spring training and its winning percentage during the regular season. The spring training record should not be expected to be a good indicator of how a team will play during the regular season.

Spring training consists of practice games between teams with wins and losses not counted in the regular season standings and not affecting the teams' chances of making the playoffs. Teams use spring training to help players regain their timing and to evaluate new players. Substitutions are frequent with regular or better players rarely playing an entire spring training game. Winning is not the primary goal in spring training games. A low correlation between spring training winning percentage and regular season winning percentage should be anticipated.

74.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

f_i	M_i	$f_i M_i$	$M_i - \overline{x}$	$(M_i - \overline{x})^2$	$f_i(M_i - \overline{x})^2$
10	47	470	–13.68	187.1424	1,871.42
40	52	2,080	–8.68	75.3424	3,013.70
150	57	8,550	–3.68	13..5424	2,031.36
175	62	10,850	+1.32	1.7424	304.92
75	67	5,025	+6.32	39.9424	2,995.68
15	72	1,080	+11.32	128.1424	1,922.14
10	77	770	+16.32	266.3424	2,663.42
475		28,825			14,802.64

a. $\overline{x} = \dfrac{28,825}{475} = 60.68$

b. $s^2 = \dfrac{14,802.64}{474} = 31.23$

$$s = \sqrt{31.23} = 5.59$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 4 Solutions

2. $$\binom{6}{3} = \frac{6!}{3!3!} = \frac{6\cdot5\cdot4\cdot3\cdot2\cdot1}{(3\cdot2\cdot1)(3\cdot2\cdot1)} = 20$$

ABC	ACE	BCD	BEF
ABD	ACF	BCE	CDE
ABE	ADE	BCF	CDF
ABF	ADF	BDE	CEF
ACD	AEF	BDF	DEF

4. a.

b. Let: H be head and T be tail:

(H,H,H)	(T,H,H)
(H,H,T)	(T,H,T)
(H,T,H)	(T,T,H)
(H,T,T)	(T,T,T)

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. The outcomes are equally likely, so the probability of each outcome is 1/8.

6. $P(E_1) = .40$, $P(E_2) = .26$, $P(E_3) = .34$

The relative frequency method was used.

8. a. There are four outcomes possible for this two-step experiment; planning commission positive—council approves; planning commission positive—council disapproves; planning commission negative—council approves; and planning commission negative—council disapproves.

b. Let p = positive, n = negative, a = approves, and d = disapproves

10. a.

Programmer	Total Lines of Code	Number of Lines of Code	Probability

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	Written	Requiring Edits	
Liwei	23,789	4,589	0.1929
Andrew	17,962	2,780	0.1548
Jaime	31,025	12,080	0.3894
Sherae	26,050	3,780	0.1451
Binny	19,586	1,890	0.0965
Roger	24,786	4,005	0.1616
Dong-Gil	24,030	5,785	0.2407
Alex	14,780	1,052	0.0712
Jay	30,875	3,872	0.1254
Vivek	21,546	4,125	0.1915

b. Probability = 4589 / 23789 = 0.1929

c. Probability = 1 – 3780 / 26050 = 1 – 0.1451 = 0.8549

d. The lowest probability is Alex at 0.0712; the highest probability is Jaime at 0.3894.

12. Initially a probability of .20 would be assigned if selection is equally likely. Data do not appear to confirm the belief of equal consumer preference. For example, using the relative frequency method we would assign a probability of 5/100 = .05 to the design 1 outcome, .15 to design 2, .30 to design 3, .40 to design 4, and .10 to design 5.

14. a. $P(E_2) = 1/4$

b. P(any 2 outcomes) = 1/4 + 1/4 = 1/2

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. P(any 3 outcomes) = 1/4 + 1/4 + 1/4 = 3/4

16. a. (6)(6) = 36 sample points

b.

Die 1 \ Die 2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Total for Both

c. 6/36 = 1/6

d. 10/36 = 5/18

e. No. P(odd) = 18/36 = P(even) = 18/36 or 1/2 for both.

f. Classical. A probability of 1/36 is assigned to each experimental outcome.

18. a. Let C = corporate headquarters located in California:

$$P(C) = 53/500 = .106$$

b. Let N = corporate headquarters located in New York

T = corporate headquarters located in Texas

$$P(N) = 50/500 = .100$$

$$P(T) = 52/500 = .104$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Located in California, New York, or Texas:

$$P(C) + P(N) + P(T) = .106 + .100 + .104 = .31$$

c. Let A = corporate headquarters located in one of the eight states

Total number of companies with corporate headquarters in the eight states = 283

$$P(A) = 283/500 = .566$$

More than half the Fortune 500 companies have corporate headquartered located in these eight states.

20. a.

Experimental Outcome	Age Financially Independent	Number of Responses	Probability	
E_1	16 to 20	191	191/944 =	0.2023
E_2	21 to 24	467	467/944 =	0.4947
E_3	25 to 27	244	244/944 =	0.2585
E_4	28 or older	42	42/944 =	0.0445
		944		

b. $P(\text{Age} < 25) = P(E_1) + P(E_2) = .2023 + .4947 = .6970$

c. $P(\text{Age} > 24) = P(E_3) + P(E_4) = .2585 + .0445 = .3030$

d. The probability of being financially independent before age 25, .6970, seems high given the general economic conditions. The teenagers who responded to this survey may have had unrealistic expectations about becoming financially independent at a

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

relatively young age.

22. a. $P(A) = .40$, $P(B) = .40$, $P(C) = .60$

b. $P(A \cup B) = P(E_1, E_2, E_3, E_4) = .80$. Yes, $P(A \cup B) = P(A) + P(B)$.

c. $A^c = \{E_3, E_4, E_5\}$ $C^c = \{E_1, E_4\}$ $P(A^c) = .60$ $P(C^c) = .40$

d. $A \cup B^c = \{E_1, E_2, E_5\}$ $P(A \cup B^c) = .60$

e. $P(B \cup C) = P(E_2, E_3, E_4, E_5) = .80$

24. Let E = experience exceeded expectations

M = experience met expectations

a. Percentage of respondents that said their experience exceeded expectations

$$= 100 - (4 + 26 + 65) = 5\%$$

$$P(E) = .05$$

b. $P(M \cup E) = P(M) + P(E) = .65 + .05 = .70$

26. a. Let D = Domestic Equity Fund

$P(D) = 16/25 = .64$

b. Let A = 4- or 5-star rating

Thirteen funds were rated 3 star or less; thus, 25 – 13 = 12 funds must be 4 star or 5 star.

$$P(A) = 12/25 = .48$$

c. Seven Domestic Equity funds were rated 4 star, and two were rated 5 star. Thus, nine funds were Domestic Equity funds and were rated 4 star or 5 star:

$$P(D \cap A) = 9/25 = .36$$

d. $P(D \cup A) = P(D) + P(A) - P(D \cap A)$

$= .64 + .48 - .36 = .76$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

28. Let: B = rented a car for business reasons

P = rented a car for personal reasons

a. $P(\text{B} \cup \text{P}) = P(\text{B}) + P(\text{P}) - P(\text{B} \cap \text{P})$

$= .54 + .458 - .30 = .698$

b. $P(\text{Neither}) = 1 - .698 = .302$

30. a. $P(\text{A}|\text{B}) = \dfrac{P(\text{A} \cap \text{B})}{P(\text{B})} = \dfrac{.40}{.60} = .6667$

b. $P(\text{B}|\text{A}) = \dfrac{P(\text{A} \cap \text{B})}{P(\text{A})} = \dfrac{.40}{.50} = .80$

c. No because $P(\text{A} \mid \text{B}) \neq P(\text{A})$

32. a. Dividing each entry in the table by 500 yields the following (rounding to two digits):

	Yes	No	Totals
Men	0.210	0.282	0.492
Women	0.186	0.322	0.508
Totals	0.396	0.604	1.000

Let M = 18-34-year-old man,

W = 18-34-year-old woman,

Y = responded yes,

N = responded no

b. $P(M) = .492$, $P(W) = .508$

$P(Y) = .396$, $P(N) = .604$

c. $P(Y|M) = .210/.492 = .4268$

d. $P(Y|W) = .186/.508 = .3661$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

e. P(Y) = .396/1 = .396

f. P(M) = .492 in the sample. Yes, this seems like a good representative sample based on gender.

34. a. Let O = flight arrives on time

L = flight arrives late

J = Jet Blue flight

N = United flight

U = US Airways flight

Given:

$$P(O \mid J) = .768 \quad P(O \mid N) = .715 \quad P(O \mid U) = .822$$

$$P(J) = .30 \quad P(N) = .32 \quad P(U) = .38$$

Joint probabilities using the multiplication law

$$P(J \cap O) = P(J)P(O \mid J) = (.30)(.768) = .2304$$

$$P(N \cap O) = P(N)P(O \mid N) = (.32)(.715) = .2288$$

$$P(U \cap O) = P(U)P(O \mid U) = (.38)(.822) = .3124$$

With the marginal probabilities $P(J)$ = .30, $P(N)$ = .32, and $P(U)$ = .38 given, the joint probability table can then be shown as follows.

	On Time	Late	Total
Jet Blue	.2304	.0696	.30
United	.2288	.0912	.32
US Airways	.3124	.0676	.38

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Total	.7716	.2284	1.00

b. Using the joint probability table, the probability of an on-time flight is the marginal probability

$$P(O) = .2304 + .2288 + .3124 = .7716$$

c. Because US Airways has the highest percentage of flights into terminal C, US Airways with $P(U) = .38$ is the most likely airline for Flight 1382.

d. From the joint probability table, $P(L) = .2284$:

$$P(J|L) = \frac{P(J \cap L)}{P(L)} = \frac{.0696}{.2284} = .3047$$

$$P(N|L) = \frac{P(N \cap L)}{P(L)} = \frac{.0912}{.2284} = .3992$$

$$P(U|L) = \frac{P(U \cap L)}{P(L)} = \frac{.0676}{.2284} = .2961$$

Most likely airline for Flight 1382 is now United with a probability of .3992. US Airways is now the least likely airline for this flight with a probability of .2961.

36. a. We have that P(Make the Shot) = .93 for each foul shot, so the probability that the player will make two consecutive foul shots is that P(Make the Shot) P(Make the Shot) = (.93)(.93) = .8649.

b. There are three unique ways in which the player can make at least one shot: He can make the first shot and miss the second shot, miss the first shot and make the second shot, or make both shots. Because the event "Miss the Shot" is the complement of the event "Make the Shot," P(Miss the Shot) = 1 – P(Make the Shot) = 1 – .93 = .07. Thus:

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

P(Make the Shot) P(Miss the Shot) = (.93)(.07) = .0651

P(Miss the Shot) P(Make the Shot) = (.07)(.93) = .0651

P(Make the Shot) P(Make the Shot) = (.93)(.93) = <u>.8649</u>

.9951

c. We can find this probability in two ways. We can calculate the probability directly:

P(Miss the Shot) P(Miss the Shot) = (.07)(.07) = .0049

Or we can recognize that the event "Miss both Shots" is the complement of the event "Make at Least One of the Two Shots," so

P(Miss the Shot) P(Miss the Shot) = 1 – .9951 = .0049

d. For the player who makes 58% of his free throws, we have:

P(Make the Shot) = .58 for each foul shot,

so the probability that this player will make two consecutive foul shots is

P(Make the Shot) P(Make the Shot) = (.58)(.58) = .3364.

Again, there are three unique ways that this player can make at least one shot: He can make the first shot and miss the second shot, miss the first shot and make the second shot, or make both shots. Because the event "Miss the Shot" is the complement of the event "Make the Shot" P(Miss the Shot) = 1 – P(Make the Shot) = 1 – .58 = .42. Thus,

P(Make the Shot) P(Miss the Shot) = (.58)(.42) = .2436

P(Miss the Shot) P(Make the Shot) = (.42)(.58) = .2436

P(Make the Shot) P(Make the Shot) = (.58)(.58) = <u>.3364</u>

.8236

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

We can again find the probability the 58% free-throw shooter will miss both shots in two ways. We can calculate the probability directly:

$$P(\text{Miss the Shot})\ P(\text{Miss the Shot}) = (.42)(.42) = .1764$$

Or we can recognize that the event "Miss Both Shots" is the complement of the event "Make at Least One of the Two Shots," so

$$P(\text{Miss the Shot})\ P(\text{Miss the Shot}) = 1 - .9951 = .1764$$

Intentionally fouling the 58% free-throw shooter is a better strategy than intentionally fouling the 93% shooter.

38. Let Y = has a college degree

N = does not have a college degree

D = a delinquent student loan

a. From the table, $P(Y) = .42$

b. From the table, $P(N) = .58$

c. $P(D \mid Y) = \dfrac{P(D \cap Y)}{P(Y)} = \dfrac{.16}{.42} = .3810$

d. $P(D \mid N) = \dfrac{P(D \cap N)}{P(N)} = \dfrac{.34}{.58} = .5862$

e. Individuals who obtained a college degree have a .3810 probability of a delinquent student loan, whereas individuals who dropped out without obtaining a college degree have a .5862 probability of a delinquent student loan. Not obtaining a college degree will lead to a greater probability of struggling to payback the student loan and will likely lead to financial problems in the future.

40. a. $P(B \cap A_1) = P(A_1)P(B \mid A_1) = (.20)(.50) = .10$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$P(B \cap A_2) = P(A_2)P(B \mid A_2) = (.50)(.40) = .20$

$P(B \cap A_3) = P(A_3)P(B \mid A_3) = (.30)(.30) = .09$

b. $P(A_2 \mid B) = \dfrac{.20}{.10 + .20 + .09} = .51$

c.

Events	$P(A_i)$	$P(B \mid A_i)$	$P(A_i \cap B)$	$P(A_i \mid B)$
A1	.20	.50	.10	.26
A2	.50	.40	.20	.51
A3	.30	.30	.09	.23
	1.00		.39	1.00

42. M = missed payment

D_1 = customer defaults

D_2 = customer does not default

$P(D_1) = .05 \quad P(D_2) = .95 \quad P(M \mid D_2) = .2 \quad P(M \mid D_1) = 1$

a. $P(D_1 \mid M) = \dfrac{P(D_1)P(M \mid D_1)}{P(D_1)P(M \mid D_1) + P(D_2)P(M \mid D_2)} = \dfrac{(.05)(1)}{(.05)(1) + (.95)(.2)} = \dfrac{.05}{.24} = .21$

b. Yes, the probability of default is greater than .20.

44. *M* = the current visitor to the ParFore website is a male

F = the current visitor to the ParFore website is a female

D = a visitor to the ParFore website previously visited the Dillard website

a. Using past history, $P(F) = .40$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $P(M) = .60$, $P(D \mid F) = .30$, and $P(D \mid M) = .10$

$$P(F \mid D) = \frac{P(F)P(D \mid F)}{P(F)P(D \mid F) + P(M)P(D \mid M)} = \frac{(.40)(.30)}{(.40)(.30) + (.60)(.10)} = .6667$$

The revised probability that the current visitor is a female is .6667.

ParFore should display the special offer that appeals to female visitors.

46. a. 422 + 181 + 80 + 121 + 201 = 1005 respondents

b. Most frequent response a day or less; probability = 422/1005 = .4199

c. 201/1005 = .20

d. Responses of two days, three days, and four or more days = 181 + 80 + 121 = 382

Probability = 382/1005 = .3801

48. a. 0.5029

b. 0.5758

c. No, from part a we have $P(F) = 0.5029$, and from part b we have $P(A|F) = 0.5758$.

Because $P(F) \neq P(A|F)$, events A and F are not independent.

50. a. Probability of the event = P(average) + P(above average) + P(excellent)

$$= \frac{11}{50} + \frac{14}{50} + \frac{13}{50} = .22 + .28 + .26 = .76$$

b. Probability of the event = P(poor) + P(below average)

$$= \frac{4}{50} + \frac{8}{50} = .24$$

52. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	Yes	No	Total
23 and Under	.1026	.0996	.2022
24 - 26	.1482	.1878	.3360
27 - 30	.0917	.1328	.2245
31 - 35	.0327	.0956	.1283
36 and Over	.0253	.0837	.1090
Total	.4005	.5995	1.0000

b. .2022

c. .2245 + .1283 + .1090 = .4618

d. .4005

54. a. .1485+.2273+.4008 = .7766

b. $P(\text{OKAY}|30-49) = \frac{P(\text{OKAY} \cap 30-49)}{P(30-49)} = \frac{0.0907}{0.3180} = 0.2852$

c. $P(50+ \mid NOT\ OKAY) = \frac{P(50\ \ \cap \text{NOT OKAY})}{P(\text{NOT OKAY})} = \frac{.4008}{.7766} = .5161$

d. The attitude about this practice is not independent of the age of the respondent. We can show this in several ways. One example is to use the result from part (b). We have

$$P\left(\text{OKAY}|30-49\right) = 0.2852$$

and

$$P\left(\text{OKAY}\right) = 0.2234$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

If the attitude about this practice were independent of the age of the respondent, we would expect these probabilities to be equal. Because these probabilities are not equal, the data suggest the attitude about this practice is not independent of the age of the respondent.

e. Respondents in the 50+ age category are far more likely to say this practice is *not okay* than are respondents in the 18–29 age category:

$$P(\text{NOT OKAY}|50+) = \frac{P(\text{NOT OKAY} \cap 50+)}{P(50+)} = \frac{0.4008}{0.4731} = 0.8472$$

$$P(\text{NOT OKAY}|18\text{-}29) = \frac{P(\text{NOT OKAY} \cap 18\text{-}29)}{P(18\text{-}29)} = \frac{0.1485}{0.2089} = 0.7109$$

56. a. $P(\text{A}) = 200/800 = .25$

b. $P(\text{B}) = 100/800 = .125$

c. $P(\text{A} \cap \text{B}) = 10/800 = .0125$

d. $P(\text{A} \mid \text{B}) = P(\text{A} \cap \text{B})/P(\text{B}) = .0125/.125 = .10$

e. No, $P(\text{A} \mid \text{B}) \neq P(\text{A}) = .25$

58. a. Let A_1 = student studied abroad

A_2 = student did not study abroad

F = female student

M = male student

$P(A_1) = .095$

$P(A_2) = 1 - P(A_1) = 1 - .095 = .905$

$P(F \mid A_1) = .60$

$P(F \mid A_2) = .49$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Tabular computations

Events	$P(A_i)$	$P(F\|A_i)$	$P(A_i \cap F)$	$P(A_i\|F)$
A_1	.095	.60	.0570	.1139
A_2	.905	.49	<u>.4435</u>	.8861

$P(F) = .5005$

$P(A_1|F) = .1139$

b.

Events	$P(A_i)$	$P(M\|A_i)$	$P(A_i \cap M)$	$P(A_i\|M)$
A_1	.095	.40	.0380	.0761
A_2	.905	.51	<u>.4615</u>	.9239

$P(M) = .4995$

$P(A_1|M) = .0761$

c. From the preceding, $P(F) = .5005$ and $P(M) = .4995$, so almost 50:50 female and male full-time students.

60. a.

$$P(\text{spam}|shipping!) = \frac{P(\text{spam})P(shipping!|\text{spam})}{P(\text{spam})P(shipping!|\text{spam}) + P(\text{ham})P(shipping!|\text{ham})}$$

$$= \frac{(0.10)(0.051)}{(0.10)(0.051) + (0.90)(0.0015)} = 0.791$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$P(\text{ham}|\textit{shipping!}) = \frac{P(\text{ham})P(\textit{shipping!}|\text{ham})}{P(\text{ham})P(\textit{shipping!}|\text{ham}) + P(\text{spam})P(\textit{shipping!}|\text{ham})}$$

$$= \frac{(0.90)(0.0015)}{(0.90)(0.0015) + (0.10)(0.051)} = 0.209$$

If a message includes the word *shipping!*, the probability the message is spam is high (.7910), so the message should be flagged as spam.

b.

$$P(\text{spam}|\textit{today!}) = \frac{P(\text{spam})P(\textit{today!}|\text{spam})}{P(\text{spam})P(\textit{today!}|\text{spam}) + P(\text{ham})P(\textit{today!}|\text{ham})}$$

$$= \frac{(0.10)(0.045)}{(0.10)(0.045) + (0.90)(0.0022)} = 0.694$$

$$P(\text{spam}|\textit{here!}) = \frac{P(\text{spam})P(\textit{here!}|\text{spam})}{P(\text{spam})P(\textit{here!}|\text{spam}) + P(\text{ham})P(\textit{here!}|\text{ham})}$$

$$= \frac{(0.10)(0.034)}{(0.10)(0.034) + (0.90)(0.0022)} = 0.632$$

A message that includes the word *today!* is more likely to be spam. *P(*spam|*today!*) is higher than *P(*spam|*here!*) because *P*(*today!*|spam) is larger than *P*(*here!*|spam) and *P*(*today!*|ham) = *P*(*here!*|ham) meaning that *today!* occurs more often in unwanted messages (spam) than *here!,* and just as often in legitimate messages (ham). Therefore, it is easier to distinguish spam from ham in messages that include *today!*.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c.

$$P(\text{spam}|\textit{available}) = \frac{P(\text{spam})P(\textit{available}|\text{spam})}{P(\text{spam})P(\textit{available}|\text{spam}) + P(\text{ham})P(\textit{available}|\text{ham})}$$

$$= \frac{(0.10)(0.014)}{(0.10)(0.014) + (0.90)(0.0041)} = 0.275$$

$$P(\text{spam}|\textit{fingertips!}) = \frac{P(\text{spam})P(\textit{fingertips!}|\text{spam})}{P(\text{spam})P(\textit{fingertips!}|\text{spam}) + P(\text{ham})P(\textit{fingertips!}|\text{ham})}$$

$$= \frac{(0.10)(0.014)}{(0.10)(0.014) + (0.90)(0.0011)} = 0.586$$

A message that includes the word *fingertips!* is more likely to be spam. *P(*spam|*fingertips!*) is larger than *P(*spam|*available*) because *P*(*available*|ham) is larger than *P*(*fingertips!*|ham) and *P*(*available*|spam) = *P*(*fingertips!*|spam) meaning that *available* occurs more often in legitimate messages (ham) than *fingertips!* and just as often in unwanted messages (spam). Therefore, it is more difficult to distinguish spam from ham in messages that include *available*.

d. It is easier to distinguish spam from ham when a word occurs more often in unwanted messages (spam), less often in legitimate messages (ham), or both.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 5 Solutions

2. a. Let x = time (in minutes) to assemble the product.

   b. It may assume any positive value: $x > 0$.

   c. Continuous

4. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

6. a. values: 0, 1, 2, . . . , 20

   discrete

   b. values: 0, 1, 2, . . .

   discrete

   c. values: 0, 1, 2, . . . , 50

   discrete

   d. values: $0 \leq x \leq 8$

   continuous

   e. values: $x > 0$

   continuous

8. a.

x	$f(x)$
1	3/20 = .15
2	5/20 = .25
3	8/20 = .40
4	4/20 = .20
	Total 1.00

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b.

c. $f(x) \geq 0$ for $x = 1,2,3,4$.

$\Sigma f(x) = 1$

10. a.

x	$f(x)$
1	0.05
2	0.09
3	0.03
4	0.42
5	0.41
	1.00

b.

x	$f(x)$
1	0.04

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

2	0.10
3	0.12
4	0.46
5	0.28
	1.00

c. $P(4 \text{ or } 5) = f(4) + f(5) = 0.42 + 0.41 = 0.83$

d. Probability of very satisfied: 0.28

e. Senior executives appear to be more satisfied than middle managers; 83% of senior executives have a score of 4 or 5, with 41% reporting a 5. Only 28% of middle managers report being very satisfied.

12. a. Yes; $f(x) \geq 0$. $\Sigma f(x) = 1$

b. $f(500{,}000) + f(600{,}000) = .10 + .05 = .15$

c. $f(100{,}000) = .10$

14. a. $f(200) = 1 - f(-100) - f(0) - f(50) - f(100) - f(150)$

$= 1 - .95 = .05$

This is the probability MRA will have a $200,000 profit.

b. $P(\text{Profit}) = f(50) + f(100) + f(150) + f(200)$

$= .30 + .25 + .10 + .05 = .70$

c. $P(\text{at least } 100) = f(100) + f(150) + f(200)$

$= .25 + .10 + .05 = .40$

16. a.

y	$f(y)$	$yf(y)$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

2	.2	.4
4	.3	1.2
7	.4	2.8
8	.1	.8
	1.0	5.2

$E(y) = \sigma = 5.2$

b.

y	$y - \mu$	$(y - \mu)^2$	$f(y)$	$(y - \mu)^2 f(y)$
2	–3.20	10.24	.20	2.048
4	–1.20	1.44	.30	.432
7	1.80	3.24	.40	1.296
8	2.80	7.84	.10	.784
				4.560

$Var(y) = 4.56$

$\sigma = \sqrt{4.56} = 2.14$

18. a/b/.

x	$f(x)$	$xf(x)$	$x - \mu$	$(x - \mu)^2$	$(x - \mu)^2 f(x)$
0	.2188	.0000	–1.1825	1.3982	.3060
1	.5484	.5484	–.1825	.0333	.0183
2	.1241	.2483	.8175	.6684	.0830

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

3	.0489	.1466	1.8175	3.3035	.1614
4	.0598	.2393	2.8175	7.9386	.4749
Total	1.0000	1.1825			1.0435
		↑			↑
		$E(x)$			$Var(x)$

c/d.

y	$f(y)$	$yf(y)$	$y-\mu$	$(y-\mu)^2$	$(y-\mu)^2 f(y)$
0	.2497	.0000	–1.2180	1.4835	.3704
1	.4816	.4816	–.2180	.0475	.0229
2	.1401	.2801	.7820	.6115	.0856
3	.0583	.1749	1.7820	3.1755	.1851
4	.0703	.2814	2.7820	7.7395	.5444
Total	1.0000	1.2180			1.2085
		↑			↑
		$E(y)$			$Var(y)$

e. The expected number of times that owner-occupied units have a water supply stoppage lasting six or more hours in the past three months is 1.1825, slightly less than the expected value of 1.2180 for renter-occupied units. And the variability is somewhat less for owner-occupied units (1.0435) as compared to renter-occupied units (1.2085).

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

20. a.

x	$f(x)$	$xf(x)$
0	.85	0
500	.04	20
1000	.04	40
3000	.03	90
5000	.02	100
8000	.01	80
10000	.01	100
Total	1.00	430

The expected value of the insurance claim is \$430. If the company charges \$430 for this type of collision coverage, it would break even.

b. From the point of view of the policyholder, the expected gain is as follows:

Expected gain = Expected claim payout – Cost of insurance coverage

= \$430 – \$520 = –\$90

The policyholder is concerned that an accident will result in a big repair bill if there is no insurance coverage. So even though the policyholder has an expected annual loss of \$90, the insurance is protecting against a large loss.

22. a. $E(x) = \Sigma\, x\, f(x) = 300\ (.20) + 400\ (.30) + 500\ (.35) + 600\ (.15) = 445$

The monthly order quantity should be 445 units.

b. Cost: 445 @ \$50 = \$22,250

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Revenue: 300 @ \$70 = 21,000

\$1,250 loss

24. a. Medium $E(x) = \Sigma x\, f(x)$

$= 50\ (.20) + 150\ (.50) + 200\ (.30) = 145$

Large: $E(x) = \Sigma x\, f(x)$

$= 0\ (.20) + 100\ (.50) + 300\ (.30) = 140$

Medium preferred.

b. Medium

x	$f(x)$	$x - \mu$	$(x - \mu)^2$	$(x - \mu)^2 f(x)$
50	.20	–95	9,025	1,805.0
150	.50	5	25	12.5
200	.30	55	3,025	907.5
				$\sigma^2 = 2{,}725.0$

Large

y	$f(y)$	$y - \mu$	$(y - \mu)^2$	$(y - \mu)^2 f(y)$
0	.20	–140	19,600	3,920
100	.50	–40	1,600	800
300	.30	160	25,600	7,680
				$\sigma^2 = 12{,}400$

Medium is preferred because of less variance.

26. a. The standard deviation for these two stocks is the square root of the variance.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\sigma_x = \sqrt{Var(x)} = \sqrt{25} = 5\% \quad \sigma_y = \sqrt{Var(y)} = \sqrt{1} = 1\%$$

Investments in stock 1 would be considered riskier than investments in stock 2 because the standard deviation is higher. Note that if the return for stock 1 falls 8.45/5 = 1.69 or more standard deviation below its expected value, an investor in that stock will experience a loss. The return for stock 2 would have to fall 3.2 standard deviations below its expected value before an investor in that stock would experience a loss.

b. Because x represents the percent return for investing in stock1, the expected return for investing \$100 in stock 1 is \$8.45 and the standard deviation is \$5.00. So to get the expected return and standard deviation for a \$500 investment, we just multiply by 5.

Expected return (\$500 investment) = 5(\$8.45) = \$42.25

Standard deviation (\$500 investment) = 5(\$5.00) = \$25.00

c. Because x represents the percent return for investing in stock 1 and y represents the percent return for investing in stock 2, we want to compute the expected value and variance for $.5x + .5y$.

$$E(.5x + .5y) = .5E(x) + .5E(y) = .5(8.45) + .5(3.2) = 4.225 + 1.6 = 5.825$$

$$\begin{aligned} Var(.5x+.5y) &= .5^2 Var(x) + .5^2 Var(y) + 2(.5)(.5)\sigma_{xy} \\ &= (.5)^2(25) + (.5)^2(1) + 2(.5)(.5)(-3) \\ &= 6.25 + .25 - 1.50 = 5 \end{aligned}$$

$$\sigma_{.5x+.5y} = \sqrt{5} = 2.236$$

d. Because x represents the percent return for investing in stock 1 and y represents the percent return for investing in stock 2, we want to compute the expected value and variance for $.7x + .3y$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$E(.7x + .3y) = .7E(x) + .3E(y) = .7(8.45) + .3(3.2) = 5.915 + .96 = 6.875$$

$$Var(.7x + .3y) = .7^2 Var(x) + .3^2 Var(y) + 2(.7)(.3)\sigma_{xy}$$

$$= .7^2(25) + .3^2(1) + 2(.7)(.3)(-3)$$

$$= 12.25 + .09 - 1.26 = 11.08$$

$$\sigma_{.7x+.3y} = \sqrt{11.08} = 3.329$$

e. The standard deviations of x and y were computed in part a. The correlation coefficient is given by

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{-3}{(5)(1)} = -.6$$

There is a fairly strong negative relationship between the variables.

28. a. Marginal distribution of Direct Labor Cost

y	$f(y)$	$y\,f(y)$	$y - E(y)$	$(y - E(y))^2$	$(y - E(y))^2 f(y)$
43	.3	12.9	–2.3	5.29	1.587
45	.4	18	–.3	.09	.036
48	.3	14.4	2.7	7.29	2.187
		45.3		$Var(y) = 3.81$	
		$E(y) = 45.3$		$\sigma_y = 1.95$	

b. Marginal distribution of Parts Cost

x	$f(x)$	$xf(x)$	$x - E(x)$	$(x–E(x))^2$	$(x–E(x))^2 f(x)$
85	.45	38.25	–5.5	30.25	13.6125

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

95	.55	52.25	4.5	20.25	11.1375
		90.5		$Var(x)=$	24.75
		$E(x) = 90.5$		$\sigma_x =$	4.97

c. Let $z = x + y$ represent total manufacturing cost (direct labor + parts).

z	$f(z)$
128	.05
130	.20
133	.20
138	.25
140	.20
143	.10
	1.00

d. The computation of the expected value, variance, and standard deviation of total manufacturing cost is shown as follows.

z	$f(z)$	$zf(z)$	$z - E(z)$	$(z - E(z))^2$	$(z - E(z))^2 f(z)$
128	.05	6.4	–7.8	60.84	3.042
130	.20	26	–5.8	33.64	6.728
133	.20	26.6	–2.8	7.84	1.568
138	.25	34.5	2.2	4.84	1.21

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

140	.20	28	4.2	17.64	3.528
143	.10	14.3	7.2	51.84	5.184
		135.8		*Var(z)=*	21.26
		E(z) = 135.8		σ_z =	4.61

e. To determine if x = parts cost and y = direct labor cost are independent, we need to compute the covariance σ_{xy}.

$$\sigma_{xy} = (\mathrm{Var}(x+y) - Var(x) - Var(y))/2 = (21.26 - 24.75 - 3.81)/2 = -3.65$$

Because the covariance is not equal to zero, we can conclude that direct labor cost is not independent of parts cost. Indeed, they are negatively correlated. When parts cost goes up, direct labor cost goes down. Maybe the parts costing $95 come from a different manufacturer and are higher quality. Working with higher quality parts may reduce labor costs.

f. The expected manufacturing cost for 1500 printers is

$$E(1500z) = 1500E(z) = 1500(135.8) = 203,700$$

The total manufacturing costs of $198,350 are less than we would have expected. Perhaps as more printers were manufactured there was a learning curve and direct labor costs went down.

30. a. Let x = percentage return for S&P 500

y = percentage return for Core Bond fund

z = percentage return for REITs

The formula for computing the correlation coefficient is given by

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

In this case, we know the correlation coefficients and the three standard deviations, so we want to rearrange the correlation coefficient formula to find the covariances.

S&P 500 and REITs: $\sigma_{xz} = \rho_{xz}\sigma_x\sigma_z = (.74)(19.45)(23.17) = 333.486$

Core Bonds and REITS: $\sigma_{yz} = \rho_{yz}\sigma_y\sigma_z = (-.04)(2.13)(23.17) = -1.974$

b. Letting r = portfolio percentage return, we have $r = .5x + .5y$. The expected return for a portfolio with 50% invested in the S&P 500 and 50% invested in REITs is

$$E(r) = .5E(x) + .5E(z) = (.5)5.04\% + (.5)13.07\% = 9.055\%$$

We are given σ_x and σ_z, so $Var(x) = 19.45^2 = 378.3025$ and $Var(z) = 23.17^2 = 536.8489$. We can now compute

$$\begin{aligned} Var(.5x + .5z) &= .5^2 Var(x) + .5^2 Var(z) + 2(.5)(.5)(\sigma_{xz}) \\ &= .25(378.3025) + .25(536.8489) + .5(333.486) \\ &= 395.53 \end{aligned}$$

Then $\sigma_{x+z} = \sqrt{395.53} = 19.89$

So, the expected return for our portfolio is 9.055%, and the standard deviation is 19.89%.

c. Letting r = portfolio percentage return, we have $r = .5y + .5z$. The expected return for a portfolio with 50% invested in Core Bonds and 50% invested in REITs is

$$E(r) = .5E(y) + .5E(z) = (.5)5.78\% + (.5)13.07\% = 9.425\%$$

We are given σ_y and σ_z, so $Var(y) = 2.13^2 = 4.5369$ and $Var(z) = 23.17^2 = 536.8489$. We can now compute

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$Var(.5y + .5z) = .5^2 Var(y) + .5^2 Var(z) = 2(.5).5)(\sigma_{yz})$$

$$= .25(4.5369) + .25(536.8489) + .5(-1.974)$$

$$= 134.3595$$

Then $\sigma_{y+z} = \sqrt{134.3595} = 11.59$

So, the expected return for our portfolio is 9.425%, and the standard deviation is 11.59%.

d. Letting r = portfolio percentage return, we have $r = .8y + .2z$. The expected return for a portfolio with 80% invested in Core Bonds and 20% invested in REITs is

$$E(r) = .8E(y) + .2E(z) = (.8)5.78\% + (.2)13.07\% = 7.238\%$$

From part c, we have $Var(y) = 4.5369$ and $Var(z) = 536.8489$. We can now compute

$$Var(.8y + .2z) = .8^2 Var(y) + .2^2 Var(z) + 2(.8)(.2)(\sigma_{yz})$$

$$= .64(4.5369) + .04(536.8489) + .32(-1.974)$$

$$= 23.7459$$

Then $\sigma_{y+z} = \sqrt{23.7459} = 4.87$

So, the expected return for our portfolio is 7.238%, and the standard deviation is 4.87%.

e. The expected returns and standard deviations for the three portfolios are summarized as follow.

Portfolio	Expected Return (%)	Standard Deviation
50% S&P 500 & 50% REITs	9.055	19.89

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

50% Core Bonds & 50% REITs	9.425	11.63
80% Core Bonds & 20% REITs	7.238	4.94

The portfolio from part c involving 50% Core Bonds and 50% REITS has the highest return. Using the standard deviation as a measure of risk, it also has less risk than the portfolio from part b involving 50% invested in an S&P 500 index fund and 50% invested in REITs. So the portfolio from part b would not be recommended for either type of investor.

The portfolio from part d involving 80% in Core Bonds and 20% in REITs has the lowest standard deviation and thus lesser risk than the portfolio in part c. We would recommend the portfolio consisting of 50% Core Bonds and 50% REITs for the aggressive investor because of its higher return and moderate amount of risk.

We would recommend the portfolio consisting of 80% Core Bonds and 20% REITS to the conservative investor because of its low risk and moderate return.

32. a. $f(0) = .3487$

b. $f(2) = .1937$

c. $P(x \leq 2) = f(0) + f(1) + f(2) = .3487 + .3874 + .1937 = .9298$

d. $P(x \geq 1) = 1 - f(0) = 1 - .3487 = .6513$

e. $E(x) = n\,p = 10\,(.1) = 1$

f. $Var(x) = n\,p\,(1 - p) = 10\,(.1)\,(.9) = .9$

$\sigma = \sqrt{.9} = .95$

34. a. Yes. Because the teenagers are selected randomly, p is the same from trial to trial and the trials are independent. The two outcomes per trial are use Pandora Media Inc.'s

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

online radio service or do not use Pandora Media Inc.'s online radio service.

Binomial $n = 10$ and $p = .35$

$$f(x) = \frac{10!}{x!(10-x)!}(.35)^x(1-.35)^{10-x}$$

b. $$f(0) = \frac{10!}{0!(10-0)!}(.35)^0(.65)^{10-0} = .0135$$

c. $$f(4) = \frac{10!}{4!(10-4)!}(.35)^4(.65)^{10-4} = .2377$$

d. Probability $(x \geq 2) = 1 - f(0) - f(1)$

From part b, $f(0) = .0135$

$$f(1) = \frac{10!}{1!(10-1)!}(.35)^1(.65)^{10-1} = .0725$$

Probability $(x \geq 2) = 1 - f(0) - f(1) = 1 - (.0135 + .0725) = .9140$

36. a. Probability of a defective part being produced must be .03 for each part selected; parts must be selected independently.

b. Let: D = defective

G = not defective

c. Two outcomes result in exactly one defect.

d. P(no defects) = (.97) (.97) = .9409

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

P (1 defect) = 2 (.03) (.97) = .0582

P (2 defects) = (.03) (.03) = .0009

38. a. .90

b. P(at least 1) = $f(1) + f(2)$

$$f(1) = \frac{2!}{1!\ 1!}(.9)^1(.1)^1$$

$$= 2(.9)(.1) = .18$$

$$f(2) = \frac{2!}{2!\ 0!}(.9)^2(.1)^0$$

$$= 1(.81)(1) = .81$$

$\therefore P$(at least 1) = .18 + .81 = .99

Alternatively

$$P(\text{at least } 1) = 1 - f(0)$$

$$f(0) = \frac{2!}{0!\ 2!}(.9)^0(.1)^2 = .01$$

Therefore, P(at least 1) = 1 – .01 = .99.

c. P(at least 1) = $1 - f(0)$

$$f(0) = \frac{3!}{0!\ 3!}(.9)^0(.1)^3 = .001$$

Therefore, P(at least 1) = 1 – .001 = .999.

d. Yes; P(at least 1) becomes very close to 1 with multiple systems, and the inability to detect an attack would be catastrophic.

40. a. Yes. Because the 18- to 34-year-olds living with their parents are selected randomly, p is the same from trial to trial and the trials are independent. The two outcomes per trial

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

either do or do not contribute to household expenses.

Binomial n = 15 and p = .75

$$f(x) = \frac{15!}{x!(15-x)!}(.75)^x(1-.75)^{15-x}$$

b. The probability that none of the 15 contribute to household expenses is

$$f(0) = \frac{15!}{0!(15-0)!}(.75)^0(1-.75)^{15-0} = .0000$$

Obtaining a sample result that shows that none of the 15 contributed to household expenses is so unlikely you would have to question whether the 75% value reported by the Pew Research Center is accurate.

c. Probability of at least 10 = $f(10) + f(11) + f(12) + f(13) + f(14) + f(15)$.

Using binomial tables:

Probability = .1651 + .2252 + .2252 + .1559 + .0668 + .0134 = .8516

42. a. $f(4) = \frac{20!}{4!(20-4)!}(.30)^4(.70)^{20-4} = .1304$

b. Probability ($x \geq 2$) = 1 – $f(0)$ – $f(1)$

$$f(0) = \frac{20!}{0!(20-0)!}(.30)^0(.70)^{20-0} = .0008$$

$$f(1) = \frac{20!}{1!(20-1)!}(.30)^1(.70)^{20-1} = .0068$$

Probability ($x \geq 2$) = 1 – $f(0)$ – $f(1)$ = 1 – (.0008+ .0068) = .9924

c. $E(x) = n\,p$ = 20(.30) = 6

d. $Var(x) = n\,p\,(1 - p)$ = 20(.30)(1 −.30) = 4.2

$\sigma = \sqrt{4.2} = 2.0499$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

44. a. $f(x) = \dfrac{3^x e^{-3}}{x!}$

b. $f(2) = \dfrac{3^2 e^{-3}}{2!} = \dfrac{9(.0498)}{2} = .2241$

c. $f(1) = \dfrac{3^1 e^{-3}}{1!} = 3(.0498) = .1494$

d. $P(x \geq 2) = 1 - f(0) - f(1) = 1 - .0498 - .1494 = .8008$

46. a. $\mu = 48\ (5/60) = 4$

$$f(3) = \frac{4^3 e^{-4}}{3\,!} = \frac{(64)\,(.0183)}{6} = .1952$$

b. $\mu = 48\ (15 / 60) = 12$

$$f(10) = \frac{12^{10} e^{-12}}{10\,!} = .1048$$

c. $\mu = 48\ (5 / 60) = 4$. I expect four callers to be waiting after five minutes.

$$f(0) = \frac{4^0 e^{-4}}{0\,!} = .0183$$

The probability none will be waiting after five minutes is .0183.

d. $\mu = 48\ (3 / 60) = 2.4$

$$f(0) = \frac{2.4^0 e^{-2.4}}{0\,!} = .0907$$

The probability of no interruptions in three minutes is .0907.

48. a. For a 15-minute period, the mean is 14.4/4 = 3.6.

$$f(0) = \frac{3.6^0 e^{-3.6}}{0!} = e^{-3.6} = .0273$$

b. Probability = $1 - f(0) = 1 - .2073 = .9727$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. Probability = 1 – [$f(0) + f(1) + f(2) + f(3)$]

= 1 – [.0273+ .0984 + .1771 + .2125] = .4847

Note: The value of $f(0)$ was computed in part a; a similar procedure was used to compute the probabilities for $f(1)$, $f(2)$, and $f(3)$.

50. a. $\mu = 18/30 = .6$ per day during June.

b. $$f(0) = \frac{.6^0 e^{-.6}}{0!} = .5488$$

c. $$f(1) = \frac{.6^1 e^{-.6}}{1!} = .3293$$

d. P(More than 1) = 1 – $f(0)$ – $f(1)$ = 1 −.5488 −.3293 = .1219

52. a. $$f(1) = \frac{\binom{3}{1}\binom{10-3}{4-1}}{\binom{10}{4}} = \frac{\left(\frac{3!}{1!2!}\right)\left(\frac{7!}{3!4!}\right)}{\frac{10!}{4!6!}} = \frac{(3)(35)}{210} = .50$$

b. $$f(2) = \frac{\binom{3}{2}\binom{10-3}{2-2}}{\binom{10}{2}} = \frac{(3)(1)}{45} = .067$$

c. $$f(0) = \frac{\binom{3}{0}\binom{10-3}{2-0}}{\binom{10}{2}} = \frac{(1)(21)}{45} = .4667$$

d. $$f(2) = \frac{\binom{3}{2}\binom{10-3}{4-2}}{\binom{10}{4}} = \frac{(3)(21)}{210} = .30$$

e. Note $x = 4$ is *greater than* $r = 3$. It is not possible to have $x = 4$ successes when there are only three successes in the population. Thus, $f(4) = 0$. In this exercise, n is greater than r. Thus, the number of successes x can only take on values up to and including r

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

= 3. Thus, $x = 0, 1, 2, 3$.

54. Hypergeometric distribution with $N = 10$ and $r = 7$

a. $$f(2) = \frac{\binom{7}{2}\binom{3}{1}}{\binom{10}{3}} = \frac{(21)(3)}{120} = .5250$$

b. Compute the probability that three prefer shopping online.

$$f(3) = \frac{\binom{7}{3}\binom{3}{0}}{\binom{10}{3}} = \frac{(35)(1)}{120} = .2917$$

P(majority prefer shopping online) = $f(2) + f(3) = .5250 + .2917 = .8167$

56. $N = 60$, $n = 10$

a. $r = 20$ $x = 0$

$$f(0) = \frac{\binom{20}{0}\binom{40}{10}}{\binom{60}{10}} = \frac{(1)\left(\frac{40!}{10!30!}\right)}{\frac{60!}{10!50!}} = \left(\frac{40!}{10!30!}\right)\left(\frac{10!50!}{60!}\right)$$

$$= \frac{40\cdot 39\cdot 38\cdot 37\cdot 36\cdot 35\cdot 34\cdot 33\cdot 32\cdot 31}{60\cdot 59\cdot 58\cdot 57\cdot 56\cdot 55\cdot 54\cdot 53\cdot 52\cdot 51} \approx .0112$$

b. $r = 20$ $x = 1$

$$f(1) = \frac{\binom{20}{1}\binom{40}{9}}{\binom{60}{10}} = 20\left(\frac{40!}{9!31!}\right)\left(\frac{10!50!}{60!}\right) \approx .0725$$

c. $1 - f(0) - f(1) = 1 - .0112 - .0725 = .9163$

d. Same as the probability one will be from Hawaii; .0725.

58. Hypergeometric with $N = 10$ and $r = 3$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

a. $n = 3, x = 0$

$$f(0) = \frac{\binom{3}{0}\binom{7}{3}}{\binom{10}{3}} = \frac{\left(\frac{3!}{0!3!}\right)\left(\frac{7!}{3!4!}\right)}{\frac{10!}{3!7!}} = \frac{(1)(35)}{120} = .2917$$

This is the probability there will be no banks with increased lending in the study.

b. $n = 3, x = 3$

$$f(3) = \frac{\binom{3}{3}\binom{7}{0}}{\binom{10}{3}} = \frac{\left(\frac{3!}{3!0!}\right)\left(\frac{7!}{0!7!}\right)}{\frac{10!}{3!7!}} = \frac{(1)(1)}{120} = .0083$$

This is the probability all three banks with increased lending will be in the study. This has a very low probability of happening.

c. $n = 3, x = 1$

$$f(1) = \frac{\binom{3}{1}\binom{7}{2}}{\binom{10}{3}} = \frac{\left(\frac{3!}{1!2!}\right)\left(\frac{7!}{2!5!}\right)}{\frac{10!}{3!7!}} = \frac{(3)(21)}{120} = .5250$$

$n = 3, x = 2$

$$f(2) = \frac{\binom{3}{2}\binom{7}{1}}{\binom{10}{3}} = \frac{\left(\frac{3!}{2!1!}\right)\left(\frac{7!}{1!6!}\right)}{\frac{10!}{3!7!}} = \frac{(3)(7)}{120} = .1750$$

x	$f(x)$
0	0.2917
1	0.5250

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

2	0.1750
3	0.0083
Total	1.0000

$f(1) = .5250$ has the highest probability showing there is more than a .50 chance there will be exactly one bank that had increased lending in the study.

d. $P(x \geq 1) = 1 - f(0) = 1 - .2917 = .7083$

There is a reasonably high probability of .7083 that at least one bank with increased lending will be in the study.

e. $E(x) = n\left(\frac{r}{N}\right) = 3\left(\frac{3}{10}\right) = .90$

$$\sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right) = 3\left(\frac{3}{10}\right)\left(1 - \frac{3}{10}\right)\left(\frac{10-3}{10-1}\right) = .49$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{.49} = .70$$

60. a.

x	$f(x)$
1	.150
2	.050
3	.075
4	.050
5	.125
6	.050

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

7	.100
8	.125
9	.125
10	.150
Total	1.000

b. Probability of outstanding service is .125 + .150 = .275

c.

x	$f(x)$	$xf(x)$	$x-\mu$	$(x-\mu)^2$	$(x-\mu)^2 f(x)$
1	.150	.150	–4.925	24.2556	3.6383
2	.050	.100	–3.925	15.4056	.7703
3	.075	.225	–2.925	8.5556	.6417
4	.050	.200	–1.925	3.7056	.1853
5	.125	.625	–.925	.8556	.1070
6	.050	.300	.075	.0056	.0003
7	.100	.700	1.075	1.1556	.1156
8	.125	1.000	2.075	4.3056	.5382
9	.125	1.125	3.075	9.4556	1.1820
10	.150	1.500	4.075	16.6056	2.4908
Total	1.000	5.925			9.6694

$E(x) = 5.925$ and $Var(x) = 9.6694$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. The probability of a new car dealership receiving an outstanding wait-time rating is 2/7 = .2857. For the remaining 40 – 7 = 33 service providers, nine received an outstanding rating; this corresponds to a probability of 9/33 = .2727. For these results, there does not appear to be much difference between the probability that a new car dealership is rated outstanding compared to the same probability for other types of service providers.

62. a. There are 600 observations involving the two variables. Dividing the entries in the table shown by 600 and summing the rows and columns we obtain the following.

	Reading Material (y)			
Snacks (x)	0	1	2	Total
0	.0	.1	.03	.13
1	.4	.15	.05	.6
2	.2	.05	.02	.27
Total	.6	.3	.1	1

The entries in the body of the table are the bivariate or joint probabilities for x and y. The entries in the right most (Total) column are the marginal probabilities for x, and the entries in the bottom (Total) row are the marginal probabilities for y.

The probability of a customer purchasing one item of reading materials and two snack items is given by $f(x = 2, y = 1) = .05$.

The probability of a customer purchasing one snack item only is given by $f(x = 1, y = 0) = .40$.

The probability $f(x = 0, y = 0) = 0$ because the point of sale terminal is only used

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

when someone makes a purchase.

b. The marginal probability distribution of x along with the calculation of the expected value and variance is shown as follows.

x	$f(x)$	$xf(x)$	$x - E(x)$	$(x - E(x))^2$	$(x - E(x))^2 f(x)$
0	0.13	0	–1.14	1.2996	0.1689
1	0.60	0.6	–0.14	0.0196	0.0118
2	0.27	0.54	0.86	0.7396	0.1997
		1.14			0.3804
		$E(x)$			$Var(x)$

We see that $E(x)$ = 1.14 snack items and $Var(x)$ = .3804.

c. The marginal probability distribution of y along with the calculation of the expected value and variance is shown as follows.

y	$f(y)$	$yf(y)$	$y - E(y)$	$(y - E(y))^2$	$(y - E(y))^2 f(y)$
0	0.60	0	–0.5	0.25	0.15
1	0.30	0.3	0.5	0.25	0.075
2	0.10	0.2	1.5	2.25	0.225
		0.5			0.45
		E(y)			Var(y)

We see that $E(y)$ = .50 reading materials and $Var(y)$ = .45.

d. The probability distribution of $t = x + y$ is shown as follows along with the calculation of its expected value and variance.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

t	$f(t)$	$tf(t)$	$T - E(t)$	$(t - E(t))^2$	$(t - E(t))^2 f(t)$
1	0.50	0.5	–0.64	0.4096	0.2048
2	0.38	0.76	0.36	0.1296	0.0492
3	0.10	0.3	1.36	1.8496	0.1850
4	0.02	0.08	2.36	5.5696	0.1114
		1.64			0.5504
		E(t)			Var(t)

We see that the expected number of items purchased is *E(t)* = 1.64, and the variance in the number of purchases is *Var(t)* = .5504.

e. From part (b), *Var(x)* = .3804. From part (c), *Var(y)* = .45. And from part (d), *Var(x + y) = Var(t)* = .5504. Therefore,

$$\sigma_{xy} = [Var(x+y) - Var(x) - Var(y)]/2$$

$$= (.5504 - .3804 - .4500)/2$$

$$= -.14$$

To compute the correlation coefficient, we must first obtain the standard deviation of *x* and *y*.

$$\sigma_x = \sqrt{Var(x)} = \sqrt{.3804} = .6168$$

$$\sigma_y = \sqrt{Var(y)} = \sqrt{.45} = .6708$$

So the correlation coefficient is given by

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{-.14}{(.6168)(.6708)} = -.3384$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The relationship between the number of reading materials purchased and the number of snacks purchased is negative. This means that the more reading materials purchased the fewer snack items purchased and vice versa.

64. a. $n = 20$ and $x = 3$

$$f(3) = \binom{20}{3}(.53)^3(.47)^{17} = .0005$$

b. $n = 20$ and $x = 0$

$$P(x \le 5) = f(0) + f(1) + f(2) + f(3) + f(4) + f(5) = .4952$$

c. $E(x) = n\,p = 2000(.49) = 980$

The expected number who would find it very hard to give up their smartphone is 980.

d. $E(x) = n\,p = 2000(.36) = 720$

The expected number who would find it extremely hard to give up their e-mail is 720.

$$\sigma^2 = np\,(1 - p) = 2000(.36)(.64) = 460.8$$

$$\sigma = \sqrt{460.8} = 21.4663$$

66. Because the shipment is large, we can assume that the probabilities do not change from trial to trial and use the binomial probability distribution.

a. $n = 5$

$$f(0) = \binom{5}{0}(0.01)^0(0.99)^5 = .9510$$

b. $f(1) = \binom{5}{1}(0.01)^1(0.99)^4 = .0480$

c. $1 - f(0) = 1 - .9510 = .0490$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. No, the probability of finding one or more items in the sample defective when only 1% of the items in the population are defective is small (only .0490). I would consider it likely that more than 1% of the items are defective.

68. a. $E(x) = 200(.235) = 47$

b. $\sigma = \sqrt{np(1-p)} = \sqrt{200(.235)(.765)} = 5.9962$

c. For this situation, $p = .765$ and $(1-p) = .235$, but the answer is the same as in part b. For a binomial probability distribution, the variance for the number of successes is the same as the variance for the number of failures. Of course, this also holds true for the standard deviation.

70. $\mu = 1.5$

P(3 or more breakdowns) = $1 - [\, f(0) + f(1) + f(2)\,]$

$1 - [\, f(0) + f(1) + f(2)\,]$

$= 1 - [\, .2231 + .3347 + .2510]$

$= 1 - .8088 = .1912$

72. a. $f(3) = \dfrac{3^3 e^{-3}}{3!} = .2240$

b. $f(3) + f(4) + \cdots = 1 - [\, f(0) + f(1) + f(2)\,]$

$f(0) = \dfrac{3^0 e^{-3}}{0!} = e^{-3} = .0498$

Similarly, $f(1) = .1494$, $f(2) = .2240$

$\therefore\ 1 - [\, .0498 + .1494 + .2241\,] = .5767$

74. a. Hypergeometric distribution with $N = 10$, $n = 2$, and $r = 7$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$f(1) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{7}{1}\binom{10-7}{2-1}}{\binom{10}{2}} = \frac{\binom{7}{1}\binom{3}{1}}{\binom{10}{2}} = \frac{(7)(3)}{45} = .4667$$

b. $$f(2) = \frac{\binom{7}{2}\binom{3}{0}}{\binom{10}{2}} = \frac{(21)(1)}{45} = .4667$$

c. $$f(0) = \frac{\binom{7}{0}\binom{3}{2}}{\binom{10}{2}} = \frac{(1)(3)}{45} = .0667$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 6 Solutions

2. a.

b. $P(x < 15) = .10(5) = .50$

c. $P(12 \le x \le 18) = .10(6) = .60$

d. $E(x) = \dfrac{10+20}{2} = 15$

e. $Var(x) = \dfrac{(20-10)^2}{12} = 8.33$

4. a.

b. $P(.25 < x < .75) = 1\ (.50) = .50$

c. $P(x \le .30) = 1\ (.30) = .30$

d. $P(x > .60) = 1\ (.40) = .40$

6. a. For a uniform probability density function $f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Thus, $\frac{1}{b-a} = .00625$.

Solving for $b-a$, we have $b-a$ = 1/.00625 = 160.

In a uniform probability distribution, ½ of this interval is below the mean and ½ of this interval is above the mean. Thus,

$$a = 136 - ½(160) = 56 \text{ and } b = 136 + ½(160) = 216$$

b. $P(100 \le x \le 200) = (200-100)(.00625) = .6250$

c. $P(x \ge 150) = (216-150)(.00625) = .4125$

d. $P(x \le 80) = (80-56)(.00625) = .1500$

$$\text{EP (\$15,000)} = 1\ (\$1000) + 0\ (0) = \$1{,}000$$

8.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

10.

a. $P(z \leq 1.5) = .9332$

b. $P(z \leq 1.0) = .8413$

c. $P(1 \leq z \leq 1.5) = P(z \leq 1.5) - P(z < 1) = .9932 - .8413 = .0919$

d. $P(0 < z < 2.5) = P(z < 2.5) - P(z \leq 0) = .9938 - .5000 = .4938$

12. a. $P(0 \leq z \leq .83) = .7967 - .5000 = .2967$

b. $P(-1.57 \leq z \leq 0) = .5000 - .0582 = .4418$

c. $P(z > .44) = 1 - .6700 = .3300$

d. $P(z \geq -.23) = 1 - .4090 = .5910$

e. $P(z < 1.20) = .8849$

f. $P(z \leq -.71) = .2389$

14. a. The z-value corresponding to a cumulative probability of .9750 is $z = 1.96$.

b. The z-value here also corresponds to a cumulative probability of .9750: $z = 1.96$.

c. The z-value corresponding to a cumulative probability of .7291 is $z = .61$.

d. Area to the left of z is $1 - .1314 = .8686$. So $z = 1.12$.

e. The z-value corresponding to a cumulative probability of .6700 is $z = .44$.

f. The area to the left of z is .6700. So $z = .44$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

16. a. The area to the left of z is 1 – .0100 = .9900. The z-value in the table with a cumulative probability closest to .9900 is $z = 2.33$.

b. The area to the left of z is .9750. So $z = 1.96$.

c. The area to the left of z is .9500. Because .9500 is exactly halfway between .9495 ($z = 1.64$) and .9505($z = 1.65$), we select $z = 1.645$. However, $z = 1.64$ or $z = 1.65$ are also acceptable answers.

d. The area to the left of z is .9000. So $z = 1.28$ is the closest z-value.

18. $\mu = 14.4$ and $\sigma = 4.4$

a. At $x = 20$, $z = \frac{20 - 14.4}{4.4} = 1.27$

$$P(z \leq 1.27) = .8980$$

$$P(x \leq 20) = 1 - .8980 = .1020$$

b. At $x = 10$, $z = \frac{10 - 14.4}{4.4} = -1.00$

$$P(z \leq -1.00) = .1587$$

So,

$$P(x \leq 10) = .1587$$

c. A z-value of 1.28 cuts off an area of approximately 10% in the upper tail.

$$x = 14.4 + 4.4(1.28) = 20.03$$

A return of 20.03% or higher will put a domestic stock fund in the top 10%.

20. a. United States: $\mu = 3.73 \quad \sigma = .25$

At $x = 3.50$,

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$z = \frac{3.5 - 3.73}{.25} = -.92$$

$$P(z < -.92) = .1788$$

So, $P(x < 3.50) = .1788$

b. Russia: $\mu = 3.40 \quad \sigma = .20$

At $x = 3.50$,

$$z = \frac{3.50 - 3.40}{.20} = .50$$

$$P(z < .50) = .6915$$

So, $P(x < 3.50) = .6915$

That is, 69.15% of the gas stations in Russia charge less than $3.50 per gallon.

c. Use mean and standard deviation for Russia.

At $x = 3.73$,

$$z = \frac{3.73 - 3.40}{.20} = 1.65$$

$$P(z > 1.65) = 1 - P(z \le 1.65) = 1 - .9505 = .0495$$

$$P(x > 3.73) = .0495$$

The probability that a randomly selected gas station in Russia charges more than the mean price in the United States is .0495. Stated another way, only 4.95% of the gas stations in Russia charge more than the average price in the United States.

22. Use $\mu = 8.35$ and $\sigma = 2.5$.

a. We want to find $P(5 \le x \le 10)$.

At $x = 10$,

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$z = \frac{x-\mu}{\sigma} = \frac{10-8.35}{2.5} = .66$$

At $x = 5$,

$$z = \frac{x-\mu}{\sigma} = \frac{5-8.35}{2.5} = -1.34$$

$$P(5 \le x \le 10) = P(-1.34 \le z \le .66) = P(z \le .66) - P(z \le -1.34)$$

$$= .7454 - .0901$$

$$= .6553$$

The probability of a household viewing television between 5 and 10 hours a day is .6553.

b. Find the z-value that cuts off an area of .03 in the upper tail. Using a cumulative probability of 1 – .03 = .97, $z = 1.88$ provides an area of .03 in the upper tail of the normal distribution.

$$x = \mu + z\sigma = 8.35 + 1.88(2.5) = 13.05 \text{ hours}$$

A household must view slightly more than 13 hours of television a day to be in the top 3% of television viewing households.

c. At $x = 3$, $z = \frac{x-\mu}{\sigma} = \frac{3-8.35}{2.5} = -2.14$

$$P(x>3) = 1 - P(z \le -2.14) = 1 - .0162 = .9838$$

The probability a household views more than three hours of television a day is .9838.

24. $\mu = 749$ and $\sigma = 225$

a. $z = \frac{x-\mu}{\sigma} = \frac{400-749}{225} = -1.55$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$P(x < 400) = P(z < -1.55) = .0606$$

The probability that expenses will be less than \$400 is .0606.

b. $z = \frac{x-\mu}{\sigma} = \frac{800-749}{225} = .23$

$$P(x \geq 800) = P(z \geq .23) = 1 - P(z \leq .23) = 1 - .5910 = .4090$$

The probability that expenses will be \$800 or more is .4090.

c. For $x = 1000$, $z = \frac{x-\mu}{\sigma} = \frac{1000-749}{225} = 1.12$

For $x = 500$, $z = \frac{x-\mu}{\sigma} = \frac{500-749}{225} = -1.11$

$$P(500 \leq x \leq 1000) = P(z \leq 1.12) - P(z \leq -1.11) = .8686 - .1335 = .7351$$

The probability that expenses will be between \$500 and \$1,000 is .7351.

d. The upper 5%, or area = $1 - .05 = .95$ occurs for $z = 1.645$

$$x = \mu + z\sigma = 749 + 1.645(225) = \$1119$$

The 5% most expensive travel plans will be \$1,119 or more.

26. a. $\mu = np = 100(.20) = 20$

$$\sigma^2 = np(1-p) = 100(.20)(.80) = 16$$

$$\sigma = \sqrt{16} = 4$$

b. Yes because $np = 20$ and $n(1-p) = 80$

c. $P(23.5 \leq x \leq 24.5)$

$$z = \frac{24.5-20}{4} = +1.13$$

$$P(z \leq 1.13) = .8708$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$z = \frac{23.5 - 20}{4} = +.88$$

$$P\ (z \le .88) = .8106$$

$$P(23.5 \le x \le 24.5) = .8708 - .8106 = .0602$$

d. $P(17.5 \le x \le 22.5)$

$$z = \frac{22.5 - 20}{4} = +.63 \quad P\ (z \le .63) = .7357$$

$$z = \frac{17.5 - 20}{4} = -.63 \quad P\ (z \le -.63) = .2643$$

$$P(17.5 \le x \le 22.5) = .7357 - .2643 = .4714$$

e. $P(x \le 15.5)$

$$z = \frac{15.5 - 20}{4} = -1.13$$

$$P(x \le 15.5) = P\ (z \le -1.13) = .1292$$

28. a. $\mu = np = (250)(.20) = 50$

b. $\sigma^2 = np(1-p) = 250(.20)(1-.20) = 40$

$$\sigma = \sqrt{40} = 6.3246$$

Allowing for the continuity correction factor, $P(x < 40) = P(x \le 39.5)$

At $x = 39.5$,

$$z = \frac{x - \mu}{\sigma} = \frac{39.5 - 50}{6.3246} = -1.66$$

$$P(x \le 39.5) = .0485$$

c. Allowing for the continuity correction factor, $P(55 \le x \le 60) = P(54.5 \le x \le 60.5)$

At $x = 54.5$,

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$z = \frac{x-\mu}{\sigma} = \frac{54.5-50}{6.3246} = .71$$

At $x = 60.5$,

$$z = \frac{x-\mu}{\sigma} = \frac{60.5-50}{6.3246} = 1.66$$

$$P(54.5 \le x \le 60.5) = .9515 - .7611 = .1904$$

d. Allowing for the continuity correction factor, $P(x \ge 70) = P(x \ge 69.5)$.

At $x = 69.5$,

$$z = \frac{x-\mu}{\sigma} = \frac{69.5-50}{6.3246} = 3.08$$

$$P(x \ge 69.5) = 1 - .9990 = .0010$$

30. a. $\mu = np = 800(.18) = 144$

b. $\mu = np = 600(.18) = 108$

$$\sigma = \sqrt{np(1-p)} = \sqrt{(600)(.18)(.82)} = 9.4106$$

For $x < 100$, use the continuity correction to find $P(x < 99.5)$

At $x = 99.5$,

$$z = \frac{x-\mu}{\sigma} = \frac{99.5-108}{9.4106} = -.90 \quad P\,(z < -.90) = .1841$$

$$P(x < 99.5) = .1841$$

The probability that fewer than 100 individuals will be less than 18 years of age is .1841.

c. $\sigma = \sqrt{np(1-p)} = \sqrt{(800)(.29)(.71)} = 12.8343$

For $x = 200$ or more, use the continuity correction to find $P(x \ge 199.5)$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

At $x = 199.5$,

$$z = \frac{x-\mu}{\sigma} = \frac{199.5-232}{12.8343} = -2.53 \quad P(z \geq -2.53) = 1 - .0057 = .9943$$

$$P(x \geq 199.5) = .9943$$

The probability that 200 or more individuals will be more than 59 is .9943.

32. a. $P(x \leq 6) = 1 - e^{-6/8} = 1 - .4724 = .5276$

b. $P(x \leq 4) = 1 - e^{-4/8} = 1 - .6065 = .3935$

c. $P(x \geq 6) = 1 - P(x \leq 6) = 1 - .5276 = .4724$

d. $P(4 \leq x \leq 6) = P(x \leq 6) - P(x \leq 4) = .5276 - .3935 = .1341$

34. a. With $\mu =$ $f(x) = \frac{1}{20}e^{-x/20}$

b. $P(x \leq 15) =$ $1 - e^{-x/\mu} = 1 - e^{-15/20}$ $= .5276$

c. $P(x > 20) = 1 - P(x \leq 20)$

$= 1 - (1 - e^{-20/20}) =$ $e^{-1} = .3679$

d. With $\mu =$ $f(x) = \frac{1}{7}e^{-x/7}$

$$P(x \leq 5) = 1 - e^{-x/\mu} = 1 - e^{-5/7} = .5105$$

36. a. $f(x) = \frac{1}{\mu}e^{-x/\mu} = \frac{1}{2}e^{-x/2}$ for $x \geq 0$

$$P(x \leq x_0) = 1 - e^{-x_0/\mu}$$

$$P(x \leq 1) = 1 - e^{-1/2} = 1 - e^{-.5} = 1 - .6065 = .3935$$

b. $P(x \leq 2) = 1 - e^{-2/2} = 1 - e^{-1.0} = 1 - .3679 = .6321$

$$P(1 \leq x \leq 2) = P(x \leq 2) - P(x \leq 1) = .6321 - .3935 = .2386$$

c. For this customer, the cable service repair would have to take longer than four hours.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$P(x > 4) = 1 - P(x \le 4) = 1 - (1 - e^{-4/2}) = e^{-2.0} = .1353$$

38. a. Because the number of calls per hour follows a Poisson distribution, the time between calls follows an exponential distribution. So, for a mean of 1.6 calls per hour, the mean time between calls is

$$\mu = \frac{60 \text{ minutes/hour}}{1.6 \text{ calls/hour}} = 37.5 \text{ minutes per call}$$

b. The exponential probability density function is $f(x) = \left(\frac{1}{37.5}\right) e^{-x/37.5}$ for $x \ge 0$, where x is the minutes between 911 calls.

c. Using time in minutes,

$$P\left(x < 60\right) = 1 - e^{-60/37.5} = 1 - .2019 = .7981$$

d. $P(x \ge 30) = 1 - P\left(x \le 30\right) = 1 - (1 - e^{-30/37.5}) = 1 - .5507 = .4493$

e. $P\left(5 \le x \le 20\right) = (1 - e^{-20/37.5}) - (1 - e^{5/37.5}) = .4134 - .1248 = .2886$

40. a. Find the z-value that cuts off an area of .10 in the lower tail. From the standard normal table $z \approx -1.28$. Solve for x:

$$z = -1.28 = \frac{x - 19{,}000}{2100}$$

$$x = 19{,}000 - 1.28(2100) = 16{,}312$$

Ten percent of athletic scholarships are valued at \$16,312 or less.

b. $z = \frac{x - \mu}{\sigma} = \frac{22{,}000 - 19{,}000}{2100} = 1.43$

$$P(x \ge 22{,}000) = 1 - P(z \le 1.43) = 1 - .9236 = .0764$$

That is, 7.64% of athletic scholarships are valued at \$22,000 or more.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. Find the z-value that cuts off an area of .03 in the upper tail: $z = 1.88$. Solve for x,

$$z = 1.88 = \frac{x - 19,000}{2100}$$

$$x = 19,000 + 1.88(2100) = 22,948$$

In other words, 3% of athletic scholarships are valued at $22,948 or more.

42. $\mu = 658$

a. $z = -1.88$ cuts off .03 in the lower tail

So,

$$z = -1.88 = \frac{610 - 658}{\sigma}$$

$$\sigma = \frac{610 - 658}{-1.88} = 25.5319$$

b. At 700, $z = \frac{x - \mu}{\sigma} = \frac{700 - 658}{25.5319} = 1.65$

At 600,

$$z = \frac{x - \mu}{\sigma} = \frac{600 - 659}{25.5319} = -2.31$$

$$P(600 < x < 700) = P(-2.31 < z < 1.65) = .9505 - .0104 = .9401$$

c. $z = 1.88$ cuts off approximately .03 in the upper tail

$$x = 658 + 1.88(25.5319) = 706$$

On the busiest 3% of days 706 or more people show up at the pawnshop.

44. a. At $x = 200$,

$$z = \frac{200 - 150}{25} = 2$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$P(x > 200) = P(z > 2) = 1 - P(z \leq 2) = 1 - .9772 = .0228$$

b. Expected profit = Expected revenue – Expected cost

$$= 200 - 150 = \$50$$

46. a. At 400,

$$z = \frac{400 - 450}{100} = -.500$$

Area to left is .3085

At 500,

$$z = \frac{500 - 450}{100} = +.500$$

Area to left is .6915

$$P(400 \leq x \leq 500) = .6915 - .3085 = .3830$$

That is, 38.3% will score between 400 and 500.

b. At 630,

$$z = \frac{630 - 450}{100} = 1.80$$

In words, 96.41% do worse and 3.59% do better.

c. At 480,

$$z = \frac{480 - 450}{100} = .30$$

Area to left is .6179

That is, 38.21% are acceptable.

48. $\sigma = .6$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

At 2%

$$z \approx -2.05 \; x = 18$$

$$z = \frac{x-\mu}{\sigma} \quad \therefore -2.05 = \frac{18-\mu}{.6}$$

$$\mu = 18 + 2.05\ (.6) = 19.23 \text{ oz.}$$

The mean filling weight must be 19.23 oz.

50. a. $\mu = np = (240)(0.49) = 117.6$

Expected number of wins is 117.6

Expected number of losses = 240(0.51) = 122.4

Expected payoff = 117.6(50) – 122.4(50) = (–4.8)(50) = –240

The player should expect to lose $240.

b. To lose $1000, the player must lose 20 more hands than he wins. With 240 hands in four hours, the player must win 110 or less in order to lose $1,000. Use normal approximation to binomial.

$$\mu = np = (240)(0.49) = 117.6$$

$$\sigma = \sqrt{240(.49)(.51)} = 7.7444$$

Find $P(x \leq 110.5)$

At $x = 110.5$,

$$z = \frac{110.5 - 117.6}{7.7444} = -.92$$

$$P(x \leq 110.5) = .1788$$

The probability he will lose $1,000 or more is .1788.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. To win, the player must have 121 or more winning hands.

Find $P(x \geq 120.5)$

At $x = 120.5$,

$$z = \frac{120.5 - 117.6}{7.7444} = .37$$

$$P(x \geq 120.5) = 1 - .6443 = .3557$$

The probability that the player will win is .3557. The odds are clearly in the house's favor.

d. To lose $1,500, the player must lose 30 hands more than he wins. This means he wins 105 or fewer hands. Find $P(x \leq 105.5)$

At $x = 105.5$

$$z = \frac{105.5 - 117.6}{7.7444} = -1.56$$

$P(x \leq 105.5) = .0594$

The probability the player will go broke is .0594.

52. a. Mean time between arrivals = 1/7 minutes

b. $f(x) = 7e^{-7x}$

c. $P(x > 1) = 1 - P(x < 1) = 1 - [1 - e^{-7(1)}] = e^{-7} = .0009$

d. 12 seconds is .2 minutes

$$P(x > .2) = 1 - P(x < .2) = 1 - [1 - e^{-7(.2)}] = e^{-1.4} = .2466$$

54. a. $\frac{1}{\mu} = 0.5$ therefore $\mu = 2$ minutes = mean time between telephone calls

b. Note: 30 seconds = .5 minutes

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$P(x \leq .5) = 1 - e^{-.5/2} = 1 - .7788 = .2212$$

c. $P(x \leq 1) = 1 - e^{-1/2} = 1 - .6065 = .3935$

d. $P(x \geq 5) = 1 - P(x < 5) = 1 - (1 - e^{-5/2}) = .0821$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 7 Solutions

2. Using the last three digits of each five-digit grouping provides the random numbers:

601, 022, 448, 147, 229, 553, 147, 289, 209

Numbers greater than 350 do not apply, and the 147 can only be used once. Thus, the simple random sample of four includes 22, 147, 229, and 289.

4. a. 5, 0, 5, 8

Vale SA, Ford, General Electric

b. $\frac{N!}{n!(N-n)!} = \frac{10!}{3!(10-3)!} = \frac{3,628,800}{(6)(5040)} = 120$

6. 2782, 493, 825, 1807, 289

8. Random numbers used: 13, 8, 27, 23, 25, 18

The second occurrence of the random number 13 is ignored.

Companies selected: Goldman Sachs, Coca Cola, UnitedHealth, Pfizer, Travelers Companies Inc, and JP Morgan Chase.

10. a. Finite population. A frame could be constructed obtaining a list of licensed drivers from the New York State driver's license bureau.

b. Infinite population. Sampling from a process. The process is the production line producing boxes of cereal.

c. Infinite population. Sampling from a process. The process is one of generating arrivals to the Golden Gate Bridge.

d. Finite population. A frame could be constructed by obtaining a listing of students enrolled in the course from the professor.

e. Infinite population. Sampling from a process. The process is one of generating orders

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

for the mail-order firm.

12. a. $\bar{p}$ = 75/150 = .50

b. $\bar{p}$ = 55/150 = .3667

14. a. Two of the 40 stocks in the sample received a five-star rating.

$$\bar{p} = \frac{2}{40} = .05$$

b. Seventeen of the 40 stocks in the sample are rated Above Average with respect to risk.

$$\bar{p} = \frac{17}{40} = .425$$

c. There are eight stocks in the sample that are rated one star or two stars.

$$\bar{p} = \frac{8}{40} = .20$$

16. a. The sampled population is U. S. adults that are 50 years of age or older.

b. We would use the sample proportion for the estimate of the population proportion.

$$\bar{p} = \frac{350}{426} = .8216$$

c. The sample proportion for this issue is .74, and the sample size is 426. The number of respondents citing education as "very important" is (.74) 426 = 315.

d. We would use the sample proportion for the estimate of the population proportion.

$$\bar{p} = \frac{354}{426} = .8310$$

e. The inferences in parts b and d are being made about the population of U.S. adults who are age 50 or older. So the population of U.S. adults who are age 50 or older is the target population. The target population is the same as the sampled population. If

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

the sampled population was restricted to members of AARP who were 50 years of age or older, the sampled population would not be the same as the target population. The inferences made in parts b and d would only be valid if the population of AARP members age 50 or older was representative of the U.S. population of adults age 50 and older.

18. a. $E(\bar{x}) = \mu = 200$

b. $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 50/\sqrt{100} = 5$

c. Normal with $E(\bar{x}) = 200$ and $\sigma_{\bar{x}} = 5$

d. It shows the probability distribution of all possible sample means that can be observed with random samples of size 100. This distribution can be used to compute the probability that $\bar{x}$ is within a specified $\pm$ from μ.

20. $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

$$\sigma_{\bar{x}} = 25/\sqrt{50} = 3.54$$

$$\sigma_{\bar{x}} = 25/\sqrt{100} = 2.50$$

$$\sigma_{\bar{x}} = 25/\sqrt{150} = 2.04$$

$$\sigma_{\bar{x}} = 25/\sqrt{200} = 1.77$$

The standard error of the mean decreases as the sample size increases.

22. a. $E(\bar{x}) = 71{,}800$ and $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 4000/\sqrt{60} = 516.40$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The normal distribution for $\bar{x}$ is based on the central limit theorem.

b. For n = 120, $E(\bar{x})$ remains \$71,800, and the sampling distribution of $\bar{x}$ can still be approximated by a normal distribution. However, $\sigma_{\bar{x}}$ is reduced to $4000/\sqrt{120}$ = 365.15.

c. As the sample size is increased, the standard error of the mean, $\sigma_{\bar{x}}$, is reduced. This appears logical from the point of view that larger samples should tend to provide sample means that are closer to the population mean. Thus, the variability in the sample mean, measured in terms of $\sigma_{\bar{x}}$, should decrease as the sample size is increased.

24. a. Normal distribution, $E(\bar{x}) = 17.5$

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 4/\sqrt{50} = .57$$

b. Within one week means $16.5 \le \bar{x} \le 18.5$.

At $\bar{x}$ = 18.5, $z = \frac{18.5-17.5}{.57} = 1.75$ $\qquad$ $P(z \le 1.75) = .9599$

At $\bar{x}$ = 16.5, $z = -1.75$. $\qquad$ $P(z < -1.75) = .0401$

So $P(16.5 \le \bar{x} \le 18.5) = .9599 - .0401 = .9198$

c. Within 1/2 week means $17.0 \le \bar{x} \le 18.0$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

At $\bar{x} = 18.0$, $z = \frac{18.0 - 17.5}{.57} = .88$ $\quad P(z \le .88) = .8106$

At $\bar{x} = 17.0$, $z = -.88$ $\quad P(z < -.88) = .1894$

$P(17.0 \le \bar{x} \le 18.0) = .8106 - .1894 = .6212$

26. a. $z = \frac{\bar{x} - 16642}{\sigma / \sqrt{n}}$

Within $\pm$ 200 means $\bar{x}$ – 16,642 must be between –200 and +200.

The z value for $\bar{x}$ – 16,642 = –200 is the negative of the z value for $\bar{x}$ – 16,642 = 200. So we just show the computation of z for $\bar{x}$ – 16,642 = 200.

$n = 30$ $\quad z = \frac{200}{2400/\sqrt{30}} = .46$ $\quad P(-.46 \le z \le .46) = .6772 - .3228 = .3544$

$n = 50$ $\quad z = \frac{200}{2400/\sqrt{50}} = .59$ $\quad P(-.59 \le z \le .59) = .7224 - .2776 = .4448$

$n = 100$ $\quad z = \frac{200}{2400/\sqrt{100}} = .83$ $\quad P(-.83 \le z \le .83) = .7967 - .2033 = .5934$

$n = 400$ $\quad z = \frac{200}{2400/\sqrt{400}} = 1.67$ $\quad P(-1.67 \le z \le 1.67) = .9525 - .0475 = .9050$

b. A larger sample increases the probability that the sample mean will be within a specified distance of the population mean. In this instance, the probability of being within $\pm$200 of μ ranges from .3544 for a sample of size 30 to .9050 for a sample of size 400.

28. a. This is a graph of a normal distribution with $E(\bar{x}) = 22$ and $\sigma_{\bar{x}} = \sigma / \sqrt{n} = 4/\sqrt{30} = .7303$.

b. Within 1 inch means $21 \le \bar{x} \le 23$

$$z = \frac{23 - 22}{.7303} = 1.37$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$z = \frac{21-22}{.7303} = -1.37$$

$$P(21 \le \bar{x} \le 23) = P(-1.37 \le z \le 1.37) = .9147 - .0853 = .8294$$

The probability the sample mean will be within 1 inch of the population mean of 22 is .8294.

c. $\sigma_{\bar{x}} = \sigma / \sqrt{n} = 4 / \sqrt{45} = .5963$

Within 1 inch means $41 \le \bar{x} \le 43$

$$z = \frac{43-42}{.5963} = 1.68 \qquad z = \frac{41-42}{.5963} = -1.68$$

$$P(41 \le \bar{x} \le 43) = P(-1.68 \le z \le 1.68) = .9535 - .0465 = .9070$$

The probability the sample mean will be within 1 inch of the population mean of 42 is .9070.

The probability of being within 1 inch is greater for New York in part c because the sample size is larger.

30. a. $n / N = 40 / 4000 = .01 < .05$; therefore, the finite population correction factor is not necessary.

b. With the finite population correction factor

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\frac{\sigma}{\sqrt{n}} = \sqrt{\frac{4000-40}{4000-1}}\frac{8.2}{\sqrt{40}} = 1.29$$

Without the finite population correction factor

$$\sigma_{\bar{x}} = \sigma / \sqrt{n} = 1.30$$

Including the finite population correction factor provides only a slightly different value for $\sigma_{\bar{x}}$ than when the correction factor is not used.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. $z = \frac{\bar{x} - \mu}{1.30} = \frac{2}{1.30} = 1.54$ $P(z \leq 1.54) = .9382$

$P(z < -1.54) = .0618$

Probability = .9382 – .0618 = .8764

32. a. $E(\bar{p}) = .40$

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.40(.60)}{200}} = .0346$$

Within ± .03 means $.37 \leq \bar{p} \leq .43$

$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.03}{.0346} = .87$ $P(z \leq .87) = .8078$

$P(z < -.87) = .1922$

$P(.37 \leq \bar{p} \leq .43) = .8078 - .1922 = .6156$

b. $z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.05}{.0346} = 1.44$ $P(z \leq 1.44) = .9251$

$P(z < -1.44) = .0749$

$P(.35 \leq \bar{p} \leq .45) = .9251 - .0749 = .8502$

34. a. $\sigma_{\bar{p}} = \sqrt{\frac{(.30)(.70)}{100}} = .0458$

Within ± .04 means $.26 \leq \bar{p} \leq .34$.

$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.04}{.0458} = .87$ $P(z \leq .87) = .8078$

$P(z < -.87) = .1922$

$P(.26 \leq \bar{p} \leq .34) = .8078 - .1922 = .6156$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $\sigma_{\bar{p}} = \sqrt{\frac{(.30)(.70)}{200}} = .0324$

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.04}{.0324} = 1.23 \qquad P(z \le 1.23) = .8907$$

$P(z < -1.23) = .1093$

$P(.26 \le \bar{p} \le .34) = .8907 - .1093 = .7814$

c. $\sigma_{\bar{p}} = \sqrt{\frac{(.30)(.70)}{500}} = .0205$

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.04}{.0205} = 1.95 \qquad P(z \le 1.95) = .9744$$

$P(z < -1.95) = .0256$

$P(.26 \le \bar{p} \le .34) = .9744 - .0256 = .9488$

d. $\sigma_{\bar{p}} = \sqrt{\frac{(.30)(.70)}{1000}} = .0145$ $\quad z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.04}{.0145} = 2.76 \qquad P(z \le 2.76) = .9971$

$P(z < -2.76) = .0029$

$P(.26 \le \bar{p} \le .34) = .9971 - .0029 = .9942$

e. With a larger sample, there is a higher probability $\bar{p}$ will be within $\pm$ .04 of the population proportion p.

36. a. This is a graph of a normal distribution with a mean of $E(\bar{p}) = .55$ and

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.55(1-.55)}{200}} = .0352$$

b. Within $\pm$ .05 means $.50 \le \bar{p} \le .60$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.60 - .55}{.0352} = 1.42 \qquad z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.50 - .55}{.0352} = -1.42$$

$$P(.50 \le \bar{p} \le .60) = P(-1.42 \le z \le 1.42) = .9222 - .0778 = .8444$$

c. This is a graph of a normal distribution with a mean of $E(\bar{p}) = .45$ and

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.45(1-.45)}{200}} = .0352$$

d. $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.45(1-.45)}{200}} = .0352$

Within ± .05 means $.40 \le \bar{p} \le .50$.

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.50 - .45}{.0352} = 1.42 \qquad z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.40 - .45}{.0352} = -1.42$$

$$P(.40 \le \bar{p} \le .50) = P(-1.42 \le z \le 1.42) = .9222 - .0778 = .8444$$

e. No, the probabilities are exactly the same. This is because $\sigma_{\bar{p}}$, the standard error, and the width of the interval are the same in both cases. Notice the formula for computing the standard error. It involves $p(1-p)$. So whenever $p = 1 - p$, the standard error will be the same. In part b, $p = .45$ and $1 - p = .55$. In part d, $p = .55$ and $1 - p = .45$.

f. For $n = 400$, $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.55(1-.55)}{400}} = .0249$.

Within ± .05 means $.50 \le \bar{p} \le .60$.

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.60 - .55}{.0249} = 2.01 \qquad z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.50 - .55}{.0249} = -2.01$$

$$P(.50 \le \bar{p} \le .60) = P(-2.01 \le z \le 2.01) = .9778 - .0222 = .9556$$

The probability is larger than in part b. This is because the larger sample size has

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

reduced the standard error from .0352 to .0249.

38. a. It is a normal distribution with $E(\bar{p}) = .42$.

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.42)(.58)}{300}} = .0285$$

b. $z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.03}{.0285} = 1.05$ $\quad P(z \le 1.05) = .8531$

$P(z < -1.05) = .1469$

$P(.39 \le \bar{p} \le .44) = .8531 - .1469 = .7062$

c. $z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.05}{.0285} = 1.75$ $\quad P(z \le 1.75) = .9599$

$P(z < -1.75) = .0401$

$P(.39 \le \bar{p} \le .44) = .9599 - .0401 = .9198$

d. The probabilities would increase. This is because the increase in the sample size makes the standard error, $\sigma_{\bar{p}}$, smaller.

40. a. $E(\bar{p}) = .76$

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.76(1-.76)}{400}} = .0214$$

Normal distribution because $np = 400(.76) = 304$ and $n(1 - p) = 400(.24) = 96$.

b. $z = \frac{.79 - .76}{.0214} = 1.40$ $\quad P(z \le 1.40) = .9192$

$z = \frac{.73 - .76}{.0214} = -1.40$ $\quad P(z < -1.40) = .0808$

$P(.73 \le \bar{p} \le .79) = P(-1.40 \le z \le 1.40) = .9192 - .0808 = .8384$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.76(1-.76)}{750}} = .0156$

$z = \frac{.79-.76}{.0156} = 1.92$ $\qquad$ $P(z \leq 1.92) = .9726$

$z = \frac{.73-.76}{.0156} = -1.92$ $\qquad$ $P(z < -1.92) = .0274$

$P(.73 \leq \bar{p} \leq .79) = P(-1.92 \leq z \leq 1.92) = .9726 - .0274 = .9452$

42. a. $E(\bar{x}) = \mu = 400$

b. $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 100/\sqrt{100{,}000} = .3162$

c. Normal with $E(\bar{x}) = 400$ and $\sigma_{\bar{x}} = .3162$.

d. It shows the probability distribution of all possible sample means that can be observed with random samples of size 100,000. This distribution can be used to compute the probability that $\bar{x}$ is within a specified $\pm$ from μ.

44. a. $E(\bar{p}) = p = .75$

b. $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.75(.25)}{100{,}000}} = .0014$

c. Normal distribution with $E(\bar{p}) = .75$ and $\sigma_{\bar{p}} = .0014$

d. It shows the probability distribution of all possible sample proportions that can be observed with random samples of size 100,000. This distribution can be used to compute the probability that $\bar{p}$ is within a specified $\pm$ from p.

46. a. Normal distribution, $E(\bar{x}) = 100$

$\sigma_{\bar{x}} = \sigma / \sqrt{n} = 48 / \sqrt{15{,}000} = .3919$

b. Within ± 1 hour means $99 \leq \bar{x} \leq 101$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

At $\bar{x}$= 101, $z = \frac{101-100}{.3919} = 2.55$ $\quad P(z \le 2.55) = .9946$

At $\bar{x}$= 99, $z = \frac{99-100}{.3919} = -2.55$ $\quad P(z < -2.55) = .0054$

So $P(99 \le \bar{x} \le 101) = .9946 - .0054 = .9892$.

c. In part a we found that $P(99 \le \bar{x} \le 101) = .9892$, so the probability the mean annual number of hours of vacation time earned for a sample of 15,000 blue collar and service employees who work for small private establishments and have at least 10 years of service differs from the population mean μ by more than 1 hour is 1 – . 9892 – .0108, or approximately 1%.

Because these sample results are unlikely if the population mean is 100, the sample results suggest either (a) the mean annual number of hours of vacation time earned by blue-collar and service employees who work for small private establishments and have at least 10 years of service is not 100 or (b) the sample is not representative of the population of blue-collar and service employees who work for small private establishments and have at least 10 years of service. The sampling procedure should be carefully reviewed.

48. a.

The normal distribution is appropriate because np = 108,700(.42) = 45,654 and $n(1 - p)$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

=108,700(.58) = 63,046 are both greater than 5.

b. $P\,(.419 \le \bar{p} \le .421) = ?$

$$z = \frac{.421 - .42}{.0015} = 0.67 \qquad P(z \le 0.67) = .7486$$

$$z = \frac{.419 - .42}{.0015} = -0.67 \qquad P(z < -0.67) = .2514$$

$$P(.419 \le \bar{p} \le .421) = .7486 - .2514 = .4972$$

c. $P\,(.4175 \ge \bar{p} \ge .4225) = ?$

$$z = \frac{.4225 - .42}{.0015} = 1.67 \qquad P(z \le 1.67) = .9525$$

$$z = \frac{.4175 - .42}{.0015} = -1.67 \qquad P(z < -1.67) = .0475$$

$$P(.25 \le \bar{p} \le .35) = .9525 - .0475 = .9050$$

For samples of the same size, the probability of being within 1% of the population proportion of repeat purchasers is much smaller than the probability of being within .25% of the population proportion of repeat purchasers.

d. $P\,(.41 \le \bar{p} \le .43) = ?$

$$z = \frac{.43 - .42}{.0015} = 6.68 \qquad P(z \le 6.68) \approx 1.0000$$

$$z = \frac{.41 - .42}{.0015} = -6.68 \qquad P(z < -6.68) \approx .0000$$

$$P(.41 \le \bar{p} \le .43) \approx 1.0000 - .0000 = .0000$$

Because the probability the proportion of orders placed by repeat customers for a sample of 108,700 orders from the past six months differs from the population proportion p by *less than 1%* is approximately 1.0000, the probability the proportion of

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

orders placed by repeat customers for a sample of 108,700 orders from the past six months differs from the population proportion p by *more than 1%* is approximately 1 – 1.0000 = .0000.

These sample results are unlikely if the population proportion is .42, so these sample results would suggest either (1) the population proportion of orders placed by repeat customers from the past six months is not 42% or (2) the sample is not representative of the population of orders placed by repeat customers from the past six months. The sampling procedure should be carefully reviewed.

50. a. Sorting the list of companies and random numbers to identify the five companies associated with the five smallest random numbers provides the following sample.

Company	Random Number
LMI Aerospace	.008012
Alpha & Omega Semiconductor	.055369
Olympic Steel	.059279
Kimball International	.144127
International Shipholding	.227759

b. Step 1: Generate a new set of random numbers in column B. Step 2: Sort the random numbers and corresponding company names into ascending order and then select the companies associated with the five smallest random numbers. It is extremely unlikely that you will get the same companies as in part a.

52. a. Normal distribution with

$$E(\bar{x}) = 406$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{80}{\sqrt{64}} = 10$$

b. $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{15}{80 / \sqrt{64}} = 1.50$ $P(z \leq 1.50) = .9332$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{-15}{80 / \sqrt{64}} = -1.50 \qquad P(z < -1.50) = .0668$$

$$P(391 \leq \bar{x} \leq 421) = P(-1.50 \leq z \leq 1.50) = .9332 - .0668 = .8664$$

c. At $\bar{x} = 380$, $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{380 - 406}{80 / \sqrt{64}} = -2.60$

$$P(\bar{x} \leq 380) = P(z \leq -2.60) = .0047$$

Yes, this is an unusually low performing group of 64 stores. The probability of a sample mean annual sales per square foot of $380 or less is only .0047.

54. $\mu = 27{,}175$ $\sigma = 7400$

a. $\sigma_{\bar{x}} = 7400 / \sqrt{60} = 955$

b. $z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{0}{955} = 0$

$P(\bar{x} > 27{,}175) = P(z > 0) = .50$

Note: This could have been answered easily without any calculations; 27,175 is the expected value of the sampling distribution of $\bar{x}$.

c. $z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{1000}{955} = 1.05$ $P(z \leq 1.05) = .8531$

$$P(z < -1.05) = .1469$$

$$P(26{,}175 \leq \bar{x} \leq 28{,}175) = P(-1.05 \leq z \leq 1.05) = .8531 - .1469 = .7062$$

d. $\sigma_{\bar{x}} = 7400 / \sqrt{100} = 740$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{1000}{740} = 1.35 \qquad P(z \le 1.35) = .9115$$

$P(z < -1.35) = .0885$

$P(26{,}175 \le \bar{x} \le 28{,}175) = P(-1.35 \le z \le 1.35) = .9115 - .0885 = .8230$

56. a. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{\sqrt{n}} = 20$

$\sqrt{n} = 500/20 = 25$ and $n = (25)^2 = 625$

b. For $\pm$ 25,

$$z = \frac{25}{20} = 1.25 \quad P(z \le 1.25) = .8944$$

$P(z < -1.25) = .1056$

Probability = $P(-1.25 \le z \le 1.25) = .8944 - .1056 = .7888$

58. $p = .15$

a. This is the graph of a normal distribution with $E(\bar{p}) = p = .15$ and

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.15(1-.15)}{240}} = .0230 .$$

b. Within $\pm$ .04 means $.11 \le \bar{p} \le .19$

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.19 - .15}{.0230} = 1.74 \qquad z = \frac{.11 - .15}{.0230} = -1.74$$

$P(.11 \le \bar{p} \le .19) = P(-1.74 \le z \le 1.74) = .9591 - .0409 = .9182$

c. Within $\pm$.02 means $.13 \le \bar{p} \le .17$.

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.17 - .15}{.0230} = .87 \qquad z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.13 - .15}{.0230} = -.87$$

$P(.13 \le \bar{p} \le .17) = P(-.87 \le z \le .87) = .8078 - .1922 = .6156$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

60. a. $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.40)(1-.40)}{380}} = .0251$

Within ± .04 means $.36 \leq \bar{p} \leq .44$.

$$z = \frac{.44-.40}{.0251} = 1.59 \qquad z = \frac{.36-.40}{.0251} = -1.59$$

$P(.36 \leq \bar{p} \leq .44) = P(-1.59 \leq z \leq 1.59) = .9441 - .0559 = .8882$

b. We want $P(\bar{p} \geq .45)$.

$$z = \frac{\bar{p}-p}{\sigma_{\bar{p}}} = \frac{.45-.40}{.0251} = 1.99$$

$P(\bar{p} \geq .45) = 1 - P(\bar{p} \leq .45) = 1 - P(z \leq 1.99) = 1 - .9767 = .0233$

62. a. $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.25(.75)}{n}} = .0625$

Solve for n

$$n = \frac{.25(.75)}{(.0625)^2} = 48$$

b. Normal distribution with $E(\bar{p}) = .25$ and $\sigma_{\bar{p}} = .0625$.

[Note: (48)(.25) = 12 > 5, and (48)(.75) = 36 > 5.]

c. $P(\bar{p} \geq .30) = ?$

$$z = \frac{.30-.25}{.0625} = .80 \qquad P(z \leq .80) = .7881$$

$P(\bar{p} \geq .30) = 1 - P(\bar{p} \leq .30) = 1 - P(z \leq .80) = 1 - .7881$ 1 – .7881 =

.2119

64. a. Normal distribution, $E(\bar{x}) = 17.6$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$\sigma_{\bar{x}} = \sigma / \sqrt{n} = 5.1/\sqrt{85{,}020} = .0175$

b. Within ±3 minutes of the mean is the same as within 3/60 – .05 hours of the mean, i.e., 17.55 $\leq \bar{x} \leq$ 17.65.

At $\bar{x}$= 17.65, $z = \dfrac{17.65 - 17.6}{.0175} = 2.86$ $\quad P(z \leq 2.86) = .9979$

At $\bar{x}$= 17.55, $z = -2.86$. $\quad P(z < -2.86) = .0021$

So $P(17.55 \leq \bar{x} \leq 17.65) = .9979 - .0021 = .9958$

c. In part b we found that if we assume the U.S. population mean and standard deviation are appropriate for Florida, then $P(17.55 \leq \bar{x} \leq 17.65) = .9958$, so the probability the mean for a sample of 85,020 Floridians will differ from the U.S. population mean by more than 3 minutes is 1 – .9958 = .0042. This result is would be highly unlikely and would suggest that the home Internet usage of Floridians differs from home Internet usage for the rest of the United States.

66. a.

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.58(.42)}{20{,}000}} = .0035$$

The normal distribution is appropriate because np = 20,000(.58) = 11,600 and $n(1 - p)$ =20,000(.42) = 8,400 are both greater than 5.

b. $P\,(.57 \leq \bar{p} \leq .59) = ?$

$z = \dfrac{.59 - .58}{.0035} = 2.87$ $\qquad P(z \leq 2.87) = .9979$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$z = \frac{.57 - .58}{.0035} = -2.87 \qquad P(z < -2.87) = .0021$$

$$P(.57 \le \bar{p} \le .59) = .9979 - .0021 = .9958$$

c. In part b we found that $P(.57 \le \bar{x} \le .59) = .9958$, so the probability the proportion of a sample of 20,000 drivers that is speeding will differ from the U.S. population proportion of drivers that is speeding by more than 1% is 1 – .9958 = .0042.

These sample results are unlikely if the population proportion is .58, so these sample results would suggest either (1) the population proportion of drivers who are speeding is not 58% or (2) the sample is not representative of the population of U.S. drivers. The sampling procedure should be carefully reviewed.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 8 Solutions

2. a. $32 \pm 1.645\ (6/\sqrt{50})$

   32 ± 1.4 or 30.6 to 33.4

   b. $32 \pm 1.96\ (6/\sqrt{50})$

   32 ± 1.66 or 30.34 to 33.66

   c. $32 \pm 2.576\ (6/\sqrt{50})$

   32 ± 2.19 or 29.81 to 34.19

6. A 95% confidence interval is of the form $\bar{x} \pm z_{.025}(\sigma/\sqrt{n})$.

   Using Excel and the web file TravelTax, the sample mean is $\bar{x} = 40.31$ and the sample size is $n = 400$. The sample-standard deviation $\sigma = 85$ is known. The confidence interval is

   $$40.31 \pm 1.96(8.5/\sqrt{200})$$

   40.31 ± 1.18 $40.31 \underline{\pm 1.18}$ or 39.13 to 41.49

8. a. Because n is small, an assumption that the population is at least approximately normal is required so that the sampling distribution of $\bar{x}$ can be approximated by a normal distribution.

   b. Margin of error: $z_{.025}(\sigma/\sqrt{n}) = 1.96(5.5/\sqrt{10}) = 3.41$

   c. Margin of error: $z_{.005}(\sigma/\sqrt{n}) = 2.576(5.5/\sqrt{10}) = 4.48$

10. a. $\bar{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

   $3{,}486 \pm 1.645\ (650/\sqrt{120})$

   $3{,}486 \pm 98$ or \$3388 to \$3584

   b. $3{,}486 \pm 1.96\ (650/\sqrt{120})$

   $3{,}486 \pm 116$ or \$3370 to \$3602

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. $3{,}486 \pm 2.576\ (650/\sqrt{120})$

$3{,}486 \pm 153$ or \$3,333 to \$3,639

The confidence interval gets wider as we increase our confidence level. We need a wider interval to be more confident that the interval will contain the population mean.

12. a. 2.179

b. –1.676

c. 2.457

d. Use .05 column, –1.708 and 1.708

e. Use .025 column, –2.014 and 2.014

14. $\bar{x} \pm t_{\alpha/2}(s/\sqrt{n})$, $df = 53$

a. $22.5 \pm 1.674\,(4.4/\sqrt{54})$

22.5 ± 1 or 21.5 to 23.5

b. $22.5 \pm 2.006\,(4.4/\sqrt{54})$

22.5 ± 1.2 or 21.3 to 23.7

c. $22.5 \pm 2.672\,(4.4/\sqrt{54})$

22.5 ± 1.6 or 20.9 to 24.1

d. As the confidence level increases, there is a larger margin of error and a wider confidence interval.

16. a. Using JMP or Excel, $\bar{x} = 9.7063$ and $s = 7.9805$

The sample mean years to maturity is 9.7063 years with a standard deviation of 7.9805.

b. $\bar{x} \pm t_{.025}(s/\sqrt{n})$, $df = 39$, $t_{.025} = 2.023$

$9.7063 \pm 2.023\,(7.9805/\sqrt{40})$

9.7063 ± 2.5527 or 7.1536 to 12.2590

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The 95% confidence interval for the population mean years to maturity is 7.1536 to 12.2590 years.

c. Using JMP or Excel, $\bar{x}$= 3.8854 and s = 1.6194

The sample mean yield on corporate bonds is 3.8854% with a standard deviation of 1.6194.

d. $\bar{x} \pm t_{.025}(s/\sqrt{n})$ $df = 39$ $t_{.025} = 2.023$

$3.8854 \pm 2.023\,(1.6194/\sqrt{40})$

3.8854 ± .5180 or 3.3674 to 4.4034

The 95% confidence interval for the population mean yield is 3.3674 to 4.4034%.

18. For the JobSearch data set, $\bar{x}$=22 and s = 11.8862

a. $\bar{x}$ = 22 weeks

b. margin of error = $t_{.025}s/\sqrt{n} = 2.023(11.8862)/\sqrt{40} = 3.8020$

c. The 95% confidence interval is $\bar{x} \pm$ margin of error

22 ± 3.8020 or 18.20 to 25.80

d. Skewness = 1.0062, data are skewed to the right. Use a larger sample next time.

20. a. $\bar{x} = \Sigma x_i / n = 51{,}020 / 20 = 2551$

The point estimate of the mean annual auto insurance premium in Michigan is $2,551.

b. $$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\Sigma(x_i - 2551)^2}{20-1}} = 301.3077$$

95% confidence interval: $\bar{x} \pm t_{.025}(s/\sqrt{n})$, $df = 19$

$2{,}551 \pm 2.093\,(301.3077/\sqrt{20})$

$2,551 ± 141.01 or $2,409.99 to $2,692.01

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. The 95% confidence interval for Michigan does not include the national average of $1,503 for the United States. We would be 95% confident that auto insurance premiums in Michigan are above the national average.

22. a. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{693{,}000}{30} = \$23{,}100$

The point estimate of the population mean ticket sales revenue per theater is $23,100.

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = 3981.89$$

95% confidence interval: $\bar{x} \pm t_{.025}(s/\sqrt{n})$ with $df = 29$ $\quad t_{.025} = 2.045$

$$23{,}100 \pm 2.045\,\frac{3981.89}{\sqrt{30}}$$

$$23{,}100 \pm 1{,}487$$

The 95% confidence interval for the population mean is $21,613 to $24,587.

b. Mean number of customers per theater = 23,100/8.11 = 2,848 customers per theater

c. Total number of customers = 4,080(2,848) = 11,619,840

Total box office ticket sales for the weekend = 4,080(23,100) = $94,248,000

24. a. Planning value of σ = Range/4 = 36/4 = 9

b. $n = \frac{z_{.025}^2\sigma^2}{E^2} = \frac{(1.96)^2(9)^2}{3^2} = 34.57$ Use $n = 35$

c. $n = \frac{(1.96)^2(9)^2}{2^2} = 77.79$ Use $n = 78$

26. a. $n = \frac{z_{.025}^2\sigma^2}{E^2} = \frac{(1.96)^2(.25)^2}{(.10)^2} = 24.01$ Use 25.

If the normality assumption for the population appears questionable, this should be adjusted upward to at least 30.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $n = \frac{(1.96)^2(.25)^2}{(.07)^2} = 49$

Use 49 to guarantee a margin of error no greater than .07. However, the U.S. EIA may choose to increase the sample size to a round number of 50.

c. $n = \frac{(1.96)^2(.25)^2}{(.05)^2} = 96.04$ Use 97

For reporting purposes, the U.S. EIA might decide to round up to a sample size of 100.

28. a. $n = \frac{z_{\alpha/2}^2\sigma^2}{E^2} = \frac{(1.645)^2(25)^2}{(3)^2} = 187.92$ Use $n = 188$

b. $n = \frac{(1.96)^2(25)^2}{(3)^2} = 266.78$ Use $n = 267$

c. $n = \frac{(2.576)^2(25)^2}{(3)^2} = 460.82$ Use $n = 461$

d. The sample size gets larger as the confidence level is increased. We would not recommend 99% confidence. The sample size must be increased by 79 respondents (267 – 188) to go from 90% to 95%. This may be reasonable. However, increasing the sample size by 194 respondents (461 – 267) to go from 95% to 99% would probably be viewed as too expensive and time consuming for the 4% gain in confidence level.

30. Planning value from previous study: $\sigma = 2000$

$$n = \frac{z_{.025}^2\sigma^2}{E^2} = \frac{(1.96)^2(2000)^2}{(100)^2} = 1536.64$$

Use n = 1537 to guarantee the margin of error will not exceed 100.

32. a. $.70 \pm 1.645\sqrt{\frac{.70(.30)}{800}}$

$.70 \pm .0267$ or .6733 to .7267

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $.70 \pm 1.96\sqrt{\frac{.70(.30)}{800}}$

$.70 \pm .0318$ or .6682 to .7318

34. Use planning value $p^* = .50$

$$n = \frac{(1.96)^2(.50)(.50)}{(.03)^2} = 1067.11 \quad \text{Use } n = 1068$$

36. a. $\bar{p} = 46/200 = .23$

b. $\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{.23(1-.23)}{200}} = .0298$

$$\bar{p} \pm z_{.025}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$.23 \pm 1.96(.0298)$

$.23 \pm .0584$ or .1716 to .2884

38. a. $\bar{p} = 81/142 = .5704$

b. Margin of error = $1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 1.96\sqrt{\frac{(.5704)(.4296)}{142}} = .0814$

Confidence interval: $.5704 \pm .0814$ or .4890 to .6518

c. $n = \frac{1.96^2(.5704)(.4296)}{(.03)^2} = 1046$

40. Margin of error: $z_{.025}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 1.96\sqrt{\frac{.52(1-.52)}{800}} = .0346$

95% confidence interval: $\bar{p} \pm .0346$

$.52 \pm .0346$ or .4854 to .5546

42. a. $\sqrt{\frac{p^*(1-p^*)}{n}} = \sqrt{\frac{.50(1-.50)}{491}} = .0226$

$z_{.025}\sqrt{\frac{p^*(1-p^*)}{n}} = 1.96(.0226) = .0442$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $n = \dfrac{z_{.025}^2 p^*(1-p^*)}{E^2}$

September $n = \dfrac{1.96^2(.50)(1-.50)}{.04^2} = 600.25$ Use 601

October $n = \dfrac{1.96^2(.50)(1-.50)}{.03^2} = 1067.11$ Use 1068

November $n = \dfrac{1.96^2(.50)(1-.50)}{.02^2} = 2401$

Preelection $n = \dfrac{1.96^2(.50)(1-.50)}{.01^2} = 9604$

44. a. $\bar{x}$ = 326.6674

b. degrees of freedom = 10,000 and $s_{\bar{x}}$ = 12,318.01, so $t_{\alpha/2}s_{\bar{x}}$ =

1.9600(12,318.01/√10,001) = 241.4165.

c. $\bar{x} \pm t_{\alpha/2}s_{\bar{x}}$ = 326.6674 ± 241.4165 = (85.25, 568.08)

46. a. $\bar{p}$ = 65120/102519 = .6352

b. $z_{\alpha/2}\sigma_{\bar{p}}$ = 1.96(.0015) = .0029

c. $\bar{x} \pm z_{\alpha/2}\sigma_{\bar{p}}$ = .6352 ± .0029 = (.6323, .6381)

48. a. Margin of error: $z_{.025}\dfrac{\sigma}{\sqrt{n}} = 1.96\dfrac{15}{\sqrt{54}} = 4.00$

b. Confidence interval: $\bar{x} \pm$ margin of error

33.77 ± 4.00 or \$29.77 to \$37.77

50. a. Margin of error = $t_{.025}(s/\sqrt{n})$

df = 79 $t_{.025}$ = 1.990 s = 550

$1.990(550/\sqrt{80})$ = 122

b. $\bar{x}$ ± margin of error

1873 ± 122 or \$1,751 to \$1,995

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. Because 92 million Americans were age 50 and older, the estimate of total expenditures = 92(1873) = 172,316.

In dollars, we estimate that $172,316 million dollars are spent annually by Americans of age 50 and older on restaurants and carryout food.

d. We would expect the median to be less than the mean of $1,873. The few individuals who spend much more than the average cause the mean to be larger than the median. This is typical for data of this type.

52. a. Using Excel shows that the sample mean and sample standard deviation are $\bar{x} = 773$ and $s = 738.3835$.

$$\text{Margin of error} = t_{.05}\left(\frac{s}{\sqrt{n}}\right) = 1.645\left(\frac{738.3835}{\sqrt{400}}\right) = 60.73$$

90% Confidence Interval: 773 ± 60.73 or $712.27 to $833.73

b. Using Excel shows that the sample mean and sample standard deviation are $\bar{x} = 187$ and $s = 178.6207$.

$$\text{Margin of error} = t_{.05}\left(\frac{s}{\sqrt{n}}\right) = 1.645\left(\frac{178.6207}{\sqrt{400}}\right) = 14.69$$

90% confidence interval: 187 ± 14.69 or $172.31 to $201.69

c. There were 136 employees who had no prescription medication cost for the year: $\bar{p} = 136/400 = .34$.

d. The margin of error in part a is 60.73; the margin of error in part c is 14.69. The margin of error in part a is larger because the sample standard deviation in part a is larger. The sample size and confidence level are the same in both parts.

54. $n = \dfrac{(2.33)^2(2.6)^2}{1^2} = 36.7$ Use $n = 37$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

56. $n = \frac{(1.96)^2(675)^2}{100^2} = 175.03$ Use $n = 176$

58. a. $\bar{p}$ = 200/369 = .5420

b. $1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 1.96\sqrt{\frac{(.5420)(.4580)}{369}} = .0508$

c. .5420 $\pm$ .0508 or .4912 to .5928

60. a. With 165 out of 750 respondents rating the economy as good or excellent,

$\bar{p}$= 165/750 = .22

b. Margin of error $= 1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 1.96\sqrt{\frac{.22(1-.22)}{750}} = .0296$

95% Confidence interval: .22 $\pm$ .0296 or .1904 to .2496

c. With 315 out of 750 respondents rating the economy as poor,

$\bar{p}$= 315/750 = .42

Margin of error $= 1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 1.96\sqrt{\frac{.42(1-.42)}{750}} = .0353$

95% Confidence interval: .42 $\pm$ .0353 or .3847 to .4553

d. The confidence interval in part c is wider. This is because the sample proportion is closer to .5 in part c.

62. a. $n = \frac{(2.33)^2(.70)(.30)}{(.03)^2} = 1266.74$ Use $n = 1267$

b. $n = \frac{(2.33)^2(.50)(.50)}{(.03)^2} = 1508.03$ Use $n = 1509$

64. a. $\bar{p}$ = 618/1993 = .3101

b. $\bar{p} \pm 1.96\sqrt{\frac{\bar{p}(1-\bar{p})}{1993}}$

.3101 $\pm$ 1.96 $\sqrt{\frac{(.3101)(.6899)}{1993}}$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

.3101 $\pm$ .0203 or .2898 to .3304

c. $$n = \frac{z^2 p^*(1-p^*)}{E^2}$$

$$z = \frac{(1.96)^2(.3101)(.6899)}{(.01)^2} = 8218.64 \quad \text{Use } n = 8219$$

No, the sample appears unnecessarily large. The .02 margin of error reported in part b should provide adequate precision.

66. a. $\bar{x}$ = 34.9872

b. degrees of freedom = 28,584 and s = 2.9246,

so $t_{\alpha/2}s_{\bar{x}}$ = 2.58(2.9246/$\sqrt{28{,}584}$) = .0446

c. $\bar{x} \pm t_{\alpha/2}s_{\bar{x}}$ = 34.9872 ± .0446 = (34.9426, 35.0318)

d. The 99% confidence interval for the mean hours worked hours during the past week does not include the value for the mean hours worked during the same week one year ago. This suggests that the mean hours worked taken changed from last year to this year.

68. a. $\bar{p}$ = 490/8749 = .0560

b. $z_{\alpha/2}\sigma_{\bar{p}}$ = 1.64(.0025) = .0040

c. $\bar{x} \pm z_{\alpha/2}\sigma_{\bar{p}}$ = .0560 ± .0040 = (.0520, .0600)

d. The 90% confidence interval for the proportion of California bridges that are deficient does not include the value for the proportion of deficient bridges in the entire country. This suggests that the proportion of California bridges that is deficient differs from the proportion for the U.S.

Even though the IRC report indicates that California has a large proportion of the nation's deficient bridges, California has a large total number of bridges, so the

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

proportion of bridges in California that are deficient is smaller than the proportion of deficient bridges nationwide.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 9 Solutions

2. a. H_0: $\mu \leq 14$

   H_a: $\mu > 14$ Research hypothesis

   b. There is no statistical evidence that the new bonus plan increases sales volume.

   c. The research hypothesis that $\mu > 14$ is supported. We can conclude that the new bonus plan increases the mean sales volume.

4. a. H_0: $\mu \geq 220$

   H_a: $\mu < 220$

   Research hypothesis to see if mean cost is less than $220.

   b. We are unable to conclude that the new method reduces costs.

   c. Conclude $\mu < 220$. Consider implementing the new method based on the conclusion that it lowers the mean cost per hour.

6. a. H_0: $\mu \leq 1$ The label claim or assumption

   H_a: $\mu > 1$

   b. Claiming $\mu > 1$ when it is not. This is the error of rejecting the product's claim when the claim is true.

   c. Concluding $\mu \leq 1$ when it is not. In this case, we miss the fact that the product is not meeting its label specification.

8. a. H_0: $\mu \geq 220$

   H_a: $\mu < 220$

   b. Claiming $\mu < 220$ when the new method does not lower costs. A mistake could be implementing the method when it does not help.

   c. Concluding $\mu \geq 220$ when the method really would lower costs. This could lead to not

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

implementing a method that would lower costs.

10. a. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{26.4 - 25}{6 / \sqrt{40}} = 1.48$

b. Using normal table with z = 1.48: p-value = 1.0000 – .9306 = .0694.

c. p-value > .01, do not reject H_0.

d. Reject H_0 if $z \geq 2.33$

1.48 < 2.33, do not reject H_0.

12. a. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{78.5 - 80}{12 / \sqrt{100}} = -1.25$

Lower tail p-value is the area to the left of the test statistic.

Using normal table with z = –1.25: p-value =.1056.

p-value > .01, do not reject H_0.

b. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{77 - 80}{12 / \sqrt{100}} = -2.50$

Lower tail p-value is the area to the left of the test statistic.

Using normal table with z = –2.50: p-value =.0062.

p-value ≤ .01, reject H_0.

c. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{75.5 - 80}{12 / \sqrt{100}} = -3.75$

Lower tail p-value is the area to the left of the test statistic.

Using normal table with z = –3.75: p-value ≈ 0.

p-value ≤ .01, reject H_0.

d. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{81 - 80}{12 / \sqrt{100}} = .83$

Lower tail p-value is the area to the left of the test statistic.

Using normal table with z = .83: p-value =.7967.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

p-value > .01, do not reject H_0.

14. a. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{23 - 22}{10 / \sqrt{75}} = .87$

Because $z > 0$, p-value is two times the upper tail area.

Using normal table with $z = .87$: p-value = 2(1 – .8078) = .3844.

p-value > .01, do not reject H_0.

b. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{25.1 - 22}{10 / \sqrt{75}} = 2.68$

Because $z > 0$, p-value is two times the upper tail area.

Using normal table with $z = 2.68$: p-value = 2(1 – .9963) = .0074.

p-value ≤ .01, reject H_0.

c. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{20 - 22}{10 / \sqrt{75}} = -1.73$

Because $z < 0$, p-value is two times the lower tail area.

Using normal table with $z = –1.73$: p-value = 2(.0418) = .0836.

p-value > .01, do not reject H_0.

16. a. H_0: $\mu \leq 3{,}173$

H_a: $\mu > 3{,}173$.

b. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{3325 - 3173}{1000 / \sqrt{180}} = 2.04$

p-value = 1.0000 – .9793 = .0207.

c. p-value ≤ .05. Reject H_0. The current population mean credit card balance for undergraduate students has increased compared to the previous all-time high of $3,173 reported in the original report.

18. a. H_0: $\mu \geq 60$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

H_a: $\mu < 60$

b. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{55 - 60}{27.4 / \sqrt{150}} = -2.23$

Lower tail p-value is the area to the left of the test statistic.

Using normal table with $z = -2.23$: p-value = .0129.

c. p-value = .0129 $\leq \alpha = .05$.

Reject H_0 and conclude that the mean number of hours worked per week during tax season by CPAs in states with flat state income tax rates is less than the mean hours CPAs work during tax season throughout the United States.

20. a. H_0: $\mu \geq 838$

H_a: $\mu < 838$

b. $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{745 - 838}{300/\sqrt{60}} = -2.40$

c. Lower tail p-value is area to left of the test statistic.

Using normal table with $z = -2.40$: p-value = .0082.

d. p-value $\leq .01$; reject H_0. Conclude that the annual expenditure per person on prescription drugs is less in the Midwest than in the Northeast.

22. a. H_0: $\mu = 8$

H_a: $\mu \neq 8$

b. $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{8.4 - 8.0}{3.2 / \sqrt{120}} = 1.37$

Because $z > 0$, p-value is two times the upper tail area.

Using normal table with $z = 1.37$: p-value = 2(1 – .9147) = .1706.

c. p-value > .05; do not reject H_0. Cannot conclude that the population mean waiting time differs from eight minutes.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. $\bar{x} \pm z_{.025}(\sigma / \sqrt{n})$

$8.4 \pm 1.96(3.2/\sqrt{120})$

$8.4 \pm .57$ (7.83 to 8.97)

Yes; $\mu = 8$ is in the interval. Do not reject H_0.

24. a. $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{17-18}{4.5/\sqrt{48}} = -1.54$

b. Degrees of freedom = $n - 1 = 47$.

Because $t < 0$, p-value is two times the lower tail area.

Using t table: area in lower tail is between .05 and .10; therefore, p-value is between .10 and .20.

Exact p-value corresponding to $t = -1.54$ is .1303.

c. p-value > .05, do not reject H_0.

d. With $df = 47$, $t_{.025} = 2.012$

Reject H_0 if $t \le -2.012$ or $t \ge 2.012$.

$t = -1.54$; do not reject H_0.

26. a. $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{103-100}{11.5/\sqrt{65}} = 2.10$

Degrees of freedom = $n - 1 = 64$.

Because $t > 0$, p-value is two times the upper tail area.

Using t table; area in upper tail is between .01 and .025; therefore, p-value is between .02 and .05.

Exact p-value corresponding to $t = 2.10$ is .0397.

p-value ≤ .05, reject H_0.

b. $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{96.5-100}{11/\sqrt{65}} = -2.57$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Because $t < 0$, p-value is two times the lower tail area.

Using t table: area in lower tail is between .005 and .01; therefore, p-value is between .01 and .02.

Exact p-value corresponding to $t = -2.57$ is .0125.

p-value $\leq .05$, reject H_0.

c. $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{102 - 100}{10.5/\sqrt{65}} = 1.54$

Because $t > 0$, p-value is two times the upper tail area.

Using t table: area in upper tail is between .05 and .10; therefore, p-value is between .10 and .20.

Exact p-value corresponding to $t = 1.54$ is .1285.

p-value $> .05$, do not reject H_0.

28. a. H_0: $\mu \geq 9$

H_a: $\mu < 9$

b. $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{7.27 - 9}{6.38/\sqrt{85}} = -2.50$

Degrees of freedom $= n - 1 = 84$

Lower tail p-value is $P(t \leq -2.50)$.

Using t table: p-value is between .005 and .01.

Exact p-value corresponding to $t = -2.50$ is .0072.

c. p-value $\leq .01$; reject H_0. The mean tenure of a CEO is significantly lower than nine years. The claim of the shareholders group is not valid.

30. a. H_0: $\mu = 6.4$

H_a: $\mu \neq 6.4$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. Using Excel or JMP, we find $\bar{x} = 7.0$ and s = 2.4276.

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{7.0 - 6.4}{2.4276 / \sqrt{40}} = 1.56$$

$$df = n - 1 = 39$$

Because $t > 0$, p-value is two times the upper tail area at t = 1.56.

Using t table: area in upper tail is between .05 and .10; therefore, p-value is between .10 and .20.

Exact p-value corresponding to t = 1.56 is .1268.

c. Most researchers would choose $\alpha = .10$ or less. If you chose α = .10 or less, you cannot reject H_0. You are unable to conclude that the population mean number of hours married men with children in your area spend in child care differs from the mean reported by *Time*.

32. a. H_0: μ = 10,192

H_a: $\mu \neq$ 10,192

b. $t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{9750 - 10{,}192}{1400 / \sqrt{50}} = -2.23$

Degrees of freedom = n – 1 = 49.

Because $t < 0$, p-value is two times the lower tail area.

Using t table: area in lower tail is between .01 and .025; therefore, p-value is between .02 and .05.

Exact p-value corresponding to t = –2.23 is .0304.

c. p-value ≤ .05; reject H_0. The population mean price at this dealership differs from the national mean price $10,192.

34. a. H_0: μ = 2

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

H_a: $\mu \neq 2$

b. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{22}{10} = 2.2$

c. $s = \sqrt{\frac{\Sigma\left(x_i - \bar{x}\right)^2}{n-1}} = .516$

d. $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2.2 - 2}{.516/\sqrt{10}} = 1.22$

Degrees of freedom = n – 1 = 9.

Because $t > 0$, p-value is two times the upper tail area.

Using t table: area in upper tail is between .10 and .20; therefore, p-value is between .20 and .40.

Exact p-value corresponding to t = 1.22 is .2535.

e. p-value > .05; do not reject H_0. No reason to change from the 2 hours for cost estimating purposes.

36. a. $z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.68 - .75}{\sqrt{\frac{.75(1-.75)}{300}}} = -2.80$

Lower tail p-value is the area to the left of the test statistic.

Using normal table with z = –2.80: p-value =.0026.

p-value ≤ .05; Reject H_0.

b. $z = \frac{.72 - .75}{\sqrt{\frac{.75(1-.75)}{300}}} = -1.20$

Lower tail p-value is the area to the left of the test statistic.

Using normal table with z = –1.20: p-value =.1151.

p-value > .05; Do not reject H_0.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. $z = \dfrac{.70 - .75}{\sqrt{\dfrac{.75(1-.75)}{300}}} = -2.00$

Lower tail p-value is the area to the left of the test statistic.

Using normal table with $z = -2.00$: p-value =.0228.

p-value $\leq .05$

Reject H_0.

d. $z = \dfrac{.77 - .75}{\sqrt{\dfrac{.75(1-.75)}{300}}} = .80$.

Lower tail p-value is the area to the left of the test statistic.

Using normal table with $z = .80$: p-value =.7881.

p-value > .05

Do not reject H_0.

38. a. H_0: p μ .64

H_a: $p \neq .64$

b. $\bar{p} = \dfrac{52}{100} = .52$

$$z = \frac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{.52 - .64}{\sqrt{\dfrac{.64(1-.64)}{100}}} = -2.50$$

Because $z < 0$, p-value is two times the lower tail area.

Using normal table with $z = -2.50$: p-value = 2(.0062) = .0124.

c. p-value $\leq .05$; reject H_0. Proportion differs from the reported .64.

d. The results of the hypothesis test provide evidence that the proportion differs from .64, and the sample proportion $\bar{p} = .52$ suggests that fewer than 64% of the shoppers

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

believe the supermarket brand is as good as the name brand. So the manufacturer has some evidence that the percentage of supermarket shoppers who believe the supermarket ketchup is as good as the national brand is less than 64%. However, the manufacturer would have more appropriate evidence if a lower tailed hypothesis test had been used.

40. a. Sample proportion: $\bar{p} = .35$

Number planning to provide holiday gifts: $n\bar{p} = 60(.35) = 21$

b. H_0: $p \geq .46$

H_a: $p < .46$

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.35 - .46}{\sqrt{\frac{.46(1-.46)}{60}}} = -1.71$$

p-value is area in lower tail.

Using normal table with $z = -1.71$: *p*-value = .0436.

c. Using a .05 level of significance, we can conclude that the proportion of business owners providing gifts has decreased from last year to this year. The smallest level of significance for which we could draw this conclusion is .0436; this corresponds to the *p*-value = .0436. This is why the *p*-value is often called the *observed level of significance*.

42. a. $\bar{p}$ = 12/80 = .15

b. $\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{.15(.85)}{80}} = .0399$

$$\bar{p} \pm z_{.025}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

.15 $\pm$ 1.96 (.0399)

.15 $\pm$ .0782 or .0718 to .2282

c. H_0: $p = .06$

H_a: $p \neq .06$

$$\bar{p} = .15$$

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.15 - .06}{\sqrt{\frac{.06(.94)}{80}}} = 3.38$$

$$p\text{-value} \approx 0$$

We conclude that the return rate for the Houston store is different than the U.S. national return rate.

44. a. H_0: $p \leq .50$

H_a: $p > .50$

b. Using Excel we find that 92 of the 150 physicians in the sample have been sued. So,

$$\bar{p} = \frac{92}{150} = .6133$$

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.6133 - .50}{\sqrt{\frac{(.50)(.50)}{150}}} = 2.78$$

p-value is the area in the upper tail at $z = 2.78$.

Using normal table with $z = 2.78$: p-value = 1 – .9973 = .0027.

c. Because p-value = .0027 $\leq$ .01, we reject H_0 and conclude that the proportion of physicians over age 55 who have been sued at least once is greater than .50.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

46. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{120}} = .46$

$$c = 10 - 1.645\ (5 / \sqrt{120}\ ) = 9.25$$

Reject H_0 if $\bar{x} \leq 9.25$.

a. When $\mu = 9$,

$$z = \frac{9.25 - 9}{5/\sqrt{120}} = .55$$

P(Reject H_0) = (1.0000 – .7088) = .2912.

b. Type II error

c. When $\mu = 8$,

$$z = \frac{9.25 - 8}{5/\sqrt{120}} = 2.74$$

β = (1.0000 – .9969) = .0031

48. a. H_0: $\mu \leq 15$

H_a: $\mu > 15$

Concluding $\mu \leq 15$ when this is not true. Fowle would not charge the premium rate even

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

though the rate should be charged.

b. Reject H_0 if $z \geq 2.33$.

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{\bar{x} - 15}{4 / \sqrt{35}} = 2.33$$

Solve for $\bar{x}$ = 16.58.

Decision rule:

Accept H_0 if $\bar{x} < 16.58$.

Reject H_0 if $\bar{x} \geq 16.58$.

For $\mu = 17$,

$$z = \frac{16.58 - 17}{4 / \sqrt{35}} = -.62$$

$$\beta = .2676$$

c. For $\mu = 18$,

$$z = \frac{16.58 - 18}{4 / \sqrt{35}} = -2.10$$

$$\beta = .0179$$

50. a. Accepting H_0 and concluding the mean average age was 28 years when it was not.

b. Reject H_0 if $z \leq -1.96$ or if $z \geq 1.96$.

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{\bar{x} - 28}{6 / \sqrt{100}}$$

Solving for $\bar{x}$, we find

at $z = -1.96$, $\bar{x} = 26.82$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

at $z = +1.96$, $\bar{x} = 29.18$

Decision rule:

Accept H_0 if $26.82 < \bar{x} < 29.18$

Reject H_0 if $\bar{x} \leq 26.82$ or if $\bar{x} \geq 29.18$

At $\mu = 26$,

$$z = \frac{26.82 - 26}{6/\sqrt{100}} = 1.37$$

$\beta = 1.0000 - .9147 = .0853$

At $\mu = 27$,

$$z = \frac{26.82 - 27}{6/\sqrt{100}} = -.30$$

$\beta = 1.0000 - .3821 = .6179$

At $\mu = 29$,

$$z = \frac{29.18 - 29}{6/\sqrt{100}} = .30$$

$\beta = .6179$

At $\mu = 30$,

$$z = \frac{29.18 - 30}{6/\sqrt{100}} = -1.37$$

$\beta = .0853$

c. Power $= 1 - \beta$.

At $\mu = 26$,

Power $= 1 - .0853 = .9147$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

When $\mu = 26$, there is a .9147 probability that the test will correctly reject the null hypothesis that $\mu = 28$.

52. $c = \mu_0 + z_{.01}\frac{\sigma}{\sqrt{n}} = 15 + 2.33\frac{4}{\sqrt{50}} = 16.32$

At $\mu = 17$ $z = \frac{16.32 - 17}{4/\sqrt{50}} = -1.20$

$\beta = .1151$

At $\mu = 18$ $z = \frac{16.32 - 18}{4/\sqrt{50}} = -2.97$

$\beta = .0015$

Increasing the sample size reduces the probability of making a type II error.

54. $n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(1.645 + 1.28)^2(5)^2}{(10-9)^2} = 214$

56. At $\mu_0 = 3$, $\alpha = .01$. $z_{.01} = 2.33$

At $\mu_a = 2.9375$, $\beta = .10$. $z_{.10} = 1.28$

$\sigma = .18$

$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(2.33 + 1.28)^2(.18)^2}{(3 - 2.9375)^2} = 108.09$ Use 109.

58. At $\mu_0 = 28$, $\alpha = .05$. Note, however, for this two-tailed test, $z_{\square/2} = z_{.025} = 1.96$.

At $\mu_a = 29$, $\beta = .15$. $z_{.15} = 1.04$

$\sigma = 6$

$n = \frac{(z_{\alpha/2} + z_\beta)^2\sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(1.96 + 1.04)^2(6)^2}{(28-29)^2} = 324$

60. H_0: $\mu = 101.5$

H_a: $\mu \neq 101.5$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{100.47 - 101.5}{25/\sqrt{10163}} = -4.15$$

Because $z < 0$, p-value is two times the lower tail area.

Using normal table with $z = -4.15$: p-value $\approx 2(.0000) = .0000$.

p-value $\leq .01$, reject H_0. Conclude that the actual mean number of business e-mails sent and received per business day by employees of this department of the federal government differs from corporate employees.

Although the difference between the sample mean number of business emails sent and received per business day by employees of this department of the Federal Government and the mean number of business emails sent and received by corporate employees is statistically significant, this difference of 1.03 emails is relatively small and so may be of little or no practical significance.

62. H_0: $p \leq .62$

H_a: $p > .62$

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{.626 - .62}{\sqrt{\frac{.62(1 - .62)}{49581}}} = 2.75$$

p-value is the area in the upper tail.

Using normal table with $z = 2.75$: p-value $= 1 - .9970 = .0030$.

p-value $\leq .05$, reject H_0. Conclude that the actual proportion of fast-food orders this year that includes French fries exceeds the proportion of fast food orders that included french fries last year.

Although the difference between the sample proportion of fast-food orders this year that includes french fries and the proportion of fast-food orders that included

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

french fries last year is statistically significant, APGA should be concerned about whether this .006 or .6% increase is large enough to be effective in an advertising campaign.

64. a. H_0: $\mu = 16$

H_a: $\mu \neq 16$

b. $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{16.32 - 16}{.8 / \sqrt{30}} = 2.19$

Because $z > 0$, p-value is two times the upper tail area.

Using normal table with $z = 2.19$: p-value = 2(.0143) = .0286.

p-value ≤ .05; reject H_0. Readjust production line.

c. $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{15.82 - 16}{.8 / \sqrt{30}} = -1.23$

Because $z < 0$, p-value is two times the lower tail area.

Using normal table with $z = -1.23$: p-value = 2(.1093) = .2186.

p-value > .05; do not reject H_0. Continue the production line.

d. Reject H_0 if $z \leq -1.96$ or $z \geq 1.96$.

For $\bar{x} = 16.32$, $z = 2.19$; reject H_0.

For $\bar{x} = 15.82$, $z = -1.23$; do not reject H_0.

Yes, same conclusion.

66. a. H_0: $\mu \leq 4$

H_a: $\mu > 4$

b. $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{4.5 - 4}{1.5 / \sqrt{60}} = 2.58$

p-value is the upper tail area at z =2.58.

Using normal table with $z = 2.58$: p-value = 1.0000 – .9951 = .0049.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. p-value $\leq$ .01, reject H_0. Conclude that the mean daily background television that children from low-income families are exposed to is greater than four hours.

68. H_0: $\mu \leq 30.8$

H_a: $\mu > 30.8$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{32.7234 - 30.8}{12.1041/\sqrt{47}} = 1.0894$$

Degrees of freedom = $n - 1 = 46$.

Using t table, area in the upper tail is between .20 and .10 and p-value is between .10 and .20.

Exact p-value is .1408.

Because p-value > $\alpha = .05$, do not reject H_0. There is no evidence to conclude that the mean age of recently wed British men exceeds the mean age for those wed in 2013.

70. H_0: $\mu \leq 125{,}000$

H_a: $\mu > 125{,}000$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{130{,}000 - 125{,}000}{12{,}500/\sqrt{32}} = 2.26$$

Degrees of freedom = 32 – 1 = 31.

Upper tail p-value is the area to the right of the test statistic.

Using t table: p-value is between .01 and .025.

Exact p-value corresponding to $t = 2.26$ is .0155.

p-value $\leq$ .05; reject H_0. Conclude that the mean cost is greater than \$125,000 per lot.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

72. a. H_0: $p \le .52$

H_a: $p > .52$

$$\bar{p} = \frac{285}{510} = .5588$$

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.5588 - .52}{\sqrt{\frac{.52(1-.52)}{510}}} = 1.75$$

p-value is the area in the upper tail.

Using normal table with $z = 1.75$: p-value = 1.0000 – .9599 = .0401.

p-value ≤ .05; reject H_0. We conclude that people who fly frequently are more likely to be able to sleep during flights.

b. H_0: $p \le .52$

H_a: $p > .52$

$$\bar{p} = \frac{285}{510} = .5588$$

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.5588 - .52}{\sqrt{\frac{.52(1-.52)}{510}}} = 1.75$$

p-value is the area in the upper tail.

Using normal table with $z = 1.75$: p-value = 1.0000 – .9599 = .0401.

p-value > .01; we cannot reject H_0. Thus, we cannot conclude that that people who fly frequently better more likely to be able to sleep during flights.

74. a. H_0: $p \le .30$

H_a: $p > .30$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $\bar{p} = \frac{136}{400} = .34$ (34%)

c. $z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.34 - .30}{\sqrt{\frac{.30(1-.30)}{400}}} = 1.75$

p-value is the upper-tail area.

Using normal table with $z = 1.75$: *p*-value = 1.0000 – .9599 = .0401.

d. *p*-value ≤ .05; reject H_0. Conclude that more than 30% of the millennials either live at home with their parents or are otherwise dependent on their parents.

76. H_0: $p \geq .90$

H_a: $p < .90$

$$\bar{p} = \frac{49}{58} = .8448$$

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.8448 - .90}{\sqrt{\frac{.90(1-.90)}{58}}} = -1.40$$

Lower tail *p*-value is the area to the left of the test statistic.

Using normal table with $z = -1.40$: *p*-value =.0808.

p-value > .05; do not reject H_0. Claim of at least 90% cannot be rejected.

78. a. H_0: $\mu \leq 72$

H_a: $\mu > 72$

Reject H_0 if $z \geq 1.645$

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{\bar{x} - 72}{20 / \sqrt{30}} = 1.645$$

Solve for $\bar{x} = 78$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Decision rule:

Reject H_0 if $\bar{x} \geq 78$.

b. For $\mu = 80$

$$z = \frac{78 - 80}{20/\sqrt{30}} = -.55$$

$\beta = .2912$

c. For $\mu = 75$,

$$z = \frac{78 - 75}{20/\sqrt{30}} = .82$$

$\beta = .7939$

d. For $\mu = 70$, H_0 is true. In this case the type II error cannot be made.

e. Power = $1 - \beta$

80. a. H_0: $\mu = 120$

H_a: $\mu \neq 120$

At $\mu_0 = 120$ $\alpha = .05$.

With a two-tailed test, $z_{\alpha/2} = z_{.025} = 1.96$.

At $\mu_a = 117$, $\beta = .02$. $z_{.02} = 2.05$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$n = \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(1.96 + 2.05)^2 (5)^2}{(120 - 117)^2} = 44.7$$ Use 45.

b. Example calculation for $\mu = 118$.

Reject H_0 if $z \le -1.96$ or if $z \ge 1.96$.

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{\bar{x} - 120}{5 / \sqrt{45}}$$

Solve for $\bar{x}$.

At $z = -1.96$ $\bar{x} = 118.54$

At $z = +1.96$ $\bar{x} = 121.46$

Decision rule:

Reject H_0 if $\bar{x} \le 118.54$ or if $\bar{x} \ge 121.46$.

For $\mu = 118$,

$$z = \frac{118.54 - 118}{5 / \sqrt{45}} = .72$$

$\beta = .2358$

Other Results

If μ is	z	β
117	2.07	.0192
118	.72	.2358
119	–.62	.7291
121	+.62	.7291

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

122	+.72	.2358
123	–2.07	.0192

82. H_0: $p \geq .001$

H_a: $p < .001$

$$\bar{p} = \frac{98}{11,667} = .000878$$

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.00878 - .001}{\sqrt{\frac{(.001)(1-.001)}{111667}}} = -1.29$$

Lower tail p-value is the area to the left of the test statistic.

Using normal table with $z = -1.29$: p-value = .0985.

Because p-value = .0985 $> \alpha = .05$, do not reject H_0. We cannot conclude that the manufacturer is meeting its goal for overfried chips at $\alpha = .05$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 10 Solutions

2. a. $$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(25.2 - 22.8) - 0}{\sqrt{\frac{(5.2)^2}{40} + \frac{6^2}{50}}} = 2.03$$

b. p-value = 1.0000 – .9788 = .0212

c. p-value $\leq$ .05, reject H_0.

4. a. μ_1 = population mean for smaller cruise ships

μ_2 = population mean for larger cruise ships

$\bar{x}_1 - \bar{x}_2$ = 85.36 – 81.40 = 3.96

b. $$z_{.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$1.96\sqrt{\frac{(4.55)^2}{37} + \frac{(3.97)^2}{44}} = 1.88$$

c. 3.96 ± 1.88 (2.08 to 5.84)

6. μ_1 = mean hotel price in Atlanta

μ_2 = mean hotel price in Houston

H_0: $\mu_1 - \mu_2 \geq 0$

H_a: $\mu_1 - \mu_2 < 0$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(91.71 - 101.13) - 0}{\sqrt{\frac{20^2}{35} + \frac{25^2}{40}}} = -1.81$$

p-value = .0351

p-value $\leq$ .05; reject H_0. The mean price of a hotel room in Atlanta is lower than the mean price of a hotel room in Houston.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

8. a. This is an upper-tail hypothesis test.

$$H_0: \mu_1 - \mu_2 \le 0$$

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

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(76-73)}{\sqrt{\frac{6^2}{60} + \frac{6^2}{60}}} = 2.74$$

p-value = area in upper tail at $z = 2.74$

p-value = 1.0000 - .9969 = .0031

Because .0031 $\le$ α = .05, we reject the null hypothesis. The difference is significant. We can conclude that customer service has improved for Rite Aid.

b. This is another upper-tail test, but it only involves one population.

$H_0: \mu \le 75.7$

$H_a: \mu > 75.7$

$$z = \frac{(\bar{x}_1 - 75.7)}{\sqrt{\frac{\sigma^2}{n_1}}} = \frac{(76-75.7)}{\sqrt{\frac{6^2}{60}}} = .39$$

p-value = area in upper tail at $z = .39$

p-value = 1.0000 – .6517 = .3483

Because .3483 > α = .05, we cannot reject the null hypothesis. The difference is not statistically significant.

c. This is an upper-tail test similar to the one in part a.

$$H_0: \mu_1 - \mu_2 \le 0$$

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

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(77-75)}{\sqrt{\frac{6^2}{60} + \frac{6^2}{60}}} = 1.83$$

p-value = area in upper tail at $z = 1.83$

p-value = 1.0000 – .9664 = .0336

Because .0336 ≤ α = .05, we reject the null hypothesis. The difference is significant. We can conclude that customer service has improved for Expedia.

d. We will reject the null hypothesis of "no increase" if the *p*-value ≤ .05. For an upper-tail hypothesis test, the *p*-value is the area in the upper tail at the value of the test statistic. A value of $z = 1.645$ provides an upper-tail area of .05. So, we must solve the following equation for $\bar{x}_1 - \bar{x}_2$.

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{6^2}{60} + \frac{6^2}{60}}} = 1.645$$

$$\bar{x}_1 - \bar{x}_2 = 1.645\sqrt{\frac{6^2}{60} + \frac{6^2}{60}} = 1.80$$

This tells us that as long as the year 2 score for a company exceeds the year 1 score by 1.80 or more the difference will be statistically significant.

e. The increase from year 1 to year 2 for J.C. Penney is not statistically significant because it is less than 1.80. We cannot conclude that customer service has improved for J.C. Penney.

10. a. $$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(13.6 - 10.1) - 0}{\sqrt{\frac{5.2^2}{35} + \frac{8.5^2}{40}}} = 2.18$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{5.2^2}{35} + \frac{8.5^2}{40}\right)^2}{\frac{1}{34}\left(\frac{5.2^2}{35}\right)^2 + \frac{1}{39}\left(\frac{8.5^2}{40}\right)^2} = 65.7$$

Use $df = 65$.

c. Using t table, area in tail is between .01 and .025 $\therefore$ two-tail p-value is between .02 and .05.

Exact p-value corresponding to $t = 2.18$ is .0329

d. p-value $\leq$ .05, reject H_0.

12. a. $\bar{x}_1 - \bar{x}_2 = 22.5 - 18.6 = 3.9$

b. $$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{8.4^2}{50} + \frac{7.4^2}{40}\right)^2}{\frac{1}{49}\left(\frac{8.4^2}{50}\right)^2 + \frac{1}{39}\left(\frac{7.4^2}{40}\right)^2} = 87.1$$

Use $df = 87$, $t_{.025} = 1.988$

$$3.9 \pm 1.988\sqrt{\frac{8.4^2}{50} + \frac{7.4^2}{40}}$$

3.9 ± 3.3 (.6 to 7.2)

14. a. H_0: $\mu_1 - \mu_2 \geq 0$

H_a: $\mu_1 - \mu_2 < 0$

b.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(48{,}537 - 55{,}317) - 0}{\sqrt{\frac{18{,}000^2}{110} + \frac{10{,}000^2}{30}}} = -2.71$$

c.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{18{,}000^2}{110} + \frac{10{,}000^2}{30}\right)^2}{\frac{1}{110 - 1}\left(\frac{18{,}000^2}{110}\right)^2 + \frac{1}{30 - 1}\left(\frac{10{,}000^2}{30}\right)^2}$$

$$= 85.20$$

Rounding down, we will use a t distribution with 85 degrees of freedom. From the t table we see that $t = -2.71$ corresponds to a p-value between .005 and 0.

Exact p-value corresponding to $t = -2.71$ is .004.

d. p-value $\leq$ .05, reject H_0. We conclude that the salaries of finance majors are lower than the salaries of business analytics majors.

16. a. μ_1 = population mean verbal score parents college grads

μ_2 = population mean verbal score parents high school grads

H_0: $\mu_1 - \mu_2 \leq 0$

H_a: $\mu_1 - \mu_2 > 0$

b. $\bar{x}_1 = \frac{\Sigma x_i}{n} = \frac{8400}{16} = 525$

$\bar{x}_2 = \frac{\Sigma x_i}{n} = \frac{5844}{12} = 487$

$\bar{x}_1 - \bar{x}_2 = 525 - 487 = 38$ points higher if parents are college grads

c. $s_1 = \sqrt{\frac{\Sigma(x_i - \bar{x}_1)^2}{n_1 - 1}} = \sqrt{\frac{52962}{16 - 1}} = \sqrt{3530.8} = 59.42$

$s_2 = \sqrt{\frac{\Sigma(x_i - \bar{x}_2)^2}{n_2 - 1}} = \sqrt{\frac{29456}{12 - 1}} = \sqrt{2677.82} = 51.75$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(525 - 487) - 0}{\sqrt{\frac{59.42^2}{16} + \frac{51.75^2}{12}}} = 1.80$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{59.42^2}{16} + \frac{51.75^2}{12}\right)^2}{\frac{1}{15}\left(\frac{59.42^2}{16}\right)^2 + \frac{1}{11}\left(\frac{51.75^2}{12}\right)^2} = 25.3$$

Use $df = 25$.

Using t table, p-value is between .025 and .05.

Exact p-value corresponding to $t = 1.80$ is .0420/

d. p-value $\leq$ .05, reject H_0. Conclude higher population mean verbal scores for students whose parents are college grads.

18. a. Let μ_1 = population mean minutes late for delayed Delta flights

μ_2 = population mean minutes late for delayed Southwest flights

H_0: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_{22} \neq 0$

b. $\bar{x}_1 = \frac{\sum_{i=1}^{n} x_i}{n_1} = \frac{1265}{25} = 50.6$ minutes

$\bar{x}_2 = \frac{\sum_{i=1}^{n} x_i}{n_2} = \frac{1056}{20} = 52.8$ minutes

The difference between sample mean delay times is 50.6 – 52.8 = –2.2 minutes, which indicates the sample mean delay time is 2.2 minutes less for Delta.

c. Sample standard deviations: $s_1 = 26.57$ and $s_2 = 20.11$

$$t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(50.6 - 52.8) - 0}{\sqrt{\frac{26.57^2}{25} + \frac{20.11^2}{20}}} = -.32$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{26.57^2}{25} + \frac{20.11^2}{20}\right)^2}{\frac{1}{24}\left(\frac{26.57^2}{25}\right)^2 + \frac{1}{19}\left(\frac{20.11^2}{20}\right)^2} = 42.9$$

Use $df = 42$.

p-value for this two-tailed test is two times the lower-tail area for $t = -.32$.

Using t table, p-value is greater than 2(.20) = .40.

Exact p-value corresponding to $t = -.32$ with 42 df is .7506.

p-value > .05, do not reject H_0. We cannot reject the assumption that the population mean delay times are the same at Delta and Southwest Airlines. There is no statistical evidence that one airline does better than the other in terms of its population mean delay time.

20. a. 3, –1, 3, 5, 3, 0, 1

b. $\bar{d} = \sum d_i / n = 14/7 = 2$

c. $s_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{26}{7-1}} = 2.08$

d. $\bar{d} = 2$

e. With 6 degrees of freedom $t_{.025} = 2.447$

$2 \pm 2.447\left(2.082/\sqrt{7}\right)$

2 ± 1.93 (.07 to 3.93)

22. a. Let

$$d_i = \frac{end\ of\ first\ quarter\ price - beginning\ of\ first\ quarter\ price}{beginning\ of\ first\ quarter\ price}$$

$$\bar{d} = \frac{\sum_i d_i}{n} = \frac{-1.74}{25} = -0.07$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b.

$$s_d = \sqrt{\frac{\sum_i (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{0.2666}{25-1}} = 0.105$$

With $df = 24$, $t_{.025} = 2.064$

$$\bar{d} \pm t_{.025} \frac{s_d}{\sqrt{n}} = -0.07 \pm 2.064 \left(\frac{0.105}{\sqrt{25}}\right) = -0.07 \pm 0.04$$

Confidence interval: (–0.11 to –0.03)

The 95% confidence interval shows that the population mean percentage change in the price per share of stock is a decrease of 3% to 11%. This may be a harbinger of a further stock market swoon.

24. a. Difference = Current year airfare – Previous year airfare

$$H_0: \mu_d \leq 0$$

$$H_a: \mu_d > 0$$

Differences 30, 63, –42, 10, 10, –27, 50, 60, 60, –30, 62, 30

$$\bar{d} = \frac{\sum d_i}{n} = \frac{276}{12} = 23$$

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{16{,}558}{12-1}} = 38.80$$

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{23 - 0}{38.80 / \sqrt{12}} = 2.05$$

$$df = n - 1 = 11$$

Using t table, p-value is between .05 and .025.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Exact p-value corresponding to $t = 2.05$ is .0325.

Because p-value < .05, reject H_0. We can conclude that there has been a significance increase in business travel airfares over the one-year period.

b. Current year: $\bar{x} = \sum x_i / n = 5844/12 = \$487$

Previous year: $\bar{x} = \sum x_i / n = 5568/12 = \$464$

c. One-year increase = \$487 – \$464 = \$23

\$23/\$464 = .05, or a 5% increase in business travel airfares for the one-year period.

26. a. H_0: $\mu_d = 0$

H_a: $\mu_d \neq 0$

Differences: –2, –1, –5, 1, 1, 0, 4, –7, –6, 1, 0, 2, –3, –7, –2, 3, 1, 2, 1, –4

$$\bar{d} = \sum d_i / n = -21/20 = -1.05$$

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = 3.3162$$

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{-1.05 - 0}{3.3162/\sqrt{20}} = -1.42$$

$$df = n - 1 = 19$$

Using t table, area in tail is between .05 and .10.

Two-tail p-value must be between .10 and .20.

Exact p-value corresponding to $t = -1.42$ is .1718.

Cannot reject H_0. There is no significant difference between the mean scores for the first and fourth rounds.

b. $\bar{d} = -1.05$; first round scores were lower than fourth round scores.

c. $\alpha = .05$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$df = 19 \qquad t = 1.729$

Margin of error $= t_{.025}\dfrac{s_d}{\sqrt{n}} = 1.729\dfrac{3.3162}{\sqrt{20}} = 1.28$

Yes, just check to see if the 90% confidence interval includes a difference of zero. If it does, the difference is not statistically significant.

90% Confidence interval: –1.05 ± 1.28 (–2.33, .23)

The interval does include 0, so the difference is not statistically significant.

28. a. $\bar{p}_1 - \bar{p}_2 = .48 - .36 = .12$

b. $$\bar{p}_1 - \bar{p}_2 \pm z_{.05}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$$

$$.12 \pm 1.645\sqrt{\frac{.48(1-.48)}{400} + \frac{.36(1-.36)}{300}}$$

$.12 \pm .0614 \qquad (.0586 \text{ to } .1814)$

c. $$.12 \pm 1.96\sqrt{\frac{.48(1-.48)}{400} + \frac{.36(1-.36)}{300}}$$

$.12 \pm .0731 \qquad (.0469 \text{ to } .1931)$

30. $\bar{p}_1 = 220/400 = .55 \quad \bar{p}_2 = 192/400 = .48$

$$\bar{p}_1 - \bar{p}_2 \pm z_{.025}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$$

$$.55 - .48 \pm 1.96\sqrt{\frac{.55(1-.55)}{400} + \frac{.48(1-.48)}{400}}$$

$.07 \pm .0691 \qquad (.0009 \text{ to } .1391)$

Seven percent more executives are predicting an increase in full-time jobs. The confidence interval shows the difference may be from 0.09% to 13.91%. Because 0%

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

does not lie strictly within this confidence interval (i.e., the lower limit of the confidence interval is not negative), we can state with 95% confidence that the percentage of executives optimistic about hiring is larger in the current year than in the previous year.

32. Let p_1 = the population proportion of tuna that is mislabeled

p_2 = the population proportion of mahi mahi that is mislabeled

a. The point estimate of the proportion of tuna that is mislabeled is $\bar{p}_1$ = 99/220 = .45.

b. The point estimate of the proportion of mahi mahi that is mislabeled is $\bar{p}_2$ = 56/160 = .35.

c. $\bar{p}_1 - \bar{p}_2$ = .45 – .35 = .10

$$.10 \pm z_{.025}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$$

$$.10 \pm 1.96\sqrt{\frac{.45(.55)}{220} + \frac{.35(.65)}{160}}$$

.10 ± .0989 (.0011 to .1989)

The 95% confidence interval estimate of the difference between the proportion of tuna and mahi mahi that is mislabeled is .10 ± .0989 or (.0011 to .1989).

34. Let p_1 = the population proportion of wells drilled in 2012 that were dry

p_2 = the population proportion of wells drilled in 2018 that were dry

a. H_0: $p_1 - p_2 \le 0$

H_a: $p_1 - p_2 > 0$

b. $\bar{p}_1$ = 24/119 = .2017

c. $\bar{p}_2$ = 21/162 = .1111

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. $\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{24+18}{119+162} = .1495$

$$z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{.2017 - .1111}{\sqrt{.1495(1-.1495)\left(\frac{1}{119}+\frac{1}{162}\right)}} = 2.10$$

Upper-tail p-value is the area to the right of the test statistic.

Using normal table with $z = 2.10$: p-value = 1 – .9821 = .0179.

p-value < .05, so reject H_0 and conclude that wells drilled in 2012 were dry more frequently than wells drilled in 2018—that is, the frequency of dry wells has decreased over the eight years from 2012 to 2018.

36. a. Let p_1= population proportion of men expecting to get a raise or promotion this year

p_2= population proportion of women expecting to get a raise or promotion this year

H_0: $p_1 - p_2 \leq 0$

H_a: $p_1 - p_2 > 0$

b. $\bar{p}_1$ = 104/200 = .52 (52%)

$\bar{p}_2$ = 74/200 = .37 (37%)

c. $\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{104+74}{200+200} = .445$

$$z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{.52 - .37}{\sqrt{.445(1-.445)\left(\frac{1}{200}+\frac{1}{200}\right)}} = 3.02$$

p-value = 1.0000 – .9987 = .0013

Reject H_0. There is a significant difference between the population proportions with a great proportion of men expecting to get a raise or a promotion this year.

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

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

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

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(4.1 - 3.4) - 0}{\sqrt{\frac{(2.2)^2}{120} + \frac{(1.5)^2}{100}}} = 2.79$$

p-value = 2(1.0000 – .9974) = .0052

p-value $\leq$ .05, reject H_0. A difference exists with system B having the lower mean checkout time.

40. a. $H_0 : \mu_1 - \mu_2 \leq 0$

$H_a : \mu_1 - \mu_2 > 0$

b. $n_1 = 30$ $n_2 = 30$

$\bar{x}_1 = 16.23$ $\bar{x}_2 = 15.70$

$s_1 = 3.52$ $s_2 = 3.31$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(16.23 - 15.70) - 0}{\sqrt{\frac{(3.52)^2}{30} + \frac{(3.31)^2}{30}}} = .60$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{3.52^2}{30} + \frac{3.31^2}{30}\right)}{\frac{1}{29}\left(\frac{3.52^2}{30}\right)^2 + \frac{1}{29}\left(\frac{3.31^2}{30}\right)^2} = 57.8$$

Use $df = 57$.

Using t table, p-value is greater than .20.

Exact p-value corresponding to t = .60 is .2754.

p-value > .05, do not reject H_0. Cannot conclude that the mutual funds with a load have a greater mean rate of return.

42. a. $\bar{d} = \frac{\sum d_i}{n} = \frac{280}{20} = 14$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $s_d = \sqrt{\frac{\Sigma(d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{54{,}880}{19}} = 53.744$

$df = n - 1 = 19$, $t_{.05} = 1.729$

$\bar{d} \pm t_{.05} \frac{s_d}{\sqrt{n}} = 14 \pm 1.729 \frac{53.744}{\sqrt{20}}$

14 ± 20.78 (–6.78 to 34.78)

c. H_0: $\mu_d = 0$

H_a: $\mu_d \neq 0$

$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{14 - 0}{53.744 / \sqrt{20}} = 1.165$

Using t table with $df = 19$, the area in upper tail is between .20 and .10. Thus, for the two-tailed test, the p-value is between .20 and .40.

Using software, the exact p-value for $t = 1.165$ is .258.

Cannot reject H_0; cannot concluded that there is a difference between the mean scores for the no sibling and with sibling groups.

44. a. $\bar{p}_1 = 76/400 = .19$

$\bar{p}_2 = 90/900 = .10$

$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{76 + 90}{400 + 900} = .1277$

$z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{.19 - .10}{\sqrt{.1277(1-.1277)\left(\frac{1}{400} + \frac{1}{900}\right)}} = 4.49$

p-value ≈ 0

Reject H_0; there is a difference between claim rates.

b. $\bar{p}_1 - \bar{p}_2 \pm z_{.025}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$.19 - .10 \pm 1.96\sqrt{\frac{.19(1-.19)}{400} + \frac{.10(1-.10)}{900}}$$

.09 $\pm$ .0432 (.0468 to .1332)

Claim rates are higher for single males.

46. Let p_1 = the population proportion of American adults under 30 years old

p_2 = the population proportion of Americans over 30 years old

a. From the file *Computer News*, there are 109 Yes responses for each age group. The total number of respondents of the under 30 years group is n_1 = 150, whereas the 30 and over group had is n_2 = 200 total respondents.

American adults under 30 years old: $\bar{p}_1$ = 109/150 = .727

Americans over 30 years old: $\bar{p}_2$ = 109/200 = .545

b. $\bar{p}_1 - \bar{p}_2 = .727 - .545 = .182$

$$s_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{.727(1-.727)}{150} + \frac{.545(1-.545)}{200}} = .0506$$

Confidence interval: .182 ± 1.96(.0506) or.182 ± .0992 (.0824 to .2809)

c. Because the confidence interval in part b does not include 0 and both values are negative, conclude that the proportion of American adults under 30 years old who use a computer to gain access to news is greater than the proportion of Americans over 30 years old that use a computer to gain access to news.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 11 Solutions

2. $s^2 = 25$

a. With 19 degrees of freedom $\chi^2_{.05} = 30.144$ and $\chi^2_{.95} = 10.117$:

$$\frac{19(25)}{30.144} \le \sigma^2 \le \frac{19(25)}{10.117}$$

$$15.76 \le \sigma^2 \le 46.95$$

b. With 19 degrees of freedom $\chi^2_{.025} = 32.852$ and $\chi^2_{.975} = 8.907$:

$$\frac{19(25)}{32.852} \le \sigma^2 \le \frac{19(25)}{8.907}$$

$$14.46 \le \sigma^2 \le 53.33$$

c. $3.8 \le \sigma \le 7.3$

4. a. $n = 24$

$s^2 = .81$

With 24 – 1 = 23 degrees of freedom, $\chi^2_{.05} = 35.172$ and $\chi^2_{.95} = 13.091$.

$$\frac{23(.81)}{35.172} \le \sigma^2 \le \frac{23(.81)}{13.091}$$

$$.53 \le \sigma^2 \le 1.42$$

b. $.73 \le \sigma \le 1.19$

6. a. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{25.6}{8} = 3.2$

The sample mean quarterly total return for General Electric is 3.2%. This is the estimate of the population mean percent total return per quarter for General Electric.

b. $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{1773.6}{8-1} = 253.37$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$s = \sqrt{253.37} = 15.92$$

c. With $df = (n - 1) = 7$ $\chi^2_{.025} = 16.013$ $\chi^2_{.975} = 1.690$

$$\frac{(n-1)s^2}{\chi_{.025}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi_{.975}}$$

$$\frac{(8-1)(253.37)}{16.013} \le \sigma^2 \le \frac{(8-1)(253.37)}{1.690}$$

$$110.76 \le \sigma^2 \le 1{,}049.47$$

d. $10.52 \le \sigma \le 32.40$

8. a. $\bar{x} = .78$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{5.2230}{12-1} = .4748$$

b. $s = \sqrt{.4748} = .6891$

c. 11 degrees of freedom

$\chi^2_{.025} = 21.920$ $\chi^2_{.975} = 3.816$

$$\frac{(12-1).4748}{21.920} \le \sigma^2 \le \frac{(12-1).4748}{3.816}$$

$$.2383 \le \sigma^2 \le 1.3687$$

$$.4882 \le \sigma \le 1.1699$$

10. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{1260}{15} = 84$

b. $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{1662}{15-1} = 118.71$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. $s = \sqrt{s^2} = \sqrt{118.71} = 10.90$

d. Hypothesis for $\sigma = 12$ is for $\sigma^2 = (12)^2 = 144$

H_0: $\sigma^2 = 144$

H_a: $\sigma^2 \neq 144$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(15-1)(118.71)}{144} = 11.54$$

Degrees of freedom = n – 1 = 14.

Using χ^2 table, p–value is greater than 2(1 – .90) = .20.

Exact p–value corresponding to $\chi^2 = 11.54$ is 2(1 – .6431) = .7138.

p–value > .05; do not reject H_0. The hypothesis that the population standard deviation is 12; cannot be rejected.

12. a. $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = .8106$

b. H_0: $\sigma^2 = .94$

H_a: $\sigma^2 \neq .94$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(12-1)(.8106)}{.94} = 9.49$$

Degrees of freedom = n – 1 = 11.

Using χ^2 table, area in tail is greater than .10.

Two–tail p–value is greater than .20.

Exact p–value corresponding to $\chi^2 = 9.49$ is .8465.

p–value > .05, cannot reject H_0.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

14. a. $F = \frac{s_1^2}{s_2^2} = \frac{5.8}{2.4} = 2.4$

Degrees of freedom 15 and 20.

Using F table, p–value is between .025 and .05.

Exact p–value corresponding to $F = 2.4$ is .0345.

p–value $\leq$ .05; reject H_0. Conclude $\sigma_1^2 > \sigma_2^2$.

b. $F_{.05} = 2.20$

Reject H_0 if $F \geq 2.20$

$2.4 \geq 2.20$; reject H_0. Conclude $\sigma_1^2 > \sigma_2^2$

16. For this type of hypothesis test, we place the larger variance in the numerator. So the Fidelity variance is given the subscript of 1.

$$H_0: \sigma_1^2 \leq \sigma_2^2$$

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

$$F = \frac{s_1^2}{s_2^2} = \frac{18.9^2}{15.0^2} = 1.59$$

Degrees of freedom in the numerator and denominator are both 59.

Using the F table, p–value is between .05 and .025.

Exact p–value corresponding to $F = 1.59$ is .0387.

p–value $\leq$ .05; reject H_0. We conclude that the Fidelity fund has a greater variance than the American Century fund.

18. We place the larger sample variance in the numerator. So, the Merrill Lynch variance is given the subscript of 1.

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

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

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

$$F = \frac{s_1^2}{s_2^2} = \frac{587^2}{489^2} = 1.44$$

Degrees of freedom 15 and 9.

Using *F* table, area in tail is greater than .10.

Two–tail *p*–value is greater than .20 .

Exact *p*–value corresponding to $F = 1.44$ is .5906.

p–value > .10; do not reject H_0. We cannot conclude there is a statistically significant difference between the variances for the two companies.

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

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

$$F = \frac{s_1^2}{s_2^2} = \frac{11.1}{2.1} = 5.29$$

Degrees of freedom 25 and 24.

Using *F* table, area in tail is less than .01.

Two–tail *p*–value is less than .02 .

Exact *p*–value $\approx$ 0.

p–value $\leq$ .05; reject H_0. The population variances are not equal for seniors and managers.

22. a. Population 1—Wet pavement.

$H_0: \sigma_1^2 \leq \sigma_2^2$

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

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$F = \frac{s_1^2}{s_2^2} = \frac{32^2}{16^2} = 4.00$$

Degrees of freedom 15 and 15.

Using F table, p–value is less than .01.

Exact p–value corresponding to $F = 4.00$ is .0054.

p–value $\leq$ .05; reject H_0. Conclude that there is greater variability in stopping distances on wet pavement.

b. Drive carefully on wet pavement because of the variability in stopping distances.

24. With 12 degrees of freedom:

$$\chi^2_{.025} = 23.337 \quad \chi^2_{.975} = 4.404$$

$$\frac{(12)(14.95)^2}{23.337} \leq \sigma^2 \leq \frac{(12)(14.95)^2}{4.404}$$

$$114.9 \leq \sigma^2 \leq 609$$

$$10.72 \leq \sigma \leq 24.68$$

26. a. H_0: $\sigma^2 \leq .0001$

H_a: $\sigma^2 \leq .0001$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(15-1)(.014)^2}{.0001} = 27.44$$

Degrees of freedom = $n - 1 = 14$.

Using χ^2 table, p– value is between .01 and .025.

Exact p–value corresponding to $\chi^2 = 27.44$ is .0169.

p–value $\leq$ .10; reject H_0. Variance exceeds maximum variance requirement.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $\chi^2_{.05} = 23.685$

$\chi^2_{.95} = 6.571$ (14 degrees of freedom)

$$\frac{(14)(.014)^2}{23.685} \le \sigma^2 \le \frac{(14)(.014)^2}{6.571}$$

$.00012 \le \sigma^2 .00042$

28. H_0: $\sigma^2 \le 1$

H_a: $\sigma^2 > 1$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(22-1)(1.5)}{1} = 31.50$$

Degrees of freedom = $n - 1 = 21$

Using χ^2 table, p–value is between .05 and .10

Exact p–value corresponding to $\chi^2 = 31.50$ is .0657

p–value $\le$ .10; reject H_0. Conclude that $\sigma^2 > 1$.

30. a. Try $n = 15$

$\chi^2_{.025} = 26.119$

$\chi^2_{.975} = 5.629$ (14 degrees of freedom)

$$\frac{(14)(64)}{26.119} \le \sigma^2 \le \frac{(14)(64)}{5.629}$$

$34.3 \le \sigma^2 \le 159.2$

$5.86 \le \sigma \le 12.62$

$\therefore$ A sample size of 15 was used.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $n = 25$; expect the width of the interval to be smaller.

$\chi^2_{.05} = 39.364$

$\chi^2_{.975} = 12.401$ (24 degrees of freedom)

$$\frac{(24)(8)^2}{39.364} \le \sigma^2 \le \frac{(24)(8)^2}{12.401}$$

$39.02 \le \sigma^2 \le 123.86$

$6.25 \le \sigma \le 11.13$

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

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

Use critical value approach because F tables do not have 351 and 72 degrees of freedom.

$F_{.025} = 1.47$

Reject H_0 if $F \ge 1.47$.

$$F = \frac{s_1^2}{s_2^2} = \frac{.940^2}{.797^2} = 1.39$$

$F < 1.47$; do not reject H_0. We are not able to conclude students who complete the course and students who drop out have different variances of grade point averages.

34. H_0: $\sigma_1^2 \le \sigma_2^2$

H_1: $\sigma_1^2 > \sigma_2^2$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$F = \frac{s_1^2}{s_2^2} = \frac{25}{12} = 2.08$$

Degrees of freedom 30 and 24

Using the F table, the area in the tail (which corresponds to the p-value) is between .025 and .05.

Exact p–value corresponding to $F = 2.08$ is .0348.

p–value $\leq$ .10; reject H_0. Conclude that the population variances has decreased because of the lean process improvement.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 12 Solutions

2. a. $\bar{p}_1 = 150/250 = .60$

$\bar{p}_2 = 150/300 = .50$

$\bar{p}_3 = 96/200 = .48$

b. Multiple comparisons

For 1 versus 2

$$CV_{12} = \sqrt{\chi^2_\alpha}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} = \sqrt{5.991}\sqrt{\frac{.60(1-.60)}{250} + \frac{.50(1-.50)}{300}} = .1037$$

$df = k - 1 = 3 - 1 = z\ \chi^2_{.05} = 5.991$

Comparison	p_i	p_j	Difference	n_i	n_j	Critical Value	Significant Diff > CV
1 vs. 2	.60	.50	.10	250	300	.1037	
1 vs. 3	.60	.48	.12	250	200	.1150	Yes
2 vs. 3	.50	.48	.02	300	200	.1117	

Only one comparison is significant, 1 versus 3. The others are not significant. We can conclude that the population proportions differ for populations 1 and 3.

4. a. H_0: $p_1 = p_2 = p_3$

H_a: Not all population proportions are equal

b. Observed Frequencies (f_{ij})

Component	A	B	C	Total
Defective	15	20	40	75

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Good	485	480	460	1425
Total	500	500	500	1500

Expected Frequencies (e_{ij})

Component	A	B	C	Total
Defective	25	25	25	75
Good	475	475	475	1425
Total	500	500	500	1500

Chi-Square Calculations $(f_{ij} - e_{ij})^2 / e_{ij}$

Component	A	B	C	Total
Defective	4.00	1.00	9.00	14.00
Good	.21	.05	.47	0.74

$\chi^2 = 14.74$

Degrees of freedom = $k - 1 = (3 - 1) = 2$

Using the χ^2 table with $df = 2$, $\chi^2 = 14.74$ shows the p-value is less than .005.

Using software, the p-value corresponding to $\chi^2 = 14.74$ is .0006.

Because the p-value < .05; reject H_0. Conclude that the three suppliers do not provide equal proportions of defective components.

c. $\bar{p}_1 = 15/500 = .03$

$\bar{p}_2 = 20/500 = .04$

$\bar{p}_3 = 40/500 = .08$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Multiple comparisons

For Supplier A Versus Supplier B

$df = k - 1 = 3 - 1 = z\ \chi^2_{.05} = 5.991$

$$CV_{ij} = \sqrt{\chi^2_\alpha}\sqrt{\frac{\bar{p}_i(1-\bar{p}_i)}{n_i} + \frac{\bar{p}_j(1-\bar{p}_j)}{n_j}} = \sqrt{5.991}\sqrt{\frac{.03(1-.03)}{500} + \frac{.04(1-.04)}{500}} = .0284$$

Comparison	p_i	p_j	Difference	n_i	n_j	Critical Value	Significant Diff > CV
A vs. B	.03	.04	.01	500	500	.0284	
A vs. C	.03	.08	.05	500	500	.0351	Yes
B vs. C	.04	.08	.04	500	500	.0366	Yes

Supplier A and supplier B are both significantly different from supplier C. Supplier C can be eliminated on the basis of a significantly higher proportion of defective components. Since suppliers A and supplier B are not significantly different in terms of the proportion defective components, both suppliers should remain candidates for use by Benson.

6. a. $\bar{p}_1 = 35/250 = .14$ 14% error rate

$\bar{p}_2 = 27/300 = .09$ 9% error rate

b. H_0: $p_1 - p_2 = 0$

H_a: $p_1 - p_2 \neq 0$

Observed Frequencies (f_{ij})

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Return	Office 1	Office 2	Total
Error	35	27	62
Correct	215	273	488
	250	300	550

Expected Frequencies (e_{ij})

Return	Office 1	Office 2	Total
Error	28.18	33.82	62
Correct	221.82	266.18	488
	250	300	550

Chi-Square Calculations $(f_{ij} - e_{ij})^2 / e_{ij}$

Return	Office 1	Office 2	Total
Error	1.65	1.37	3.02
Correct	.21	.17	.38

$\chi^2 = 3.41$

$df = k - 1 = (2 - 1) = 1.$

Using the χ^2 table with $df = 1$, $\chi^2 = 3.41$ shows the p-value is between .10 and .05.

Using software, the p-value corresponding to $\chi^2 = 3.41$ is .0648.

p-value < .10; reject H_0. Conclude that the two offices do not have the same population proportion error rates.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. With two populations, a chi-square test for equal population proportions has one degree of freedom. In this case, the test statistic χ^2 is always equal to z^2. This relationship between the two test statistics always provides the same p-value and the same conclusion when the null hypothesis involves equal population proportions. However, the use of the z-test statistic provides options for one–tailed hypothesis tests about two population proportions while the chi-square test is limited to two–tailed hypothesis tests about the equality of the two population proportions.

8. H_0: The distribution of defects is the same for all suppliers

H_a: The distribution of defects is not the same all suppliers

Observed Frequencies (f_{ij})

Part Tested	A	B	C	Total
Minor defect	15	13	21	49
Major defect	5	11	5	21
Good	130	126	124	380
Total	150	150	150	450

Expected Frequencies (e_{ij})

Part tested	A	B	C	Total
Minor defect	16.33	16.33	16.33	49
Major defect	7.00	7.00	7.00	21

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Good	126.67	126.67	126.67	380
Total	150	150	150	450

Chi-Square Calculations $(f_{ij} - e_{ij})^2 / e_{ij}$

Part tested	A	B	C	Total
Minor defect	.11	.68	1.33	2.12
Major defect	.57	2.29	.57	3.43
Good	.09	.00	.06	.15

$\chi^2 = 5.70$

Degrees of freedom = $(r - 1)(k - 1) = (3 - 1)(3 - 1) = 4$

Using the χ^2 table with $df = 4$, $\chi^2 = 5.70$ shows the p-value is greater than .10

Using software, the p-value corresponding to $\chi^2 = 5.70$ is .2227

p-value > .05; do not reject H_0. Conclude that we are unable to reject the hypothesis that the population distribution of defects is the same for all three suppliers. There is no evidence that quality of parts from one suppliers is better than either of the others two suppliers.

10. H_0: The column variable is independent of the row variable

H_a: The column variable is dependent on the row variable

Observed Frequencies (f_{ij})

	A	B	C	Total
P	20	30	20	70

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Q	30	60	25	115
R	10	15	30	55
Total	60	105	75	240

Expected Frequencies (e_{ij})

	A	B	C	Total
P	17.50	30.63	21.88	70
Q	28.75	50.31	35.94	115
R	13.75	24.06	17.19	55
Total	60	105	75	240

Chi-Square Calculations $(f_{ij} - e_{ij})^2 / e_{ij}$

	A	B	C	Total
P	.36	.01	.16	.53
Q	.05	1.87	3.33	5.25
R	1.02	3.41	9.55	13.99

$\chi^2 = 19.77$

Degrees of freedom = $(r - 1)(c - 1) = (3 - 1)(3 - 1) = 4$.

Using the χ^2 table with $df = 4$, $\chi^2 = 19.77$ shows the p-value is less than .005.

Using software, the p-value corresponding to $\chi^2 = 19.77$ is .0006.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

p-value $\leq$ .05; reject H_0. Conclude that the column variable is not independent of the row variable.

12. a. H_0: Employment plan is independent of the type of company

H_a: Employment plan is not independent of the type of company

Observed Frequency (f_{ij})

Employment Plan	Private	Public	Total
Add employees	37	32	69
No change	19	34	53
Lay off employees	16	42	58
Total	72	108	180

Expected Frequency (e_{ij})

Employment plan	Private	Public	Total
Add employees	27.6	41.4	69
No change	21.2	31.8	53
Lay off employees	23.2	34.8	58
Total	72.0	108.0	180

Chi-Square Calculations $(f_{ij} - e_{ij})^2 / e_{ij}$

Employment Plan	Private	Public	Total
Add employees	3.20	2.13	5.34

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

No change	0.23	0.15	0.38
Lay off employees	2.23	1.49	3.72
		$\chi^2 =$	9.44

Degrees of freedom = $(r - 1)(c - 1) = (3 - 1)(2 - 1) = 2$

Using the χ^2 table with $df = 2$, $\chi^2 = 9.44$ shows the p-value is between .01 and 0.005.

Using software, the p-value corresponding to $\chi^2 = 9.44$ is .0089.

Because the p-value $\leq$ .05; reject H_0. Conclude the employment plan is not independent of the type of company. Thus, we expect employment plan to differ for private and public companies.

b. Column probabilities: For example, 37/72 = .5139

Employment plan	Private	Public
Add employees	.5139	.2963
No change	.2639	.3148
Lay off employees	.2222	.3889

Employment opportunities look to be much better for private companies with over 50% of private companies planning to add employees (51.39%). Public companies have the greater proportions of no change and lay off employees planned. 38.89% of public companies are planning to lay off employees over the next 12 months. 69/180 = .3833, or 38.33% of the companies in the survey are planning to hire and add employees during the next 12 months.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

14. a. H_0: Quality rating is independent of the education of the owner

H_a: Quality rating is not independent of the education of the owner

Observed Frequencies (f_{ij}) Quality Rating	Some HS	HS Grad	Some College	College Grad	Total
Average	35	30	20	60	145
Outstanding	45	45	50	90	230
Exceptional	20	25	30	50	125
Total	100	100	100	200	500

Expected Frequencies (e_{ij})

Quality Rating	Some HS	HS Grad	Some College	College Grad	Total
Average	29	29	29	58	145
Outstanding	46	46	46	92	230
Exceptional	25	25	25	50	125
Total	100	100	100	200	500

Chi-Square Calculations $(f_{ij} - e_{ij})^2 / e_{ij}$

Quality Rating	Some HS	HS Grad	Some College	College Grad	Total
Average	1.24	.03	2.79	.07	4.14
Outstanding	.02	.02	.35	.04	.43
Exceptional	1.00	.00	1.00	.00	2.00

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$\chi^2 = 6.57$

Degrees of freedom = $(r - 1)(c - 1) = (3 - 1)(4 - 1) = 6$.

Using the χ^2 table with $df = 6$, $\chi^2 = 6.57$ shows the p-value is greater than .10.

Using software, the p-value corresponding to $\chi^2 = 6.57$ is .3624.

p-value > .05; do not reject H_0. We are unable to conclude that the quality rating is not independent of the education of the owner. Thus, quality ratings are not expected to differ with the education of the owner.

b. Average: 145/500 = 29%

Outstanding 230/500 = 46%

Exceptional 125/500 = 25%

New owners look to be pretty satisfied with their new automobiles with almost 50% rating the quality outstanding and over 70% rating the quality outstanding or exceptional.

16. a. Observed Frequency (fij)

	Age of Respondent				
Actress	18–30	31–44	45–58	Over 58	Totals
Jessica Chastain	51	50	41	42	184
Jennifer Lawrence	63	55	37	50	205
Emmanuelle Riva	15	44	56	74	189
Quvenzhané Wallis	48	25	22	31	126
Naomi Watts	36	65	62	33	196

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Totals	213	239	218	230	900

The sample size is 900.

b. The sample proportion of movie fans who prefer each actress is:

$$\bar{p}_1 = \frac{184}{900} = .2044$$

$$\bar{p}_2 = \frac{205}{900} = .2278$$

$$\bar{p}_3 = \frac{189}{900} = .2100$$

$$\bar{p}_4 = \frac{126}{900} = .1400$$

$$\bar{p}_5 = \frac{196}{900} = .2178$$

The movie fans favored Jennifer Lawrence, but three other nominees (Jessica Chastain, Emmanuelle Riva, and Naomi Watts) each were favored by almost as many of the fans.

c. Expected Frequency (e_{ij})

	Age of Respondent				
Actress	18–30	31–44	45–58	Over 58	Totals
Jessica Chastain	43.5	48.9	44.6	47.0	184
Jennifer Lawrence	48.5	54.4	49.7	52.4	205
Emmanuelle Riva	44.7	50.2	45.8	48.3	189
Quvenzhané Wallis	29.8	33.5	30.5	32.2	126
Naomi Watts	46.4	52.0	47.5	50.1	196

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Totals	213	239	218	230	900

Calculate $\frac{(f_{ij} - e_{ij})^2}{e_{ij}}$ for each cell in the table.

	Age of Respondent				
Actress	18–30	31–44	45–58	More Than 58	Totals
Jessica Chastain	1.28	0.03	0.29	0.54	2.12
Jennifer Lawrence	4.32	0.01	3.23	0.11	7.66
Emmanuelle Riva	19.76	0.76	2.28	13.67	36.48
Quvenzhané Wallis	11.08	2.14	2.38	0.04	15.65
Naomi Watts	2.33	3.22	4.44	5.83	15.82
Totals	38.77	6.16	12.61	20.20	77.74

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}} = 77.74$$

With (5 – 1)(4 – 1) = 12 degrees of freedom, the *p*-value is approximately 0. *p*-value $\leq$ .05; reject H_0. Attitude toward the actress who was most deserving of the 2013 Academy Award for actress in a leading role is not independent of age.

18. Expected frequencies:

$e_{11} = 11.81$ $e_{12} = 8.44$ $e_{13} = 24.75$

$e_{21} = 8.40$ $e_{22} = 6.00$ $e_{23} = 17.60$

$e_{31} = 21.79$ $e_{32} = 15.56$ $e_{33} = 45.65$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Host A	Host B	Observed Frequency (f_i)	Expected Frequency (e_i)	Chi-Square $(f_i - e_i)^2 / e_i$
Con	Con	24	11.81	12.57
Con	Mixed	8	8.44	.02
Con	Pro	13	24.75	5.58
Mixed	Con	8	8.40	.02
Mixed	Mixed	13	6.00	8.17
Mixed	Pro	11	17.60	2.48
Pro	Con	10	21.79	6.38
Pro	Mixed	9	15.56	2.77
Pro	Pro	64	45.65	7.38

$\chi^2 = 45.36$

Degrees of freedom = $(r - 1)(c - 1) = (3 - 1)(3 - 1) = 4$

Using the χ^2 table with $df = 2$, $\chi^2 = 45.36$ shows the p-value is less than .005.

Using software, the p-value corresponding to $\chi^2 = 45.36$ is .0000.

p-value $\leq$ .01; reject H_0. Conclude that the ratings of the two hosts are not independent. The host responses are more similar than different and they tend to agree or be close in their ratings.

20. With $n = 30$ we will use six classes, each with the probability of .1667.

$\bar{x} = 22.8$ $s = 6.27$

The z values that create six intervals, each with probability .1667, are –.97, –

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

.43, 0, .43, .97.

z	Cutoff Value of x
–.97	22.8 – .97 (6.27) = 16.74
–.43	22.8 – .43 (6.27) = 20.10
0	22.8 + 0 (6.27) = 22.80
.43	22.8 + .43 (6.27) = 25.50
.97	22.8 + .97 (6.27) = 28.86

Interval	Observed Frequency	Expected Frequency	Difference
Less than 16.74	3	5	–2
16.74–20.10	7	5	2
20.10–22.80	5	5	0
22.80–25.50	7	5	2
25.50–28.86	3	5	–2
28.86 and up	5	5	0

$$\chi^2 = \frac{(-2)^2}{5} + \frac{(2)^2}{5} + \frac{(0)^2}{5} + \frac{(2)^2}{5} + \frac{(-2)^2}{5} + \frac{(0)^2}{5} = \frac{16}{5} = 3.20$$

Degrees of freedom = $k - p - 1 = 6 - 2 - 1 = 3$

Using the χ^2 table with $df = 3$, $\chi^2 = 3.20$ shows the p-value is greater than .10.

Using software, the p-value corresponding to $\chi^2 = 3.20$ is .3618.

p-value > .05; do not reject H_0. The claim that the data come from a normal

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

distribution cannot be rejected.

22.

Category	Hypothesized Proportion	Observed Frequency (f_i)	Expected Frequency (e_i)	Chi-Square $(f_i - e_i)^2 / e_i$
Blue	.24	105	120	1.88
Brown	.13	72	65	.75
Green	.20	89	100	1.21
Orange	.16	84	80	.20
Red	.13	70	65	.38
Yellow	.14	80	70	1.43
	Total:	500		$\chi^2 = 5.85$

$k - 1 = 6 - 1 = 5$ degrees of freedom

Using the χ^2 table with $df = 5$, $\chi^2 = 5.85$ shows the p-value is greater than .10

Using software, the p-value corresponding to $\chi^2 = 5.85$ is .3211

p-value > .05; do not reject H_0. We cannot reject the hypothesis that the overall percentages of colors in the population of M&M milk chocolate candies are .24 blue, .13 brown, .20 green, .16 orange, .13 red and .14 yellow.

24. a. H_0: $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = 1/7$

H_a: Not all proportions are equal

Observed Frequency (f_i)

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
66	50	53	47	55	69	80

Expected Frequency (e_i) $e_i = 1/7(420) = 60$

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
60	60	60	60	60	60	60

Chi-Square Calculations $(f_i - e_i)^2 / e_i$

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
.60	1.67	.82	2.82	.42	1.35	6.67

$\chi^2 = 14.33$

Degrees of freedom = $(k - 1) = (7 - 1) = 6$

Using the χ^2 table with $df = 6$, $\chi^2 = 14.33$ shows the p-value is between .05 and .025.

Using software, the p-value corresponding to $\chi^2 = 14.33$ is .0262

p-value $\leq$ .05; reject H_0. Conclude the proportion of traffic accidents is not the same for each day of the week.

b. Percentage of traffic accidents by day of the week

Sunday	66/420 = .1571	15.71%
Monday	50/420 = .1190	11.90%
Tuesday	53/420 = .1262	12.62%
Wednesday	47/420 = .1119	11.19%

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Thursday	55/420 = .1310	13.10%
Friday	69/420 = .1643	16.43%
Saturday	80/420 = .1905	19.05%

Saturday has the highest percentage of traffic accident (19%). Saturday is typically the late night and more social day/evening of the week. Alcohol, speeding and distractions are more likely to affect driving on Saturdays. Friday is the second highest with 16.43%.

26. $\bar{x} = 24.5$ $s = 3$ $n = 30$

Use six classes,

Percentage	z	Data Value
16.67%	–.97	24.5–.97(3) = 21.59
33.33%	–.43	24.5–.43(3) = 23.21
50.00%	.00	24.5+.00(3) = 24.50
66.67%	.43	24.5+.43(3) = 25.79
83.33%	.97	24.5+.97(3) = 27.41

Interval	Observed Frequency	Expected Frequency
Less than 21.59	5	5
21.59–23.21	4	5
23.21–24.50	3	5
24.50–25.79	7	5

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

25.7927.41	7	5
27.41 up	4	5

$\chi^2 = 2.80$

Degrees of freedom = $(k - p - 1) = 6 - 2 - 1 = 3$.

Using the χ^2 table with $df = 3$, $\chi^2 = 2.80$ shows the p-value is greater than .10.

Using software, the p-value corresponding to $\chi^2 = 2.80$ is .4235.

p-value > .10; do not reject H_0. The assumption of a normal distribution cannot be rejected.

28. a. H_0: $p_1 = p_2 = p_3$

H_a: Not all population proportions are equal

Observed Frequencies (f_{ij})

Quality	First	Second	Third	Total
Good	285	368	176	829
Defective	15	32	24	71
Total	300	400	200	900

Expected Frequencies (e_{ij})

Quality	First	Second	Third	Total
Good	276.33	368.44	184.22	829

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Defective	23.67	31.56	15.78	71
Total	300	400	200	900

Chi-Square Calculations $(f_{ij} - e_{ij})^2 / e_{ij}$

Quality	First	Second	Third	Total
Good	.27	.00	.37	.64
Defective	3.17	.01	4.28	7.46

$\chi^2 = 8.10$

Degrees of freedom = $k - 1 = (3 - 1) = 2$.

Using the χ^2 table with $df = 2$, $\chi^2 = 8.10$ shows the p-value is between .025 and .01.

Using software, the p-value corresponding to $\chi^2 = 8.10$ is .0174.

p-value $\leq$ .05; reject H_0. Conclude the population proportion of good parts is not equal for all three shifts. The shifts differ in terms of production quality.

b. $\bar{p}_1 = 285 / 300 = .95$

$\bar{p}_2 = 368 / 400 = .92$

$\bar{p}_3 = 176 / 200 = .88$

$df = k - 1 = 3 - 1 = 2$ $\chi^2_{.05} = 5.991$

Comparison	p_i	p_j	Difference	n_i	n_j	Critical Value	Significant Diff > CV

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

1 vs. 2	.95	.92	.03	300	400	.0453	
1 vs. 3	.95	.88	.07	300	200	.0641	Yes
2 vs. 3	.92	.88	.04	400	200	.0653	

Shifts 1 and 3 differ significantly with shift 1 producing better quality (95%) than shift 3 (88%). The study cannot identify shift 2 (92%) as better or worse quality than the other two shifts. Shift 3, at 7% more defectives than shift 1 should be studied to determine how to improve its production quality.

30. a. H_0: The preferred pace of life is independent of gender

H_a: The preferred pace of life is not independent of gender

Observed Frequency (f_{ij})

	Gender		
Preferred Pace of Life	Male	Female	Total
Slower	230	218	448
No Preference	20	24	44
Faster	90	48	138
Total	340	290	630

Expected Frequency (e_{ij})

Gender

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Preferred Pace of Life	Male	Female	Total
Slower	241.78	206.22	448
No Preference	23.75	20.25	44
Faster	74.48	63.52	138
Total	340	290	630

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chi-Square Calculations $(f_{ij} - e_{ij})^2 / e_{ij}$

	Gender		
Preferred Pace of Life	Male	Female	Total
Slower	.57	.67	1.25
No preference	.59	.69	1.28
Faster	3.24	3.79	7.03

$\chi^2 = 9.56$

Degrees of freedom = $(r - 1)(c - 1) = (3 - 1)(2 - 1) = 2$.

Using the χ^2 table with $df = 2$, $\chi^2 = 9.56$ shows the p-value is less than .01.

Using software, the p-value corresponding to $\chi^2 = 9.56$ is .0084.

p-value < .05; reject H_0. The preferred pace of life is not independent of gender. Thus, we expect men and women differ with respect to the preferred pace of life.

b. Percentage responses for each gender

	Gender	
Preferred Pace of Life	Male	Female
Slower	67.65	75.17
No preference	5.88	8.28
Faster	26.47	16.55

The highest percentages are for a slower pace of life by both men and women.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

However, 75.17% of women prefer a slower pace compared to 67.65% of men and 26.47% of men prefer a faster pace compared to 16.55% of women. More women prefer a slower pace while more men prefer a faster pace.

32. H_0: The county with the emergency call is independent of the day of week

H_a: The county with the emergency call is not independent of the day of week

Observed Frequencies (f_{ij})

	Day of Week							
County	Sun	Mon	Tues	Wed	Thu	Fri	Sat	Total
Urban	61	48	50	55	63	73	43	393
Rural	7	9	16	13	9	14	10	78
Total	68	57	66	68	72	87	53	471

Expected Frequencies (e_{ij})

	Day of Week							
County	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
Urban	56.74	47.56	55.07	56.74	60.08	72.59	44.22	393
Rural	11.26	9.44	10.93	11.26	11.92	14.41	8.78	78
Total	68	57	66	68	72	87	53	471

Chi-Square $(f_{ij} - e_{ij})^2 / e_{ij}$

Day of Week

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

County	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
Urban	.32	.00	.47	.05	.14	.00	.03	1.02
Rural	1.61	.02	2.35	.27	.72	.01	.17	5.15

$\chi^2 = 6.17$

Degrees of freedom = $(r - 1)(c - 1) = (2 - 1)(7 - 1) = 6$.

Using the χ^2 table with $df = 6$, $\chi^2 = 6.17$ shows the p-value is greater than .10.

Using software, the p-value corresponding to $\chi^2 = 6.17$ is .4044.

p-value > .05; do not reject H_0. The assumption of independence cannot be rejected. The county with the emergency call does not vary or depend upon the day of the week.

34.

Expected Frequencies for $n = 344$

Professional football	$e_1 = .33*344 = 113.52$
Baseball	$e_2 = .15*344 = 51.60$
Men's college football	$e_3 = .10*344 = 34.40$
Auto racing	$e_4 = .06*344 = 20.64$
Men's professional basketball	$e_5 = .05*344 = 17.20$
Ice hockey	$e_6 = .05*344 = 17.20$
Other sports	$e_7 = .26*344 = 89.44$

Actual Frequencies

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Professional football	$f_1 = 111$
Baseball	$f_2 = 39$
Men's college football	$f_3 = 46$
Auto racing	$f_4 = 14$
Men's professional basketball	$f_5 = 6$
Ice hockey	$f_6 = 20$
Other sports	$f_7 = 108$

$$\chi^2 = \frac{(111-113.52)^2}{113.52} + \frac{(39-51.6)^2}{51.6} + \frac{(46-34.4)^2}{34.4} + \frac{(14-20.64)^2}{20.64} + \frac{(6-17.2)^2}{17.2} + \frac{(20-17.2)^2}{17.2} + \frac{(108-89.44)^2}{89.44}$$
$$= 20.78$$

$k - 1 = 6$ degrees of freedom

Using the χ^2 table with $df = 6$, $\chi^2 = 20.78$ shows the p-value is less than .005.

Using Excel, the p-value corresponding to $\chi^2 = 20.78$ with $df = 6$ is .0020.

p-value < .05; reject H_0. Yes, undergraduate students differ from the general public with regard to their favorite sports. Men's college football is more popular among undergraduate students, and baseball and auto racing are less popular among undergraduate students. The difference between expected and actual number of undergraduate students who responded *other sports* also suggests that undergraduate students are interested in a broader range of sports.

36. $\bar{x} = 76.83$ $s = 12.43$

Interval	Observed Frequency	Expected Frequency

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Less than 62.54	5	5
62.54–68.50	3	5
68.50–72.85	6	5
72.85–76.83	5	5
76.83–80.81	5	5
80.81–85.16	7	5
85.16–91.12	4	5
91.12 up	5	5

$\chi^2 = 2$

Degrees of freedom = $k - p - 1 = 8 - 2 - 1 = 5$.

Using the χ^2 table with $df = 5$, $\chi^2 = 2.00$ shows the p-value is greater than .10.

Using software, the p-value corresponding to $\chi^2 = 2.00$ is .8491.

p-value > .05; do not reject H_0. The assumption of a normal distribution cannot be rejected.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 13 Solutions

2.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Treatments	300	4	75	14.07	.0000
Error	160	30	5.33		
Total	460	34			

4.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Treatments	150	2	75	4.80	.0233
Error	250	16	15.63		
Total	400	18			

Using F table (2 degrees of freedom numerator and 16 denominator), p-value is between .01 and .025

Using Excel, the p-value corresponding to $F = 4.80$ is .0233.

Because p-value $\leq \alpha = .05$, we reject the null hypothesis that the means of the three treatments are equal.

6.

A B C

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Sample Mean	119	107	100
Sample Variance	146.86	96.44	173.78

$$\bar{\bar{x}} = \frac{8(119)+10(107)+10(100)}{28} = 107.93$$

$$\text{SSTR} = \sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2 = 8(119 - 107.93)^2 + 10(107 - 107.93)^2 + 10(100 - 107.93)^2$$

$= 1617.857$

MSTR = SSTR /(k – 1) = 1617.857 /2 = 808.93

$$\text{SSE} = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 7(146.86) + 9(96.44) + 9(173.78) = 3{,}460$$

MSE = SSE /(n_T – k) = 3,460 /(28 – 3) = 138.4

F = MSTR /MSE = 809.95 /138.4 = 5.85

Using F table (2 degrees of freedom numerator and 25 denominator), p-value is less than .01

Using Excel, the p-value corresponding to F = 5.85 is .0082.

Because p-value $\leq \alpha$ = .05, we reject the null hypothesis that the means of the three treatments are equal.

8. $\bar{\bar{x}}$ = (79 + 74 + 66)/3 = 73

$$\text{SSTR} = \sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2 = 6(79 - 73)^2 + 6(74 - 73)^2 + 6(66 - 73)^2 = 516$$

MSTR = SSTR /(k – 1) = 516/2 = 258

$s_1^2 = 34$ $s_2^2 = 20$ $s_3^2 = 32$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\text{SSE} = \sum_{j=1}^{k}(n_j - 1)s_j^2 = 5(34) + 5(20) + 5(32) = 430$$

$$\text{MSE} = \text{SSE}/(n_T - k) = 430/(18 - 3) = 28.67$$

$$F = \text{MSTR}/\text{MSE} = 258/28.67 = 9.00$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Treatments	516	2	258	9.00	.003
Error	430	15	28.67		
Total	946	17			

Using F table (2 degrees of freedom numerator and 15 denominator), p-value is less than .01.

Using Excel the p-value corresponding to $F = 9.00$ is .003.

Because p-value $\leq \alpha = .05$, we reject the null hypothesis that the means for the three plants are equal. In other words, analysis of variance supports the conclusion that the population mean examination score at the three NCP plants are not equal.

10.

	Direct Experience	Indirect Experience	Combination
Sample Mean	17.0	20.4	25.0
Sample Variance	5.01	6.26	4.01

$$\bar{\bar{x}} = (17 + 20.4 + 25)/3 = 20.8$$

$$\text{SSTR} = \sum_{j=1}^{k} n_j\left(\bar{x}_j - \bar{\bar{x}}\right)^2 = 7(17 - 20.8)^2 + 7(20.4 - 20.8)^2 + 7(25 - 20.8)^2 = 225.68$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

MSTR = SSTR /(k – 1) = 225.68 /2 = 112.84

$$\text{SSE} = \sum_{j=1}^{k}(n_j - 1)s_j^2 = 6(5.01) + 6(6.26) + 6(4.01) = 91.68$$

MSE = SSE /(n_T – k) = 91.68 /(21 – 3) = 5.09

F = MSTR /MSE = 112.84 /5.09 = 22.17

Using F table (2 degrees of freedom numerator and 18 denominator), p-value is less than .01.

Using Excel the p-value corresponding to F = 22.17 is .0000.

Because p-value $\leq \alpha$ = .05, we reject the null hypothesis that the means for the three groups are equal.

12.

	Italian	Seafood	Steakhouse
Sample Mean	17	19	24
Sample Variance	14.857	13.714	14.000

$\bar{\bar{x}}$ = (17 + 19 + 24)/3 = 20

$$\text{SSTR} = \sum_{j=1}^{k} n_j\left(\bar{x}_j - \bar{\bar{x}}\right)^2 = 8(17 - 20)^2 + 8(19 - 20)^2 + 8(24 - 20)^2 = 208$$

MSTR = SSTR /(k – 1) = 208/2 = 104

$$\text{SSE} = \sum_{j=1}^{k}(n_j - 1)s_j^2 = 7(14.857) + 7(13.714) + 7(14.000) = 298$$

MSE = SSE /(n_T – k) = 298 /(24 – 3) = 14.19

F = MSTR /MSE = 104 /14.19 = 7.33

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Using the F table (2 degrees of freedom numerator and 21 denominator), the p-value is less than .01.

Using Excel the p-value corresponding to $F = 7.33$ is .0038.

Because p-value $\leq \alpha = .05$, we reject the null hypothesis that the mean meal prices are the same for the three types of restaurants.

14. a.

	Sample 1	Sample 2	Sample 3
Sample Mean	51	77	58
Sample Variance	96.67	97.34	81.99

$\bar{\bar{x}} = (51 + 77 + 58)/3 = 62$

$$\text{SSTR} = \sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2 = 4(51 - 62)^2 + 4(77 - 62)^2 + 4(58 - 62)^2 = 1{,}448$$

$\text{MSTR} = \text{SSTR}/(k - 1) = 1{,}448/2 = 724$

$$\text{SSE} = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 3(96.67) + 3(97.34) + 3(81.99) = 828$$

$\text{MSE} = \text{SSE}/(n_T - k) = 828/(12 - 3) = 92$

$F = \text{MSTR}/\text{MSE} = 724/92 = 7.87$

Using F table (2 degrees of freedom numerator and 9 denominator), p-value is between .01 and .025.

Actual p-value = .0106

Because p-value $\leq \alpha = .05$, we reject the null hypothesis that the means of the three populations are equal.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $\text{LSD} = t_{\alpha/2}\sqrt{\text{MSE}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} = t_{.025}\sqrt{92\left(\frac{1}{4}+\frac{1}{4}\right)} = 2.262\sqrt{46} = 15.34$

$|\bar{x}_1 - \bar{x}_2| = |51-77| = 26 >$ LSD; significant difference

$|\bar{x}_1 - \bar{x}_3| = |51-58| = 7 <$ LSD; no significant difference

$|\bar{x}_2 - \bar{x}_3| = |77-58| = 19 >$ LSD; significant difference

16. $\bar{x}_1 - \bar{x}_2 \pm \text{LSD}$

23 – 28 $\pm$ 3.54

–5 $\pm$ 3.54 = –8.54 to –1.46

18. a.

	Machine 1	Machine 2	Machine 3	Machine 4
Sample Mean	7.1	9.1	9.9	11.4
Sample Variance	1.21	.93	.70	1.02

$\bar{\bar{x}}$ = (7.1 + 9.1 + 9.9 + 11.4)/4 = 9.38

$\text{SSTR} = \sum_{j=1}^{k} n_j\left(\bar{x}_j - \bar{\bar{x}}\right)^2$ = 6(7.1 – 9.38) ^{2} + 6(9.1 – 9.38) ^{2} + 6(9.9 – 9.38) ^{2} + 6(11.4 – 9.38) ^{2} = 57.77

MSTR = SSTR /(k – 1) = 57.77/3 = 19.26

$\text{SSE} = \sum_{j=1}^{k}(n_j - 1)s_j^2$ = 5(1.21) + 5(.93) + 5(.70) + 5(1.02) = 19.30

MSE = SSE /(n_T – k) = 19.30/(24 – 4) = .97

F = MSTR /MSE = 19.26/.97 = 19.86

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Using F table (3 degrees of freedom numerator and 20 denominator), p-value is less than .01.

Using Excel, the p-value corresponding to $F = 19.86$ is .0000.

Because p-value $\leq \alpha = .05$, we reject the null hypothesis that the mean time between breakdowns is the same for the four machines.

b. Note: $t_{\alpha/2}$ is based on 20 degrees of freedom.

$$\text{LSD} = t_{\alpha/2}\sqrt{\text{MSE}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} = t_{.025}\sqrt{0.97\left(\frac{1}{6}+\frac{1}{6}\right)} = 2.086\sqrt{.3233} = 1.19$$

$$|\bar{x}_2 - \bar{x}_4| = |9.1 - 11.4| = 2.3 > \text{LSD; significant difference}$$

20. a. Partial output is shown below:

One-way ANOVA: Attendance versus Division

```
Analysis of Variance for Attendance

 Source          DF      Adj SS       Adj MS       F-Value    P-Value

 Division        2       18109727     9054863      6.96       0.011

 Error           11      14315319     1301393

 Total           13      32425045

      Model Summary

      S               R-sq         R-sq(adj)     R-sq(pred)

      1140.79         55.85%       47.82%        30.33%

      Means

      Division          N    Mean    StDev    95% CI
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
North            6    7702    1301    (6677, 8727)

South            4    5566    1275    (4310, 6821)

West             4    8430    570     (7174, 9685)

Pooled StDev = 1140.79
```

Because p-value = .011 $\leq \alpha$ = .05, we reject the null hypothesis that the mean attendance values are equal.

b. $n_1 = 6$ $n_2 = 4$ $n_3 = 4$

$t_{\alpha/2}$ is based upon 11 degrees of freedom

Comparing North and South

$$\text{LSD} = t_{.025}\sqrt{1{,}301{,}393\left(\frac{1}{6}+\frac{1}{4}\right)} = 2.201\sqrt{1{,}301{,}393\left(\frac{1}{6}+\frac{1}{4}\right)} = 1620.76$$

$|7702 - 5566| = 2136 >$ LSD; significant difference

Comparing North and West

$$\text{LSD} = t_{.025}\sqrt{1{,}301{,}393\left(\frac{1}{6}+\frac{1}{4}\right)} = 2.201\sqrt{1{,}301{,}393\left(\frac{1}{6}+\frac{1}{4}\right)} = 1620.76$$

$|7702 - 8430| = 728 <$ LSD; no significant difference

Comparing South and West

$$\text{LSD} = t_{.025}\sqrt{1{,}301{,}393\left(\frac{1}{4}+\frac{1}{4}\right)} = 2.201\sqrt{1{,}301{,}393\left(\frac{1}{4}+\frac{1}{4}\right)} = 1775.45$$

$|5566 - 8430| = 2864 >$ LSD; significant difference

The difference in the mean attendance among the three divisions is the result of low attendance in the South division.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

22.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Treatments	310	4	77.5	17.69	.0005
Blocks	85	2	42.5		
Error	35	8	4.38		
Total	430	14			

Using F table (4 degrees of freedom numerator and 8 denominator), p-value is less than .01.

Using Excel, the p-value corresponding to $F = 17.69$ is .0005.

Because p-value $\leq \alpha = .05$, we reject the null hypothesis that the means of the treatments are equal.

24. Treatment Means:

$\bar{x}_{.1} = 56$ $\bar{x}_{.2} = 44$

Block Means:

$\bar{x}_{1.} = 46$ $\bar{x}_{2.} = 49.5$ $\bar{x}_{3.} = 54.5$

Overall Mean:

$\bar{\bar{x}} = 300/6 = 50$

Step 1

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\text{SST} = \sum_i \sum_j \left(x_{ij} - \bar{\bar{x}}\right)^2 = (50 - 50)^2 + (42 - 50)^2 + \cdot \cdot \cdot + (46 - 50)^2 = 310$$

Step 2

$$\text{SSTR} = b\sum_j \left(\bar{x}_{\cdot j} - \bar{\bar{x}}\right)^2 = 3\ [\ (56 - 50)^2 + (44 - 50)^2\ ] = 216$$

Step 3

$$\text{SSBL} = k\sum_i \left(\bar{x}_{i\cdot} - \bar{\bar{x}}\right)^2 = 2\ [\ (46 - 50)^2 + (49.5 - 50)^2 + (54.5 - 50)^2\ ] = 73$$

Step 4

$$\text{SSE} = \text{SST} - \text{SSTR} - \text{SSBL} = 310 - 216 - 73 = 21$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Treatments	216	1	216	20.57	.0453
Blocks	73	2	36.5		
Error	21	2	10.5		
Total	310	5			

Using F table (1 degree of freedom numerator and 2 denominator), p-value is between .025 and .05.

Using Excel, the p-value corresponding to $F = 20.57$ is .0453.

Because p-value $\leq \alpha = .05$, we reject the null hypothesis that the mean tune-up times are the same for both analyzers.

26. a. Treatment Means:

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$\bar{x}_{\cdot 1} = 502$ $\bar{x}_{\cdot 2} = 515$ $\bar{x}_{\cdot 3} = 494$

Block Means:

$\bar{x}_{1\cdot} = 530$ $\bar{x}_{2\cdot} = 590$ $\bar{x}_{3\cdot} = 458$ $\bar{x}_{4\cdot} = 560$ $\bar{x}_{5\cdot} = 448$ $\bar{x}_{6\cdot} = 436$

Overall Mean:

$\bar{\bar{x}} = 9066/18 = 503.67$

Step 1

$$\text{SST} = \sum_i \sum_j \left(x_{ij} - \bar{\bar{x}}\right)^2 = (526 - 503.67)^2 + (534 - 503.67)^2 + \cdots + (420 - 503.67)^2$$

$= 65{,}798$

Step 2

$$\text{SSTR} = b\sum_j \left(\bar{x}_{\cdot j} - \bar{\bar{x}}\right)^2 = 6[\,(502 - 503.67)^2 + (515 - 503.67)^2 + (494 - 503.67)^2\,] =$$

1348

Step 3

$$\text{SSBL} = k\sum_i \left(\bar{x}_{i\cdot} - \bar{\bar{x}}\right)^2 = 3\,[\,(530 - 503.67)^2 + (590 - 503.67)^2 + \cdots + (436 -$$

$503.67)^2\,] = 63{,}250$

Step 4

SSE = SST – SSTR – SSBL = 65,798 – 1348 – 63,250= 1200

Source of Variation	Sum of	Degrees of	Mean	*F*	*p*-value

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	Squares	Freedom	Square		
Treatments	1348	2	674	5.62	.0231
Blocks	63,250	5	12,650		
Error	1200	10	120		
Total	65,798	17			

Using F table (2 degrees of freedom numerator and 10 denominator), p-value is between .01 and .025.

Using Excel, the p-value corresponding to $F = 5.62$ is .0231.

Because p-value $\leq \alpha = .05$, we reject the null hypothesis that the mean scores for the three parts of the SAT are equal.

b. The mean test scores for the three sections are 502 for critical reading, 515 for mathematics, and 494 for writing. Because the writing section has the lowest average score, this section appears to give the students the most trouble.

28.

		Factor B			Factor A
		Level 1	Level 2	Level 3	Means
Factor A	Level 1	$\bar{x}_{11} = 150$	$\bar{x}_{12} = 78$	$\bar{x}_{13} = 84$	$\bar{x}_{1.} = 104$
	Level 2	$\bar{x}_{21} = 110$	$\bar{x}_{22} = 116$	$\bar{x}_{23} = 128$	$\bar{x}_{2.} = 118$
Factor B	Means	$\bar{x}_{.1} = 130$	$\bar{x}_{.2} = 97$	$\bar{x}_{.3} = 106$	$\bar{\bar{x}} = 111$

Step 1

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\text{SST} = \sum_i \sum_j \sum_k \left(x_{ijk} - \bar{\bar{x}}\right)^2 = (135 - 111)^2 + (165 - 111)^2 + \cdots + (136 - 111)^2 =$$

9,028

Step 2

$$\text{SSA} = br\sum_i \left(\bar{x}_{i.} - \bar{\bar{x}}\right)^2 = 3\ (2)\ [\ (104 - 111)^2 + (118 - 111)^2\ ] = 588$$

Step 3

$$\text{SSB} = ar\sum_j \left(\bar{x}_{.j} - \bar{\bar{x}}\right)^2 = 2\ (2)\ [\ (130 - 111)^2 + (97 - 111)^2 + (106 - 111)^2\ ] = 2{,}328$$

Step 4

$$\text{SSAB} = r\sum_i \sum_j \left(\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}}\right)^2 = 2\ [\ (150 - 104 - 130 + 111)^2 + (78 - 104 - 97 +$$

$$111)^2 + \cdots + (128 - 118 - 106 + 111)^2\ ] = 4{,}392$$

Step 5

SSE = SST – SSA – SSB – SSAB = 9,028 – 588 – 2,328 – 4,392 = 1,720

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*-value
Factor A	588	1	588	2.05	.2022
Factor B	2328	2	1164	4.06	.0767
Interaction	4392	2	2196	7.66	.0223
Error	1720	6	286.67		

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Total	9028	11

Factor A: $F = 2.05$

Using F table (1 degree of freedom numerator and 6 denominator), p-value is greater than .10.

Using Excel, the p-value corresponding to $F = 2.05$ is .2022.

Because p-value $> \alpha = .05$, Factor A is not significant.

Factor B: $F = 4.06$

Using F table (2 degrees of freedom numerator and 6 denominator), p-value is between .05 and .10.

Using Excel, the p-value corresponding to $F = 4.06$ is .0767.

Because p-value $> \alpha = .05$, Factor B is not significant.

Interaction: $F = 7.66$

Using F table (2 degrees of freedom numerator and 6 denominator), p-value is between .01 and .025.

Using Excel, the p-value corresponding to $F = 7.66$ is .0223.

Because p-value $\leq \alpha = .05$, Interaction is significant.

30. Factor A is navigation menu position; Factor B is amount of text entry required.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

		Factor B		Factor A
		Small	Large	Means
	A	$\bar{x}_{11} = 10$	$\bar{x}_{12} = 10$	$\bar{x}_{1.} = 10$
Factor A	B	$\bar{x}_{21} = 18$	$\bar{x}_{22} = 28$	$\bar{x}_{2.} = 23$
	C	$\bar{x}_{31} = 14$	$\bar{x}_{32} = 16$	$\bar{x}_{3.} = 15$
Factor B	Means	$\bar{x}_{.1} = 14$	$\bar{x}_{.2} = 18$	$\bar{\bar{x}} = 16$

Step 1

$$\text{SST} = \sum_i \sum_j \sum_k \left(x_{ijk} - \bar{\bar{x}}\right)^2 = (8 - 16)^2 + (12 - 16)^2 + (12 - 16)^2 + \cdot\cdot\cdot + (14 - 16)^2 =$$

664

Step 2

$$\text{SSA} = br\sum_i \left(\bar{x}_{i.} - \bar{\bar{x}}\right)^2 = 2\ (2)\ [\ (10 - 16)^2 + (23 - 16)^2 + (15 - 16)^2\ ] = 344$$

Step 3

$$\text{SSB} = ar\sum_j \left(\bar{x}_{.j} - \bar{\bar{x}}\right)^2 = 3\ (2)\ [\ (14 - 16)^2 + (18 - 16)^2\ ] = 48$$

Step 4

$$\text{SSAB} = r\sum_i \sum_j \left(\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}}\right)^2 = 2\ [\ (10 - 10 - 14 + 16)^2 + \cdot\cdot\cdot + (16 - 15 - 18$$

$+16)^2\ ] = 56$

Step 5

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

SSE = SST – SSA – SSB – SSAB = 664 – 344 – 48 – 56 = 216

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*-value
Factor A	344	2	172	172/36 = 4.78	.0574
Factor B	48	1	48	48/36 = 1.33	.2921
Interaction	56	2	28	28/36 = 0.78	.5008
Error	216	6	36		
Total	664	11			

Using *F* table for Factor A (2 degrees of freedom numerator and 6 denominator), *p*-value is between .05 and .10.

Using Excel the *p*-value corresponding to $F = 4.78$ is .0574.

Because *p*-value > α= .05, Factor A is not significant; there is not sufficient evident to suggest a difference due to the navigation menu position.

Using *F* table for Factor B (1 degree of freedom numerator and 6 denominator), *p*-value is greater than .10.

Using Excel the *p*-value corresponding to F =1.33 is .2921.

Because *p*-value > α = .05, Factor B is not significant; there is not a significant difference due to amount of required text entry.

Using *F* table for Interaction (2 degrees of freedom numerator and 6 denominator), *p*-value is greater than .10.

Using Excel, the *p*-value corresponding to $F = 0.78$ is .5008.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Because p-value $> \alpha = .05$, Interaction is not significant.

32. Factor A is class of vehicle tested (small car, midsize car, small SUV, and midsize SUV) and Factor B is Type (hybrid or conventional). The data in tabular format follow.

	Hybrid	Conventional
Small Car	37	28
	44	32
Midsize Car	27	23
	32	25
Small SUV	27	21
	28	22
Midsize SUV	23	19
	24	18

Summary statistics for the preceding data follow.

	Hybrid	Conventional	
Small Car	$\bar{x}_{11} = 40.5$	$\bar{x}_{12} = 30.0$	$\bar{x}_{1.} = 35.25$
Midsize Car	$\bar{x}_{21} = 29.5$	$\bar{x}_{22} = 24.0$	$\bar{x}_{2.} = 26.75$
Small SUV	$\bar{x}_{31} = 27.5$	$\bar{x}_{32} = 21.5$	$\bar{x}_{3.} = 24.50$
Midsize SUV	$\bar{x}_{41} = 23.5$	$\bar{x}_{42} = 18.5$	$\bar{x}_{4.} = 21.00$
	$\bar{x}_{.1} = 30.25$	$\bar{x}_{.2} = 23.5$	$\bar{\bar{x}} = 26.875$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Step 1

$$\text{SST} = \sum_i \sum_j \sum_k \left(x_{ijk} - \bar{\bar{x}}\right)^2 = (37 - 26.875)^2 + (44 - 26.875)^2 + \cdot\cdot\cdot + (18 - 26.875)^2 =$$

691.75

Step 2

$$\text{SSA} = br\sum_i \left(\bar{x}_{i\cdot} - \bar{\bar{x}}\right)^2 = 2(2)\ [(35.25 - 26.875)^2 + (26.75 - 26.875)^2 + (24.5 - 26.875)$$

$^2 + (21.0 - 26.875)^2] = 441.25$

Step 3

$$\text{SSB} = ar\sum_j \left(\bar{x}_{\cdot j} - \bar{\bar{x}}\right)^2 = 4(2)\ [(30.25 - 26.875)^2 + (23.5 - 26.875)^2\ ] = 182.25$$

Step 4

$$\text{SSAB} = r\sum_i \sum_j \left(\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{\bar{x}}\right)^2 = 2[(40.5 - 35.25 - 30.25 + 26.875)^2 + (30 - 35.25 -$$

$23.5 + 26.875)^2 + \cdot\cdot\cdot + (18.5 - 21.0 - 23.5 + 26.875)^2] = 19.25$

Step 5

SSE = SST – SSA – SSB – SSAB = 691.75 – 441.25 – 182.25 – 19.25 = 49

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*-value
Factor A	441.25	3	147.083	24.01	.0002
Factor B	182.25	1	182.250	29.76	.0006

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Interaction	19.25	3	6.417	1.05	.4229
Error	49.00	8	6.125		
Total	691.75	15			

Conclusions

Factor A: Because p-value = .0002 < α = .05, Factor A (Class) is significant.

Factor B: Because p-value = .0006 < α = .05, Factor B (Type) is significant.

Interaction: Because p-value = .4229 > α = .05, Interaction is not significant.

The class of vehicles has a significant effect on miles per gallon with cars showing more miles per gallon than SUVs. The type of vehicle also has a significant effect, with hybrids having more miles per gallon than conventional vehicles. There is no evidence of a significant interaction effect.

34.

	x	y	z
Sample Mean	92	97	84
Sample Variance	30	6	35.33

$\bar{\bar{x}} = (92 + 97 + 44)/3 = 91$

$$\text{SSTR} = \sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2 = 4(92 - 91)^2 + 4(97 - 91)^2 + 4(84 - 91)^2 = 344$$

$\text{MSTR} = \text{SSTR}/(k - 1) = 344/2 = 172$

$$\text{SSE} = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 3(30) + 3(6) + 3(35.33) = 213.99$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

MSE = SSE /($n_T - k$) = 213.99 /(12 – 3) = 23.78

F = MSTR /MSE = 172 /23.78 = 7.23

Using F table (2 degrees of freedom numerator and 9 denominator), p-value is between .01 and .025

Using Excel, the p-value corresponding to $F = 7.23$ is .0134.

Because p-value $\leq \alpha = .05$, we reject the null hypothesis that the mean absorbency ratings for the three brands are equal.

36. The blocks correspond to the 10 dates on which the data were collected (Date) and the treatments correspond to the four cities (City).

Partial ANOVA output follows.

```
Analysis of Variance for Ozone Level

Source          DF        SS        MS        F        P

Date            9         903.02    100.34    4.55     0.001

City            3         160.08    53.36     2.42     0.088

Error           27        595.68    22.06

Total           39        1658.78

S = 4.69702     R-Sq = 64.09%       R-Sq(adj) = 48.13%
```

Because the p-value for City (.088) is greater than $\alpha = .05$, there is no significant difference in the mean ozone level among the four cities. But, if the level of significance was $\alpha = .10$, the difference would have been significant.

38.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	Method A	Method B	Method C
Sample Mean	90	84	81
Sample Variance	98.00	168.44	159.78

$\bar{\bar{x}} = (90 + 84 + 81)/3 = 85$

$\text{SSTR} = \sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2 = 10(90 - 85)^2 + 10(84 - 85)^2 + 10(81 - 85)^2 = 420$

$\text{MSTR} = \text{SSTR}/(k - 1) = 420/2 = 210$

$\text{SSE} = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 9(98.00) + 9(168.44) + 9(159.78) = 3{,}836$

$\text{MSE} = \text{SSE}/(n_T - k) = 3{,}836/(30 - 3) = 142.07$

$F = \text{MSTR}/\text{MSE} = 210/142.07 = 1.48$

Using F table (2 degrees of freedom numerator and 27 denominator), p-value is greater than .10.

Using Excel, the p-value corresponding to $F = 1.48$ is .2455.

Because p-value $> \alpha = .05$, we can not reject the null hypothesis that the means are equal.

40. a. Treatment Means:

$\bar{x}_{.1} = 22.8$ $\bar{x}_{.2} = 24.8$ $\bar{x}_{.3} = 25.80$

Block Means:

$\bar{x}_{1.} = 19.67$ $\bar{x}_{2.} = 25.67$ $\bar{x}_{3.} = 31$ $\bar{x}_{4.} = 23.67$ $\bar{x}_{5.} = 22.33$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Overall Mean:

$\bar{\bar{x}} = 367 / 15 = 24.47$

Step 1

$$\text{SST} = \sum_i \sum_j \left(x_{ij} - \bar{\bar{x}}\right)^2 = (18 - 24.47)^2 + (21 - 24.47)^2 + \cdots + (24 - 24.47)^2 =$$

253.73

Step 2

$$\text{SSTR} = b\sum_j \left(\bar{x}_{\cdot j} - \bar{\bar{x}}\right)^2 = 5\ [\ (22.8 - 24.47)^2 + (24.8 - 24.47)^2 + (25.8 - 24.47)^2\ ] =$$

23.33

Step 3

$$\text{SSBL} = k\sum_i \left(\bar{x}_{i\cdot} - \bar{\bar{x}}\right)^2 = 3\ [\ (19.67 - 24.47)^2 + (25.67 - 24.47)^2 + \cdots + (22.33 -$$

$24.47)^2\ ] = 217.02$

Step 4

SSE = SST – SSTR – SSBL = 253.73 – 23.33 – 217.02 = 13.38

Source of Variation	Sum of Squares	Degreesof Freedom	Mean Square	*F*	*p*-value
Treatment	23.33	2	11.67	6.99	.0175
Blocks	217.02	4	54.26	32.49	

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Error	13.38	8	1.67
Total	253.73	14	

Using F table (2 degrees of freedom numerator and 8 denominator), p-value is between .01 and .025.

Using Excel, the p-value corresponding to $F = 6.99$ is .0175.

Because p-value $\leq \alpha = .05$, we reject the null hypothesis that the mean miles per gallon ratings for the three brands of gasoline are equal.

b.

	I	II	III
Sample Mean	22.8	24.8	25.8
Sample Variance	21.2	9.2	27.2

$\bar{\bar{x}} = (22.8 + 24.8 + 25.8)/3 = 24.47$

$\text{SSTR} = \sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2 = 5(22.8 - 24.47)^2 + 5(24.8 - 24.47)^2 + 5(25.8 - 24.47)^2 =$

23.33

MSTR = SSTR /(k – 1) = 23.33 /2 = 11.67

$\text{SSE} = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 4(21.2) + 4(9.2) + 4(27.2) = 230.4$

MSE = SSE /(n_T – k) = 230.4 /(15 – 3) = 19.2

F = MSTR /MSE = 11.67 /19.2 = .61

Using F table (2 degrees of freedom numerator and 12 denominator), p-value

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

is greater than .10.

Using Excel, the p-value corresponding to $F = .61$ is .561.

Because p-value $> \alpha = .05$, we cannot reject the null hypothesis that the mean miles per gallon ratings for the three brands of gasoline are equal.

Thus, we must remove the block effect in order to detect a significant difference due to the brand of gasoline. The following table illustrates the relationship between the randomized block design and the completely randomized design.

Sum of Squares	Randomized Block Design	Completely Randomized Design
SST	253.73	253.73
SSTR	23.33	23.33
SSBL	217.02	does not exist
SSE	13.38	230.4

Note that SSE for the completely randomized design is the sum of SSBL (217.02) and SSE (13.38) for the randomized block design. This illustrates that the effect of blocking is to remove the block effect from the error sum of squares; thus, the estimate of σ^2 for the randomized block design is substantially smaller than it is for the completely randomized design.

42. The blocks correspond to the 12 golfers (Golfer) and the treatments correspond to the three designs (Design).

The ANOVA output follows.

```
ANOVA: Distance versus Design, Golfer
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Analysis of Variance for Distance

Source          DF     SS        MS        F         P

Design          2      3032.0    1516.0    12.89     0.000

Golfer          11     5003.3    454.8     3.87      0.003

Error           22     2586.7    117.6

Total           35     10622.0


S = 10.8432     R-Sq = 75.65%    R-Sq(adj) = 61.26%
```

Because the p-value for Design (.000) is less than $\alpha = .05$, there is a significant difference in the mean driving distance for the three designs.

44.

		Factor B		Factor B
		Manual	Automatic	Means
Factor A	Machine 1	$\bar{x}_{11} = 32$	$\bar{x}_{12} = 28$	$\bar{x}_{1.} = 30$
	Machine 2	$\bar{x}_{21} = 21$	$\bar{x}_{22} = 26$	$\bar{x}_{2.} = 23.5$
Factor B	Means	$\bar{x}_{.1} = 26.5$	$\bar{x}_{.2} = 27$	$\bar{\bar{x}} = 26.75$

Step 1

$$\text{SST} = \sum_i \sum_j \sum_k \left(x_{ijk} - \bar{\bar{x}}\right)^2 = (30 - 26.75)^2 + (34 - 26.75)^2 + \cdots + (28 - 26.75)^2 =$$

151.5

Step 2

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\text{SSA} = br\sum_i \left(\bar{x}_{i.} - \bar{\bar{x}}\right)^2 = 2\ (2)\ [\ (30 - 26.75)^2 + (23.5 - 26.75)^2\ ] = 84.5$$

Step 3

$$\text{SSB} = ar\sum_j \left(\bar{x}_{.j} - \bar{\bar{x}}\right)^2 = 2\ (2)\ [\ (26.5 - 26.75)^2 + (27 - 26.75)^2\ ] = 0.5$$

Step 4

$$\text{SSAB} = r\sum_i \sum_j \left(\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}}\right)^2 = 2[(32 - 30 - 26.5 + 26.75)^2 + \cdot\cdot\cdot + (26 - 23.5 - 27 + 26.75)^2] = 40.5$$

Step 5

SSE = SST – SSA – SSB – SSAB = 151.5 – 84.5 – 0.5 – 40.5 = 26

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*-value
Factor A	84.5	1	84.5	13	.0226
Factor B	.5	1	.5	.08	.7913
Interaction	40.5	1	40.5	6.23	.0671
Error	26	4	6.5		
Total	151.5	7			

Factor A: Using Excel, the *p*-value corresponding to $F = 13$ is .0226. Because *p*-value $\leq \alpha$ = .05, Factor A (machine) is significant.

Factor B: Using Excel, the *p*-value corresponding to $F = .08$ is .7913. Because

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

p-value $> \alpha = .05$, Factor B (loading system) is not significant.

Interaction: Using Excel, the p-value corresponding to $F = 6.23$ is .0671. Because p-value $> \alpha =$ .05, Interaction is not significant.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 14 Solutions

2. a.

b. There appears to be a negative linear relationship between x and y.

c. Many different straight lines can be drawn to provide a linear approximation of the relationship between x *and* y; in part d we will determine the equation of a straight line that "best" represents the relationship according to the least squares criterion.

d. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{55}{5} = 11 \quad \bar{y} = \frac{\Sigma y_i}{n} = \frac{175}{5} = 35$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -540 \quad \Sigma(x_i - \bar{x})^2 = 180$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{-540}{180} = -3$$

$$b_0 = \bar{y} - b_1\bar{x} = 35 - (-3)(11) = 68$$

$$\hat{y} = 68 - 3x$$

e. $\hat{y} = 68 - 3(10) = 38$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

4. a.

b. There appears to be a positive linear relationship between the percentage of women working in the five companies (x) and the percentage of management jobs held by women in that company (y).

c. Many different straight lines can be drawn to provide a linear approximation of the relationship between x and y; in part d we will determine the equation of a straight line that "best" represents the relationship according to the least squares criterion.

d. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{300}{5} = 60 \quad \bar{y} = \frac{\Sigma y_i}{n} = \frac{215}{5} = 43$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 624 \qquad \Sigma(x_i - \bar{x})^2 = 480$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{624}{480} = 1.3$$

$$b_0 = \bar{y} - b_1\bar{x} = 43 - 1.3(60) = -35$$

$$\hat{y} = -35 + 1.3x$$

e. $\hat{y} = -35 + 1.3x = -35 + 1.3(60) = 43\%$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

6. a.

b. The scatter diagram indicates a positive linear relationship between x = average number of passing yards per attempt and y = the percentage of games won by the team.

c. $\bar{x} = \Sigma x_i / n = 680/10 = 6.8 \quad \bar{y} = \Sigma y_i / n = 464/10 = 46.4$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 121.6 \quad \Sigma(x_i - \bar{x})^2 = 7.08$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{121.6}{7.08} = 17.1751$$

$$b_0 = \bar{y} - b_1\bar{x} = 46.4 - (17.1751)(6.8) = -70.391$$

$$\hat{y} = -70.391 + 17.1751x$$

d. The slope of the estimated regression line is approximately 17.2. So, for every increase of one yard in the average number of passes per attempt, the percentage of

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

games won by the team increases by 17.2%.

e. With an average number of passing yards per attempt of 6.2, the predicted percentage of games won is $\hat{y} = -70.391 + 17.175(6.2) = 36\%$. With a record of seven wins and nine losses, the percentage of Kansas City Chiefs wins is 43.8 or approximately 44%. Considering the small data size, the prediction made using the estimated regression equation is not too bad.

8. a.

b. The scatter diagram indicates a positive linear relationship between x = speed of execution rating and y = overall satisfaction rating for electronic trades.

c. $\bar{x} = \Sigma x_i / n = 36.3/11 = 3.3 \quad \bar{y} = \Sigma y_i / n = 35.2/11 = 3.2$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 2.4 \quad \Sigma(x_i - \bar{x})^2 = 2.6$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{2.4}{2.6} = .9077$$

$$b_0 = \bar{y} - b_1\bar{x} = 3.2 - (.9077)(3.3) = .2046$$

$$\hat{y} = .2046 + .9077x$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. The slope of the estimated regression line is approximately .9077. So, a one unit increase in the speed of execution rating will increase the overall satisfaction rating by approximately .9 points.

e. The average speed of execution rating for the other brokerage firms is 3.4. Using this as the new value of x for Zecco.com, we can use the estimated regression equation developed in part c to estimate the overall satisfaction rating corresponding to $x = 3.4$.

$$\hat{y} = .2046 + .9077x = .2046 + .9077(3.4) = 3.29$$

Thus, an estimate of the overall satisfaction rating when $x = 3.4$ is approximately 3.3.

10. a.

b. The scatter diagram indicates a positive linear relationship between x = age of wine and y = price of a 750-ml bottle of wine. In other words, the price of the wine increases with age.

c. $\bar{x} = \Sigma x_i / n = 317/10 = 31.7 \quad \bar{y} = \Sigma y_i / n = 2294/10 = 229.4$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 3838.2 \quad \Sigma(x_i - \bar{x})^2 = 552.1$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{3838.2}{552.1} = 6.952$$

$$b_0 = \bar{y} - b_1\bar{x} = 229.4 - (6.952)(31.7) = 9.0216$$

$$\hat{y} = 9.0216 + 6.952x$$

d. The slope of the estimated regression line is approximately 6.95. So, for every additional year of age, the price of the wine increases by $6.95.

12. a.

b. The scatter diagram indicates a somewhat positive linear relationship between x = percentage return of the S&P 500 and y = percentage return for Coca-Cola.

c. $\bar{x} = \Sigma x_i / n = 2.00/10 = .2 \quad \bar{y} = \Sigma y_i / n = 15/10 = 1.5$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 95 \quad \Sigma(x_i - \bar{x})^2 = 179.6$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{95}{179.6} = .5290$$

$$b_0 = \bar{y} - b_1\bar{x} = 1.5 - (.5290)(.2) = 1.3942$$

$$\hat{y} = 1.3942 + .529x$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. A 1-percent increase in the percentage return of the S&P 500 will result in a .529 increase in the percentage return for Coca-Cola.

e. The beta of .529 for Coca-Cola differs somewhat from the beta of .82 reported by Yahoo Finance. This is likely to the result of differences in the period over which the data were collected and the amount of data used to calculate the beta. Note: Yahoo uses the last five years of monthly returns to calculate beta.

14. a.

The scatter diagram indicates a negative linear relationship between x = distance to work and y = number of days absent.

b. $\bar{x} = \Sigma x_i / n = 90/10 = 9 \quad \bar{y} = \Sigma y_i / n = 50/10 = 5$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -95 \quad \Sigma(x_i - \bar{x})^2 = 276$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{-95}{276} = -.3442$$

$$b_0 = \bar{y} - b_1\bar{x} = 5 - (-.3442)(9) = 8.0978$$

$$\hat{y} = 8.0978 - .3442x$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. A prediction of the number of days absent is $\hat{y} = 8.0978 - .3442(5) = 6.4$ or approximately six days.

16. a. The estimated regression equation and the mean for the dependent variable are:

$$\hat{y}_i = 68 - 3x \qquad \bar{y} = 35$$

The sum of squares due to error and the total sum of squares are

$$\text{SSE} = \Sigma(y_i - \hat{y}_i)^2 = 230 \qquad \text{SST} = \Sigma(y_i - \bar{y})^2 = 1850$$

Thus, SSR = SST – SSE = 1850 – 230 = 1620

b. r^2 = SSR/SST = 1620/1850 = .876

The least squares line provided an excellent fit; 87.6% of the variability in y has been explained by the estimated regression equation.

c. $r_{xy} = -\sqrt{.876} = -.936$

Note: the sign for r is negative because the slope of the estimated regression equation is negative.

($b_1 = -3$)

18. a. $\bar{x} = \Sigma x_i / n = 600/6 = 100 \quad \bar{y} = \Sigma y_i / n = 330/6 = 55$

$$\text{SST} = \Sigma(y_i - \bar{y})^2 = 1800 \quad \text{SSE} = \Sigma(y_i - \hat{y}_i)^2 = 287.624$$

SSR = SST – SSE = 1800 – 287.624 = 1512.376

b. $r^2 = \dfrac{\text{SSR}}{\text{SST}} = \dfrac{1512.376}{1800} = .84$

c. $r = \sqrt{r^2} = \sqrt{.84} = .917$

20. a. $\bar{x} = \Sigma x_i / n = 160/10 = 16 \quad \bar{y} = \Sigma y_i / n = 55{,}500/10 = 5550$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -31,284 \quad \Sigma(x_i - \bar{x})^2 = 21.74$$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{-31,284}{21.74} = -1439$$

$$b_0 = \bar{y} - b_1\bar{x} = 5550 - (-1439)(16) = 28,574$$

$$\hat{y} = 28,574 - 1439x$$

b. SST = 52,120,800 SSE = 7,102,922.54

SSR = SST – SSE = 52,120,800 – 7,102,922.54 = 45,017,877

r^2 = SSR/SST = 45,017,877/52,120,800 = .864

The estimated regression equation provided a good fit.

c. $\hat{y} = 28,574 - 1439x = 28,574 - 1439(15) = 6989$

Thus, an estimate of the price for a bike that weighs 15 pounds is $6,989.

22. a. SSE = 1043.03

$$\bar{y} = \Sigma y_i / n = 462/6 = 77 \quad \text{SST} = \Sigma(y_i - \bar{y})^2 = 10,568$$

SSR = SST – SSR = 10,568 – 1043.03 = 9524.97

$$r^2 = \frac{\text{SSR}}{\text{SST}} = \frac{9524.97}{10,568} = .9013$$

b. The estimated regression equation provided an especially good fit; approximately 90% of the variability in the dependent variable was explained by the linear relationship between the two variables.

c. $r = \sqrt{r^2} = \sqrt{.9013} = .95$

This reflects a strong linear relationship between the two variables.

24. a. s^2 = MSE = SSE/(n – 2) = 230/3 = 76.6667

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $s = \sqrt{\text{MSE}} = \sqrt{76.6667} = 8.7560$

c. $\Sigma(x_i - \bar{x})^2 = 180$

$$s_{b_1} = \frac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \frac{8.7560}{\sqrt{180}} = 0.6526$$

d. $t = \dfrac{b_1}{s_{b_1}} = \dfrac{-3}{.653} = -4.59$

Using t table (3 degrees of freedom), area in tail is less than .01; p-value is less than .02.

Using Excel, the p-value corresponding to $t = -4.59$ is .0193.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$.

e. MSR = SSR/1 = 1,620

F = MSR/MSE = 1,620/76.6667 = 21.13

Using F table (1 degree of freedom numerator and 3 denominator), p-value is less than .025.

Using Excel, the p-value corresponding to $F = 21.13$ is .0193.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Regression	1620	1	1620	21.13	.0193
Error	230	3	76.6667		
Total	1850	4			

26. a. In the statement of exercise 18, $\hat{y} = 23.194 + .318x$.

In solving exercise 18, we found SSE = 287.624.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

s^2 = MSE = SSE/(n-2) =287.624 / 4 = 71.906 $s = \sqrt{\text{MSE}} = \sqrt{71.906} = 8.4797$

$$\sum (x - \bar{x})^2 = 14{,}950$$

$$s_{b_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{8.4797}{\sqrt{14{,}950}} = .0694$$

$$t = \frac{b_1}{s_{b_1}} = \frac{.318}{.0694} = 4.58$$

Using t table (4 degrees of freedom), area in tail is between .005 and .01.

p-value is between .01 and .02.

Using Excel, the p-value corresponding to t = 4.58 is .010.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$; there is a significant relationship between price and overall score.

b. In exercise 18, we found SSR = 1,512.376

MSR = SSR/1 = 1,512.376/1 = 1,512.376

F = MSR/MSE = 1,512.376/71.906 = 21.03

Using F table (1 degree of freedom numerator and 4 denominator), p-value is between .025 and .01.

Using Excel, the p-value corresponding to F = 11.74 is .010.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$

c.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Regression	1,512.376	1	1,512.376	21.03	.010

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Error	287.624	4	71.906
Total	1,800	5	

28. The sum of squares due to error and the total sum of squares are:

$$\text{SSE} = \Sigma(y_i - \hat{y}_i)^2 = 1.4379 \qquad \text{SST} = \Sigma(y_i - \bar{y})^2 = 3.5800$$

Thus,

SSR = SST - SSE = 3.5800 – 1.4379 = 2.1421

s^2 = MSE = SSE / (n - 2) = 1.4379 / 9 = .1598

$$s = \sqrt{\text{MSE}} = \sqrt{.1598} = .3997$$

We can use either the t test or F test to determine whether speed of execution and overall satisfaction are related.

We will first illustrate the use of the t test.

$$\Sigma(x_i - \bar{x})^2 = 2.6$$

$$s_{b_1} = \frac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \frac{.3997}{\sqrt{2.6}} = .2479$$

$$t = \frac{b_1}{s_{b_1}} = \frac{.9077}{.2479} = 3.66$$

Using t table (9 degrees of freedom), the area in the tail is less than .005; the p-value is less than .01.

Using Excel, the p-value corresponding to t = 3.66 is .000.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$.

Because we can reject H_0: $\beta_1 = 0$ we conclude that speed of execution and

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

overall satisfaction are related.

Next we illustrate the use of the F test.

MSR = SSR / 1 = 2.1421

F = MSR / MSE = 2.1421 / .1598 = 13.4

Using the F table (1 degree of freedom numerator and 9 denominator), the p-value is less than .01.

Using Excel, the p-value corresponding to F = 13.4 is .005.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$.

Because we can reject H_0: $\beta_1 = 0$ we conclude that speed of execution and overall satisfaction are related.

The ANOVA table follows.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Regression	2.1421	1	2.1421	13.4	.000
Error	1.4379	9	.1598		
Total	3.5800	10			

30. SSE $= \Sigma(y_i - \hat{y}_i)^2 =$ 1043.03 SST $= \Sigma(y_i - \bar{y})^2 =$ 10,568

Thus,

SSR = SST – SSE = 10,568 – 1043.03 = 9524.97

s^2 = MSE = SSE/(n-2) = 1043.03/4 = 260.7575

$s = \sqrt{260.7575} = 16.1480$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\Sigma(x_i - \bar{x})^2 = 56.655$$

$$s_{b_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}} = \frac{16.148}{\sqrt{56.655}} = 2.145$$

$$t = \frac{b_1}{s_{b_1}} = \frac{12.966}{2.145} = 6.045$$

Using t table (4 degrees of freedom), area in tail is less than .005.

p-value is less than .01.

Using Excel, the p-value corresponding to $t = 6.045$ is .004.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$.

There is a significant relationship between cars in service and annual revenue.

32. a. $s = 2.033$

$\bar{x} = 3 \quad \Sigma(x_i - \bar{x})^2 = 10$

$$s_{\hat{y}^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 2.033\sqrt{\frac{1}{5} + \frac{(4-3)^2}{10}} = 1.11$$

b. $\hat{y}^* = .2 + 2.6\,x^* = .2 + 2.6(4) = 10.6$

$\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$

$$10.6 \pm 3.182\ (1.11) = 10.6 \pm 3.53$$

or 7.07 to 14.13.

c. $$s_{\text{pred}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 2.033\sqrt{1 + \frac{1}{5} + \frac{(4-3)^2}{10}} = 2.32$$

d. $\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$

$$10.6 \pm 3.182\ (2.32) = 10.6 \pm 7.38$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

or 3.22 to 17.98.

34. $s = 6.5141$

$$\bar{x} = 10 \quad \Sigma(x_i - \bar{x})^2 = 190$$

$$s_{\hat{y}^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 6.5141\sqrt{\frac{1}{5} + \frac{(12-10)^2}{190}} = 3.0627$$

$$\hat{y}^* = 7.6 + .9x^* = 7.6 + .9(12) = 18.40$$

$$\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$$

18.40 $\pm$ 3.182(3.0627) = 18.40 $\pm$ 9.75

or 8.65 to 28.15.

$$s_{\text{pred}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 6.5141\sqrt{1 + \frac{1}{5} + \frac{(12-10)^2}{190}} = 7.1982$$

$$\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$$

18.40 $\pm$ 3.182(7.1982) = 18.40 $\pm$ 22.90

or -4.50 to 41.30.

The two intervals are different because there is more variability associated with predicting an individual value than there is a mean value.

36. a. $$s_{\hat{y}^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 4.6098\sqrt{\frac{1}{10} + \frac{(9-7)^2}{142}} = 1.6503$$

$$\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$$

$$\hat{y}^* = 80 + 4x^* = 80 + 4(9) = 116$$

116 $\pm$ 2.306(1.6503) = 116 $\pm$ 3.8056

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

or 112.19 to 119.81 ($112,190 to $119,810).

b. $$s_{\text{pred}} = s\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}} = 4.6098\sqrt{1+\frac{1}{10}+\frac{(9-7)^2}{142}} = 4.8963$$

$$\hat{y}^* \pm t_{\alpha/2}s_{\text{pred}}$$

$$116 \pm 2.306(4.8963) = 116 \pm 11.2909$$

or 104.71 to 127.29 ($104,710 to $127,290).

c. As expected, the prediction interval is much wider than the confidence interval. This is because it is more difficult to predict annual sales for one new salesperson with nine years of experience than it is to estimate the mean annual sales for all salespersons with nine years of experience.

38. a. $\hat{y}^* = 1{,}246.67 + 7.6(500) = \$5{,}046.67$

b. $\bar{x} = 575 \quad \Sigma(x_i-\bar{x})^2 = 93{,}750$

s^2 = MSE = 58,333.33 s = 241.52

$$s_{\text{pred}} = s\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}} = 241.52\sqrt{1+\frac{1}{6}+\frac{(500-575)^2}{93{,}750}} = 267.50$$

$$\hat{y}^* \pm t_{\alpha/2}s_{\text{pred}}$$

$$5{,}046.67 \pm 4.604\ (267.50) = 5{,}046.67 \pm 1231.57$$

or $3,815.10 to $6,278.24.

c. Based on one month, $6,000 is not out of line because $3,815.10 to $6,278.24 is the prediction interval. However, a sequence of five to seven months with consistently high costs should cause concern.

40. a. 9

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $\hat{y} = 20.0 + 7.21x$

c. 1.3626

d. SSE = SST – SSR = 51,984.1 – 41,587.3 = 10,396.8

MSE = 10,396.8/7 = 1,485.3

F = MSR / MSE = 41,587.3 /1,485.3 = 28.00

Using F table (1 degree of freedom numerator and 7 denominator), p-value is less than .01.

Using Excel, the p-value corresponding to F = 28.00 is .0011.

Because p-value $\leq \alpha$ = .05, we reject H_0: $\beta_1 = 0$.

Selling price is related to annual gross rents.

e. $\hat{y} = 20.0 + 7.21(50) = 380.5$ or \$380,500

42. a. $\hat{y} = 80.0 + 50.0x$

b. 30

c. F = MSR / MSE = 6828.6/82.1 = 83.17

Using the F table (1 degree of freedom numerator and 28 denominator), the p-value is less than .01.

Using Excel, the p-value corresponding to F = 83.17 is .000.

Because p-value $< \alpha$ = .05, we reject H_0: $B_1 = 0$.

Annual sales is related to the number of salespersons.

d. $\hat{y} = 80 + 50\ (12) = 680$ or \$680,000

44. a. Scatter diagram:

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. There appears to be a negative linear relationship between the two variables. The heavier helmets tend to be less expensive.

c. The output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	462761	462761	54.90	0.000
Error	16	134865	8429		
Total	17	597626			

Model Summary

S	R-sq	R-sq(adj)
91.8098	77.43%	76.02%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2044	226	9.03	0.000	
Weight	-28.35	3.83	-7.41	0.000	1.00

Regression Equation

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Price = 2044 - 28.35 Weight

Fits and Diagnostics for Unusual Observations

Obs          Price    Fit      Resid    Std Resid

7            900.0    655.2    244.8    3.03 R

R Large residual
```

d. Significant relationship: p-value = .000 < α = .05.

e. $r^2 = 0.774$; a good fit.

46. a. $\hat{y} = 2.32 + .64x$

b.

The assumption that the variance is the same for all values of x is questionable. The variance appears to increase for larger values of x.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

48. a. $\hat{y} = 80 + 4x$

b. The assumptions concerning the error term appear reasonable.

50. a. The output follows:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.620115502							
R Square	0.384543235							
Adjusted R Square	0.261451882							
Standard Error	12.61505192							
Observations	7							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	497.1594684	497.1594684	3.124047515	0.137389804			
Residual	5	795.6976744	159.1395349					
Total	6	1292.857143						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	66.10465116	32.06135318	2.061817254	0.09421512	-16.31168095	148.5209833	-16.31168095	148.5209833
x	0.402325581	0.22762441	1.767497529	0.137389804	-0.182801591	0.987452754	-0.182801591	0.987452754

The standardized residuals are calculated as follows (as in Table14.8):

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

x	y		x	x-xbar	(x-xbar)^2	(x-xbar)^2/ sum((x-xbr)^2)	h	s(y-yhat)	y-yhat	Std Res
135	145		135	-4	18	0.0060	0.1488	11.6385	24.5814	2.1121
110	100		110	-29	858	0.2792	0.4221	9.5900	-10.3605	-1.0803
130	120		130	-9	86	0.0281	0.1709	11.4864	1.5930	0.1387
145	120		145	6	33	0.0106	0.1535	11.6066	-4.4419	-0.3827
175	130		175	36	1276	0.4153	0.5581	8.3856	-6.5116	-0.7765
160	130		160	21	429	0.1397	0.2826	10.6852	-0.4767	-0.0446
120	110		120	-19	372	0.1211	0.2640	10.8229	-4.3837	-0.4050
		sum	975		3071.428571					
		xbar	139							
		n	7							
		S	12.6151							

The first data point (y= 145, x=135) has a large standardized residual.

b.

The standardized residual plot indicates that the observation $x = 135$, $y = 145$ may be an outlier; note that this observation has a standardized residual of 2.11.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. The scatter diagram follows.

The scatter diagram also indicates that the observation $x = 135$, $y = 145$ may be an outlier; the implication is that for simple linear regression an outlier can be identified by looking at the scatter diagram.

52. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The scatter diagram does indicate potential influential observations. For example, the 22.2% fund-raising expense for the American Cancer Society and the 16.9% fund-raising expense for the St. Jude Children's Research Hospital look like they may each have a large influence on the slope of the estimated regression line. And with a fund-raising expense of 2.6%, the percentage spent on programs and services by the Smithsonian Institution (73.7%) seems to be somewhat lower than would be expected; thus, this observeraton may need to be considered a possible outlier.

b. A portion of the output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	408.4	408.35	7.31	0.027
Error	8	446.9	55.86		
Total	9	855.2			

Model Summary

S	R-sq	R-sq(adj)
7.47387	47.75%	41.22%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	90.98	3.18	28.64	0.000	
Fundraising Expenses (%)	-0.917	0.339	-2.70	0.027	1.00

Regression Equation

Program Expenses (%) = 90.98 - 0.917 Fundraising Expenses (%)

Fits and Diagnostics for Unusual Observations

Obs	Program Expenses(%)	Fit	Resid	Std Resid		
3	73.70	88.60	-14.90	-2.13	R	
5	71.60	70.62	0.98	0.21		X

R Large residual

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
X Unusual X

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large leverage.
```

c. The slope of the estimated regression equation is –0.917. Thus, for every 1% increase in the amount spent on fund-raising, the percentage spent on program expenses will decrease by .917%; in other words, just a little less than 1%. The negative slope and value seem to make sense in the context of this problem situation.

d. The output in part b indicates that there are two unusual observations:

- Observation 3 (Smithsonian Institution) is an outlier because it has a large standardized residual.
- Observation 5 (American Cancer Society) is an influential observation because it has high leverage.

Although fund-raising expenses for the Smithsonian Institution are on the low side compared to most of the other supersized charities, the percentage spent on program expenses appears to be much lower than one would expect. It appears that the Smithsonian's administrative expenses are too high. But thinking about the expenses of running a large museum like the Smithsonian, the percentage spent on administrative expenses may not be unreasonable; in general, operating costs for a museum are higher than for some other types of organizations. The especially large value of fund-raising expenses for the American Cancer Society suggests that this obervation has a large influence on the estimated regression equation. The following output shows the results if this observation is deleted from the original data.

The regression equation is:

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
 Program Expenses (%) = 91.3 - 1.00 Fundraising Expenses (%)

Predictor                  Coef       SE Coef     T        P

Constant                   91.256     3.654       24.98    0.000

Fundraising Expenses (%)   -1.0026    0.5590      -1.79    0.116

 S = 7.96708    R-Sq = 31.5%    R-Sq(adj) = 21.7%
```

The *y*-intercept has changed slightly, but the slope has changed from –.917 to –1.00.

54. a.

The scatter diagram does indicate potential outliers or influential observations (or both). For example, the New York Yankees have both the highest revenue and value and appear to be an influential observation. The Los Angeles Dodgers have the second highest value and appear to be an outlier.

b. A portion of the output follows:

Regression Statistics

Multiple R	0.9062
R Square	0.8211

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Adjusted R Square	0.8148
Standard Error	165.6581
Observations	30

ANOVA

	df	SS	MS	F	Significance F
Regression	1	3527616.598	3527616.6	128.5453	5.616E-12
Residual	28	768392.7687	27442.599		
Total	29	4296009.367			

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	Coefficients	Standard Error	*t* Stat	*P*-value	Lower 95%	Upper 95%
Intercept	–601.4814	122.4288	–4.9129	3.519E-05	–852.2655	–350.6973
Revenue ($ millions)	5.9271	0.5228	11.3378	5.616E-12	4.8562	6.9979

Thus, the estimated regression equation that can be used to predict the team's value given the value of annual revenue is $\hat{y} = -601.4814 + 5.9271$ revenue.

c. The standard residual value for the Los Angeles Dodgers is 4.7 and should be treated as an outlier. To determine if the New York Yankees point is an influential observation, we can remove the observation and compute a new estimated regression equation. The results show that the estimated regresssion equation is $\hat{y} = -449.061 + 5.2122$ revenue. The following two scatter diagrams illustrate the small change in the estimated regression equation after removing the observation for the New York Yankees. These diagrams show that the effect of the New York Yankees observation on the regression results is not that dramatic.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Scatter Diagram Including the New York Yankees Observation

Scatter Diagram Excluding the New York Yankees Observation

56. The estimate of a mean value is an estimate of the average of all *y* values associated with the same *x*. The estimate of an individual *y* value is an estimate of only one of the *y* values associated with a particular *x*.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

58. a.

b. A portion of the output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	22146	22145	6 239.89	0.000
Error	13	1200	92.3		
Total	14	23346			

Model Summary

S	R-sq	R-sq(adj)
9.60811	94.86%	94.46%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-669	131	-5.12	0.000	
DJIA	0.1573	0.0102	15.49	0.000	1.00

Regression Equation

S&P = -669 + 0.1573 DJIA

c. Using the F test, the p-value corresponding to $F = 239.89$ is .000. Because the p-value $\leq \alpha = .05$, we reject $H_0 : \beta_1 = 0$; there is a significant relationship.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. With R-Sq = 94.9%, the estimated regression equation provided an excellent fit.

e. $\hat{y} = -669.0 + .15727(\text{DJIA}) = -669.0 + .15727(13{,}500) = 1454$

f. The DJIA is not that far beyond the range of the data. With the excellent fit provided by the estimated regression equation, we should not be too concerned about using the estimated regression equation to predict the S&P 500.

60. a.

The scatter diagram indicates a positive linear relationship between the two variables. Online universities with higher retention rates tend to have higher graduation rates.

b. The output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1224.3	1224.29	22.02	0.000
Error	27	1501.0	55.59		
Total	28	2725.3			

Model Summary

S	R-sq	R-sq(adj)

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
7.45610          44.92%      42.88%

 Coefficients

Term            Coef      SE Coef   T-Value   P-Value   VIF

Constant        25.42     3.75      6.79      0.000

RR(%)           0.2845    0.0606    4.69      0.000     1.00

 Regression Equation

 GR(%) = 25.42 + 0.2845 RR(%)

 Fits and Diagnostics for Unusual Observations

Obs           GR(%)    Fit Resid   Std Resid

2             25.00    39.93       -14.93      -2.04   R

3             28.00    26.56       1.44        0.22    X

 R Large residual

 X Unusual X

 R denotes an observation with a large standardized residual.

 X denotes an observation whose X value gives it large leverage.
```

c. Because the p-value = .000 < α =.05, the relationship is significant.

d. The estimated regression equation is able to explain 44.9% of the variability in the graduation rate based on the linear relationship with the retention rate. It is not a great fit, but the fit is reasonably good given the type of data.

e. In the output in part b, South University is identified as an observation with a large standardized residual. With a retention rate of 51% it does appear that the graduation rate of 25% is low compared to the results for other online universities. The president of South University should be concerned after looking at the data. Using the estimated regression equation, we estimate that the graduation rate at South University should be 25.4 + .285(51) = 40%.

f. In the output in part b, the University of Phoenix is identified as an observation whose

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

x value gives it large influence. With a retention rate of only 4%, the president of the University of Phoenix should be concerned after looking at the data.

62. a. The output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	25.130	25.130	11.33	0.028
Error	4	8.870	2.217		
Total	5	34.000			

Model Summary

S	R-sq	R-sq(adj)
1.48909	73.91%	67.39%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	22.17	1.65	13.42	0.000	
Speed	-0.1478	0.0439	-3.37	0.028	1.00

Regression Equation

Defects = 22.17 - 0.1478 Speed

Variable	Setting
Speed	50

Fit	SE Fit	95% CI	95% PI
14.7826	0.896327	(12.2940, 17.2712)	(9.95703, 19.6082)

b. Because the p-value corresponding to $F = 11.33 = .028 < \alpha = .05$, the relationship is significant.

c. $r^2 = .739$; a good fit. The least squares line explained 73.9% of the variability in the number of defects.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. Using the output in part a, the 95% confidence interval is 12.294 to 17.2712.

64. a. The output follows.

```
 Analysis of Variance

Source            DF     Adj SS     Adj MS    F-Value    P-Value

Regression        1      312050     312050    54.75      0.000

Error             8      45600      5700

Total             9      357650

 Model Summary

S                R-sq       R-sq(adj)

75.4983          87.25%     85.66%

 Coefficients

Term             Coef      SE Coef    T-Value   P-Value   VIF

Constant         220.0     58.5       3.76      0.006

Age              131.7     17.8       7.40      0.000     1.00

 Regression Equation

 Cost = 220.0 + 131.7 Age

Variable            Setting

Age                 4


 Fit            SE Fit     95% CI                95% PI

 746.667        29.7769    (678.001, 815.332)   (559.515, 933.818)
```

b. Because the p-value corresponding to $F = 54.75$ is $.000 < \alpha = .05$, we reject H_0: $\beta_1 = 0$. Maintenance cost and age of bus are related.

c. $r^2 = .873$. The least squares line provided a good fit.

d. The 95% prediction interval is 559.515 to 933.818 or \$559.52 to \$933.82

66. a. The output follows.

```
Analysis of Variance
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	50.26	50.255	7.08	0.029
Error	8	56.78	7.098		
Total	9	107.04			

Model Summary

S	R-sq	R-sq(adj)
2.66413	46.95%	40.32%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.275	0.900	0.31	0.768	
S&P 500	0.950	0.357	2.66	0.029	1.00

Regression Equation

Horizon = 0.275 + 0.950 S&P 500

The market beta for Horizon is $b_1 = .95$

b. Because the p-value = 0.029 is less than $\alpha = .05$, the relationship is significant.

c. $r^2 = .470$. The least squares line does not provide an especially good fit.

d. Xerox has higher risk with a market beta of 1.22.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

68. a.

b. There appears to be a negative relationship between the two variables that can be approximated by a straight line. An argument could also be made that the relationship is perhaps curvilinear because at some point a car has so many miles that its value becomes very small.

c. The output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	47.158	47.158	19.85	0.000
Error	17	40.389	2.376		
Total	18	87.547			

Model Summary

S	R-sq	R-sq(adj)
1.54138	53.87%	51.15%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	16.470	0.949	17.36	0.000	
Miles (1000s)	-0.0588	0.0132	-4.46	0.000	1.00

Regression Equation

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Price ($1000s) = 16.470 - 0.0588 Miles (1000s)
```

d. Significant relationship: p-value = 0.000 < α = .05.

e. r^2 = .5387; a reasonably good fit considering that the condition of the car is also an important factor in what the price is.

f. The slope of the estimated regression equation is –.0558. Thus, a one-unit increase in the value of x coincides with a decrease in the value of y equal to .0558. Because the data were recorded in thousands, every additional 1,000 miles on the car's odometer will result in a $55.80 decrease in the predicted price.

g. The predicted price for a 2007 Camry with 60,000 miles is $\hat{y}$ = 16.47 –.0588(60) = 12.942 or $12,942. Because of other factors such as condition and whether the seller is a private party or a dealer, this is probably not the price you would offer for the car. But it should be a good starting point in figuring out what to offer the seller.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 15 Solutions

2. a. Partial output follows:

Analysis of Variance

Source	DF Adj	SS Adj	MS	F-Value	P-Value
Regression	1	10021.2	10021.2	15.53	0.004
Error	8	5161.7	645.2		
Total	9	15182.9			

Model Summary

S	R-sq	R-sq(adj)
25.4009	66.00%	61.75%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	45.1	25.4	1.77	0.114	
X1	1.944	0.493	3.94	0.004	1.00

Regression Equation

Y = 45.1 + 1.944 X1

The estimated regression equation is $\hat{y} = 45.1 + 1.944x_1$.

An estimate of y when $x_1 = 47$ is $\hat{y} = 45.1 + 1.944(47) = 136.468$.

b. Partial output follows:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	3363.4	3363.4	2.28	0.170

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Error	8	11819.5	1477.4
Total	9	15182.9	

Model Summary

S	R-sq	R-sq(adj)
38.4374	22.15%	12.42%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	85.2	38.4	2.22	0.057	
X2	4.32	2.86	1.51	0.170	1.00

Regression Equation

Y = 85.2 + 4.32 X2

The estimated regression equation is

$$\hat{y} = 85.2 + 4.32x_2$$

An estimate of y when $x_2 = 10$ is

$$\hat{y} = 85.2 + 4.32(10) = 128.4$$

c. Partial output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	14052	7026.1	43.50	0.000
Error	7	1131	161.5		

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Total              9    1518

3

      Model Summary

S              R-sq     R-sq(adj)

12.7096        92.55%   90.42%

      Coefficients

Term             Coef    SE Coef  T-Value  P-Value   VIF

Constant         -18.4   18.0     -1.02    0.341

X1               2.010   0.247    8.13     0.000     1.00

X2               4.738   0.948    5.00     0.002     1.00

      Regression Equation

Y = -18.4 + 2.010 X1 + 4.738 X2
```

The estimated regression equation is

$$\hat{y} = -18.4 + 2.01x_1 + 4.738x_2$$

An estimate of y when $x_1 = 47$ and $x_2 = 10$ is

$$\hat{y} = -18.4 + 2.01(47) + 4.738(10) = 123.45$$

4. a. $\hat{y} = 25 + 10(15) + 8(10) = 255$. Sales estimate: \$255,000.

   b. Sales can be expected to increase by \$10 for every dollar increase in inventory investment when advertising expenditure is held constant. Sales can be expected to increase by \$8 for every dollar increase in advertising expenditure when inventory investment is held constant.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

6. a. Partial output follows:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	4814.3	4814.3	19.11	0.001
Error	14	3527.4	252.0		
Total	15	8341.7			

Model Summary

S	R-sq	R-sq(adj)
15.8732	57.71%	54.69%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-58.8	26.2	-2.25	0.041	
Yds/Att	16.39	3.75	4.37	0.001	1.00

Regression Equation

Win% = -58.8 + 16.39 Yds/Att

Fits and Diagnostics for Unusual Observations

Std	Obs	Win%	Fit	Resid	Resid	
14	81.30	47.77	33.53		2.19	R

R Large residual

b. Partial output follows:

Analysis of Variance

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	3653	3652.8	10.91	0.005
Error	14	4689	334.9		
Total	15	8342			

Model Summary

S	R-sq	R-sq(adj)
18.3008	43.79%	39.77%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	97.5	13.9	7.04	0.000	
Int/Att	-1600	485	-3.30	0.005	1.00

Regression Equation

Win% = 97.5 - 1600 Int/Att

Fits and Diagnostics for Unusual Observations

Obs	Win%	Fit	Resid	Std Resid	
8	12.50	55.93	-43.43	-2.45	R

RLarge residual

c. Partial output follows:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	6277	3138.5	19.76	0.000

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Error              13     2065      158.8

Total              15     8342

      Model Summary

S               R-sq       R-sq(adj)

12.6024         75.25%     71.44%

      Coefficients

Term  Coef  SE Coef     T-Value     P-Value     VIF

Constant        -5.8     27.1    -0.21   0.835

Yds/Att         12.95    3.19    4.06    0.001   1.15

Int/Att         -1084    357     -3.03   0.010   1.15

      Regression Equation

Win% = -5.8 + 12.95 Yds/Att - 1084 Int/Att

      Fits and Diagnostics for Unusual Observations

Obs          Win%    Fit     Resid     Std Resid

8            12.50   38.57   -26.07    -2.28       R

R     Large residual
```

d. The predicted value of Win% for the Kansas City Chiefs is

$$\text{Win\%} = -5.8 + 12.95(6.2) - 1{,}084(.036) = 35.47\%$$

With seven wins and nine loses, the Kansas City Chiefs won 43.75% of the games they played. The predicted value is somewhat lower than the actual value.

8. a. Partial output follows.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	60.202	60.2022	17.21	0.001
Error	18	62.963	3.4980		
Total	19	123.166			

Model Summary

S	R-sq	R-sq(adj)
1.87028	48.88%	46.04%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	69.30	4.80	14.44	0.000	
Shore Excursions	0.2348	0.0566	4.15	0.001	1.00

Regression Equation

Overall = 69.30 + 0.2348 Shore Excursions

b. Partial output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	90.95	45.477	24.00	0.000
Error	17	32.21	1.895		
Total	19	123.17			

Model Summary

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

S	R-sq	R-sq(adj)
1.37650	73.85%	70.77%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	45.18	6.95	6.50	0.000	
Shore Excursions	0.2529	0.0419	6.04	0.000	1.01
Food/Dining	0.2482	0.0616	4.03	0.001	1.01

Regression Equation

Overall = 45.18 + 0.2529 Shore Excursions + 0.2482 Food/Dining

c. $\hat{y} = 45.18 + .2529(\text{Shore Excursions}) + .2482(\text{Food/Dining}) = 45.18 + .2529(80) + .2482(90) = 87.75$

Thus, an estimate of the overall score is approximately 88.

10. a. Partial output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.047263	0.047263	13.01	0.002
Error	18	0.065392	0.003633		
Total	19	0.112655			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0602733	41.95%	38.73%	29.38%

Coefficients

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.6758	0.0631	10.71	0.000	
SO/IP	-0.2838	0.0787	-3.61	0.002	1.00

Regression Equation

R/IP = 0.6758 - 0.2838 SO/IP

b. Partial output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.02887	0.028874	6.20	0.023
Error	18	0.08378	0.004655		
Total	19	0.11265			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0682239	25.63%	21.50%	14.87%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.3081	0.0604	5.10	0.000	
HR/IP	1.347	0.541	2.49	0.023	1.00

Regression Equation

R/IP = 0.3081 + 1.347 HR/IP

c. Partial output follows.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	0.06348	0.031738	10.97	0.001
Error	17	0.04918	0.002893		
Total	19	0.11266			

Model Summary

S	R-sq	R-sq(adj)
0.0537850	56.35%	51.21%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.5365	0.0814	6.59	0.000	
SO/IP	-0.2483	0.0718	-3.46	0.003	1.05
HR/IP	1.032	0.436	2.37	0.030	1.05

Regression Equation

R/IP = 0.5365 - 0.2483 SO/IP + 1.032 HR/IP

d. Using the estimated regression equation in part c we obtain

$$R/IP = 0.5365 - 0.2483\ SO/IP + 1.032\ HR/IP = 0.5365 - 0.2483(.91) + 1.032(.16) = .4757$$

The predicted value for R/IP was less than the actual value.

e. This suggestion does not make sense. If a pitcher gives up more runs per inning pitched, then earned run average also has to increase. For these data, the sample correlation coefficient between ERA and R/IP is .964. The following partial output

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

shows the results for part c using ERA as the dependent variable.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	5.174	2.5870	14.17	0.000
SO/IP	1	1.905	1.9052	10.44	0.005
HR/IP	1	2.190	2.1901	12.00	0.003
Error	17	3.103	0.1825		
Total	19	8.276			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.427204	62.51%	58.10%	50.29%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.878	0.647	6.00	0.000	
SO/IP	-1.843	0.570	-3.23	0.005	1.05
HR/IP	11.99	3.46	3.46	0.003	1.05

Regression Equation

ERA = 3.878 - 1.843 SO/IP + 11.99 HR/IP

12. a. $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{14,052.2}{15,182.9} = .926$

b. $R_a^2 = 1-(1-R^2)\frac{n-1}{n-p-1} = 1-(1-.926)\frac{10-1}{10-2-1} = .905$

c. Yes; after adjusting for the number of independent variables in the model, we see that

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

90.5% of the variability in y has been accounted for.

14. a. $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{12,000}{16,000} = .75$

b. $R_a^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} = 1 - .25\frac{9}{7} = .68$

c. The adjusted coefficient of determination shows that 68% of the variability has been explained by the two independent variables; thus, we conclude that the model does not explain a large amount of variability.

16. a. $r^2 = .577$. Thus, the average number of passing yards per attempt is able to explain 57.7% of the variability in the percentage of games won. Considering the nature of the data and all the other factors that might be related to the number of games won, this is not too bad a fit.

b. The value of the coefficient of determination increased to $R^2 = .752$, and the adjusted coefficient of determination is $R_a^2 = .714$. Thus, using both independent variables provides a much better fit.

18. a. A portion of the output follows:

```
Model Summary

S                R-sq       R-sq(adj)   R-sq(pred)

0.0537850        56.35%     51.21%      41.25%
```

The output in part c of exercise 10 shows that R-sq = .5635 and R-sq(adj) = .5121.

b. The fit is not great, but considering the nature of the data being able to explain slightly more than 50% of the variability in the number of runs given up per inning pitched using just two independent variables is not too bad.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. Partial output using ERA as the dependent variable follows.

```
      Analysis of Variance

Source            DF      Adj SS     Adj MS     F-Value   P-Value

Regression        2       5.174      2.5870     14.17     0.000

Error             17      3.103      0.1825

Total             19      8.276

      Model Summary

S               R-sq      R-sq(adj)

0.427204        62.51%    58.10%

      Coefficients

Term            Coef      SE Coef    T-Value    P-Value    VIF

Constant        3.878     0.647      6.00       0.000

SO/IP           -1.843    0.570      -3.23      0.005      1.05

HR/IP           11.99     3.46       3.46       0.003      1.05

      Regression Equation

ERA = 3.878 - 1.843 SO/IP + 11.99 HR/IP
```

The output shows that R-sq = .6251 and R-sq(adj) = .5810

Approximately 60% of the variability in the ERA can be explained by the linear effect of HR/IP and SO/IP. This is not too bad considering the complexity of predicting pitching performance.

20. A portion of the output follows.

```
      Analysis of Variance
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	14052	7026.1	43.50	0.000
Error	7	1131	161.5		
Total	9	15183			

Model Summary

S	R-sq	R-sq(adj)
12.7096	92.55%	90.42%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-18.4	18.0	-1.02	0.341	
X1	2.010	0.247	8.13	0.000	1.00
X2	4.738	0.948	5.00	0.002	1.00

Regression Equation

Y = -18.4 + 2.010 X1 + 4.738 X2

a. Because the p-value corresponding to $F = 43.50$ is $.000 < \alpha = .05$, we reject H_0: $\beta_1 = \beta_2 = 0$; there is a significant relationship.

b. Because the p-value corresponding to $t = 8.13$ is $.000 < \alpha = .05$, we reject H_0: $\beta_1 = \beta_2 = 0$; β_1 is significant.

c. Because the p-value corresponding to $t = 5.00$ is $.002 < \alpha = .05$, we reject H_0: $\beta_1 = \beta_2 = 0$; β_2 is significant.

22. a. SSE = SST – SSR = 16,000 – 12,000 = 4,000

$$s^2 = \frac{\text{SSE}}{n - p - 1} = \frac{4000}{7} = 571.43$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\text{MSR} = \frac{\text{SSR}}{p} = \frac{12000}{2} = 6000$$

b. F = MSR/MSE = 6,000/571.43 = 10.50

Using F table (two degrees of freedom numerator and seven denominator), p-value is less than .01.

Actual p-value = .008

Because p-value $\leq \alpha$, we reject H_0. There is a significant relationship among the variables.

24. a. Partial output follows:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	6179	3089.6	13.18	0.000
Error	29	6797	234.4		
Total	31	12976			

Model Summary

S	R-sq	R-sq(adj)
15.3096	47.62%	44.01%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	60.5	28.4	2.14	0.041	
OffPassYds/G	0.3186	0.0626	5.09	0.000	1.21
DefYds/G	-0.2413	0.0893	-2.70	0.011	1.21

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Regression Equation

Win% = 60.5 + 0.3186 OffPassYds/G - 0.2413 DefYds/G
```

$$\hat{y} = 60.5 + 0.3186 \text{ OffPassYds/G} - 0.2413 \text{ DefYds/G}$$

b. Because the p-value for the F test = .000 < α = .05, there is a significant relationship.

c. For OffPassYds/G: Because the p-value = .000 < α = .05, OffPassYds/G is significant.

For DefYds/G: Because the p-value = .011 < = .05, DefYds/G is significant.

26. The partial output from part c of exercise 10 follows.

```
Analysis of Variance

Source          DF     Adj SS     Adj MS      F-Value    P-Value

Regression      2      0.06348    0.031738    10.97      0.001

Error           17     0.04918    0.002893

Total           19     0.11266

Model Summary

S               R-sq       R-sq(adj)

0.0537850       56.35%     51.21%

Coefficients

Term            Coef     SE Coef    T-Value    P-Value    VIF

Constant        0.5365   0.0814     6.59       0.000

SO/IP           -0.2483  0.0718     -3.46      0.003      1.05

HR/IP           1.032    0.436      2.37       0.030      1.05

Regression Equation
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
R/IP = 0.5365 - 0.2483 SO/IP + 1.032 HR/IP
```

a. The *p*-value associated with $F = 10.97$ is .001. Because the *p*-value < .05, there is a significant overall relationship.

b. For SO/IP, the *p*-value associated with $t = -3.46$ is .003. Because the *p*-value < .05, SO/IP is significant. For HR/IP, the *p*-value associated with $t = 2.37$ is .030. Because the *p*-value < .05, HR/IP is also significant.

28. Partial output follows:

```
      Regression Equation

Y = -18.4 + 2.010 X1 + 4.738 X2

      Variable Setting

X1         47

X2         10
```

a. Using JMP, the 95% confidence interval is 112.13 to 134.84.

b. Using JMP, the 95% prediction interval is 91.36 to 155.62.

30 a. Partial output follows:

```
      Analysis of Variance

Source           DF   Seq SS  Contribution    Adj SS  Adj MS   F-Value   P-
                                                                         Value

Regression       2    6179    47.62%          6179    3089.6   13.18     0.000

Error            29   6797    52.38%          6797    234.4

Total            31   12976   100.00%

      Model Summary
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
S                R-sq      R-sq(adj)   PRESS

15.3096          47.62%    44.01%      8636.13

Coefficients

Term             Coef      SE Coef   95% CI       T-Value   P-Value   VIF

Constant         60.5405   28.4      ( 2.5,       2.14      0.041
                                      118.5)

OffPassYds/G     0.3186    0.0626    ( 0.1907,    5.09      0.000     1.21
                                      0.4466)

DefYds/G         -0.2413   0.0893    (-0.4239,    -2.70     0.011     1.21
                                      -0.0587)

      Regression Equation

Win% = 60.5405 + 0.3186 OffPassYds/G - 0.2413 DefYds/G
```

The estimated regression equation is

$$\hat{y} = 60.5405 + 0.3186 \text{ OffPassYds/G} - 0.2413 \text{ DefYds/G}$$

For OffPassYds/G = 225 and DefYds/G = 300, the predicted value of the percentage of games won is

$$\hat{y} = 60.5405 + 0.3186(225) - 0.2413(300) = 59.8$$

b. Using JMP, the 95% prediction interval is 26.96 to 92.70.

32 a. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ where

$x_2 = 0$ if level 1 and 1 if level 2

b. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 (0) = \beta_0 + \beta_1 x_1$

c. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 (1) = \beta_0 + \beta_1 x_1 + \beta_2$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. $\beta_2 = E(y \mid \text{level } 2) - E(y \mid \text{level } 1)$

β_1 is the change in E(y) for a one-unit change in x_1 holding x_2 constant.

34 a. $15,300

b. Estimate of sales = 10.1 – 4.2(2) + 6.8(8) + 15.3(0) = 56.1 or $56,100

c. Estimate of sales = 10.1 – 4.2(1) + 6.8(3) + 15.3(1) = 41.6 or $41,600

36 a. The output follows:

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	3	9.4305	90.02%	9.43049	3.14350	18.04	0.002
Error	6	1.0455	9.98%	1.04551	0.17425		
Total	9	10.4760	100.00%				

Model Summary

S	R-sq	R-sq(adj)	PRESS
0.417434	90.02%	85.03%	3.38309

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	1.860	0.729	(0.077, 3.643)	2.55	0.043	
Months	0.2914	0.0836	(0.0869, 0.4960)	3.49	0.013	2.43
Type	1.102	0.303	(0.360, 1.845)	3.63	0.011	1.27
Person	-0.609	0.388	(-1.558, 0.340)	-1.57	0.167	2.16

Regression Equation

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Repair Time_(hours) = 1.860 + 0.2914 Months + 1.102 Type - 0.609 Person
```

b. Because the p-value corresponding to $F = 18.04$ is $.002 < \alpha = .05$, the overall model is statistically significant.

c. The p-value corresponding to $t = -1.57$ is $.167 > \alpha = .05$; thus, the addition of Person is not statistically significant. Person is highly correlated with Months (the sample correlation coefficient is –.691); thus, once the effect of Months has been accounted for, Person will not add much to the model.

38 a. Partial output follows:

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	3	3660.7	87.35%	3660.7	1220.25	36.82	0.000
Error	16	530.2	12.65%	530.2	33.14		
Total	19	4190.9	100.00%				

Model Summary

S	R-sq	R-sq(adj)	PRESS
5.75657	87.35%	84.98%	799.476

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	-91.8	15.2	(-124.0, -59.5)	-6.03	0.000	
Age	1.077	0.166	(0.725, 1.429)	6.49	0.000	1.46
Pressure	0.2518	0.0452	(0.1559, 0.3477)	5.57	0.000	1.25

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Smokers	8.74	3.00	(2.38, 15.10)	2.91	0.010	1.36

```
Regression Equation

Risk = -91.8 + 1.077 Age + 0.2518 Pressure + 8.74 Smokers
```

b. Because the p-value corresponding to $t = 2.91$ is $.010 < \alpha = .05$, smoking is a significant factor.

c. Partial output follows.

```
Regression Equation

Risk = -91.8 + 1.077 Age + 0.2518 Pressure + 8.74 Smokers

Variable Setting

Age          68

Pressure     175

Smokers 1

Fit          SE Fit    95% CI              95% PI

34.2661      1.99785   (30.0309, 38.5014)  (21.3487, 47.1836)
```

The point estimate is 34.2661; the 95% prediction interval is 21.3487 to 47.1836. Thus, the probability of a stroke (.213487 to .471836 at the 95% confidence level) appears to be quite high. The physician would probably recommend that Art quit smoking and begin some type of treatment designed to reduce his blood pressure.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

40 a. Partial output follows:

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1934.42	98.76%	1934.42	1934.42	238.03	0.001
Error	3	24.38	1.24%	24.38	8.13		
Total	4	1958.80	100.00%				

Model Summary

S	R-sq	R-sq(adj)	PRESS
2.85073	98.76%	98.34%	243.374

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	-53.28	5.79	(-71.69, -34.87)	-9.21	0.003	
x	3.110	0.202	(2.468, 3.752)	15.43	0.001	1.00

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Regression Equation

y = -53.28 + 3.110 x
```

b. We obtained the following values:

x_i	y_i	Studentized Deleted Residual
22	12	–1.94
24	21	–.12
26	31	1.79
28	35	.40
40	70	–1.90

$t_{.025} = 4.303$ ($n - p - 2 = 5 - 1 - 2 = 2$ degrees of freedom)

Because none of the studentized deleted residuals are less than –4.303 or greater than 4.303, none of the observations can be classified as an outlier.

c. We obtained the following values:

x_i	y_i	h_i
22	12	.38
24	21	.28
26	31	.22
28	35	.20
40	70	.92

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The critical value is $\frac{3(p+1)}{n} = \frac{3(1+1)}{5} = 1.2$.

Because none of the values exceed 1.2, we conclude that there are no influential observations in the data.

d. We obtained the following values:

x_i	y_i	D_i
22	12	.60
24	21	.00
26	31	.26
28	35	.03
40	70	11.09

Because $D_5 = 11.09 > 1$ (rule of thumb critical value), we conclude that the fifth observation is influential.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

42 a. Partial output follows:

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	2	915.66	91.94%	915.66	457.828	74.12	0.000
Error	13	80.30	8.06%	80.30	6.177		
Total	15	995.95	100.00%				

Model Summary

S	R-sq	R-sq(adj)	PRESS
2.48532	91.94%	90.70%	142.697

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	71.33	2.25	(66.47, 76.18)	31.73	0.000	
Price($1000s)	0.1072	0.0392	(0.0225, 0.1918)	2.74	0.017	1.30
Horsepower	0.08450	0.00931	(0.06439, 0.10460)	9.08	0.000	1.30

Regression Equation

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Speed at 1/4 mile (mph) = 71.33 + 0.1072 Price ($1000s) + 0.08450 Horsepower r
```

Fits and Diagnostics for All Observations

Obs	Speed at 1/4 mile (mph)	Fit	SE Fit	95% CI	Resid	Std Resid	Del Resid	HI
1	90.70	90.49	0.89	(88.56, 92.41)	0.21	0.09	0.09	0.128399
2	108.00	105.88	2.01	(101.55, 110.22)	2.12	1.45	1.52	0.652221
3	93.20	91.68	0.88	(89.78, 93.59)	1.52	0.65	0.64	0.126054
4	103.20	99.76	1.14	(97.30, 102.23)	3.44	1.56	1.66	0.210277
5	102.10	105.85	0.94	(103.81, 107.89)	-3.75	-1.63	-1.76	0.144325
6	116.20	116.83	1.70	(113.16, 120.50)	-0.63	-0.35	-0.33	0.467435
7	91.70	92.83	0.89	(90.90, 94.76)	-1.13	-0.49	-0.47	0.129294
8	89.70	90.63	0.87	(88.75, 92.51)	-0.93	-0.40	-0.39	0.122422
9	93.00	94.32	0.78	(92.63, 96.00)	-1.32	-0.56	-0.54	0.098752
10	92.30	91.54	0.94	(89.52, 93.56)	0.76	0.33	0.32	0.141539
11	99.00	103.46	0.80	(101.73, 105.19)	-4.46	-1.90	-2.14	0.104138

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

12	84.60	87.11	1.06	(84.81, 89.41)	-2.51	-1.12	-1.13	0.183349
13	103.20	100.08	1.05	(97.80, 102.35)	3.12	1.39	1.44	0.180096
14	93.20	93.20	0.87	(91.31, 95.08)	0.00	0.00	0.00	0.123771
15	105.00	102.76	0.86	(100.91, 104.62)	2.24	0.96	0.96	0.119325
16	97.00	95.68	0.65	(94.27, 97.08)	1.32	0.55	0.54	0.068602

Obs	Cook's D	DFITS	
1	0.00	0.03363	
2	1.31	2.07558	X
3	0.02	0.24241	
4	0.21	0.85470	
5	0.15	-0.72273	
6	0.03	-0.31262	
7	0.01	-0.18152	

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

8	0.01	-0.14467
9	0.01	-0.17970
10	0.01	0.12868
11	0.14	-0.73025
12	0.09	-0.53557
13	0.14	0.67723
14	0.00	0.00071
15	0.04	0.35229
16	0.01	0.14555

X Unusual X

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. The standardized residual plot follows. There appears to be a highly unusual trend in the standardized residuals.

c. The output shown in part a did not identify any observations with a large standardized residual; thus, there do not appear to be any outliers in the data.

d. The output shown in part a identifies observation 2 as an influential observation.

44 a. $E(y) = \dfrac{e^{\beta_0 + \beta x}}{1 + e^{\beta_0 + \beta x}}$

b. It is an estimate of the probability that a customer who does not have a Simmons credit card will make a purchase.

c. A portion of the binary logistic regression output follows:

Coefficients

Term	Coef	SE Coef	VIF
Constant	-0.944	0.315	
Card	1.025	0.423	1.00

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Odds Ratios for Continuous Predictors
```

Odds	Ratio	95% CI
Card	2.7857	(1.2147, 6.3886)

```
Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))

Y' = -0.944 + 1.025 Card
```

Thus, the estimated logit is $\hat{g}(x) = -0.944 + 1.025x$.

d. For customers who do not have a Simmons credit card ($x = 0$):

$$\hat{g}(0) = -0.945 + 1.25(0) = 0.945$$

and

$$\hat{y} = \frac{e^{\hat{g}(0)}}{1+e^{\hat{g}(0)}} = \frac{e^{-0.945}}{1+e^{-0.945}} = \frac{0.38868}{1+0.38868} = 0.279$$

For customers who have a Simmons credit card ($x = 1$)

$$\hat{g}(1) = -0.945 + 1.025(1) = 0.0800$$

and

$$\hat{y} = \frac{e^{\hat{g}(1)}}{1+e^{\hat{g}(1)}} = \frac{e^{0.08}}{1+e^{0.08}} = \frac{1.0833}{1+1.0833} = 0.52$$

e. Using the output shown in part c, the estimated odds ratio is 2.7857. We can conclude that the estimated odds of making a purchase for customers who have a Simmons credit card are 2.7857 times greater than the estimated odds of making a purchase for customers who do not have a Simmons credit card.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

46. a. $E(y) = \frac{e^{\beta_0 + \beta x}}{1 + e^{\beta_0 + \beta x}}$

b. A portion of the binary logistic regression output follows:

Coefficients

Term	Coef	SE Coef	VIF
Constant	-2.633	0.799	
Balance	0.2202	0.0900	1.00

Odds Ratios for Continuous Predictors

Odds	Ratio	95% CI
Balance	1.2463	(1.0447, 1.4868)

Regression Equation

```
P(1) = exp(Y')/(1 + exp(Y'))

Y' = -2.633 + 0.2202 Balance
```

Thus, the estimated logistic regression equation is

$$E(y) = \frac{e^{-2.633 + 0.2202x}}{1 + e^{-2.633 + 0.2202x}}$$

c. A portion of the binary logistic regression output follows:

Deviance Table

Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	1	9.460	9.460	9.46	0.002
Error	48	51.626	1.076		
Total	49	61.086			

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Significant result: the p-value corresponding to the x^2 test statistic is 0.002.

d. For an average monthly balance of \$1,000, $x = 10$.

$$E(y) = \frac{e^{-2.633+0.2202x}}{1+e^{-2.633+0.2202x}} = \frac{e^{-2.633+0.2202(10)}}{1+e^{-2.633+0.2202(10)}} = \frac{e^{-0.431}}{1+e^{-0.431}} = \frac{0.6499}{1.6499} = 0.3939$$

Thus, an estimate of the probability that customers with an average monthly balance of \$1,000 will sign up for direct payroll deposit is 0.3939.

e. Repeating the calculations in part d using various values for x, a value of $x = 12$ or an average monthly balance of approximately \$1,200 is required to achieve this level of probability.

f. Using the output shown in part b, the estimated odds ratio is 1.2463. Because values of x are measured in hundreds of dollars, the estimated odds of signing up for payroll direct deposit for customers who have an average monthly balance of \$600 is 1.2463 times greater than the estimated odds of signing up for payroll direct deposit for customers who have an average monthly balance of \$500. Moreover, this interpretation is true for any 100-dollar increment in the average monthly balance.

48 a. $E(y) = \dfrac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}$

b. A portion of the binary logistic regression output follows:

Coefficients

Term	Coef	SE Coef	VIF
Constant	-39.5	12.5	
Wet	3.37	1.26	1.03
Noise	1.816	0.831	1.03

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
      Odds Ratios for Continuous Predictors

Odds              Ratio        95% CI

Wet               29.2095      (2.4521, 347.9413)

Noise             6.1489       (1.2059, 31.3540)

      Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))

Y' = -39.5 + 3.37 Wet + 1.816 Noise
```

Thus, the estimated logit is $\hat{g}(x) = -39.5 + 3.37\text{Wet} + 1.816\text{Noise}$

c. For tires that have a Wet performance rating of 8 and a Noise performance rating of 8:

$$\hat{g}(x) = -39.5 + 3.37\text{Wet} + 1.816\text{Noise}$$

$$\hat{g}(x) = -39.5 + 3.37(8) + 1.816(8) = 1.988$$

$$\hat{y} = \frac{e^{1.988}}{1+e^{1.988}} = \frac{7.30092}{1+7.30092} = 0.8795$$

The probability that a customer will probably or definitely purchase a particular tire again with these performance characteristics is .8795.

d. For tires that have a Wet performance rating of 7 and a Noise performance rating of 7:

$$\hat{g}(x) = -39.5 + 3.37\text{Wet} + 1.816\text{Noise}$$

$$\hat{g}(x) = -39.5 + 3.37(7) + 1.816(7) = -3.198$$

$$\hat{y} = \frac{e^{-3.198}}{1+e^{-3.198}} = \frac{.04084}{1+.04084} = 0.0392$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The probability that a customer will probably or definitely purchase a particular tire again with these performance characteristics is .0392.

e. Wet and Noise performance ratings of 7 are both considered excellent performance ratings using the Tire Rack performance scale. Nonetheless, the probability that the customer will repurchase a tire with these characteristics is extremely low. But a one-point increase in both ratings increases the probability to .8795, so achieving the highest possible levels of performance is essential if the manufacturer wants to have the greatest chance of having an existing customer buy its tire again.

50 a. Job satisfaction can be expected to decrease by 8.69 units with a one-unit increase in length of service if the wage rate does not change. A dollar increase in the wage rate is associated with a 13.5-point increase in the job satisfaction score when the length of service does not change.

b. $\hat{y} = 14.4 - 8.69(4) + 13.5(13) = 155.14$

52 a. The computer output with the missing values filled in follows.

```
Regression Equation

y = -1.41 + 0.0235 X1 + 0.00486 X2

Coefficients
```

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-1.41	0.4848	-2.91	0.023	
x1	0.0235	0.0087	2.71	0.030	1.54
x2	0.00486	0.0011	4.51	0.003	1.54

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

S	R-sq	R-sq(adj)	R-sq(pred)
0.1298	93.73%	91.93%	74.53%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	1.76209	0.88105	52.29	0.000
Error	7	0.11794	0.01685		
Total	9	1.88003			

b. $F_{.05} = 4.74$

$F = 52.29 > F_{.05}$; significant relationship

Actual p-value = .000

Because p-value $\leq \alpha = .05$, the overall relationship is significant.

c. For β_1: p-value = .030; reject H_0: $\beta_1 = 0$

for β_2: p-value = .003; reject H_0: $\beta_2 = 0$

d. $R^2 = \frac{\text{SSR}}{\text{SST}} = .9373$

$R_a^2 = 1 - (1 - .9373)\frac{9}{7} = .9193$ Good fit

54. a. A portion of the output follows

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	60.787	60.7866	85.93	0.000

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Error	16	11.318	0.7074
Total	17	72.105	

Model Summary

S	R-sq	R-sq(adj)
0.841071	84.30%	83.32%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-7.52	1.47	-5.13	0.000	
Steering	1.815	0.196	9.27	0.000	1.00

Regression Equation

Buy Again = -7.52 + 1.815 Steering

Because the p-value = .000 < α = .05, there is a significant relationship.

b. The estimated regression equation provided a good fit; 84.3 % of the variability in the Buy Again rating was explained by the linear effect of the Steering rating.

c. A portion of the output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	67.185	33.5924	102.41	0.000
Error	15	4.920	0.3280		
Total	17	72.105			

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Model Summary

S	R-sq	R-sq(adj)
0.572723	93.18%	92.27%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-5.39	1.11	-4.86	0.000	
Steering	0.690	0.288	2.40	0.030	4.65
Tread Wear	0.911	0.206	4.42	0.001	4.65

Regression Equation

Buy Again = -5.39 + 0.690 Steering + 0.911 Tread Wear

d. For the Treadwear independent variable, the *p*-value = .001 < α = .05; thus, the addition of Treadwear is significant.

56. a. Type of Fund is a categorical variable with three levels. Let FundDE = 1 for a domestic equity fund and FundIE = 1 for an international fund. The output follows:

Regression Statistics

Multiple R	0.7838
R Square	0.6144
Adjusted R Square	0.5960
Standard Error	5.5978
Observations	45

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

ANOVA

	Df	SS	MS	*F*	Significance *F*
Regression	2	2096.8489	1048.4245	33.4584	2.03818E-09
Residual	42	1316.0771	31.3352		
Total	44	3412.9260			

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	Coefficients	Standard Error	*t* Stat	*P*-Value	Lower 95%	Upper 95%
Intercept	4.9090	1.7702	2.7732	0.0082	1.3366	8.4814
FundDE	10.4658	2.0722	5.0505	9.033E-06	6.2839	14.6477
FundIE	21.6823	2.6553	8.1658	3.288E-10	16.3237	27.0408

$\hat{y}$ = 4.9090+ 10.4658 FundDE + 21.6823 FundIE

Because the *p*-value corresponding to $F = 33.4584$ is $.0000 < \alpha = .05$, there is a significant relationship.

b. *R* square = .6144. A reasonably good fit using only Type of Fund.

c. The output follows:

Regression Statistics

Multiple R	0.8135
R Square	0.6617
Adjusted R Square	0.6279
Standard Error	5.3726

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Observations	45

ANOVA

	Df	SS	MS	*F*	Significance *F*
Regression	4	2258.3432	564.5858	19.5598	5.48647E-09
Residual	40	1154.5827	28.8646		
Total	44	3412.9260			

	Coefficients	Standard Error	*t* Stat	*P*-Value	Lower 95%	Upper 95%
Intercept	1.1899	2.3781	0.5004	0.6196	–3.6164	5.9961
FundDE	6.8969	2.7651	2.4942	0.0169	1.3083	12.4854
FundIE	17.6800	3.3161	5.3315	4.096E-06	10.9778	24.3821
Net Asset Value ($)	0.0265	0.0670	0.3950	0.6950	–0.1089	0.1619
Expense Ratio (%)	6.4564	2.7593	2.3399	0.0244	0.8798	12.0331

Because the p-value corresponding to $F = 19.5558$ is $.0000 < \alpha = .05$, there is a significant relationship.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

For Net Asset Value ($), the p-value corresponding to $t = .3950$ is $.6950 > \alpha = .05$, Net Asset Value ($) is not significant and can be deleted from the model.

d. Morningstar Rank is a categorical variable. The data set only contains funds with four ranks (two-star through –five star), so three dummy variables are needed. Let 3StarRank = 1 for a 3-StarRank, 4StarRank = 1 for a 4-StarRank, and 5StarRank = 1 for a 5-StarRank. The output follows:

Regression Statistics

Multiple R	0.8501
R Square	0.7227
Adjusted R Square	0.6789
Standard Error	4.9904
Observations	45

ANOVA

	Df	SS	MS	F	Significance F
Regression	6	2466.5721	411.0954	16.5072	2.96759E-09

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Residual	38	946.3539	24.9040
Total	44	3412.9260	

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	–4.6074	3.2909	–1.4000	0.1696	–11.2694	2.0547
FundDE	8.1713	2.2754	3.5912	0.0009	3.5650	12.7776
FundIE	19.5194	2.7795	7.0227	2.292E-08	13.8926	25.1461
Expense Ratio (%)	5.5197	2.5862	2.1343	0.0393	0.2843	10.7552
3StarRank	5.9237	2.8250	2.0969	0.0427	0.2048	11.6426
4StarRank	8.2367	2.8474	2.8927	0.0063	2.4725	14.0009
5StarRank	6.6241	3.1425	2.1079	0.0417	0.2624	12.9858

$\hat{y}$ = -4.6074 + 8.1713 FundDE + 19.5194 FundIE +5.5197 Expense Ratio (%) + 5.9237 3StarRank + 8.2367 4StarRank + 6.6241 5StarRank

At the .05 level of significance, all the independent variables are significant.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

e. $\hat{y}$ = -4.6074 + 8.1713(1) + 19.5194(0) +5.5197(1.05) + 5.9237(1) + 8.2367(0)

+6.62415(0) = 15.28%

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 16 Solutions

2. a. The output follows.

```
Analysis of Variance

      Source       DF    Adj SS    Adj MS    F-Value   P-Value

      Regression   1     59.39     59.39     4.76      0.117

      Error        3     37.41     12.47

      Total        4     96.80

Model Summary

      S         R-sq      R-sq(adj)

      3.53110   61.36%    48.48%

Coefficients

      Term         Coef   SE Coef   T-Value   P-Value   VIF

      Constant     9.32   4.20      2.22      0.113

      x            0.424  0.194     2.18      0.117     1.00

Regression Equation

      y = 9.32 + 0.424 x
```

The high *p*-value (.117) indicates a weak relationship; note that 61.4% of the variability in *y* has been explained by *x*.

b. The output follows.

```
Analysis of Variance

      Source       DF    Adj SS    Adj MS    F-Value   P-Value
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Regression    2     93.529    46.765    28.60     0.034

Error         2     3.271     1.635

Total         4     96.800

Model Summary

S           R-sq      R-sq(adj)

1.27880     96.62%    93.24%

Coefficients

Term        Coef      SE Coef    T-Value   P-Value   VIF

Constant    -8.10     4.10       -1.97     0.187

x           2.413     0.441      5.47      0.032     39.22

xsq         -0.0480   0.0105     -4.57     0.045     39.22

Regression Equation

y = -8.10 + 2.413 x - 0.0480 xsq
```

At the .05 level of significance, the relationship is significant; the fit is excellent.

c. $\hat{y} = -8.101 + 2.413(20) - 0.048(20)^2 = 20.959$

4. a. The output follows.

```
Analysis of Variance

Source      DF    Adj      SS Adj    MS        F-Value   P-Value

Regression  1     33223    33223     31.86     0.005

Error       4     4172     1043

Total       5     37395
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Model Summary

      S              R-sq        R-sq(adj)

      32.2940        88.84%      86.06%

Coefficients

      Term          Coef     SE Coef   T-Value   P-Value   VIF

      Constant      943.0    59.4      15.88     0.000

      Vehicle       8.71     1.54      5.64      0.005     1.00
      Speed

Regression Equation

      Traffic Flow = 943.0 + 8.71 Vehicle Speed
```

b. p-value = .005 < α = .01; reject H_0.

6. a. The scatter diagram follows.

b. No; the relationship appears to be curvilinear.

c. Several possible models can be fitted to these data as follow:

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\hat{y} = 2.90 - 0.185x + .00351x^2 \quad R_a^2 = .91$$

$$\hat{y} = -0.0468 + 14.4\left(\frac{1}{x}\right) \quad R_a^2 = .91$$

8. a. The scatter diagram follows.

A simple linear regression model does not appear to be appropriate. There appears to be a curvilinear relationship between the two variables.

b. The output follows.

```
Analysis of Variance

      Source      DF  Adj SS    Adj MS   F-Value  P-Value

      Regression  2   12643604  6321802  14.15    0.001

      Error       12  5359686   446641

      Total       14  18003290
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Model Summary

      S        R-sq    R-sq(adj)

      668.312  70.23%  65.27%

Coefficients

      Term      Coef   SE Coef  T-Value  P-Value  VIF

      Constant  33829  13657    2.48     0.029

      Rating    -4571  1688     -2.71    0.019    296.05

      RatingSq  153.6  51.7     2.97     0.012    296.05

Regression Equation

      Price ($1000) = 33829 - 4571 Rating + 153.6 RatingSq
```

c. The output for the model using the base 10 log of Price as the dependent variable follows.

```
Analysis of Variance

      Source      DF  Adj SS  Adj MS   F-Value  P-Value

      Regression  1   3.7885  3.78851  53.81    0.000

      Error       13  0.9153  0.07041

      Total       14  4.7038

Model Summary

      S         R-sq    R-sq(adj)

      0.265348  80.54%  79.04%

Coefficients

      Term      Coef    SE Coef  T-Value  P-Value  VIF
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Constant  -2.208  0.658    -3.36    0.005

Rating    0.2857  0.0390   7.34     0.000    1.00
```

```
Regression Equation

logPrice = -2.208 + 0.2857 Rating
```

d. The model in part c is preferred because it provides a better fit.

10. a. SSR = SST – SSE = 1,030

MSR = 1,030 MSE = 520/25 = 20.8 F = 1,030/20.8 = 49.52

Using statistical software, the p-value corresponding to F = 49.52 is .000.

Because p-value $\leq \alpha$, x_1 is significant.

b. $F = \dfrac{(520-100)/2}{100/23} = 48.3$

Using statistical software, the p-value corresponding to F = 48.3 is .000.

Because p-value $\leq \alpha$, the addition of variables x_2 and x_3 is significant.

12. a. A portion of the output follows.

```
Analysis of Variance

Source      DF   Adj SS  Adj MS   F-Value  P-Value

Regression  1    78.25   78.2452  61.86    0.000

Error       132  166.95  1.2648

Total       133  245.20
```

```
Model Summary

S        R-sq    R-sq(adj)

1.12463  31.91%  31.40%
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	47.95	3.12	15.36	0.000	
Putting Ave.	0.806	0.102	7.87	0.000	1.00

Regression Equation

Scoring Ave. = 47.95 + 0.806 Putting Ave.

b. A portion of the output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	229.677	76.559	641.25	0.000
Error	130	15.521	0.119		
Total	133	245.197			

Model Summary

S	R-sq	R-sq(adj)
0.345528	93.67%	93.52%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	57.16	1.06	53.95	0.000	
Putting Ave.	1.0319	0.0322	32.07	0.000	1.04
Greens in Reg.	-23.102	0.653	-35.37	0.000	1.05
Drive Accuracy	-0.022	0.481	-0.05	0.964	1.02

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Regression Equation

      Scoring Ave. = 57.16 + 1.0319 Putting Ave. - 23.102 Greens in
       Reg. - 0.022 Drive Accuracy
```

c. SSE(reduced) = 166.95

SSE(full) = 15.521

MSE(full) = .119

$$F = \frac{\dfrac{\text{SSE(reduced) - SSE(full)}}{\text{number of extra terms}}}{\text{MSE(full)}} = \frac{\dfrac{166.95 - 15.521}{2}}{.119} = 634.19$$

The p-value associated with $F = 634.19$ (two degrees of freedom numerator and 130 denominator) is 0.00. With a p-value < α =.05, the addition of the two independent variables is statistically significant.

14. a. The output follows.

```
Analysis of Variance

      Source      DF  Adj SS  Adj MS   F-Value  P-Value

      Regression  2   3379.6  1689.82  35.41    0.000

      Error       17  811.3   47.72

      Total       19  4190.9

Model Summary

      S        R-sq    R-sq(adj)

      6.90826  80.64%  78.36%

Coefficients
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-110.9	16.5	-6.74	0.000	
Age	1.315	0.173	7.59	0.000	1.11
Blood_Pressure	0.2964	0.0511	5.80	0.000	1.11

Regression Equation

Risk = -110.9 + 1.315 Age + 0.2964 Blood_Pressure

b. The output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	3672.11	918.03	26.54	0.000
Error	15	518.84	34.59		
Total	19	4190.95			

Model Summary

S	R-sq	R-sq(adj)
5.88127	87.62%	84.32%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-123.2	56.9	-2.16	0.047	
Age	1.513	0.780	1.94	0.071	30.87
Blood_Pressure	0.448	0.346	1.30	0.214	69.91
Smoker	8.87	3.07	2.88	0.011	1.37

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
AgePress        -0.00276  0.00481  -0.57    0.575    71.91
```

```
Regression Equation

Risk = -123.2 + 1.513 Age + 0.448 Blood_Pressure + 8.87 Smoker -
 0.00276 AgePress
```

c. $$F = \frac{\dfrac{\text{SSE(reduced) - SSE(full)}}{\text{number of extra terms}}}{\text{MSE(full)}} = \frac{\dfrac{811.3 - 518.84}{2}}{34.59} = 4.23$$

The p-value associated with $F = 4.23$ (2 numerator and 15 denominator DF) is .0349.

Because p-value $\leq \alpha = .05$, the addition of the two terms is significant.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

16. a. The sample correlation coefficients are as follows.

	Weeks	Age	Educ	Married	Head	Tenure	Manager
Age	0.577						
	0.000						
Educ	0.007	0.100					
	0.962	0.490					
Married	-0.130	-0.209	-0.151				
	0.370	0.145	0.296				
Head	-0.205	0.027	-0.156	-0.449			
	0.153	0.854	0.280	0.001			
Tenure	0.398	0.459	0.174	-0.057	-0.046		
	0.004	0.001	0.228	0.692	0.750		
Manager	-0.198	0.097	0.160	0.073	-0.200	-0.113	
	0.167	0.504	0.266	0.616	0.164	0.435	
Sales	-0.134	0.137	0.124	-0.148	-0.013	0.097	-0.156
	0.354	0.343	0.393	0.306	0.926	0.504	0.279

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Cell Contents: Pearson correlation

P-Value
```

The independent variable most highly correlated with Weeks is Age. The output corresponding to using Age as the independent variable follows.

```
Analysis of Variance

      Source      DF  Adj SS  Adj MS  F-Value  P-Value

      Regression  1   9161    9161.4  24.01    0.000

      Error       48  18316   381.6

      Total       49  27478

Model Summary

      S        R-sq    R-sq(adj)

      19.5342  33.34%  31.95%

Coefficients

      Term      Coef   SE Coef  T-Value  P-Value  VIF

      Constant  -8.9   11.0     -0.80    0.425

      Age       1.509  0.308    4.90     0.000    1.00

Regression Equation

      Weeks = -8.9 + 1.509 Age
```

b. Stepwise regression output follows.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Stepwise Selection of Terms

Candidate terms: Age, Educ, Married, Head, Tenure, Manager, Sales

	----Step 1----		-----Step 2----		-----Step 3----		-----Step 4----	
	Coef	P	Coef	P	Coef	P	Coef	P
Constant	-8.9		-9.1		-0.1		-0.07	
Age	1.509	0.000	1.574	0.000	1.609	0.000	1.725	0.000
Manager			-20.06	0.029	-24.61	0.006	-28.67	0.001
Head					-14.31	0.012	-15.09	0.005
Sales							-17.42	0.008

S	19.5342	18.7494	17.6857	16.5069
R-sq	33.34%	39.87%	47.64%	55.38%
R-sq(adj)	31.95%	37.31%	44.22%	51.41%
Mallows' Cp	22.52	17.81	11.83	5.87

α-to-enter = 0.05, α-to-leave = 0.05

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The results suggest a model using four independent variables: Age, Manager, Head, and Sales. The corresponding output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	15216	3804.0	13.96	0.000
Error	45	12261	272.5		
Total	49	27478			

Model Summary

S	R-sq	R-sq(adj)
16.5069	55.38%	51.41%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.07	9.84	-0.01	0.994	
Age	1.725	0.265	6.51	0.000	1.04
Head	-15.09	5.12	-2.95	0.005	1.05
Manager	-28.67	8.12	-3.53	0.001	1.09
Sales	-17.42	6.24	-2.79	0.008	1.05

Regression Equation

Weeks = -0.07 + 1.725 Age - 15.09 Head - 28.67 Manager - 17.42 Sales

c. The results using the forward selection procedure are the same as the results using the stepwise procedure in part b.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

d. The results using the backward elimination procedure follow:

```
Backward Elimination of Terms

      Candidate terms: Age, Educ, Married, Head, Tenure, Manager, Sales

                  -----Step 1----   -----Step 2----   -----Step 3----  -----Step 4----

                  Coef     P        Coef     P        Coef     P       Coef     P

      Constant    22.9     13.6     13.1     -0.07

      Age         1.509    0.000    1.521    0.000    1.637    0.000   1.725    0.000

      Educ        -0.613   0.516

      Married     -10.74   0.081    -9.85    0.098    -9.76    0.099

      Head        -19.78   0.002    -19.01   0.002    -19.41   0.001   -15.09   0.005

      Tenure      0.426    0.366    0.373    0.418

      Manager     -26.74   0.003    -27.71   0.001    -28.99   0.001   -28.67   0.001

      Sales       -18.56   0.005    -18.99   0.004    -18.97   0.004   -17.42   0.008

      S           16.3497  16.2408  16.1794  16.5069

      R-sq        59.14%   58.72%   58.08%   55.38%

      R-sq(adj)   52.33%   52.96%   53.32%   51.41%
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Mallows' Cp 8.00 6.43 5.09 5.87

α-to-leave = 0.05

These results also suggest using the model with four independent variables: Age, Head, Manager, and Sales.

e. The results using Minitab's Best-Subset procedure follow:

Response is Weeks

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	Age	Educ	Married	Head	Tenure	Manager	Sales
1	33.3	32.0	28.7	22.5	19.534	X						
1	15.8	14.0	10.0	40.6	21.954	X						
2	39.9	37.3	33.3	17.8	18.749	X	X					
2	38.2	35.6	30.2	19.5	19.005	X	X					
3	47.6	44.2	37.8	11.8	17.686	X	X	X				
3	46.8	43.3	38.1	12.7	17.831	X	X	X				
4	55.4	51.4	44.9	5.9	16.507	X	X	X	X			
4	49.1	44.6	37.4	12.3	17.628	X	X	X	X			

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

5	58.1	53.3	46.6	5.1	16.179	X	X	X	X	X		
5	56.0	51.0	43.9	7.3	16.582	X	X	X	X	X		
6	58.7	53.0	46.0	6.4	16.241	X	X	X	X	X	X	
6	58.3	52.5	44.6	6.8	16.318	X	X	X	X	X	X	
7	59.1	52.3	43.8	8.0	16.350	X	X	X	X	X	X	X

The results suggest a model using five independent variables: Age, Married, Head, Manager, and Sales.

The corresponding output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	15959.5	3191.9	12.19	0.000
Error	44	11518.0	261.8		
Total	49	27477.5			

Model Summary

S	R-sq	R-sq(adj)
16.1794	58.08%	53.32%

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	13.1	12.4	1.05	0.298	
Age	1.637	0.265	6.18	0.000	1.08
Married	-9.76	5.79	-1.69	0.099	1.35
Head	-19.41	5.64	-3.44	0.001	1.32
Manager	-28.99	7.96	-3.64	0.001	1.09
Sales	-18.97	6.18	-3.07	0.004	1.08

Regression Equation

Weeks = 13.1 + 1.637 Age - 9.76 Married - 19.41 Head - 28.99 Manager - 18.97 Sales

18. a. Because the independent variable most highly correlated with RPG is OBP, it will provide the best one-variable estimated regression equation. The output using OBP to predict RPG follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	72.11	72.1076	78.85	0.000
Error	18	16.46	0.9145		
Total	19	88.57			

Model Summary

S	R-sq	R-sq(adj)
0.956308	81.41%	80.38%

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-4.05	1.01	-4.02	0.001	
OBP	27.55	3.10	8.88	0.000	1.00

Regression Equation

RPG = -4.05 + 27.55 OBP

b. The output using the stepwise regression procedure using α to enter = 0.05 and α to leave = 0.05 follows.

Candidate terms: H, 2B, 3B, HR, RBI, BB, SO, SB, CS, OBP, SLG, AVG

	-----Step 1----		-----Step 2-----		-----Step 3-----	
	Coef	P	Coef	P	Coef	P
Constant	-4.05	-1.595	-0.981			
OBP	27.55	0.000	17.16	0.000	25.14	0.000
HR	0.0711	0.001	0.0695	0.000		
AVG	-12.62	0.005				
S	0.956308	0.692987	0.555522			
R-sq	81.41%	90.78%	94.43%			
R-sq(adj)	80.38%	89.70%	93.38%			
Mallows' Cp	66.65	26.99	12.79			

α-to-enter = 0.05, α-to-leave = 0.05

Analysis of Variance

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	83.631	27.8771	90.33	0.000
Error	16	4.938	0.3086		
Total	19	88.569			

Model Summary

S	R-sq	R-sq(adj)
0.555522	94.43%	93.38%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.981	0.776	-1.26	0.224	
HR	0.0695	0.0137	5.06	0.000	2.24
OBP	25.14	3.65	6.88	0.000	4.11
AVG	-12.62	3.90	-3.23	0.005	2.76

Regression Equation

RPG = -0.981 + 0.0695 HR + 25.14 OBP - 12.62 AVG

Using less-sensitive values for α to enter and α to leave will provide a model with additional independent variables. For example, the output using the stepwise regression procedure using α to enter = 0.10 and α to leave = 0.10 follows.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Stepwise Selection of Terms

Candidate terms: H, 2B, 3B, HR, RBI, BB, SO, SB, CS, OBP, SLG, AVG

	-----Step 1----		-----Step 2-----		-----Step 3-----		-----Step 4-----	
	Coef	P	Coef	P	Coef	P	Coef	P
Constant	-4.05		-1.595		-0.981		-0.616	
OBP	27.55	0.000	17.16	0.000	25.14	0.000	26.61	0.000
HR			0.0711	0.001	0.0695	0.000	0.0681	0.000
AVG					-12.62	0.005	-16.50	0.001
3B							0.1823	0.079
BB								
S	0.956308		0.692987		0.555522		0.515891	
R-sq	81.41%		90.78%		94.43%		95.49%	
R-sq(adj)	80.38%		89.70%		93.38%		94.29%	
Mallows' Cp	66.65		26.99		12.79		10.04	

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
                ------Step 5------

                Coef          P

      Constant  -0.909

      OBP       32.18         0.000

      HR        0.1088        0.000

      AVG       -21.51        0.000

      3B        0.2439        0.010

      BB        -0.02231      0.011


      S            0.420960

      R-sq         97.20%

      R-sq(adj)    96.20%

      R-sq(pred)   93.86%

      Mallows' Cp  4.46

      α-to-enter = 0.1, α-to-leave = 0.1

Analysis of Variance

      Source      DF  Adj SS  Adj MS   F-Value  P-Value

      Regression  5   86.088  17.2176  97.16    0.000

      Error       14  2.481   0.1772

      Total       19  88.569
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Model Summary

S	R-sq	R-sq(adj)
0.420960	97.20%	96.20%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.909	0.617	-1.47	0.163	
3B	0.2439	0.0817	2.99	0.010	1.55
HR	0.1088	0.0174	6.26	0.000	6.26
BB	-0.02231	0.00764	-2.92	0.011	8.02
OBP	32.18	3.42	9.40	0.000	6.28
AVG	-21.51	3.81	-5.65	0.000	4.58

Regression Equation

RPG = -0.909 + 0.2439 3B + 0.1088 HR - 0.02231 BB + 32.18 OBP - 21.51 AVG

The following output using the best subset procedure also confirms that a variety of models will provide a good fit.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Response is RPG

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	H	2B	3B	HR	RBI	BB	SO	SB	CS	OBP	SLG	AVG
1	81.4	80.4	75.4	66.6	0.95631	X											
1	78.9	77.7	73.1	77.9	1.0192	X											
2	90.8	89.7	87.4	27.0	0.69299	X	X										
2	88.4	87.0	83.5	37.8	0.77872	X	X										
3	94.4	93.4	89.2	12.8	0.55552	X	X	X									
3	94.4	93.3	87.9	13.0	0.55820	X	X	X									
4	95.8	94.6	89.9	8.8	0.50014	X	X	X	X								
4	95.5	94.3	90.0	10.0	0.51589	X	X	X	X								
5	97.2	96.2	93.9	4.5	0.42096	X	X	X	X	X							
5	97.2	96.2	93.8	4.6	0.42336	X	X	X	X	X							
6	97.6	96.6	92.6	4.5	0.40042	X	X	X	X	X	X						
6	97.5	96.4	94.1	5.1	0.41198	X	X	X	X	X	X						
7	98.2	97.2	94.4	3.9	0.36245	X	X	X	X	X	X	X					

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

7	98.2	97.1	94.4	4.1	0.36664	X	X	X	X	X	X	X					
8	98.3	97.1	93.3	5.3	0.36471	X	X	X	X	X	X	X	X				
8	98.3	97.0	93.3	5.7	0.37269	X	X	X	X	X	X	X	X				
9	98.4	97.0	92.4	7.1	0.37506	X	X	X	X	X	X	X	X	X			
9	98.4	96.9	91.3	7.3	0.38077	X	X	X	X	X	X	X	X	X			
10	98.4	96.7	91.8	9.0	0.39477	X	X	X	X	X	X	X	X	X	X		
10	98.4	96.7	88.1	9.0	0.39496	X	X	X	X	X	X	X	X	X	X		
11	98.4	96.3	86.0	11.0	0.41758	X	X	X	X	X	X	X	X	X	X	X	
11	98.4	96.2	88.6	11.0	0.41848	X	X	X	X	X	X	X	X	X	X	X	
12	98.4	95.7	83.8	13.0	0.44629	X	X	X	X	X	X	X	X	X	X	X	X

It would be difficult to make an argument that there is one best model given these results. The five-variable model identified using the stepwise regression procedure with α to enter = 0.10 and α to leave = 0.10 seems like a reasonable choice. The regression output corresponding to this model follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	86.088	17.2176	97.16	0.000

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Error	14	2.481	0.1772
Total	19	88.569	

Model Summary

S	R-sq	R-sq(adj)
0.420960	97.20%	96.20%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.909	0.617	-1.47	0.163	
3B	0.2439	0.0817	2.99	0.010	1.55
HR	0.1088	0.0174	6.26	0.000	6.26
BB	-0.02231	0.00764	-2.92	0.011	8.02
OBP	32.18	3.42	9.40	0.000	6.28
AVG	-21.51	3.81	-5.65	0.000	4.58

Regression Equation

RPG = -0.909 + 0.2439 3B + 0.1088 HR - 0.02231 BB + 32.18 OBP - 21.51 AVG

The corresponding standardized residual plot follows.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The standardized residual plot does not indicate any reason to question the usual assumptions regarding the error term. Thus, the estimated regression equation using OBP, HR, AVG, 3B, and BB appears to be a good choice.

20. The dummy variables are defined as follow.

x_1	x_2	x_3	Treatment
0	0	0	A
1	0	0	B
0	1	0	C

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

0	0	1	D

$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

22. Factor A

$x_1 = 0$ if level 1 and 1 if level 2

Factor B

x_2	x_3	Level
0	0	1
1	0	2
0	1	3

$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1 x_3$

24. a. The dummy variables are defined as follows.

D1	D2	D3	Paint
0	0	0	1

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

1	0	0	2
0	1	0	3
0	0	1	4

The output follows.

```
Analysis of Variance

      Source      DF  Adj SS   Adj MS  F-Value  P-Value

      Regression  3   330.00   110.00  2.54     0.093

      Error       16  692.00   43.25

      Total       19  1022.00

Model Summary

      S        R-sq    R-sq(adj)

      6.57647  32.29%  19.59%

Coefficients

      Term      Coef    SE Coef  T-Value  P-Value  VIF

      Constant  133.00  2.94     45.22    0.000
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
D1        6.00    4.16     1.44     0.168    1.50

D2        3.00    4.16     0.72     0.481    1.50

D3        11.00   4.16     2.64     0.018    1.50

Regression Equation

      Time = 133.00 + 6.00 D1 + 3.00 D2 + 11.00 D3
```

The appropriate hypothesis test is:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

The p-value of .093 is greater than $\alpha = .05$; therefore, at the 5% level of significance we cannot reject H_0.

b. Estimating the mean drying for paint 2 using the estimated regression equations developed in part a may not be the best approach because at the 5% level of significance, we cannot reject H_0. However, if we want to use the output, we would proceed as follows:

$$\text{Time} = 133 + 6(1) + 3(0) + 11(0) = 139$$

Note that this is equivalent of just computing the average of the five drying times for paint 2—that is, (144 + 133 + 142 + 146 + 130) / 5 = 139.

Alternatively, we observe from the regression output that paints 1 and 2 drying times (as represented by the D1 coefficient)

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

and paints 1 and 3 drying times (as represented by the D2 coefficient) do not seem to significantly differ.

Regression Statistics								
Multiple R	0.485							
R Square	0.235							
Adjusted R Square	0.192							
Standard Error	6.591							
Observations	20							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	240	240	5.524	0.030			
Residual	18	782	43.444					
Total	19	1022						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.0%*	*Upper 99.0%*
Intercept	136	1.702	79.913	0.000	132.425	139.575	131.101	140.899
D3	8	3.404	2.350	0.030	0.849	15.151	-1.797	17.797

Using this regression model, the paint 2 drying time can be estimated by:

Time = 136+ 8(0) = 136.

Note that this is equivalent to just computing the average of the 15 drying times for paint 1, paint 2, and paint 3. If there is no difference between paint 1, paint 2, and paint 3, this estimate of 136 minutes is preferable to the previous estimate of 139 minutes because it is based on more data (15 observations versus just 5 observations)

26. Size = 0 if a small advertisement and 1 if a large advertisement.

DesignB and DesignC are defined as follows.

DesignB DesignC Advertisement Design

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

0	0	A
1	0	B
0	1	C

LargeDesignB denotes the interaction between Large and DesignB

LargeDesignC denotes the interaction between Large and DesignC

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The complete data set and the output follow.

Number	Size	DesignB	DesignC	LargeDesignB	LargeDesignC
8	0	0	0	0	0
12	0	0	0	0	0
22	0	1	0	0	0
14	0	1	0	0	0
10	0	0	1	0	0
18	0	0	1	0	0
12	1	0	0	0	0
8	1	0	0	0	0
26	1	1	0	1	0
30	1	1	0	1	0
18	1	0	1	0	1

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

14	1	0	1	0	1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	448.000	89.6000	5.60	0.029
Error	6	96.000	16.0000		
Total	11	544.000			

Model Summary

S	R-sq	R-sq(adj)
4	82.35%	67.65%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	10.00	2.83	3.54	0.012	
Size	0.00	4.00	0.00	1.000	3.00
DesignB	8.00	4.00	2.00	0.092	2.67

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
DesignC       4.00   4.00     1.00     0.356    2.67

LargeDesignB  10.00  5.66     1.77     0.128    3.33

LargeDesignC  2.00   5.66     0.35     0.736    3.33

Regression Equation

Number = 10.00 + 0.00 Size + 8.00 DesignB + 4.00 DesignC + 10.00 LargeDesignB + 2.00
LargeDesignC
```

Overall the model is significant because the p-value corresponding to $F = 5.60 < \alpha = .05$.

Individually, none of the variables are significant using $\alpha = .05$.

The output using only Design B follows

```
Analysis of Variance

Source      DF  Adj SS  Adj MS  F-Value  P-Value

Regression  1   294.0   294.00  11.76    0.006

Error       10  250.0   25.00

Total       11  544.0

Model Summary
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
S         R-sq    R-sq(adj)

5 54.04%  49.45%  27.84%

Coefficients

Term      Coef    SE Coef     T-Value     P-Value     VIF

Constant          12.50       1.77        7.07        0.000

DesignB   10.50    3.06       3.43        0.006       1.00

Regression Equation

Number = 12.50 + 10.50 DesignB
```

Thus, DesignB is significant using $\alpha = .05$. However, the model involving just the interaction between Large and DesignB also provides some interesting results:

```
Analysis of Variance

Source      DF  Adj SS  Adj MS  F-Value  P-Value

Regression  1   345.6   345.60  17.42    0.002

Error       10  198.4   19.84

Total       11  544.0
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Model Summary

      S        R-sq    R-sq(adj)

      4.45421  63.53%  59.88%



Coefficients

      Term          Coef   SE Coef  T-Value  P-Value  VIF

      Constant      13.60  1.41     9.66     0.000

      LargeDesignB  14.40  3.45     4.17     0.002    1.00

Regression Equation

      Number = 13.60 + 14.40 LargeDesignB
```

Here we see that the interaction term is significant. Thus, one might consider that the differences are to the result of both design and a large advertisement. But it is difficult to reach any definite conclusions given the size of the data set. A larger sample size is needed to make any stronger conclusions about the relationships among the variables.

28. From statistical software, $d = 1.59617$. At the .05 level of significance, $d_L = 1.04$ and $d_U = 1.77$. Because $d_L \le d \le d_U$, the test is inconclusive. However, these data are not a time series, so autocorrelation is not a concern and it is not appropriate to perform a Durbin-Watson test.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

30. a.

There appears to be a curvilinear relationship between weight and price.

b. A portion of the output follows.

```
Analysis of Variance

      Source      DF  Adj SS   Adj MS   F-Value  P-Value

      Regression  2   3161747  1580874  26.82    0.000

      Error       16  943263   58954

      Total       18  4105011
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
Model Summary

      S        R-sq    R-sq(adj)

      242.804  77.02%  74.15%

Coefficients

      Term      Coef   SE Coef  T-Value  P-Value  VIF

      Constant  11376  2565     4.43     0.000

      Weight    -728   194      -3.76    0.002    287.41

      WeightSq  11.97  3.54     3.38     0.004    287.41

Regression Equation

      Price = 11376 - 728 Weight + 11.97 WeightSq
```

The results obtained support the conclusion that there is a curvilinear relationship between weight and price.

c. A portion of the output follows.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	2944410	1472205	20.30	0.000
Error	16	1160601	72538		
Total	18	4105011			

Model Summary

S	R-sq	R-sq(adj)
269.328	71.73%	68.19%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1283.8	95.2	13.48	0.000	
Type_Fitness	-572	154	-3.72	0.002	1.20
Type_Comfort	-907	145	-6.24	0.000	1.20

Regression Equation

Price = 1283.8 - 572 Type_Fitness - 907 Type_Comfort

Type of bike appears to be a significant factor in predicting price, but the estimated regression equation developed in part b

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

appears to provide a slightly better fit.

d. A portion of the output follows. In this output, WxF denotes the interaction between the weight of the bike, and the dummy variable Type_Fitness and WxC denote the interaction between the weight of the bike and the dummy variable Type_Comfort.

```
Regression Analysis: Price versus Weight, Type_Fitness, Type_Comfort, WxF, WxC

Analysis of Variance

Source      DF  Adj SS   Adj MS  F-Value  P-Value

Regression  5   3450170  690034  13.70    0.000

Error       13  654841   50372

Total       18  4105011

Model Summary

S        R-sq    R-sq(adj)

224.438  84.05%  77.91%

Coefficients

Term          Coef    SE Coef  T-Value  P-Value  VIF
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Constant	5924	1547	3.83	0.002	
Weight	-214.6	71.4	-3.00	0.010	45.74
Type_Fitness	-6343	2596	-2.44	0.030	492.96
Type_Comfort	-7232	2518	-2.87	0.013	516.81
WxF	261	112	2.34	0.036	537.48
WxC	266.4	94.0	2.83	0.014	762.67

Regression Equation

Price = 5924 - 214.6 Weight - 6343 Type_Fitness - 7232 Type_Comfort + 261 WxF + 266.4 WxC

By taking into account the type of bike, the weight, and the interaction between these two factors this estimated regression equation provides an excellent fit.

32. a. The output follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	2587.7	646.9	5.42	0.002
Error	35	4176.3	119.3		

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Total	39	6764.0

Model Summary

S	R-sq	R-sq(adj)
10.9235	38.26%	31.20%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	80.43	5.92	13.60	0.000	
Industry	11.94	3.80	3.15	0.003	1.06
Public	-4.82	4.23	-1.14	0.263	1.05
Quality	-2.62	1.18	-2.22	0.033	1.06
Finished	-4.07	1.85	-2.20	0.035	1.00

Regression Equation

Delay = 80.43 + 11.94 Industry - 4.82 Public - 2.62 Quality - 4.07 Finished

b. The low value of the adjusted coefficient of determination (31.2%) does not indicate a good fit.

c. The scatter diagram follows.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The scatter diagram suggests a curvilinear relationship between these two variables.

d. The output from the best subsets procedure follows, where FinishedSq is the square of Finished.

```
Response is Delay

                                                                  F
                                                  I               F   i
                                                  n         Q     i   n
                                                  d   P     u     n   i
                                                  u   u     a     i   s
                                                  s   b     l     s   h
                                                  t   l     i     h   e
                     R-Sq   R-Sq    Mallows        r   i     t     e   d
 Vars         R-Sq  (adj)  (pred)  Cp       S     y   c     y     d   q
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

										Sq
1	15.9	13.7	6.5	34.0	12.234	X				
1	9.9	7.6	0.0	39.0	12.662				X	
2	37.8	34.4	27.0	17.8	10.667				X	X
2	26.9	22.9	14.9	26.9	11.561	X		X		
3	51.1	47.1	39.7	8.7	9.5813	X			X	X
3	42.2	37.3	30.4	16.1	10.425			X	X	X
4	59.1	54.4	48.3	4.1	8.8960	X		X	X	X
4	51.6	46.1	35.3	10.3	9.6712	X	X		X	X
5	59.1	53.1	43.0	6.0	9.0149	X	X	X	X	x

The estimated regression equation using Industry, Quality, Finished, and FinishedSq has the largest adjusted coefficient of determination of 54.4%.

The following regression output summarizes this model given by the following equation:

Delay = 112.79 + 11.63*Industry – 2.49*Quality – 36.55*Finished + 6.63*FinishedSq

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Regression Statistics								
Multiple R	0.768							
R Square	0.591							
Adjusted R Square	0.544							
Standard Error	8.896							
Observations	40							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	4	3994.139	998.535	12.618	0.000			
Residual	35	2769.836	79.138					
Total	39	6763.975						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.0%*	*Upper 99.0%*
Intercept	112.790	8.852	12.742	0.000	94.820	130.761	88.679	136.901
Industry	11.631	3.060	3.800	0.001	5.418	17.844	3.295	19.967
Quality	-2.487	0.957	-2.600	0.014	-4.429	-0.545	-5.092	0.118
Finished	-36.551	7.439	-4.914	0.000	-51.652	-21.449	-56.813	-16.289
FinishedSq	6.630	1.493	4.442	0.000	3.600	9.661	2.564	10.696

34. a. The output follows.

```
Analysis of Variance

      Source      DF  Adj SS  Adj MS  F-Value  P-Value

      Regression  2   1818.6  909.3   6.80     0.003

      Error       37  4945.4  133.7

      Total       39  6764.0

Model Summary
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

S	R-sq	R-sq(adj)
11.5611	26.89%	22.93%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	70.63	4.56	15.50	0.000	
Industry	12.74	3.97	3.21	0.003	1.03
Quality	-2.92	1.24	-2.36	0.024	1.03

Regression Equation

Delay = 70.63 + 12.74 Industry - 2.92 Quality

Durbin-Watson Statistic = 1.42611

b. The residual plot as a function of the order in which the data are presented follows.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

There is no obvious pattern in the data indicative of positive autocorrelation.

c. At the .05 level of significance, $d_L = 1.39$ and $d_U = 1.60$. Because $d_L \leq d \leq d_U$, the test is inconclusive. However, these data are not a time series, so autocorrelation is not a concern and it is not appropriate to perform a Durbin-Watson test.

36. The dummy variables are defined as follow:

D1	D2	Type
0	0	Non
1	0	Light
0	1	Heavy

The output follows.

```
Analysis of Variance
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	9.333	4.667	4.00	0.034
Error	21	24.500	1.167		
Total	23	33.833			

Model Summary

S	R-sq	R-sq(adj)
1.08012	27.59%	20.69%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4.250	0.382	11.13	0.000	
D1	1.000	0.540	1.85	0.078	1.33
D2	1.500	0.540	2.78	0.011	1.33

Regression Equation

Score = 4.250 + 1.000 D1 + 1.500 D2

Because the p-value = .034 is less than $\alpha = .05$, there are significant differences between comfort levels for the three types of browsers.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 17 Solutions

2. The following table shows the calculations for parts a, b, and c.

Week	Time-Series Value	Forecast	Forecast Error	Absolute Value of Forecast Error	Squared Forecast Error	Percentage Error	Absolute Value of Percentage Error
1	18						
2	13	18.00	–5.00	5.00	25.00	–38.46	38.46
3	16	15.50	0.50	0.50	0.25	3.13	3.13
4	11	15.67	–4.67	4.67	21.81	–42.45	42.45
5	17	14.50	2.50	2.50	6.25	14.71	14.71
6	14	15.00	–1.00	1.00	1.00	–7.14	7.14
			Totals	13.67	54.31	–70.21	105.86

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

a. MAE = 13.67/5 = 2.73

b. MSE = 54.31/5 = 10.86

c. MAPE = 105.89/5 = 21.18

d. Forecast for week 7 is (18 + 13 + 16 + 11 + 17 + 14) / 6 = 14.83

4. a.

Month	Time-Series Value	Forecast	Forecast Error	Squared Forecast Error
1	24			
2	13	24	–11	121
3	20	13	7	49
4	12	20	–8	64
5	19	12	7	49
6	23	19	4	16
7	15	23	–8	64
			Total	363

MSE = 363/6 = 60.5

Forecast for month 8 = 15.

b.

Week	Time-Series Value	Forecast	Forecast Error	Squared Forecast Error

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

1	24			
2	13	24.00	–11.00	121.00
3	20	18.50	1.50	2.25
4	12	19.00	–7.00	49.00
5	19	17.25	1.75	3.06
6	23	17.60	5.40	29.16
7	15	18.50	–3.50	12.25
			Total	216.72

MSE = 216.72/6 = 36.12

Forecast for month 8 = (24 + 13 + 20 + 12 + 19 + 23 + 15) / 7 = 18.

c. The average of all the previous values is better because MSE is smaller.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

6. a.

The data appear to follow a horizontal pattern.

Three-Week Moving Average

Week	Time-Series Value	Forecast	Forecast Error	Squared Forecast Error
1	24			
2	13			
3	20			
4	12	19.00	–7.00	49.00
5	19	15.00	4.00	16.00
6	23	17.00	6.00	36.00
7	15	18.00	–3.00	9.00
			Total	110.00

MSE = 110/4 = 27.5.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The forecast for week 8 = (19 + 23 + 15) / 3 = 19.

b. Smoothing constant = .2

Week	Time-Series Value	Forecast	Forecast Error	Squared Forecast Error
1	24			
2	13	24.00	–11.00	121.00
3	20	21.80	–1.80	3.24
4	12	21.44	–9.44	89.11
5	19	19.55	–0.55	0.30
6	23	19.44	3.56	12.66
7	15	20.15	–5.15	26.56
			Total	252.87

MSE = 252.87/6 = 42.15

The forecast for week 8 is .2(15) + (1 – .2)20.15 = 19.12.

c. The three-week moving average provides a better forecast because it has a smaller MSE.

d. Smoothing constant = .4.

Week	Time-Series Value	Forecast	Forecast Error	Squared Value of Forecast Error
1	24			

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

2	13	24.00	–11.00	121.00
3	20	19.60	0.40	0.16
4	12	19.76	–7.76	60.22
5	19	16.66	2.34	5.49
6	23	17.59	5.41	29.23
7	15	19.76	–4.76	22.62
			Total	238.72

MSE = 238.72/6 = 39.79

The exponential smoothing forecast using $\alpha = .4$ provides a better forecast than the exponential smoothing forecast using $\alpha = .2$ because it has a smaller MSE.

8. a.

Week	Time-Series Value	Weighted Moving Average Forecast	Forecast Error	(Error)^2
1	17			
2	21			
3	19			
4	23	19.33	3.67	13.47
5	18	21.33	–3.33	11.09
6	16	19.83	–3.83	14.67
7	20	17.83	2.17	4.71
8	18	18.33	–0.33	0.11

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

9	22	18.33	3.67	13.47
10	20	20.33	–0.33	0.11
11	15	20.33	–5.33	28.41
12	22	17.83	4.17	17.39
			Total	103.43

b. MSE = 103.43 / 9 = 11.49

Prefer the unweighted moving average here; it has a smaller MSE.

c. You could always find a weighted moving average at least as good as the unweighted one. Actually the unweighted moving average is a special case of the weighted ones where the weights are equal.

10.

a.

$$F_{13} = .2Y_{12} + .16Y_{11} + .64(.2Y_{10} + .8F_{10}) = .2Y_{12} + .16Y_{11} + .128Y_{10} + .512F_{10}$$

$$F_{13} = .2Y_{12} + .16Y_{11} + .128Y_{10} + .512(.2Y_9 + .8F_9) = .2Y_{12} + .16Y_{11} + .128Y_{10} + .1024Y_9 + .4096F_9$$

$$F_{13} = .2Y_{12} + .16Y_{11} + .128Y_{10} + .1024Y_9 + .4096(.2Y_8 + .8F_8) = .2Y_{12} + .16Y_{11} + .128Y_{10} + .1024Y_9 + .08192Y_8 + .32768F_8$$

b. The more recent data receive the greater weight or importance in determining the forecast. The moving averages method weights the last *n* data values equally in determining the forecast.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

12. a.

The data appear to follow a horizontal pattern.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b.

Month	Time-Series Value	Three-Month Moving Average Forecast	$(\text{Error})^2$	Four-Month Moving Average Forecast	$(\text{Error})^2$
1	9.5				
2	9.3				
3	9.4				
4	9.6	9.40	0.04		
5	9.8	9.43	0.14	9.45	0.12
6	9.7	9.60	0.01	9.53	0.03
7	9.8	9.70	0.01	9.63	0.03
8	10.5	9.77	0.53	9.73	0.59
9	9.9	10.00	0.01	9.95	0.00
10	9.7	10.07	0.14	9.98	0.08
11	9.6	10.03	0.18	9.97	0.14

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

12	9.6	9.73	0.02	9.92	0.10
		Totals	1.08		1.09

MSE(3-month) = 1.08 / 9 = .12

MSE(4-month) = 1.09 / 8 = .14

Use three-month moving averages.

c. Forecast = (9.7 + 9.6 + 9.6) / 3 = 9.63

14. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The data appear to follow a horizontal pattern.

b. Smoothing constant = .3.

Month t	Time-Series Value Y_t	Forecast F_t	Forecast Error $Y_t - F_t$	Squared Error $(Y_t - F_t)^2$
1	105			
2	135	105.00	30.00	900.00
3	120	114.00	6.00	36.00
4	105	115.80	–10.80	116.64
5	90	112.56	–22.56	508.95
6	120	105.79	14.21	201.92
7	145	110.05	34.95	1,221.50
8	140	120.54	19.46	378.69
9	100	126.38	–26.38	695.90
10	80	118.46	–38.46	1,479.17
11	100	106.92	–6.92	47.89
12	110	104.85	5.15	26.52
			Total	5,613.18

MSE = 5,613.18 / 11 = 510.29

Forecast for month 13: F_{13} = .3(110) + .7(104.85) = 106.4.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. Smoothing constant = .5

Month t	Time-Series Value Y_t	Forecast F_t	Forecast Error $Y_t - F_t$	Squared Error $(Y_t - F_t)^2$
1	105			
2	135	105	30.00	900.00
3	120	120	0.00	0.00
4	105	120	–15.00	225.00
5	90	112.50	–22.50	506.25
6	120	101.25	18.75	351.56
7	145	110.63	34.37	1,181.30
8	140	127.81	12.19	148.60
9	100	133.91	–33.91	1,149.89
10	80	116.95	–36.95	1,365.30
11	100	98.48	1.52	2.31
12	110	99.24	10.76	115.78
				5,945.99

MSE = 5945.99 / 11 = 540.55

Forecast for month 13: F_{13} = .5(110) + .5(99.24) = 104.62

Conclusion: a smoothing constant of .3 is better than a smoothing constant of .5 because the MSE is less for 0.3.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

16. a.

The time-series plot exhibits a trend pattern. Although the recession of 2008 led to a downturn in prices, the median price rose form 2010 to 2011.

b. The methods disucssed in this section are only applicable for a time series that has a horizontal pattern. Because the time-series plot exhibits a trend pattern, the methods disucssed in this section are not appropriate.

c. In 2003 the median price was $189,500, and in 2004 the median price was $222,300, so it appears that the time series shifted to a new level in 2004. The time-series plot using just the data for 2004 and later follows.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

This time-series plot exhibits a horizontal pattern. Therefore, the methods discussed in this section are approporiate.

18. a.

The time-series plot shows a linear trend.

b.

$$\bar{t} = \frac{\sum_{t=1}^{n} t}{n} = \frac{28}{7} = 4 \qquad \bar{Y} = \frac{\sum_{t=1}^{n} Y_t}{n} = \frac{48}{7} = 6.857$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\Sigma(t-\bar{t})(Y_t-\bar{Y}) = -29 \qquad \Sigma(t-\bar{t})^2 = 32.857$$

$$b_1 = \frac{\sum_{t=1}^{n}(t-\bar{t})(Y_t-\bar{Y})}{\sum_{t=1}^{n}(t-\bar{t})^2} = \frac{-29}{28} = -1.0357$$

$$b_0 = \bar{y} - b_1\bar{x} = 6.857 - (-1.0357)(4) = 11$$

$$T_t = 11 - 1.0357(t)$$

c. $T_8 = 11 - 1.0357(8) = 2.7$

20. a.

The time-series plot exhibits a curvilinear trend.

b. The linear trend equation is T_t =107.857 – 28.9881 t + 2.65476 r^2.

c. T_t =107.857 – 28.9881(8) +2.65476 $(8)^2$ = 45.86

22. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The time-series plot shows a downward linear trend.

b.

$$\bar{t} = \frac{\sum_{t=1}^{n} t}{n} = \frac{28}{7} = 4 \qquad \bar{Y} = \frac{\sum_{t=1}^{n} Y_t}{n} = \frac{77}{7} = 11$$

$$\Sigma(t-\bar{t})(Y_t-\bar{Y}) = -19.6 \qquad \Sigma(t-\bar{t})^2 = 28$$

$$b_1 = \frac{\sum_{t=1}^{n}(t-\bar{t})(Y_t-\bar{Y})}{\sum_{t=1}^{n}(t-\bar{t})^2} = \frac{-19.6}{28} = -.7$$

$$b_0 = \bar{Y} - b_1\bar{t} = 11-(-.7)(4) = 13.8$$

$$T_t = 13.8 - .7t$$

c. time period $t = 8$. $T_8 = 13.8 - .7(8) = 8.2$

d. If SCF can continue to decrease the percentage of funds spent on administrative and fund-raising by .7% per year, the forecast of expenses for 2018 is 4.70%.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

24. a.

After declining for several years, the interest rate apparently leveled off in the last few years. Thus, the time-series plot suggests that a quadratic trend may provide a better fit for this time series.

b.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The forecasted value for period 11 is –.3204(11) + 6.624 = 3.10%.

c.

The forecasted value for period 11 is $.0375(11)^2 – .7333(11) + 7.4498 = 3.92\%$.

d. The quadratic trend equation fits the time-series data more closely than does the linear trend equation. Furthermore, the linear trend equation forecasts that the percentage interest rate will continue to decrease indefinitely, which the time-series plot of the data suggests will not happen and is unrealistic. Therefore, the quadratic trend equation appears to be superior to the linear trend equation.

26. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Part a

A linear trend is not appropriate.

b. The following output shows the results of fitting a quadratic trend to the time series.

Part b

The quadratic trend equation is $T_t = 5.702 + 2.889t - .1618t^2$

c. $T_{11} = 5.702 + 2.889(11) - .1618(11)^2 = 17.90$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

28. a.

The time-series plot shows a horizontal pattern, but there is a seasonal pattern in the data. For instance, in each year the lowest value occurs in quarter 2 and the highest value occurs in quarter 4.

b. The regression output follows.

```
The regression equation is

Value = 77.00 - 10.00 Qtr1 - 30.00 Qtr2 - 20.00 Qtr3
```

c. The quarterly forecasts for next year are as follow:

Quarter 1 forecast = 77.0 – 10.0(1) – 30.0(0) – 20.0(0) = 67

Quarter 2 forecast = 77.0 – 10.0(0) – 30.0(1) – 20.0(0) = 47

Quarter 3 forecast = 77.0 – 10.0(0) – 30.0(0) – 20.0(1) = 57

Quarter 4 forecast = 77.0 – 10.0(0) – 30.0(0) – 20.0(0) = 77

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

30. a.

There appears to be a seasonal pattern in the data and perhaps a moderate upward linear trend.

b. The regression output follows.

```
Regression Equation

Sales = 2491.7 - 712 Qtr1 - 1512 Qtr2 + 327 Qtr3
```

c. The quarterly forecasts for next year follow:

Quarter 1 forecast = 2,491.7 – 712(1) – 1,512(0) + 327(0) = 1,779.7 or 1,780

Quarter 2 forecast = 2,491.7 – 712(0) – 1,512(1) + 327(0) = 979.7 or 980

Quarter 3 forecast = 2,491.7 – 712(0) – 1,512(0) + 327(1) = 2,818.7 or 2,819

Quarter 4 forecast = 2,491.7 – 712(0) – 1,512(0) + 327(0) = 2,491.7 or 2,492

d. The regression output follows.

```
Regression Equation

Sales = 2306.7 - 642.3 Qtr1 - 1465.4 Qtr2 + 349.8 Qtr3 + 23.13 t
```

The quarterly forecasts for next year follow:

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Quarter 1 forecast = 2,306.7 – 642.3(1) – 1,465.4(0) + 349.8(0) + 23.13(17) = 2,057.61 or 2,058

Quarter 2 forecast = 2,306.7 – 642.3(0) – 1,465.4(1) + 349.8(0) + 23.13(18) = 1,257.64 or 1,258

Quarter 3 forecast = 2,306.7 – 642.3(0) – 1,465.4(0) + 349.8(1) + 23.13(19) = 3,095.97 or 3,096

Quarter 4 forecast = 2,306.7 – 642.3(0) – 1,465.4(0) + 349.8(0) + 23.13(20) = 2,769.3 or 2,769

32. a.

The time-series plot shows both a linear trend and seasonal effects.

b. A portion of the regression output follows.

```
The regression equation is

      Revenue = 70.0 + 10.0 Qtr1 + 105 Qtr2 + 245 Qtr3
```

Quarter 1 forecast = 70.0 + 10.0(1) + 105(0) + 245(0) = 80

Quarter 2 forecast = 70.0 + 10.0(0) + 105(1) + 245(0) = 175

Quarter 3 forecast = 70.0 + 10.0(0) + 105(0) + 245(1) = 315

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Quarter 4 forecast = 70.0 + 10.0(0) + 105(0) + 245(0) = 70

c. A portion of the regression output follows.

```
The regression equation is

      Revenue = - 70.1 + 45.0 Qtr1 + 128.3 Qtr2 + 256.7 Qtr3 + 11.68

       Period
```

Quarter 1 forecast = –70.1 + 45.0(1) + 128.3(0) + 256.7(0) + 11.68(21) = 220.18

Quarter 2 forecast = –70.1 + 45.0(0) + 128.3(1) + 256.7(0) + 11.68(22) = 315.16

Quarter 3 forecast = –70.1 + 45.0(0) + 128.3(0) + 256.7(1) + 11.68(23) = 455.24

Quarter 4 forecast = –70.1 + 45.0(0) + 128.3(0) + 256.7(0) + 11.68(24) = 210.22

34. a.

The time-series plot shows seasonal and linear trend effects.

b. *Note:* Jan = 1 if January, 0 otherwise; Feb = 1 if February, 0 otherwise; and so on.

The regression output follows.

```
Expense = 174.58 - 18.42 January - 3.72 February
 + 12.66 March + 45.69 April + 57.07 May + 135.10 June
 + 181.48 July + 104.51 August + 47.55 September
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
+ 50.59 October + 35.30 November + 1.962 Period
```

c. *Note:* The next time period in the time series is period = 37 (January of year 4).

January forecast = 174.58 – 18.42(1) – 3.72(0) + 12.66(0) + 45.69(0) + 57.07(0) + 135.1(0) + 181.48(0) + 104.51(0) + 47.59(0) + 50.59(0) + 35.3(0) + 1.962(37) = 229

February forecast = 174.58 – 18.42(0) – 3.72(1) + 12.66(0) + 45.69(0) + 57.07(0) + 135.1(0) + 181.48(0) + 104.51(0) + 47.59(0) + 50.59(0) + 35.3(0) + 1.962(38) = 245

March forecast = 174.58 – 18.42(0) – 3.72(0) + 12.66(1) + 45.69(0) + 57.07(0) + 135.1(0) + 181.48(0) + 104.51(0) + 47.59(0) + 50.59(0) + 35.3(0) + 1.962(39) = 264

April forecast = 174.58 – 18.42(0) – 3.72(0) + 12.66(0) + 45.69(1) + 57.07(0) + 135.1(0) + 181.48(0) + 104.51(0) + 47.59(0) + 50.59(0) + 35.3(0) + 1.962(40) = 299

May forecast = 174.58 – 18.42(0) – 3.72(0) + 12.66(0) + 45.69(0) + 57.07(1) + 135.1(0) + 181.48(0) + 104.51(0) + 47.59(0) + 50.59(0) + 35.3(0) + 1.962(41) = 312

June forecast = 174.58 – 18.42(0) – 3.72(0) + 12.66(0) + 45.69(0) + 57.07(0) + 135.1(1) + 181.48(0) + 104.51(0) + 47.59(0) + 50.59(0) + 35.3(0) + 1.962(42) = 392

July forecast = 174.58 – 18.42(0) – 3.72(0) + 12.66(0) + 45.69(0) + 57.07(0) + 135.1(0) + 181.48(1) + 104.51(0) + 47.59(0) + 50.59(0) + 35.3(0) + 1.962(43) = 440

August forecast = 174.58 – 18.42(0) – 3.72(0) + 12.66(0) + 45.69(0) + 57.07(0) + 135.1(0) + 181.48(0) + 104.51(1) + 47.59(0) + 50.59(0) + 35.3(0) + 1.962(44) = 365

September forecast = 174.58 – 18.42(0) – 3.72(0) + 12.66(0) + 45.69(0) + 57.07(0) + 135.1(0) + 181.48(0) + 104.51(0) + 47.59(1) + 50.59(0) + 35.3(0) + 1.962(45) =

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

310

October forecast = 174.58 – 18.42(0) – 3.72(0) + 12.66(0) + 45.69(0) + 57.07(0) + 135.1(0) + 181.48(0) + 104.51(0) + 47.59(0) + 50.59(1) + 35.3(0) + 1.962(46) = 315

November forecast = 174.58 – 18.42(0) – 3.72(0) + 12.66(0) + 45.69(0) + 57.07(0) + 135.1(0) + 181.48(0) + 104.51(0) + 47.59(0) + 50.59(0) + 35.3(1) + 1.962(47) = 302

December forecast = 174.58 – 18.42(0) – 3.72(0) + 12.66(0) + 45.69(0) + 57.07(0) + 135.1(0) + 181.48(0) + 104.51(0) + 47.59(0) + 50.59(0) + 35.3(0) + 1.962(48) = 269

36. a.

Year	Quarter	Time-Series Value	Adjusted Seasonal Index	Deseasonalized Value
1	1	4	1.205	3.320
	2	2	0.746	2.681
	3	3	0.858	3.501
	4	5	1.191	4.198
2	1	6	1.205	4.979
	2	3	0.746	4.021
	3	5	0.858	5.834
	4	7	1.191	5.877

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

3	1	7	1.205	5.809
	2	6	0.746	8.043
	3	6	0.858	7.001
	4	8	1.191	6.717

b. Let period = 1 denote the time-series value in year 1, quarter 1; period = 2 denote the time-series value in year 1, quarter 2; and so on. The regression output treating period as the independent variable and the deseasonlized values as the values of the dependent variable follows.

```
The regression equation is

      Deseasonalized Value = 2.424 + 0.4217 Period
```

c. The quarterly deseasonalized trend forecasts for year 4 (periods 13, 14, 15, and 16) follow:

Forecast for quarter 1 = 2.424 + 0.4217(13) = 7.906

Forecast for quarter 2 = 2.424 + 0.4217(14) = 8.328

Forecast for quarter 3 = 2.424 + 0.4217(15) = 8.750

Forecast for quarter 4 = 2.424 + 0.4217(16) = 9.171

d. Adjusting the quarterly deseasonalized trend forecasts provides the following quarterly estimates:

Forecast for quarter 1 = 7.906(1.205) = 9.527

Forecast for quarter 2 = 8.328(.746) = 6.213

Forecast for quarter 3 = 8.750(.857) = 7.499

Forecast for quarter 4 = 9.171(1.191) = 10.923

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

38. a.

The time-series plot shows a linear trend and seasonal effects.

b.

Month	Expense	Centered Moving Average	Seasonal Irregular Value
1	170		
2	180		
3	205		
4	230		
5	240		
6	315		
7	360	241.67	1.49
8	290	243.13	1.19

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

9	240	244.58	0.98
10	240	245.63	0.98
11	230	247.29	0.93
12	195	248.96	0.78
13	180	251.25	0.72
14	205	254.79	0.80
15	215	257.50	0.83
16	245	259.58	0.94
17	265	261.88	1.01
18	330	263.96	1.25
19	400	265.63	1.51
20	335	266.46	1.26
21	260	267.29	0.97
22	270	269.38	1.00
23	255	271.88	0.94
24	220	275.42	0.80
25	195	278.75	0.70
26	210	279.38	0.75
27	230	280.42	0.82

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

28	280	282.71	0.99
29	290	284.79	1.02
30	390	287.08	1.36
31	420		
32	330		
33	290		
34	295		
35	280		
36	250		

Month	Seasonal Irregular Values		Seasonal Index
1	0.72	0.70	0.71
2	0.80	0.75	0.78
3	0.83	0.82	0.83
4	0.94	0.99	0.97
5	1.01	1.02	1.02
6	1.25	1.36	1.30
7	1.49	1.51	1.50
8	1.19	1.26	1.23
9	0.98	0.97	0.98

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

10	0.98	1.00	0.99
11	0.93	0.94	0.93
12	0.78	0.80	0.79
		Total	12.03

Notes: (1) Adjustment for seasonal index = 12 / 12.03 = 0.998. (2) Because the seasonal irregular values and the seasonal index values were rounded to two decimal places to simplify the presentation, the adjustment is really not necessary in this problem because it implies more accuracy than is warranted.

c.

Month	Expense	Seasonal Index	Deseasonalized Expense
1	170	0.71	239.44
2	180	0.78	230.77
3	205	0.83	246.99
4	230	0.97	237.11
5	240	1.02	235.29
6	315	1.3	242.31
7	360	1.5	240.00
8	290	1.23	235.77
9	240	0.98	244.90
10	240	0.99	242.42

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

11	230	0.93	247.31
12	195	0.79	246.84
13	180	0.71	253.52
14	205	0.78	262.82
15	215	0.83	259.04
16	245	0.97	252.58
17	265	1.02	259.80
18	330	1.3	253.85
19	400	1.5	266.67
20	335	1.23	272.36
21	260	0.98	265.31
22	270	0.99	272.73
23	255	0.93	274.19
24	220	0.79	278.48
25	195	0.71	274.65
26	210	0.78	269.23
27	230	0.83	277.11
28	280	0.97	288.66
29	290	1.02	284.31

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

30	390	1.3	300.00
31	420	1.5	280.00
32	330	1.23	268.29
33	290	0.98	295.92
34	295	0.99	297.98
35	280	0.93	301.08
36	250	0.79	316.46

d. Let period = 1 denote the time-series value in January, year 1; period = 2 denote the time-series value in February, year 2; and so on. The regression output treating period as the independent variable and the deseasonlized values as the values of the dependent variable follows.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression Period	1	14932	14932.2	265.05	0.000
Error	34	1915	56.3		
Total	35	16848			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
7.50580	88.63%	88.30%	86.77%

Coefficients

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	228.01	2.55	89.24	0.000	
Period	1.960	0.120	16.28	0.000	1.00

Regression Equation

Deseasonalized Expense = 228.01 + 1.960 Period

Fits and Diagnostics for Unusual Observations

Obs	Deseasonalized Expense	Fit	Resid	Std Resid
32	268.29	290.75	-22.46	-3.11 R
36	316.46	298.59	17.87	2.52 R

e. The linear trend estimates for the deseasonalized time series and the adjustment based on the seasonal effects follow.

Month	Deseasonalized Trend Forecast	Seasonal Index	Monthly Forecast
January	300.53	0.71	213.38
February	302.49	0.78	235.94
March	304.45	0.83	252.69
April	306.41	0.97	297.22
May	308.37	1.02	314.54
June	310.33	1.3	403.43
July	312.29	1.5	468.44

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

August	314.25	1.23	386.53
September	316.21	0.98	309.89
October	318.17	0.99	314.99
November	320.13	0.93	297.72
December	322.09	0.79	254.45

For instance, using the estimated regression equation the deseasonalized trend forecast for January in year 4 (period = 37) is 228.01 + 1.96(37) = 300.53. Because the January seasonal index is .71, the forecast for January is .71(300.53) = $213.38, or approximately $213. The other monthly forecasts were computed in a similar fashion.

40. a.

The time-series plot indicates a seasonal effect. Power consumption is lowest in the time period 12–4 AM, steadily increases to the highest value in the 12–4 PM time period, and then decreases again. There may also be some linear trend in the data.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b.

Day	Time Period	Power	Centered Moving Average	Seasonal Irregular Values
Monday	12–4 PM	124,299		
Monday	4–8 PM	113,545		
Monday	8–12 midnight	41,300		
Tuesday	12–4 AM	19,281	71,803.3	0.2685
Tuesday	4–8 AM	33,195	71,598.2	0.4636
Tuesday	8–12 noon	99,516	72,013.5	1.3819
Tuesday	12–4 PM	123,666	73,575.2	1.6808
Tuesday	4–8 PM	111,717	74,887.4	1.4918
Tuesday	8–12 midnight	48,112	76,910.0	0.6256
Wednesday	12–4 AM	31,209	81,311.6	0.3838
Wednesday	4–8 AM	37,014	85,439.8	0.4332
Wednesday	8–12 noon	119,968	89,021.8	1.3476
Wednesday	12–4 PM	156,033	90,849.4	1.7175
Wednesday	4–8 PM	128,889	90,167.9	1.4294
Wednesday	8–12 midnight	73,923	92,517.8	0.7990
Thursday	12–4 AM	27,330		

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Thursday	4–8 AM	32,715
Thursday	8–12 noon	152,465

Time Period	Seasonal Irregular Values		Seasonal Index	Adjusted Seasonal Index
12–4 AM	0.3838	0.2685	0.3262	0.3256
4–8 AM	0.4332	0.4636	0.4484	0.4476
8–12 noon	1.3476	1.3819	1.3648	1.3622
12–4 PM	1.6808	1.7175	1.6992	1.6959
4–8 PM	1.4918	1.4294	1.4606	1.4578
8–12 midnight	0.6256	0.7990	0.7123	0.7109
		Total	6.0114	

c.

Day	Time Period	Power	Adjusted Seasonal Index	Deseasonalized Power
Monday	12–4 PM	124,299	1.6959	73,292.80
Monday	4–8 PM	113,545	1.4578	77,885.67
Monday	8–12 midnight	41,300	0.7109	58,092.48
Tuesday	12–4 AM	19,281	0.3256	59,225.36

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Tuesday	4–8 AM	33,195	0.4476	74,166.93
Tuesday	8–12 noon	99,516	1.3622	73,056.72
Tuesday	12–4 PM	123,666	1.6959	72,919.55
Tuesday	4–8 PM	111,717	1.4578	76,631.76
Tuesday	8–12 midnight	48,112	0.7109	67,674.22
Wednesday	12–4 AM	31,209	0.3256	95,864.54
Wednesday	4–8 AM	37,014	0.4476	82,699.65
Wednesday	8–12 noon	119,968	1.3622	88,070.95
Wednesday	12–4 PM	156,033	1.6959	92,004.73
Wednesday	4–8 PM	128,889	1.4578	88,410.82
Wednesday	8–12 midnight	73,923	0.7109	103,979.91
Thursday	12–4 AM	27,330	0.3256	83,949.43
Thursday	4–8 AM	32,715	0.4476	73,094.48
Thursday	8–12 noon	152,465	1.3622	111,927.66

The following output shows the results of fitting a linear trend equation to the deseasonalized time series.

```
Coefficients

Term            Coef      SE Coef    T-Value    P-Value    VIF

Constant        63108     5185       12.17      0.000
```

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
t                1854      479        3.87       0.001      1.00

Regression Equation

      Deseasonalized Power = 63108 + 1854 t
```

Using the deseasonalized linear trend equation and the seasonal indexes computed in part b, we can forecast power consumption for the 12–4 PM and 4–8 PM time periods.

12–4 PM (corresponds to $t = 19$)

Deseasonalized power = 63,108 + 1,854(19) = 98,334.

Seasonal index for this period = 1.6959.

Forecast for 12–4 PM = 1.6959(98,334) = 166,764.63 or approximately 166,765 kWh.

4–8 PM (corresponds to $t = 20$)

Deseasonalized power = 63,108 + 1,854(20) = 100,188.

Seasonal index for this period = 1.4578.

Forecast for 4–8 PM = 1.4578(100,188) = 146,054.07 or approximately 146,054 kWh.

Thus, the forecast of power consumption from noon to 8 PM is 166,765 + 146,054 = 312,819 kWh.

42. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The time-series plot indicates a horizontal pattern.

b.

Period	Time-Series Value	α = .2 Forecasts	α = .3 Forecasts	α = .4 Forecasts
1	29.8			
2	31.0	29.80	29.80	29.80
3	29.9	30.04	30.16	30.28
4	30.1	30.01	30.08	30.13
5	32.2	30.03	30.09	30.12
6	31.5	30.46	30.72	30.95
7	32.0	30.67	30.95	31.17
8	31.9	30.94	31.27	31.50
9	30.0	31.13	31.46	31.66

MSE(α = .2) = 11.22/8 = 1.40

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

MSE(α = .3) = 10.19/8 = 1.27

MSE(α = .4) = 9.83/8 = 1.23

A smoothing constant of α = .4 provides the best forecast because it has a smaller MSE.

c. Using α = .4, F_{10} = .4(30) + .6(31.66) = 31.00

44. a.

There appears to be an increasing trend in the data through t = 16 followed by periods of decreasing and increasing cost.

b. Using Excel's Regression tool, the estimated regression equation is:

Cost = 81.29 + .58t

The forecast for t = 49 is Cost = 81.29 + .58(49) = \$109.71.

c. Using Excel's Regression tool, the estimated multiple regression equation is:

Cost = 68.82 + 2.08t – .03t^2

The forecast for t = 49 is Cost = 68.82 + 2.08(49) – .03$(49)^2$ = \$98.71.

d.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Linear trend equation: MSE = 69.6

Quadratic trend equation: MSE = 41.9

The quadratic trend equation provides the best forecast accuracy for the historical data.

46. a. The following table shows the calculations using a smoothing constant of .4.

Month	Sales ($1000s)	Forecast
January	185.72	
February	167.84	185.72
March	205.11	178.57
April	210.36	189.18
May	255.57	197.65
June	261.19	220.82

The forecast for July is .4(261.19) + .6(220.82) = 236.97.

The forecast for August, using forecast for July as the actual sales in July, is 236.97.

Exponential smoothing provides the same forecast for every period in the future, which is why it is not usually recommended for long-term forecasting.

b. Using regression, we obtained the linear trend equation:

$T_t = 149.72 + 18.451t$

Forecast for July (t = 7) is 149.72 + 18.451(7) = 278.88

Forecast for August (t = 8) is 149.72 + 18.451(8) = is 297.33

c. The proposed settlement is not fair because it does not account for the upward trend

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

in sales. Based on trend projection, the settlement should be based on forecasted lost sales of $278,880 in July and $297,330 in August.

48. a.

The time-series plot shows a linear trend.

b.

$$\bar{t} = \frac{\sum_{t=1}^{n} t}{n} = \frac{15}{5} = 3 \qquad \bar{Y} = \frac{\sum_{t=1}^{n} Y_t}{n} = \frac{200}{5} = 40$$

$$\Sigma(t-\bar{t})(Y_t-\bar{Y}) = 150 \qquad \Sigma(t-\bar{t})^2 = 10$$

$$b_1 = \frac{\sum_{t=1}^{n}(t-\bar{t})(Y_t-\bar{Y})}{\sum_{t=1}^{n}(t-\bar{t})^2} = \frac{150}{10} = 15$$

$$b_0 = \bar{Y} - b_1\bar{t} = 40 - (15)(3) = -5$$

$$T_t = -5 + 15t$$

The slope of 15 indicates that the average increase in sales is 15 pianos per year.

c. Forecast for year 6 is $T_6 = -5 + 15(6) = 85$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Forecast for year 7 is $T_7 = -5 + 15(7) = 100$.

50. a.

t	Sales	Centered Moving Average	Seasonal Irregular Value
1	4		
2	2		
3	1	3.250	0.308
4	5	3.750	1.333
5	6	4.375	1.371
6	4	5.875	0.681
7	4	7.500	0.533
8	14	7.875	1.778
9	10	7.875	1.270
10	3	8.250	0.364
11	5	8.750	0.571
12	16	9.750	1.641
13	12	10.750	1.116
14	9	11.750	0.766
15	7	13.250	0.528

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

16	22	14.125	1.558
17	18	15.000	1.200
18	10	17.375	0.576
19	13		
20	35		

Quarter	Seasonal Irregular Values	Seasonal Index
1	1.371, 1.270, 1.116, 1.200	1.2394
2	0.681, 0.364, 0.766, 0.576	0.5965
3	0.308, 0.533, 0.571, 0.528	0.4852
4	1.333, 1.778, 1.641, 1.558	1.5774
	Total	3.8985

Quarter	Adjusted Seasonal Index
1	1.2717
2	0.6120
3	0.4978
4	1.6185

Note: Adjustment for seasonal index = 4 / 3.8985 = 1.0260.

b. The largest effect is in quarter 4; this seems reasonable because retail sales are generally higher during October, November, and December.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

52. a.

A linear trend pattern appears to be present in the time-series plot.

b. The regression output follows.

```
Coefficients

Term           Coef      SE Coef    T-Value   P-Value    VIF

Constant       22.86     4.50       5.08      0.004

Year           15.54     1.01       15.43     0.000      1.00

Regression Equation

      Number Sold = 22.86 + 15.54 Year
```

c. Forecast in year 8 = 22.86 +15.54(8) = 147.18 or approximately 147 units.

54. a.

t	Sales	Centered Moving Average
1	6	
2	15	

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

3	10	9.250
4	4	10.125
5	10	11.125
6	18	12.125
7	15	13.000
8	7	14.500
9	14	16.500
10	26	18.125
11	23	19.375
12	12	20.250
13	19	20.750
14	28	21.750
15	25	22.875
16	18	24.000
17	22	25.125
18	34	25.875
19	28	26.500
20	21	27.000
21	24	27.500

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

22	36	27.625
23	30	28.000
24	20	29.000
25	28	30.125
26	40	31.625
27	35	
28	27	

b.

The centered moving average values smooth out the time series by removing seasonal effects and some of the random variability. The centered moving average time series shows the trend in the data.

c.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

t	Sales	Centered Moving Average	Seasonal Irregular Value
1	6		
2	15		
3	10	9.250	1.081
4	4	10.125	0.395
5	10	11.125	0.899
6	18	12.125	1.485
7	15	13.000	1.154
8	7	14.500	0.483
9	14	16.500	0.848
10	26	18.125	1.434
11	23	19.375	1.187
12	12	20.250	0.593
13	19	20.750	0.916
14	28	21.750	1.287
15	25	22.875	1.093
16	18	24.000	0.750
17	22	25.125	0.876
18	34	25.875	1.314
19	28	26.500	1.057
20	21	27.000	0.778
21	24	27.500	0.873
22	36	27.625	1.303
23	30	28.000	1.071

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

24	20	29.000	0.690
25	28	30.125	0.929
26	40	31.625	1.265
27	35		
28	27		

Quarter	Seasonal Irregular Component Values	Seasonal Index
1	0.899, 0.848, 0.916, 0.876, 0.873, 0.929	0.890
2	1.485, 1.434, 1.287, 1.314, 1.303, 1.265	1.348
3	1.081, 1.154, 1.187, 1.093, 1.057, 1.071	1.107
4	0.395, 0.483, 0.593, 0.750, 0.778, 0.690	0.615
	Total	3.960

Quarter	Adjusted Seasonal Index
1	0.899
2	1.362
3	1.118
4	0.621

Note: Adjustment for seasonal index = 4.00 / 3.96 = 1.0101.

d. Hudson Marine experiences the largest seasonal increase in quarter 2. Because this quarter occurs before the peak summer boating season, this result seems reasonable. But the largest seasonal effect is the seasonal decrease in quarter 4. This is also reasonable because of decreased boating in the fall and winter.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 18 Solutions

2. Let p = the probability of a preference for brand A

H_0: $p = .50$

H_a: $p \neq .50$

Dropping the individual with no preference, the binomial probabilities for $n = 9$ and $p = .50$ are as follows.

x	Probability
0	0.0020
1	0.0176
2	0.0703
3	0.1641
4	0.2461
5	0.2461
6	0.1641
7	0.0703
8	0.0176
9	0.0020

Number of plus signs = 7.

$P(x \geq 7) = P(7) + P(8) + P(9) = .0703 + .0176 + .0020 = .0899$

Two-tailed p-value = 2(.0899) = .1798.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

p-value > .05; do not reject H_0. There is no indication that a difference in preference exists. A larger sample size should be considered.

4. a. H_0: Median $\geq$ 15

   H_a: Median < 15

   b. Dropping Vanguard GNMA with net assets of $15 billion, the binomial probabilities for n = 9 and p = .50 are as follow.

x	Probability
0	.0020
1	.0176
2	.0703
3	.1641
4	.2461
5	.2461
6	.1641
7	.0703
8	.0176
9	.0020

Number of plus signs = 1 American Funds with net assets of $22.4 billion.

$P(x \leq 1) = P(0) + P(1) = .0020 + .0176 = .0196$

p-value = .0196.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

p-value $\leq$.05; reject H_0. Conclude that Bond Mutual Funds have a significantly lower median net assets than Stock Mutual Funds.

6. H_0: Median $\leq$ 56.2

H_a: Median > 56.2

There are n = 31 + 17 = 48 observations where the value is different from 56.2.

Use the normal approximation with

$\mu = .5n = .5(48) = 24$ and

$$\sigma = \sqrt{.25n} = \sqrt{.25(48)} = 3.4641$$

With the number of plus signs = 31in the upper tail, we use the continuity correction factor and the normal distribution approximation as follows

P(31or more plus signs) = $P(x \geq 30.5) = P\left(z \geq \frac{30.5 - 24}{3.4641}\right) = P(z \geq 1.88)$

Upper-tail p-value = (1.0000 – .9699) = .0301.

p-value $\leq$ .05; reject H_0. Conclude that the median annual income for families living in Chicago is greater that \$56.2 thousand.

8. a. H_0: p = .50

H_a: $p \neq$ .50

Dropping the individual with no preference, selected binomial probabilities for n = 15 and p = .50 are as follow.

x	Probability
0	.0000

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

1	.0005
2	.0032
3	.0139
4	.0417

Let the response faster pace be a plus sign. Number of plus signs = 4.

Because this is in the lower tail of binomial distribution, we compute

$P(x \leq 4) = P(0) + P(1) + P(2) + P(3) + P4) = .0000 + .0005 + .0032 + .0139 + .0417 = .0593$

Two-tailed p-value = 2(.0593) = .1186,

p-value > .05; do not reject H_0. The sample does not allow the conclusion that a difference in preference exists for the fast pace or slower pace of life.

b. Of the original 16 respondents, 4/16(100) = 25% favored a faster pace of life and 11/16(100) = 68.8% favored a slower pace of life. Almost 3 to 1 favor the slower pace of life. In addition, the p-value (.1186) is low but not low enough to detect a difference in preference. Continuing to study and taking a larger sample should be considered.

10. Let p = proportion who favor *This Is Us*.

H_0: $p = .50$

H_a: $p \neq .50$

Eliminating the responses that selected a different television show, we have

$n = 330 + 270 = 600$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Using the normal distribution, we have

$\mu = .5\ n = .5(600) = 300$

$\sigma = \sqrt{.25n} = \sqrt{.25(600)} = 12.2474$

With the number of plus signs = 330 in the upper tail, we will use the continuity correction factor and the normal distribution approximation as follows:

P(330 or more plus signs) = $P(x \geq 329.5) = P\left(z \geq \frac{329.5 - 300}{12.2474}\right) = P(z \geq 2.41)$

The two-tailed p-value = 2(1.0000 – .9920) = .0160.

p-value $\leq$ .05; reject H_0.. Conclude that there is a significant difference between the preference for *This Is Us* and *The Big Bang Theory.* Based on the data, *This Is Us* is most preferred.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

12. H_0: Median for Additive 1 – Median for Additive 2 = 0

H_a: Median for Additive 1 – Median for Additive 2 ≠ 0

	Additive					Signed Ranks	
Car	1	2	Difference	Absolute Difference	Rank	Negative	Positive
1	20.12	18.05	2.07	2.07	9		9
2	23.56	21.77	1.79	1.79	7		7
3	22.03	22.57	–0.54	0.54	3	–3	
4	19.15	17.06	2.09	2.09	10		10
5	21.23	21.22	0.01	0.01	1		1
6	24.77	23.80	0.97	0.97	4		4
7	16.16	17.20	–1.04	1.04	5	–5	
8	18.55	14.98	3.57	3.57	12		12
9	21.87	20.03	1.84	1.84	8		8

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

10	24.23	21.15	3.08	3.08	11	11
11	23.21	22.78	0.43	0.43	2	2
12	25.02	23.70	1.32	1.32	6	6
				Sum of Positive Signed Ranks		$T^+ = 70$

$$\mu_{T^+} = \frac{n(n+1)}{4} = \frac{12(13)}{4} = 39$$

$$\sigma_{T^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{12(13)(25)}{24}} = 12.7475$$

$T^+ = 70$ is in the upper tail of the sampling distribution. Using the continuity correction factor, we have:

$$P(T^+ \geq 70) = P\left(z \geq \frac{69.5 - 39}{12.7475}\right) = P(z \geq 2.39)$$

Using $z = 2.39$, the p-value = 2(1.0000 – .9916) = .0168.

p-value $\leq .05$; reject H_0. Conclude that there is a significant difference in the median miles per gallon for the two additives.

14. H_0: Median percent on-time in 2006 – Median percent on time in 2007 = 0

H_a: Median percent on-time in 2006 – Median percent on time in 2007 $\neq$ 0

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	Year			Absolute		Signed Ranks	
Airport	2006 (%)	2007 (%)	Difference	Difference	Rank	Negative	Positive
1	71.78	69.69	2.09	2.09	4		4
2	68.23	65.88	2.35	2.35	5		5
3	77.98	78.40	–0.42	0.42	2	–2	
4	78.71	75.78	2.93	2.93	6		6
5	77.59	73.45	4.14	4.14	9		9
6	77.67	78.68	–1.01	1.01	3	–3	
7	76.67	76.38	0.29	0.29	1		1
8	76.29	70.98	5.31	5.31	10		10
9	69.39	62.84	6.55	6.55	11		11
10	79.91	76.49	3.42	3.42	8		8
11	75.55	72.42	3.13	3.13	7		7

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Sum of Positive Signed Ranks $T^+ = 61$

$$\mu_{T^+} = \frac{n(n+1)}{4} = \frac{11(12)}{4} = 33$$

$$\sigma_{T^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{11(12)(23)}{24}} = 11.2472$$

$T^+ = 61$ is in the upper tail of the sampling distribution. Using the continuity correction factor, we have:

$$P(T^+ \geq 61) = P\left(z \geq \frac{60.5 - 33}{11.2472}\right) = P(z \geq 2.45)$$

Using $z = 2.45$, p-value = 2(1.0000 – .9929) = .0142.

p-value $\leq$ .05; reject H_0. Conclude that there is a significance difference between the median percentage of on-time flights for the two years. Median percentage of on-time flights was better in the earlier of the two years for which data were collected.

16. H_0: Median score for Round 1 – Median score for Round 2 = 0

H_a: Median score for Round 1 – Median score for Round 2 $\neq$ 0

	Round					Signed Ranks	
Golfer	1	2	Difference	Absolute Difference	Rank	Negative	Positive

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Brittany Lang	73	72	1	1	1.5		1.5
Amy Anderson	74	70	4	4	7		7
Meena Lee	71	73	–2	2	4	–4	
Juli Inkster	69	75	–6	6	8.5	–8.5	
Ha Na Jang	72	72	0				
Haeji Kang	71	74	–3	3	6	–6	
Ai Miyazato	68	74	–6	6	8.5	–8.5	
Stephanie Meadow	76	68	8	8	11.5		11.5
Catriona Matthew	71	69	2	2	4		4
Sandra Gal	75	68	7	7	10		10
Caroline Masson	72	73	–1	1	1.5	–1.5	
Suzann Pettersen	76	68	8	8	11.5		11.5
Mo Martin	74	72	2	2	4		4

Sum of Positive Signed Ranks $T^{+} = 49.5$.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Ha Na Jang had the same score on both rounds and is removed from the sample. Thus, $n = 12$.

$$\mu_{T^+} = \frac{n(n+1)}{4} = \frac{12(13)}{4} = 39$$

$$\sigma_{T^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{12(13)(25)}{24}} = 12.7475$$

$T^+ = 49.5$ is in the upper tail of the sampling distribution. Using the continuity correction factor, we have:

$$P(T^+ \geq 49.5) = P\left(z \geq \frac{49-39}{12.7475}\right) = P(z \geq .78)$$

Using $z = .78$, p-value = 2(1 – 7,823) = .4354.

p-value > .05; do not reject H_0. There is no significant difference between the median scores for the two rounds of golf.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

18. H_0: The two populations of additives are identical

H_a: The two populations of additives are not identical

Additive 1	Rank	Additive 2	Rank
17.3	2	18.7	8.5
18.4	6	17.8	4
19.1	10	21.3	15
16.7	1	21.0	14
18.2	5	22.1	16
18.6	7	18.7	8.5
17.5	3	19.8	11
		20.7	13
		20.2	12

$W = 34$

$$\mu_W = \frac{1}{2}n_1(n_1 + n_2 + 1) = \frac{1}{2}7(7+9+1) = 59.5$$

$$\sigma_W = \sqrt{\frac{1}{12}n_1 n_2(n_1 + n_2 + 1)} = \sqrt{\frac{1}{12}7(9)(7+9+1)} = 9.4472$$

With $W = 34$ in the lower tail, we will use the continuity correction factor and the normal distribution approximation as follows:

$$P(W \le 34) = P\left(z \le \frac{34.5 - 59.5}{9.4472}\right) = P(z \le -2.65)$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Using $z = -2.65$, the two-tailed p-value =2(.0040) = .0080.

p-value < .05; reject H_0.

Conclusion: The two populations of fuel additives are not identical. The population with additive 2 tends to provide higher miles per gallon.

20. a. Median salary for men \$54,900 (4th position when ranked)

Median salary for women \$40,400 (4th position when ranked)

b. H_0: The two populations of salaries are identical

H_a: The two populations of salaries are not identical

Men	Rank	Women	Rank
35.6	4	49.5	8
80.5	14	40.4	5
50.2	9	32.9	3
67.2	13	45.5	7
43.2	6	30.8	2
54.9	11	52.5	10
60.3	12	29.8	1
$W =$	69		

$$\mu_W = \frac{1}{2}n_1(n_1 + n_2 + 1) = \frac{1}{2}7(7 + 7 + 1) = 52.5$$

$$\sigma_W = \sqrt{\frac{1}{12}n_1 n_2(n_1 + n_2 + 1)} = \sqrt{\frac{1}{12}7(7)(7 + 7 + 1)} = 7.8262$$

With $W = 69$ in the upper tail, we will use the continuity correction factor and the

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

normal distribution approximation as follows:

$$P(W \geq 69) = P\left(z \geq \frac{68.5 - 52.5}{7.8262}\right) = P(z \geq 2.04)$$

Using $z = 2.04$, the two-tailed p-value =2(1.000–.9793) = .0414.

p-value $\leq$ .05; reject H_0.

Conclusion: The two populations of salaries are not identical. The population of men tends to have the higher salaries.

22. H_0: The two populations of P/E ratios are identical

H_a: The two populations of P/E ratios are not identical

Japan		U.S.	
P/E Ratio	Rank	P/E Ratio	Rank
153	20	19	6
21	8	24	11.5
18	5	24	11.5
125	19	43	16
31	13	22	10
213	21	14	2
64	17	21	8
666	22	14	2
33	14	21	8

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

68	18	38	15
		15	4
		14	2

$W = 157$

$$\mu_W = \frac{1}{2}n_1(n_1 + n_2 + 1) = \frac{1}{2}10(10 + 12 + 1) = 115$$

$$\sigma_W = \sqrt{\frac{1}{12}n_1 n_2(n_1 + n_2 + 1)} = \sqrt{\frac{1}{12}10(12)(10 + 12 + 1)} = 15.1658$$

With $W = 157$ in the upper tail, we will use the continuity correction factor and the normal distribution approximation as follows:

$$P(W \geq 157) = P\left(z \geq \frac{156.5 - 115}{15.1658}\right) = P(z \geq 2.74)$$

Using $z = 2.74$, the two-tailed p-value =2(1.0000 – .9969) = .0062.

p-value $\leq$ .01; reject H_0.

Conclusion: The two populations of P/E ratios are not identical. The population of Japanese companies tends to have higher P/E ratios than the population of United States companies.

24. H_0: The two populations of microwave prices are identical

H_a: The two populations of microwave prices are not identical

Dallas	Rank	San Antonio	Rank
445	14.5	460	19
489	23	451	17

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

405	1.5	435	11
485	22	479	21
439	13	475	20
449	16	445	14.5
436	12	429	7
420	4	434	10
430	8.5	410	3
405	1.5	422	5
		425	6
		459	18
		430	8.5

$W = 116$

$$\mu_W = \frac{1}{2}n_1(n_1 + n_2 + 1) = \frac{1}{2}10(10 + 13 + 1) = 120$$

$$\sigma_W = \sqrt{\frac{1}{12}n_1 n_2(n_1 + n_2 + 1)} = \sqrt{\frac{1}{12}10(13)(10 + 13 + 1)} = 16.1245$$

With $W = 116$ in the lower tail, we will use the continuity correction factor and the normal distribution approximation as follows:

$$P(W \le 116) = P\left(z \le \frac{116.5 - 120}{16.1245}\right) = P(z \le -.22)$$

Using $z = -.22$, the two-tailed p-value =2(.4129) = .8258.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

p-value > .05; do not reject H_0.

Conclusion: We cannot reject the null hypothesis that the two populations of microware prices are identical. There is no indication that there is a difference between the populations of microwave prices for the two cities.

26. H_0: All populations of product ratings are identical

H_a: Not all populations of product ratings are identical

A	Rank	B	Rank	C	Rank
50	4	80	11	60	7
62	8	95	14	45	2
75	10	98	15	30	1
48	3	87	12	58	6
65	9	90	13	57	5
Sum of Ranks	34		65		21

$$H = \left[\frac{12}{n_T(n_T+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right] - 3(n_T+1) = \left[\frac{12}{15(16)}\left(\frac{34^2}{5}+\frac{65^2}{5}+\frac{21^2}{5}\right)\right] - 3(16) = 10.22$$

Using the χ^2 table with $df = 2$, $\chi^2 = 10.22$ shows the p-value is between .005 and .01.

Using Excel, the p-value for $\chi^2 = 10.22$ with two degrees of freedom is .0060.

p-value $\leq$ .05; reject H_0. Conclude that the populations of product ratings are not identical.

28. H_0: All populations of calories burned are identical

H_a: Not all populations of calories burned are identical

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Swimming	Rank	Tennis	Rank	Cycling	Rank
408	8	415	9	385	5
380	4	485	14	250	1
425	11	450	13	295	3
400	6	420	10	402	7
427	12	530	15	268	2
Sum of Ranks	41		61		18

$$H = \left[\frac{12}{n_T(n_T+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right] - 3(n_T+1) = \left[\frac{12}{15(16)}\left(\frac{41^2}{5}+\frac{61^2}{5}+\frac{18^2}{5}\right)\right] - 3(16) = 9.26$$

Using the χ^2 table with $df = 2$, $\chi^2 = 9.26$ shows the p-value is between .005 and .01.

Using Excel, the p-value for $\chi^2 = 9.26$ with two degrees of freedom is .0098.

p-value $\leq .05$; reject H_0. Conclude that the populations of calories burned by the three activities are not identical. Cycling tends to have the lowest calories burned.

30. H_0: All populations of training courses are identical

H_a: Not all populations of training courses are identical

Course

A	B	C	D
3	2	19	20
14	7	16	4
10	1	9	15

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	12	5	18	6
	13	11	17	8
Sum of Ranks	52	26	79	53

$$H = \left[\frac{12}{n_T(n_T+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right] - 3(n_T+1) = \left[\frac{12}{20(21)}\left(\frac{52^2}{5}+\frac{26^2}{5}+\frac{79^2}{5}+\frac{53^2}{5}\right)\right] - 3(21) = 8.03$$

Using the χ^2 table with $df = 3$, $\chi^2 = 8.03$ shows the p-value is between .025 and .05.

Using Excel, the p-value for $\chi^2 = 8.03$ with three degrees of freedom is .0454.

p-value $\leq$ .05; reject H_0. Conclude that the populations are not identical. There is a significant difference in the quality of the courses offered at the four management development centers.

32. a. $\Sigma d_i^2 = 52$

$$r_s = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)} = 1 - \frac{6(52)}{10(99)} = .685$$

b. $\sigma_{r_s} = \sqrt{\frac{1}{n-1}} = \sqrt{\frac{1}{9}} = .3333$

$$z = \frac{r_s - 0}{\sigma_{r_s}} = \frac{.685}{.3333} = 2.05$$

p-value = 2(1.0000 – .9798) = .0404.

p-value $\leq$ .05; reject H_0. Conclude that significant positive rank correlation exists.

34. $\Sigma d_i^2 = 250$

$$r_s = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)} = 1 - \frac{6(250)}{11(120)} = -.136$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$\sigma_{r_s} = \sqrt{\frac{1}{n-1}} = \sqrt{\frac{1}{10}} = .3162$$

$$z = \frac{r_s - 0}{\sigma_{r_s}} = \frac{-.136}{.3162} = -.43$$

p-value = 2(.3336) = .6672

p-value > .05; do not reject H_0. We cannot conclude there is a significant relationship between ranking based on the expenditure per student and the ranking based on the student–teacher ratio.

36.

Driving Distance	Putting	d_i	d_i^2
1	5	–4	16
5	6	–1	1
4	10	–6	36
9	2	7	49
6	7	–1	1
10	3	7	49
2	8	–6	36
3	9	–6	36
7	4	3	9
8	1	7	49

$$\Sigma d_i^2 = 282$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$r_s = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)} = 1 - \frac{6(282)}{10(100-1)} = -.709$$

$$\sigma_{r_s} = \sqrt{\frac{1}{n-1}} = \sqrt{\frac{1}{9}} = .3333$$

$$z = \frac{r_s - \mu_{r_s}}{\sigma_{r_s}} = \frac{-.709-0}{.3333} = -2.13$$

p-value = 2(.0166) = .0332.

p-value $\leq$ .10; reject H_0. There is a significant rank correlation between driving distance and putting. Professional golfers who rank high in driving distance tend to rank low in putting.

38. Let p = proportion who favor the proposal.

H_0: $p = .50$

H_a: $p \neq .50$

Eliminating the 60 responses that offered no opinion, we have

$n = 905 + 1045 = 1950$

Using the normal distribution, we have

$\mu = .5\ n = .5(1950) = 975$

$$\sigma = \sqrt{.25n} = \sqrt{.25(1950)} = 22.0794$$

With the number of plus signs = 905 in the lower tail, we will use the continuity correction factor and the normal distribution approximation as follows:

$$P(\text{905 or fewer plus signs}) = P(x \leq 905.5) = P\left(z \leq \frac{905.5-975}{22.0794}\right) = P(z \leq -3.15)$$

Area in lower tail area is less than .0010.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The two-tailed p-value is less than 2(.0010) = .0020.

p-value $\leq$ .05; reject H_0. Conclude there is a significant difference between the preferences for the tax-funded vouchers or tax deductions for parents who send their children to private schools. The greater percentage opposed the proposal.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

40. H_0: Median price for Model 1 – Median price for Model 2 = 0

H_a: Median price for Model 1 – Median price for Model 2 ≠ 0

						Signed Ranks	
Homemaker	Model 1	Model 2	Difference	Absolute Difference	Rank	Negative	Positive
1	850	1100	–250	250	11	–11	
2	960	920	40	40	2		2
3	940	890	50	50	3		3
4	900	1050	–150	150	6	–6	
5	790	1120	–330	330	12	–12	
6	820	1000	–180	180	7	–7	
7	900	1090	–190	190	8.5	–8.5	
8	890	1120	–230	230	10	–10	
9	1100	1200	–100	100	5	–5	
10	700	890	–190	190	8.5	–8.5	

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

11	810	900	–90	90	4	–4	
12	920	900	20	20	1		1
				Sum of Positive Signed Ranks			$T^+ = 6$

$$\mu_{T^+} = \frac{n(n+1)}{4} = \frac{12(13)}{4} = 39$$

$$\sigma_{T^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{12(13)(25)}{24}} = 12.7475$$

$T^+ = 6$ is in the lower tail of the sampling distribution. Using the continuity correction factor, we have:

$$P(T^+ \le 6) = P\left(z \le \frac{6.5 - 39}{12.7475}\right) = P(z \le -2.55)$$

Using $z = -2.55$, the p-value = 2(.0054) = .0108.

p-value $\le$ .05; reject H_0. Conclude there is a significance difference between the median prices of the two models.

The housewives are estimating model 2 to have the higher median price.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

42. H_0: The two populations of product weights are identical

H_a: The two populations of product weights are not identical

Line 1	Rank	Line 2	Rank
13.6	6.5	13.7	8
13.8	9	14.1	13
14.0	11.5	14.2	14
13.9	10	14.0	11.5
13.4	4	14.6	19
13.2	2	13.5	5
13.3	3	14.4	16.5
13.6	6.5	14.8	20
12.9	1	14.5	18
14.4	16.5	14.3	15
		15.0	22
		14.9	21

$W = 70$

$$\mu_W = \frac{1}{2}n_1(n_1 + n_2 + 1) = \frac{1}{2}10(10 + 12 + 1) = 115$$

$$\sigma_W = \sqrt{\frac{1}{12}n_1 n_2(n_1 + n_2 + 1)} = \sqrt{\frac{1}{12}10(12)(10 + 12 + 1)} = 15.1658$$

With $W = 70$ is in the lower tail, we will use the continuity correction factor and the

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

normal distribution approximation as follows:

$$P(W \le 70) = P\left(z \le \frac{70.5 - 115}{15.1658}\right) = P(z \le -2.93)$$

Using z = –2.93, the two-tailed p-value = 2(.0017) = .0034.

p-value $\le$ .05; reject H_0.

Conclusion: The two populations of product weights are not identical. The population of product weights from production line 1 tends to be less that the population of product weights from production line 2.

44. H_0: All populations of managerial potential ratings are identical

H_a: Not all populations of managerial potential ratings are identical

	No Program	Company	Off-Site
	16	12	7
	9	20	1
	10	17	4
	15	19	2
	11	6	3
	13	18	8
		14	5
Sum of Ranks	74	106	30

$$H = \left[\frac{12}{n_T(n_T+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right] - 3(n_T+1) = \left[\frac{12}{20(21)}\left(\frac{74^2}{6} + \frac{106^2}{7} + \frac{30^2}{7}\right)\right] - 3(21) = 12.61$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Using the χ^2 table with $df = 2$, $\chi^2 = 12.61$ shows the p-value is less than .005.

Using Excel, the p-value for $\chi^2 = 12.61$ with two degrees of freedom is .0018.

Using JMP, $\chi^2 = 12.6109$ with a p-value of .0018.

p-value $\leq .025$; reject H_0. Conclude that the three populations of managerial potential ratings are not identical. Engineers who attended the off-site program tend to have the highest potential ratings.

46. $\Sigma d_i^2 = 136$

$$r_s = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)} = 1 - \frac{6(136)}{15(224)} = .757$$

$$\sigma_{r_s} = \sqrt{\frac{1}{n-1}} = \sqrt{\frac{1}{14}} = .2673$$

$$z = \frac{r_s - \mu_{r_s}}{\sigma_{r_s}} = \frac{.757}{.2673} = 2.83$$

p-value = 2(1.0000 – .9977) = .0046.

p-value $\leq$ .10; reject H_0. Conclude that there is a significant positive rank correlation between the two exams. Students who rank high on the midterm exam tend to rank high on the final exam.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 19 Solutions

2. a. $EV(d_1) = 0.5(14) + 0.2(9) + 0.2(10) + 0.1(5) = 11.3$

$EV(d_2) = 0.5(11) + 0.2(10) + 0.2(8) + 0.1(7) = 9.8$

$EV(d_3) = 0.5(9) + 0.2(10) + 0.2(10) + 0.1(11) = 9.6$

$EV(d_4) = 0.5(8) + 0.2(10) + 0.2(11) + 0.1(13) = 9.5$

Recommended decision: d_1.

b. The best decision in this case is the one with the smallest expected value; thus, d_4, with an expected cost of 9.5, is the recommended decision.

4. a. The decision to be made is to choose the type of service to provide. The chance event is the level of demand for the Myrtle Air service. The consequence is the amount of quarterly profit. There are two decision alternatives (full price and discount service). There are two outcomes for the chance event (strong demand and weak demand).

b. EV(Full) = 0.7(960) + 0.3(–490) = 525

EV(Discount) = 0.7(670) + 0.3(320) = 565

Optimal Decision: Discount service

c. EV(Full) = 0.8(960) + 0.2(–490) = 670

EV(Discount) = 0.8(670) + 0.2(320) = 600

Optimal Decision: Full price service

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

6. a. The decision is to choose what type of grapes to plant, the chance event is demand for the wine, and the consequence is the expected annual profit contribution. There are three decision alternatives: Chardonnay, Riesling, and both. There are four chance outcomes: (W,W), (W,S), (S,W), and (S,S). For instance, (W,S) denotes the outcomes corresponding to weak demand for Chardonnay and strong demand for Riesling.

b. In constructing a decision tree, it is only necessary to show two branches when only a single grape is planted. But, the branch probabilities in these cases are the sum of two probabilities. For example, the probability that demand for Chardonnay is strong is given by:

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

P(Strong demand for Chardonnay) = P(S,W) + P(S,S)

= 0.25 + 0.20

= 0.45

c. EV(Plant Chardonnay) = 0.55(20) +0.45(70) = 42.5

EV(Plant both grapes) = 0.05(22) + 0.50(40) + 0.25(26) + 0.20(60) = 39.6

EV(Plant Riesling) = 0.30(25) + 0.70(45) = 39.0

Optimal decision: Plant Chardonnay grapes only.

d. This changes the expected value in the case where both grapes are planted and when Riesling only is planted.

EV(Plant both grapes) = 0.05(22) + 0.50(40) +0.05(26) + 0.40(60) = 46.4

EV(Plant Riesling) = 0.10(25) + 0.90(45) = 43.0

We see that the optimal decision is now to plant both grapes. The optimal decision is sensitive to this change in probabilities.

e. Only the expected value for node 2 in the decision tree needs to be recomputed.

EV(Plant Chardonnay) = 0.55(20) + 0.45(50) = 33.5

This change in the payoffs makes planting Chardonnay only less attractive. It is now best to plant both types of grapes. The optimal decision is sensitive to a change in the payoff of this magnitude.

8. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. EV(node 6) = 0.57(100) + 0.43(300) = 186

EV(node 7) = 0.57(400) + 0.43(200) = 314

EV(node 8) = 0.18(100) + 0.82(300) = 264

EV(node 9) = 0.18(400) + 0.82(200) = 236

EV(node 10) = 0.40(100) + 0.60(300) = 220

EV(node 11) = 0.40(400) + 0.60(200) = 280

EV(node 3) = Max(186,314) = 314 d_2

EV(node 4) = Max(264,236) = 264 d_1

EV(node 5) = Max(220,280) = 280 d_2

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EV(node 2) = 0.56(314) + 0.44(264) = 292

EV(node 1) = Max(292,280) = 292

∴ Market Research

If Favorable, decision d_2.

If Unfavorable, decision d_1.

10. a.

Outcome 1 ($ in 000s)

Bid	–$200
Contract	–2,000
Market research	–150
High demand	+5,000
	$2,650

Outcome 2 ($ in 000s)

Bid	–$200
Contract	–2,000
Market research	–150
Moderate demand	+3,000
	$650

b. EV(node 8) = 0.85(2,650) + 0.15(650) = 2350

EV(node 5) = Max(2,350, 1,150) = 2350 Decision: build

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EV(node 9) = 0.225(2,650) + 0.775(650) = 1,100

EV(node 6) = Max(1,100, 1,150) = 1150 Decision: sell

EV(node 10) = 0.6(2,800) + 0.4(800)= 2,000

EV(node 7) = Max(2,000, 13,00) = 2,000 Decision: build

EV(node 4) = 0.6 EV(node 5) + 0.4 EV(node 6) = 0.6(2,350) + 0.4(1,150) = 1,870

EV(node 3) = MAX (EV(node 4), EV(node 7)) = Max (1,870, 2,000) = 2,000

Decision: No market research

EV(node 2) = 0.8 EV(node 3) + 0.2 (–200) = 0.8(2000) + 0.2(–200) = 1560

EV(node 1) = MAX (EV(node 2), 0) = Max (1560, 0) = 1560

Decision: Bid on contract.

Decision strategy: Bid on the contract.

Do not do the market research.

Build the complex.

Expected value is $1,560,000.

c. Compare expected values at nodes 4 and 7.

EV(node 4) = 1,870 Includes $150 cost for research

EV(node 7) = 2,000

Difference is 2,000 – 1,870 = $130

Market research cost would have to be lowered $130,000 to $20,000 or less to make undertaking the research desirable.

12. a.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. Using node 5,

EV(node 10) = 0.4(3,500) + 0.3(1,000) + 0.3(–1,500) = 1,250

EV(node 11) = 0.4(7,000) + 0.3(2,000) + 0.3(–9,000) = 700

Decision: d_1 blade attachment expected value \$1,250

c. EVwPI = 0.4(7,000) + 0.3(2,000) + 0.3(–1,500) = \$2,950

EVPI = \$2,950 – \$1,250 = \$1,700

d. EV(node 6) = 0.35(3,500) + 0.30(1,000) + 0.35(–1,500) = 1,000

EV(node 7) = 0.35(7,000) + 0.30(2,000) + 0.35(–9,000) = –100

EV(node 8) = 0.62(3,500) + 0.31(1,000) + 0.07(–1,500) = 2,375

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EV(node 9) = 0.62(7,000) + 0.31(2,000) + 0.07(–9,000) = 4,330

EV(node 3) = Max(1,000,–100) = 1,000 d_1 Blade attachment

EV(node 4) = Max(2,375, 4,330) = 4,330 d_2 New snowplow

If normal, blade attachment.

If unseasonably cold, snowplow $1,666.

The expected value of this decision strategy is the expected value of node 2.

EV(node 2) = 0.8(1,000) + 0.2(4,330) = 1,666

Recommend: Wait until September and follow the decision strategy.

14.

State of Nature	$P(s_j)$	$P(I \mid s_j)$	$P(I \cap s_j)$	$P(s_j \mid I)$
s_1	0.2	0.10	0.020	0.1905
s_2	0.5	0.05	0.025	0.2381
s_3	0.3	0.20	0.060	0.5714
	1.0	$P(I) = 0.105$		1.0000

16. a, b. The revised probabilities are shown on the branches of the decision tree.

EV(node 7) = 30

EV(node 8) = 0.98(25) + 0.02(45) = 25.4

EV(node 9) = 30

EV(node 10) = 0.79(25) + 0.21(45) = 29.2

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EV(node 11) = 30

EV(node 12) = 0.00(25) + 1.00(45) = 45.0

EV(node 13) = 30

EV(node 14) = 0.85(25) + 0.15(45) = 28.0

EV(node 3) = Min(30,25.4) = 25.4 Expressway

EV(node 4) = Min(30,29.2) = 29.2 Expressway

EV(node 5) = Min(30,45) = 30.0 Queen City

EV(node 6) = Min(30,28) = 28.0 Expressway

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EV(node 2) = 0.695(25.4) + 0.215(29.2) + 0.09(30.0) = 26.6

EV(node 1) = Min(26.6,28) = 26.6 Weather

c. Strategy:

Check the weather, take the expressway unless there is rain. If rain, take Queen City Avenue.

Expected time: 26.6 minutes.

18. a. The expected value for the large-cap stock mutual fund is as follows:

EV = 0.1(35.3) + 0.3(20.0) + 0.1(28.3) + 0.1(10.4) + 0.4(–9.3) = 9.68

Repeating this calculation for each mutual fund provides the following expected annual returns:

Mutual Fund	Expected Annual Return
Large-cap stock	9.68
Mid-cap stock	5.91
Small-cap stock	15.20
Energy/resources sector	11.74
Health sector	7.34
Technology sector	16.97
Real estate sector	15.44

The technology sector provides the maximum expected annual return of 16.97%. Using this recommendation, the minimum annual return is –20.1%. and the maximum annual return is 93.1%.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. The expected annual return for the small-cap stock mutual fund is 15.20%. The technology sector mutual fund recommended in part a has a larger expected annual return. The difference is 16.97% – 15.20% = 1.77%.

c. The annual return for the technology sector mutual fund ranges from –20.1% to 93.1%, whereas the annual return for the small-cap stock ranges from 6.0% to 33.3%. The annual return for the technology sector mutual fund shows the greater variation in annual return. It is considered the investment with the more risk. It does have a higher expected annual return, but only by 1.77%.

d. This is a judgment recommendation and opinions may vary. The higher risk technology sector mutual fund only has a 1.77% higher expected annual return. We believe the lower risk, small-cap stock mutual fund would be the preferred recommendation for most investors.

20. a. EV(node 4) = 0.5(34) + 0.3(20) + 0.2(10) = 25

EV(node 3) = Max(25,20) = 25

Decision: Build.

EV(node 2) = 0.5(25) + 0.5(–5) = 10

EV(node 1) = Max(10,0) = 10

Decision: Start R&D.

Optimal strategy:

Start the R&D project.

If it is successful, build the facility.

Expected value = $10M

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. At node 3, payoff for sell rights would have to be $25 M or more. To recover the $5 M R&D cost, the selling price would have to be $30 M or more.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 20 Solutions

2. a. From the price relative, we see the percentage increase was 32%.

   b. Divide the current cost by the price relative and multiply by 100.

$$\text{Cost in 2001} = \frac{\$10.75}{132}(100) = \$8.14$$

4. $$I = \frac{.19(500)+1.80(50)+4.20(100)+13.20(40)}{.15(500)+1.60(50)+4.50(100)+12.00(40)}(100) = 104$$

Paasche index

6.

Item	Price Relative	Base Period Price	Base Period Usage	Weight	Weighted Price Relatives
A	150	22.00	20	440	66,000
B	90	5.00	50	250	22,500
C	120	14.00	40	560	67,200
				1250	155,700

$$I = \frac{155,700}{1250} = 125$$

8.

Item	Price Relatives	Base Price	Quantity	Weight	Weighted Price Relatives
Beer	115	17.50	35,000	612,500	70,437,500
Wine	118	100.00	5,000	500,000	59,000,000

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Soft Drink	110	8.00	60,000	480,000	52,800,000
				1,592,500	182,237,500

$$I = \frac{182,237,500}{1,592,500} = 114$$

This matches the value of the weighted aggregate index computed in exercise 5.

10. a. Deflated 2007 wages: $\frac{\$30.04}{207.3}(100) = \$14.49$

Deflated 2017 wages: $\frac{\$35.36}{245.1}(100) = \$14.43$

b. $\frac{\$35.36}{\$30.04} \times 100 = 117.7$

The percentage increase in nominal wages is 17.7%.

c. $\frac{\$14.43}{\$14.49} \times 100 = 99.6$

The percentage *decrease* in real wages is .4%.

12. a.

Year 1: $\frac{\$29.1}{216.0}(100) = \$13.47$

Year 2: $\frac{\$33.3}{218.4}(100) = \$15.25$

Year 3: $\frac{\$32.9}{226.9}(100) = \$14.50$

b.

Year 1: $\frac{\$29.1}{173.4}(100) = \$16.78$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Year 2: $\frac{\$33.3}{180.2}(100) = \$18.48$

Year 3: $\frac{\$32.9}{192.5}(100) = \$17.09$

c. PPI is more appropriate than CPI because these figures reflect prices paid by retailers (rather than by consumers).

14. $I = \frac{300(18.00) + 400(4.90) + 850(15.00)}{350(18.00) + 220(4.90) + 730(15.00)}(100) = \frac{20,110}{18,328}(100) = 110$

16.

Model	Quantity Relatives	Base Quantity	Price ($)	Weight	Weighted Quantity Relatives
Sedan	85	200	15,200	3,040,000	258,400,000
Sport	80	100	17,000	1,700,000	136,000,000
Wagon	80	75	16,800	1,260,000	100,800,000
				6,000,000	495,200,000

$$I = \frac{495,200,000}{6,000,000} = 83$$

18. a. Price Relatives

A = (15.90 / 10.50) (100) = 151

B = (32.00 / 16.25) (100) = 197

C = (17.40 /12.20) (100) = 143

D = (35.50 / 20.00) (100) = 178

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

b. $$I = \frac{15.90(2000) + 32.00(5000) + 17.40(6500) + 35.50(2500)}{10.50(2000) + 16.25(5000) + 12.20(6500) + 20.00(2500)}(100) = 170$$

20.

$$I_{\text{Jan}} = \frac{22.75(100) + 49(150) + 32(75) + 6.5(50)}{31.50(100) + 65(150) + 40(75) + 18(50)}(100) = 73.5$$

$$I_{\text{Mar}} = \frac{22.50(100) + 47.5(150) + 29.5(75) + 3.75(50)}{31.50(100) + 65(150) + 40(75) + 18(50)}(100) = 70.1$$

Market is down compared to the base Year 1.

22.

	Price Relatives	Base Price	Quantity	Weight	Weighted Price Relatives
Item	$(P_{it}/P_{i0})100$	P_{i0}	Q_i	$W_i = P_{i0}Q_i$	$(P_{it}/P_{i0})100\ W_i$
Handle	126.9	7.46	1	7.46	946.67
Blades	153.7	1.90	17	32.30	4,964.51
			Totals	39.76	5,911.18

I = (5911.18 / 39.76) = 148.7

24.

Year 1	(84810/215.3)100 =	\$39,392
Year 2	(87210/214.5)100 =	\$40,657
Year 3	(87650/218.1)100 =	\$40,188
Year 4	(91060/224.9)100 =	\$40,489
	(40489/39392) =	1.028

The median salary increased 2.8% over the period.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

26. $$I = \frac{1200(30) + 500(20) + 500(25)}{800(30) + 600(20) + 200(25)}(100) = 143$$

Quantity is up 43%.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 21 Solutions

2. a. $\mu = \frac{677.5}{25(5)} = 5.42$

b. UCL = $\mu + 3(\sigma / \sqrt{n}) = 5.42 + 3(.5 / \sqrt{5}) = 6.09$

LCL = $\mu - 3(\sigma / \sqrt{n}) = 5.42 - 3(.5 / \sqrt{5}) = 4.75$

4. R Chart:

UCL = $\bar{R}D_4 = 1.6(1.864) = 2.98$

LCL = $\bar{R}D_3 = 1.6(0.136) = 0.22$

$\bar{x}$ Chart:

UCL = $\bar{\bar{x}} + A_2\bar{R} = 28.5 + 0.373(1.6) = 29.10$

LCL = $\bar{\bar{x}} - A_2\bar{R} = 28.5 - 0.373(1.6) = 27.90$

6. Process mean = $\frac{20.12 + 19.90}{2} = 20.01$

UCL = $\mu + 3(\sigma / \sqrt{n}) = 20.01 + 3(\sigma / \sqrt{5}) = 20.12$

Solve for σ:

$$\sigma = \frac{(20.12 - 20.01)\sqrt{5}}{3} = 0.082$$

8. a. $p = \frac{141}{20(150)} = 0.0470$

b. $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.0470(0.9530)}{150}} = 0.0173$

UCL = $p + 3\,\sigma_{\bar{p}} = 0.0470 + 3(0.0173) = 0.0989$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

LCL = $p - 3\,\sigma_{\bar{p}}$ = 0.0470 –3(0.0173) = –0.0049

Use LCL = 0.

c. $\bar{p} = \dfrac{12}{150} = 0.08$

Process should be considered in control.

d. p = .047, n = 150

UCL = $np + 3\sqrt{np(1-p)}$ = 150(0.047) + 3 $\sqrt{150(0.047)(0.953)}$ = 14.826

LCL = $np - 3\sqrt{np(1-p)}$ = 150(0.047) – 3 $\sqrt{150(0.047)(0.953)}$ = –0.726

Thus, the process is out of control if more than 14 defective packages are found in a sample of 150.

e. Process should be considered to be in control because 12 defective packages were found.

f. The *np* chart may be preferred because a decision can be made by simply counting the number of defective packages.

10. $f(x) = \dfrac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

When p = .02, the probability of accepting the lot is

$$f(0) = \frac{25!}{0!(25-0)!}(0.02)^0(1-0.02)^{25} = 0.6035$$

When p = .06, the probability of accepting the lot is

$$f(0) = \frac{25!}{0!(25-0)!}(0.06)^0(1-0.06)^{25} = 0.2129$$

12. At p_0 = .02, the n = 20 and c = 1 plan provides

P (Accept lot) = $f(0) + f(1)$ = .6676 + .2725 = .9401

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Producer's risk: $\alpha = 1 - .9401 = .0599$

At $p_0 = .06$, the $n = 20$ and $c = 1$ plan provides

$$P\text{ (Accept lot)} = f(0) + f(1) = .2901 + .3703 = .6604$$

Producer's risk: $\alpha = 1 - .6604 = .3396$

For a given sample size, the producer's risk decreases as the acceptance number c is increased.

14.

	c	P (Accept) p_0 = .05	Producer's Risk α	P (accept) p_1 = .30	Consumer's Risk β
(n = 10)	0	.5987	.4013	.0282	.0282
	1	.9138	.0862	.1493	.1493
	2	.9884	.0116	.3828	.3828
(n = 15)	0	.4633	.5367	.0047	.0047
	1	.8291	.1709	.0352	.0352
	2	.9639	.0361	.1268	.1268
	3	.9946	.0054	.2968	.2968
(n = 20)	0	.3585	.6415	.0008	.0008
	1	.7359	.2641	.0076	.0076
	2	.9246	.0754	.0354	.0354
	3	.9842	.0158	.1070	.1070

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The plan with $n = 15$, $c = 2$ is close with $\alpha = .0361$ and $\beta = .1268$. However, the plan with $n = 20$, $c = 3$ is necessary to meet both requirements.

16. a. $\mu = \frac{\Sigma\bar{x}}{20} = \frac{1908}{20} = 95.4$

b.

$$\text{UCL} = \mu + 3(\sigma / \sqrt{n}) = 95.4 + 3(.50 / \sqrt{5}) = 96.07$$

$$\text{LCL} = \mu - 3(\sigma / \sqrt{n}) = 95.4 - 3(.50 / \sqrt{5}) = 94.73$$

c. No; all were in control

18. *R* Chart:

UCL = $\bar{R}D_4 = 2(2.115) = 4.23$

LCL = $\bar{R}D_3 = 2(0) = 0$

$\bar{x}$Chart:

UCL = $\bar{\bar{x}} + A_2\bar{R} = 5.42 + 0.577(2) = 6.57$

LCL = $\bar{\bar{x}} - A_2\bar{R} = 5.42 - 0.577(2) = 4.27$

Estimate of Standard Deviation:

$$\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{2}{2.326} = 0.86$$

20. $\bar{R} = .053$ $\bar{\bar{x}} = 3.082$

$\bar{x}$Chart:

UCL = $\bar{\bar{x}} + A_2\bar{R} = 3.082 + 0.577(0.053) = 3.112$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

LCL = $\bar{\bar{x}} - A_2\bar{R}$ = 3.082 – 0.577(0.053) = 3.051

R Chart:

UCL = $\bar{R}D_4$ = 0.053(2.115) = 0.1121

LCL = $\bar{R}D_3$ = 0.053(0) = 0

All sample averages and sample ranges are within the control limits for both charts.

22. a. p = .04

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.04(0.96)}{200}} = 0.0139$$

UCL = $p + 3\,\sigma_{\bar{p}}$ = 0.04 + 3(0.0139) = 0.0817

LCL = $p - 3\,\sigma_{\bar{p}}$ = 0.04 – 3(0.0139) = –0.0017

Use LCL = 0.

b. For month 1 $\bar{p}$= 10/200 = 0.05. Other monthly values are .075, .03, .065, .04, and .085. Only the last month with $\bar{p}$ = 0.085 is an out-of-control situation. The seesaw (i.e., zigzag) pattern of the points is also not considered normal for an in-control process.

24. a. *P* (Accept) are shown below (using n = 15):

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	$p = .01$	$p = .02$	$p = .03$	$p = .04$	$p = .05$
$f(0)$	.8601	.7386	.6333	.5421	.4633
$f(1)$	.1303	.2261	.2938	.3388	.3658
	.9904	.9647	.9271	.8809	.8291
$\alpha = 1 - P$ (Accept)	.0096	.0353	.0729	.1191	.1709

Using $p_0 = .03$ since α is close to .075. Thus, .03 is the fraction defective where the producer will tolerate a .075 probability of rejecting a good lot (only .03 defective).

b.

	$p = .25$
$f(0)$	.0134
$f(1)$	.0668
$\beta =$	.0802

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Chapter 22 Solutions

2. a. Estimate of population total = $N\bar{x}$ = 400(75) = 30,000.

b. Estimate of standard error = $Ns_{\bar{x}}$.

$$Ns_{\bar{x}} = 400\left(\frac{8}{\sqrt{80}}\right)\sqrt{\frac{400-80}{400}} = 320$$

c. 30,000 ± 2(320) or 29,360 to 30,640

4. $B = 15$

$$n = \frac{(70)^2}{\frac{(15)^2}{4} + \frac{(70)^2}{450}} = \frac{4900}{67.1389} = 72.9830$$

A sample size of 73 will provide an approximate 95% confidence interval of width 30.

6. B = 5,000/2 = 2,500. Use the value of s for the previous year in the formula to determine the necessary sample size.

$$n = \frac{(31.3)^2}{\frac{(2.5)^2}{4} + \frac{(31.3)^2}{724}} = \frac{979.69}{2.9157} = 336.0051$$

A sample size of 337 will provide an approximate 95% confidence interval of width no larger than $5,000.

8. a.

Stratum 1: $N_1\bar{x}_1$ = 200(138) = 27,600

Stratum 2: $N_2\bar{x}_2$ = 250(103) = 25,750

Stratum 3: $N_3\bar{x}_3$ = 100(210) = 21,000

b. $N\,\bar{x}_{st}$ = 27,600 + 25,750 + 21,000 = 74,350

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Note: The sum of the estimate for each stratum total equals $N\,\bar{x}_{st}$

c. $s_{\bar{x}_{st}}$ = 550(3.4092) = 1875.06 (see 7c)

Approximate 95% confidence interval:

74,350 ± 2(1875.06)

or

70,599.88 to 78,100.12

10. a. $$n = \frac{\left[300(150)+600(75)+500(100)\right]^2}{(1400)^2\left(\frac{(20)^2}{2}\right)+\left[300(150)^2+600(75)^2+500(100)^2\right]} = \frac{(140,000)^2}{196,000,000+15,125,000} = 92.8359$$

Rounding up we choose a total sample of 93.

$$n_1 = 93\left(\frac{300(150)}{140,000}\right) = 30$$

$$n_2 = 93\left(\frac{600(75)}{140,000}\right) = 30$$

$$n_3 = 93\left(\frac{500(100)}{140,000}\right) = 33$$

b. With B = 10, the first term in the denominator in the formula for n changes.

$$n = \frac{(140,000)^2}{(1400)^2\left(\frac{(10)^2}{4}\right)+15,125,000} = \frac{(140,000)^2}{49,000,000+15,125,000} = 305.6530$$

Rounding up, we see that a sample size of 306 is needed to provide this level of precision.

$$n_1 = 306\left(\frac{300(150)}{140,000}\right) = 98$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$$n_2 = 306\left(\frac{600(75)}{140,000}\right) = 98$$

$$n_3 = 306\left(\frac{500(100)}{140,000}\right) = 109$$

Because of rounding, the total of the allocations to each strata only add to 305. Note that even though the sample size is larger, the proportion allocated to each stratum has not changed.

c. $$n = \frac{(140,000)^2}{\frac{(15,000)^2}{4} + 15,125,000} = \frac{(140,000)^2}{56,250,000 + 15,125,000} = 274.6060$$

Rounding up, we see that a sample size of 275 will provide the desired level of precision.

The allocations to the strata are in the same proportion as for parts a and b.

$$n_1 = 275\left(\frac{300(150)}{140,000}\right) = 98$$

$$n_2 = 275\left(\frac{600(75)}{140,000}\right) = 88$$

$$n_3 = 275\left(\frac{500(100)}{140,000}\right) = 98$$

Again, because of rounding, the stratum allocations do not add to the total sample size. Another item could be sampled from, say, stratum 3 if desired.

12. a. St. Louis total = $N_1\bar{x}_1$ = 80 (45.2125) = 3,617

In dollars: $3,617,000

b. Indianapolis total = $N_1\bar{x}_1$ = 38 (29.5333) = 1,122.2654

In dollars: $1,122,265

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c.

$$\bar{x}_{st} = \left(\frac{38}{233}\right)29.5333 + \left(\frac{45}{233}\right)64.775 + \left(\frac{80}{233}\right)45.2125 + \left(\frac{70}{233}\right)53.0300 = 48.7821$$

$$N_1(N_1 - n_1)\frac{s_1^2}{n_1} = 38(32)\frac{(13.3603)^2}{6} = 36{,}175.517$$

$$N_2(N_2 - n_2)\frac{s_2^2}{n_2} = 45(37)\frac{(25.0666)^2}{8} = 130{,}772.1$$

$$N_3(N_3 - n_3)\frac{s_3^2}{n_3} = 80(72)\frac{(19.4084)^2}{8} = 271{,}213.91$$

$$N_4(N_4 - n_4)\frac{s_4^2}{n_4} = 70(60)\frac{(29.6810)^2}{10} = 370{,}003.94$$

$$s_{\bar{x}_{st}} = \sqrt{\left(\frac{1}{(233)^2}\right)\left[36{,}175.517 + 130{,}772.1 + 271{,}213.91 + 370{,}003.94\right]}$$

$$= \sqrt{\frac{1}{(233)^2}(808{,}165.47)} = 3.8583$$

Approximate 95% confidence interval:

$$\bar{x}_{st} \pm 2s_{\bar{x}_{st}}$$

$$48.7821 \pm 2(3.8583)$$

or

41.0655 to 56.4987

In dollars: \$41,066 to \$56,499

d. Approximate 95% confidence interval:

$$N\bar{x}_{st} \pm 2Ns_{\bar{x}_{st}}$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$233(48.7821) \pm 2(233)(3.8583)$

$11{,}366.229 \pm 1797.9678$

or

9,568.2612 to 13,164.197

In dollars: \$9,568,261 to \$13,164,197

14. a. $\bar{x}_c = \dfrac{\sum x_i}{\sum M_i} = \dfrac{750}{50} = 15$

$\hat{X} = M\,\bar{x}_c = 300(15) = 4{,}500$

$\bar{p}_c = \dfrac{\sum a_i}{\sum M_i} = \dfrac{15}{50} = .30$

b. $\sum(x_i - \bar{x}_c M_i)^2 = [95 - 15\,(7)]^2 + [325 - 15\,(18)]^2 + [190 - 15\,(15)]^2 + [140 - 15\,(10)]^2$

$= (-10)^2 + (55)^2 + (-35)^2 + (-10)^2$

$= 4450$

$s_{\bar{x}_c} = \sqrt{\left(\dfrac{25-4}{(25)(4)(12)^2}\right)\left(\dfrac{4450}{3}\right)} = 1.4708$

$s_{\hat{X}} = M s_{\bar{x}_c} = 300(1.4708) = 441.24$

$\sum(a_i - \bar{p}_c M_i)^2 \qquad = [1 - .3\,(7)]^2 + [6 - .3\,(18)]^2 + [6 - .3\,(15)]^2 + [2 - .3\,(10)]^2$

$= (-1.1)^2 + (.6)^2 + (1.5)^2 + (-1)^2$

$= 4.82$

$s_{\bar{p}_c} = \sqrt{\left(\dfrac{25-4}{(25)(4)(12)^2}\right)\left(\dfrac{4.82}{3}\right)} = .0484$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

c. Approximate 95% confidence

Interval for population mean:

$$15 \pm 2(1.4708)$$

or

12.0584 to 17.9416

d. Approximate 95% confidence

Interval for population total:

$$4{,}500 \pm 2(441.24)$$

or

3,617.52 to 5,382.48

e. Approximate 95% confidence

Interval for population proportion:

$$.30 \pm 2(.0484)$$

or

.2032 to .3968

16. a. $\bar{x}_c = \frac{2000}{50} = 40$

Estimate of mean age of mechanical engineers: 40 years

b. $\bar{p}_c = \frac{35}{50} = .70$

Estimate of proportion attending local university: .70

c.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$\Sigma(x_i - \bar{x}_c M_i)^2 \quad = [520 - 40\ (12)]^2 + \cdot\cdot\cdot + [462 - 40\ (13)]^2$

$= (40)^2 + (-7)^2 + (-10)^2 + (-11)^2 + (30)^2 + (9)^2 + (22)^2 + (8)^2 + (-23)^2$

$+ (-58)^2$

$= 7{,}292$

$$s_{\bar{x}_c} = \sqrt{\left(\frac{120-10}{(120)(10)(50/12)^2}\right)\left(\frac{7292}{9}\right)} = 2.0683$$

Approximate 95% confidence

Interval for mean age:

$40 \pm 2(2.0683)$

or

35.8634 to 44.1366

d.

$\Sigma(a_i - \bar{p}_c M_i)^2 = [8 - .7\ (12)]^2 + \cdot\cdot\cdot + [12 - .7\ (13)]^2$

$= (-.4)^2 + (-.7)^2 + (-.4)^2 + (.3)^2 + (-1.2)^2 + (-.1)^2 + (-1.4)^2 + (.3)^2$

$+ (.7)^2 + (2.9)^2$

$= 13.3$

$$s_{\bar{p}_c} = \sqrt{\left(\frac{120-10}{(120)(10)(50/12)^2}\right)\left(\frac{13.3}{9}\right)} = .0883$$

Approximate 95% confidence

Interval for proportion attending local university:

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$.70 \pm 2(.0883)$

or

.5234 to .8766

18. a. $\bar{p} = 0.19$

$$s_{\bar{p}} = \sqrt{\frac{(0.19)(0.81)}{363}} = 0.0206$$

Approximate 95% confidence interval:

$0.19 \pm 2(0.0206)$

or

0.1488 to 0.2312

b. $\bar{p} = 0.31$

$$s_{\bar{p}} = \sqrt{\frac{(0.31)(0.69)}{363}} = 0.0243$$

Approximate 95% confidence interval:

$0.31 \pm 2(0.0243)$

or

0.2615 to 0.3585

c. $\bar{p} = 0.17$

$$s_{\bar{p}} = \sqrt{\frac{(0.17)(0.83)}{373}} = 0.0197$$

Approximate 95% confidence interval:

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$0.17 \pm 2(0.0197)$

or

0.1306 to 0.2094

d. The largest standard error is when $\bar{p}$ = .50.

At $\bar{p}$ = .50, we get

$$s_{\bar{p}} = \sqrt{\frac{(0.5)(0.5)}{363}} = 0.0262$$

Multiplying by 2, we get a bound of $B = 2(.0262) = 0.0525$.

For a sample of 363, then, they know that in the worst case ($\bar{p} = 0.50$) the bound will be approximately 5%.

e. If the poll was conducted by calling people at home during the day, the sample results would only be representative of adults not working outside the home. It is likely that the Louis Harris organization took precautions against this and other possible sources of bias.

20. a. $$s_{\bar{x}} = \sqrt{\frac{3000-200}{3000}}\frac{3000}{\sqrt{200}} = 204.9390$$

Approximate 95% confidence interval for mean annual salary:

$23{,}200 \pm 2(204.9390)$

or

\$22,790 to \$23,610

b. $N\,\bar{x}$ = 3000 (23,200) = 69,600,000

$s_{\hat{x}}$ = 3000 (204.9390) = 614,817

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Approximate 95% confidence interval for population total salary:

69,600,000 ± 2(614,817)

or

$68,370,366 to $70,829,634

c. $\bar{p} = .73$

$$s_{\bar{p}} = \sqrt{\left(\frac{3000-200}{3000}\right)\left(\frac{(.73)(.27)}{199}\right)} = .0304$$

Approximate 95% confidence interval for proportion that are generally satisfied:

.73 ± 2(.0304)

or

.6692 to .7908

d. If management administered the questionnaire and anonymity was not guaranteed, then we would expect a definite upward bias in the percent reporting of those who were "generally satisfied" with their jobs. A procedure for guaranteeing anonymity should reduce the bias.

22. a. $\hat{X}$ = 380 (9/30) + 760 (12/45) + 260 (11/25) = 431.0667

Estimate approximately 431 deaths because of beating.

b.

$$\bar{p}_{st} = \left(\frac{380}{1400}\right)\left(\frac{9}{30}\right) + \left(\frac{760}{1400}\right)\left(\frac{12}{45}\right) + \left(\frac{260}{1400}\right)\left(\frac{11}{25}\right) = .3079$$

$$\sum N_h (N_h - n_h) \frac{\left[\bar{p}_h (1-\bar{p}_h)\right]}{n_h - 1}$$

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

= (380) (380 – 30) (9/30) (21/30) / 29 + (760) (760 – 45) (12/45) (33/45) / 44 +

(260) (260 – 25) (11/25) (14/25) / 24

= 4,005.5079

$$s_{\bar{p}_{st}} = \sqrt{\left(\frac{1}{(1400)^2}\right)4005.5079} = .0452$$

Approximate 95% confidence interval:

.3079 ± 2(.0452)

or

.2175 to .3983

c.

$$\bar{p}_{st} = \left(\frac{380}{1400}\right)\left(\frac{21}{30}\right) + \left(\frac{760}{1400}\right)\left(\frac{34}{45}\right) + \left(\frac{260}{1400}\right)\left(\frac{15}{25}\right) = .7116$$

$$\sum N_h(N_h - n_h)\frac{\left[\bar{p}_h(1-\bar{p}_h)\right]}{n_h - 1}$$

= (380) (380 – 30) (21/30) (9/30)/29 + (760) (760 – 45) (34/45) (11/45) / 44 +

(260) (260 – 25) (15/25) (10/25) / 24

= 3,855.0417

$$s_{\bar{p}_{st}} = \sqrt{\left(\frac{1}{(1400)^2}\right)3855.0417} = .0443$$

Approximate 95% confidence interval:

.7116 ± 2(.0443)

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

or

.6230 to .8002

d. $\hat{X}$ = 1400 (.7116) = 996.24

Estimate of total number of victims is 996.

24. a. $$\bar{x}_c = \frac{14(61)+7(74)+96(78)+23(69)+71(73)+29(84)}{14+7+96+23+71+29} = \frac{18,066}{240} = 75.275$$

Estimate of mean age is approximately 75 years old.

b.

$$\bar{p}_c = \frac{12+2+30+8+10+22}{14+7+96+23+71+29} = \frac{84}{240} = .35$$

$$\Sigma(a_i - \bar{p}_c M_i)^2 = [12 - .35\ (14)]^2 + [2 - .35\ (7)]^2 + [30 - .35\ (96)]^2$$

$$+ [8 - .35\ (23)]^2 + [10 - .35\ (71)]^2 + [22 - .35\ (29)]^2$$

$$= (7.1)^2 + (-.45)^2 + (-3.6)^2 + (-.05)^2 + (-14.85)^2 + (11.85)^2$$

$$= 424.52$$

$$s_{\bar{p}_c} = \sqrt{\left(\frac{100-6}{(100)(6)(48)^2}\right)\left(\frac{424.52}{5}\right)} = .0760$$

Approximate 95% confidence interval:

$.35 \pm 2(.0760)$

or

.198 to .502

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

$\hat{X} = 4{,}800\ (.35) = 1{,}680$

Estimate of total number of disabled persons is 1,680.

© 2019 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.