Business Analytics 4th edition Camm Solutions Manual

SKU: 9780357131787-SOLUTIONS

Chapter 2
Descriptive Statistics

Solutions:

1. a. Quantitative

   b. Categorical

   c. Categorical

   d. Quantitative

   e. Categorical

2. a. The top 10 countries according to GDP are listed below.

Country	Continent	GDP (millions of US$)
United States	North America	15,094,025
China	Asia	7,298,147
Japan	Asia	5,869,471
Germany	Europe	3,577,031
France	Europe	2,776,324
Brazil	South America	2,492,908
United Kingdom	Europe	2,417,570
Italy	Europe	2,198,730
Russia	Asia	1,850,401
Canada	North America	1,736,869

   b. The top 5 countries by GDP located in Africa are listed below.

Country	Continent	GDP (millions of US$)
South Africa	Africa	408,074
Nigeria	Africa	238,920
Egypt	Africa	235,719
Algeria	Africa	190,709
Angola	Africa	100,948

   c. The top-five European countries by GDP are: Germany, France, the United Kingdom, Italy, and Spain.

3. a. The sorted list of carriers appears below.

Carrier	Previous Year On-time Percentage	Current Year On-time Percentage
Blue Box Shipping	88.4%	94.8%
Cheetah LLC	89.3%	91.8%

Business Analytics 4th edition Camm Solutions Manual

Smith Logistics	84.3%	88.7%
Granite State Carriers	81.8%	87.6%
Super Freight	92.1%	86.8%
Minuteman Company	91.0%	84.2%
Jones Brothers	68.9%	82.8%
Honsin Limited	74.2%	80.1%
Rapid Response	78.8%	70.9%

Blue Box Shipping is providing the best on-time service in the current year. Rapid Response is providing the worst on-time service in the current year.

b. The output from Excel with conditional formatting appears below.

	A	B	C	D
1	**Carrier**	**Previous Year On-time Percentage**	**Current Year On-time Percentage**	**Change in On-time Percentage**
2	Blue Box Shipping	88.4%	94.8%	6.4%
3	Cheetah LLC	89.3%	91.8%	2.5%
4	Smith Logistics	84.3%	88.7%	4.4%
5	Granite State Carriers	81.8%	87.6%	5.8%
6	Super Freight	92.1%	86.8%	-5.3%
7	Minuteman Company	91.0%	84.2%	-6.8%
8	Jones Brothers	68.9%	82.8%	13.9%
9	Honsin Limited	74.2%	80.1%	5.9%
10	Rapid Response	78.8%	70.9%	-7.9%

c. The output from Excel containing data bars appears below.

	A	B	C	D
1	**Carrier**	**Previous Year On-time Percentage**	**Current Year On-time Percentage**	**Change in On-time Percentage**
2	Blue Box Shipping	88.4%	94.8%	6.4%
3	Cheetah LLC	89.3%	91.8%	2.5%
4	Smith Logistics	84.3%	88.7%	4.4%
5	Granite State Carriers	81.8%	87.6%	5.8%
6	Super Freight	92.1%	86.8%	-5.3%
7	Minuteman Company	91.0%	84.2%	-6.8%
8	Jones Brothers	68.9%	82.8%	13.9%
9	Honsin Limited	74.2%	80.1%	5.9%
10	Rapid Response	78.8%	70.9%	-7.9%

d. The top 4 shippers based on current year on-time percentage (Blue Box Shipping, Cheetah LLC, Smith Logistics, and Granite State Carriers) all have positive increases from the previous year and high on-time percentages. These are good candidates for carriers to use in the future.

4. a. The relative frequency of D is 1.0 – 0.22 – 0.18 – 0.40 = 0.20.

b. If the total sample size is 200 the frequency of D is 0.20*200 = 40.

c. and d.

Class	Relative Frequency	Frequency

Business Analytics 4th edition Camm Solutions Manual

			% Frequency
A	0.22	44	22
B	0.18	36	18
C	0.40	80	40
D	0.20	40	20
Total	1.0	200	100

5. a. These data are categorical.

b.

Web site	Frequency	% Frequency
FB	7	14
GOOG	14	28
WIKI	9	18
YAH	13	26
YT	7	14
Total	50	100

c. The most frequent most-visited-web site is google.com (GOOG); second is yahoo.com (YAH).

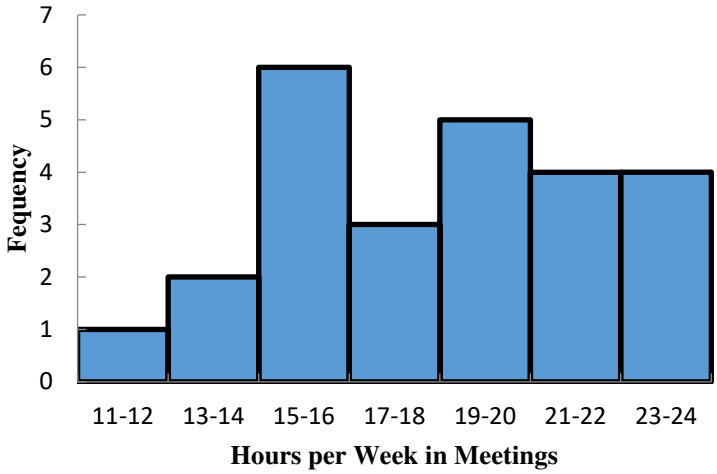
6. a. Least = 12, Highest = 23

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
Total	25	100

Business Analytics 4th edition Camm Solutions Manual

c.



The distribution is slightly skewed to the left.

7. a.

Industry	Frequency	Percent Frequency
Bank	26	13
Cable	44	22
Car	42	21
Cell	60	30
Collection	28	1
Total	200	100

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

c. 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.

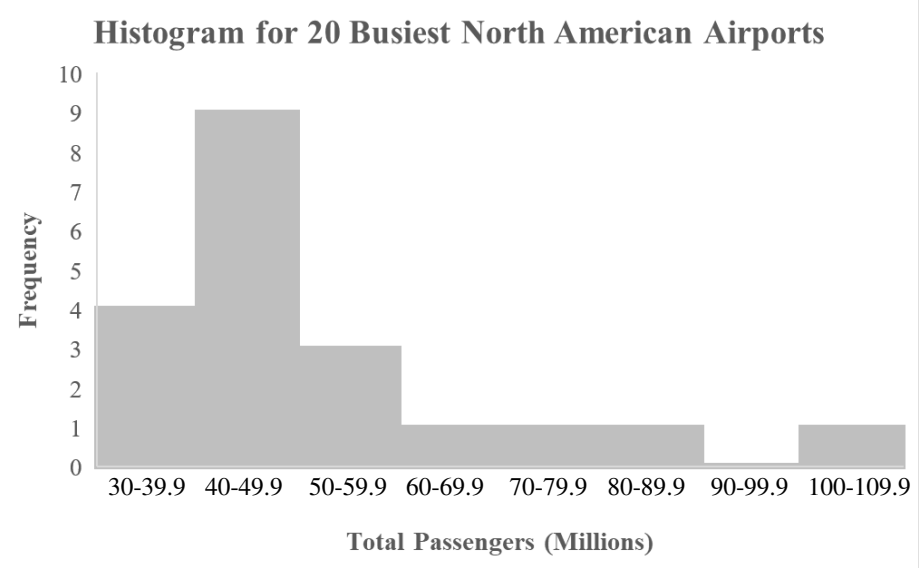
Business Analytics 4th edition Camm Solutions Manual

8. a. The busiest airport is Hartsfield-Jackson Atlanta (ATL) with 104.2 million total passengers. The least busy airport is Detroit Metropolitan (DTW) with 34.4 million total passengers.

b.

Total Passengers (Millions)	Frequency
30–39.9	4
40–49.9	9
50–59.9	3
60–69.9	1
70–79.9	1
80–89.9	1
90–99.9	0
100–109.9	1

c.



Most of the top 20 busiest North American airports service fewer than 60 million passengers. Only four of the 20 airports have more than 60 million passengers.

Business Analytics 4th edition Camm Solutions Manual

9. a.

Class	Frequency
12–14	2
15–17	8
18–20	11
21–23	10
24–26	9
Total	40

b.

Class	Relative Frequency	Percent Frequency
12–14	0.050	5.0
15–17	0.200	20.0
18–20	0.275	27.5
21–23	0.250	25.0
24–26	0.225	22.5
Total	1.000	100.0

10.

Class	Frequency	Cumulative Frequency
10–19	10	10
20–29	14	24
30–39	17	41
40–49	7	48
50–59	2	50

11. a–d.

Class	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
0–4	4	0.20	4	0.20
5–9	8	0.40	12	0.60
10–14	5	0.25	17	0.85
15–19	2	0.10	19	0.95
20–24	1	0.05	20	1.00
Total	20	1.00		

e. From the cumulative relative frequency distribution, 60% of customers wait 9 minutes or less.

Business Analytics 4th edition Camm Solutions Manual

12. a–d.

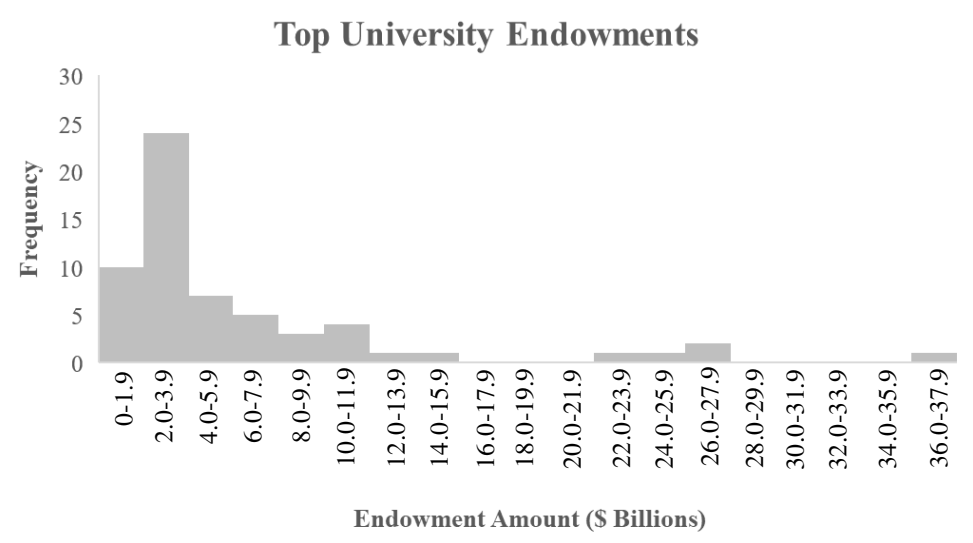
Endowment Amount ($ Billions)	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
0–1.9	10	0.17	10	0.17
2.0–3.9	24	0.40	34	0.57
4.0–5.9	7	0.12	41	0.68
6.0–7.9	5	0.08	46	0.77
8.0–9.9	3	0.05	49	0.82
10.0–11.9	4	0.07	53	0.88
12.0–13.9	1	0.02	54	0.90
14.0–15.9	1	0.02	55	0.92
16.0–17.9	0	0.00	55	0.92
18.0–19.9	0	0.00	55	0.92
20.0–21.9	0	0.00	55	0.92
22.0–23.9	1	0.02	56	0.93
24.0–25.9	1	0.02	57	0.95
26.0–27.9	2	0.03	59	0.98
28.0–29.9	0	0.00	59	0.98
30.0–31.9	0	0.00	59	0.98
32.0–33.9	0	0.00	59	0.98
34.0–35.9	0	0.00	59	0.98

Business Analytics 4th edition Camm Solutions Manual

36.0–37.9	1	0.02	60	1.00
Total	60	1.00		

e. Most universities (55) have endowments of less than $16 billion. Only five have endowments larger than $16 billion. We see that .92, or 92%, of the universities have endowments of less than $16 billion, and only .08, or 8%, of the universities have endowments larger than $16 billion.

f.



The histogram shows the distribution is skewed to the right with five university endowments in the $22 billion to $38 billion range.

g. Harvard University has the largest endowment at $36 billion. All other universities have endowments less than $28 billion. Most (92%) have endowments less than $16 billion.

13. a. Mean = $\frac{10+20+12+17+16}{5} = 15$ or use the Excel function AVERAGE.
To calculate the median, we arrange the data in ascending order:
10 12 16 17 20
Because we have n = 5 values which is an odd number, the median is the middle value which is 16 or use the Excel function MEDIAN.

b. Because the additional data point, 12, is lower than the mean and median computed in part a, we expect the mean and median to decrease. Calculating the new mean and median gives us mean = 14.5 and median = 14.

14. Without Excel, to calculate the 20th percentile, we first arrange the data in ascending order:
15 20 25 25 27 28 30 34
The location of the pth percentile is given by the formula $L_p = \frac{p}{100}(n+1)$

Business Analytics 4th edition Camm Solutions Manual

For our date set, $L_{20} = \frac{20}{100}(8+1) = 1.8$. Thus, the 20th percentile is 80% of the way between the value in position 1 and the value in position 2. In other words, the 20th percentile is the value in position 1 (15) plus 0.80 time the difference between the value in position 2 (20) and position 1 (15). Therefore, the 20th percentile is
15 + 0.80*(20 – 15) = 19.

We can repeat the steps above to calculate the 25th, 65th, and 75th percentiles. Or using Excel, we can use the function PERCENTILE.EXC to get:
25th percentile = 21.25
65th percentile = 27.85
75th percentile = 29.5

15. Mean = $\frac{53+55+70+58+64+57+53+69+57+68+53}{11} = 59.727$ or use the Excel function AVERAGE.
To calculate the median arrange the values in ascending order
53 53 53 55 57 57 58 64 68 69 70
Because we have $n = 11$, an odd number of values, the median is the middle value which is 57 or use the Excel function MEDIAN.
The mode is the most often occurring value which is 53 because 53 appears three times in the data set, or use the Excel function MODE.SNGL because there is only a single mode in this data set.

16. To find the mean annual growth rate, we must use the geometric mean. First we note that

$3{,}500 = 5{,}000^{\left[(x_1)(x_2)\cdots(x_9)\right]}$, so $\left[(x_1)(x_2)\cdots(x_9)\right] = 0.700$
where $x_1, x_2, \ldots$ are the growth factors for years, 1, 2, etc. through year 9.
Next, we calculate $\bar{x}_g = \sqrt[n]{(x_1)(x_2)\cdots(x_n)} = \sqrt[9]{0.70} = 0.961144$
So the mean annual growth rate is (0.961144 – 1)100% = –0.38856%

17. For the Stivers mutual fund,
$18{,}000 = 10{,}000^{\left[(x_1)(x_2)\cdots(x_8)\right]}$, so $\left[(x_1)(x_2)\cdots(x_8)\right] = 1.8$
where $x_1, x_2, \ldots$ are the growth factors for years, 1, 2, etc. through year 8.
Next, we calculate $\bar{x}_g = \sqrt[n]{(x_1)(x_2)\cdots(x_8)} = \sqrt[8]{1.80} = 1.07624$
So the mean annual return for the Stivers mutual fund is (1.07624 – 1)100% = 7.624%.

For the Trippi mutual fund we have:
$10{,}600 = 5{,}000^{\left[(x_1)(x_2)\cdots(x_8)\right]}$, so $\left[(x_1)(x_2)\cdots(x_8)\right] = 2.12$ and
$\bar{x}_g = \sqrt[n]{(x_1)(x_2)\cdots(x_8)} = \sqrt[8]{2.12} = 1.09848$
So the mean annual return for the Trippi mutual fund is (1.09848 – 1)100% = 9.848%.

While 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.

Business Analytics 4th edition Camm Solutions Manual

Alternatively, we can use Excel and the function GEOMEAN as shown below:

	A	B	C	D	E
1		Stivers		Trippi	
2	Year	End of Year Value	Growth Factor	End of Year Value	Growth Factor
3	0	$10,000		$5,000	
4	1	$11,000	1.100	$5,600	1.120
5	2	$12,000	1.091	$6,300	1.125
6	3	$13,000	1.083	$6,900	1.095
7	4	$14,000	1.077	$7,600	1.101
8	5	$15,000	1.071	$8,500	1.118
9	6	$16,000	1.067	$9,200	1.082
10	7	$17,000	1.063	$9,900	1.076
11	8	$18,000	1.059	$10,600	1.071
12					
13					
14		Stivers Geometric Mean:		1.07623984	
15		Trippi Geometric Mean:		1.09847957	

	A	B	C	D	E
1		Stivers		Trippi	
2	Year	End of Year Value	Growth Factor	End of Year Value	Growth Factor
3	0	10000		5000	
4	1	11000	=B4/B3	5600	=D4/D3
5	2	12000	=B5/B4	6300	=D5/D4
6	3	13000	=B6/B5	6900	=D6/D5
7	4	14000	=B7/B6	7600	=D7/D6
8	5	15000	=B8/B7	8500	=D8/D7
9	6	16000	=B9/B8	9200	=D9/D8
10	7	17000	=B10/B9	9900	=D10/D9
11	8	18000	=B11/B10	10600	=D11/D10
12					
13					
14			Geometric Mean:	=GEOMEAN(C3:C11)	
15			Geometric Mean:	=GEOMEAN(E4:E11)	

18. a. Mean = $\frac{\sum_{i=1}^{n} x_i}{n} = \frac{1{,}291.5}{48} = 26.906$

b. To calculate the median, we first sort all 48 commute times in ascending order. Because there are an even number of values (48), the median is between the 24th and 25th largest values. The 24th largest value is 25.8 and the 25th largest value is 26.1.
(25.8 + 26.1)/2 = 25.95
Or we can use the Excel function MEDIAN.

c. The values 23.4 and 24.8 both appear three times in the data set, so these two values are the modes of the commute times. To find this using Excel, we must use the MODE.MULT function.

d. Standard deviation = 4.6152. In Excel, we can find this value using the function STDEV.S.
Variance = 4.6152^2 = 21.2998. In Excel, we can find this value using the function VAR.S.

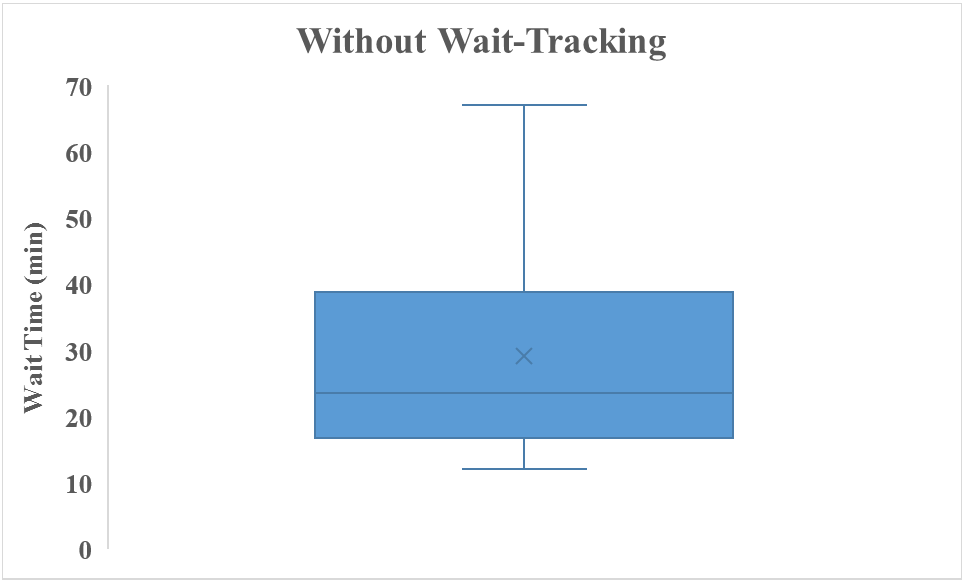
e. The third quartile is the 75th percentile of the data. To find the 75th percentile without Excel, we first arrange the data in ascending order. Next we calculate $L_p = \frac{p}{100}(n+1) = L_{75} = \frac{75}{100}(48+1) = 36.75$.
In other words, this value is 75% of the way between the 36th and 37th positions. However, in our data the values in both the 36th and 37th positions are 28.5. Therefore, the 75th percentile is 28.5. Or using Excel, we can use the function PERCENTILE.EXC.

19. a. The mean waiting time for patients with the wait-tracking system is 17.2 minutes and the median waiting time is 13.5 minutes. The mean waiting time for patients without the wait-tracking system is 29.1 minutes and the median is 23.5 minutes.

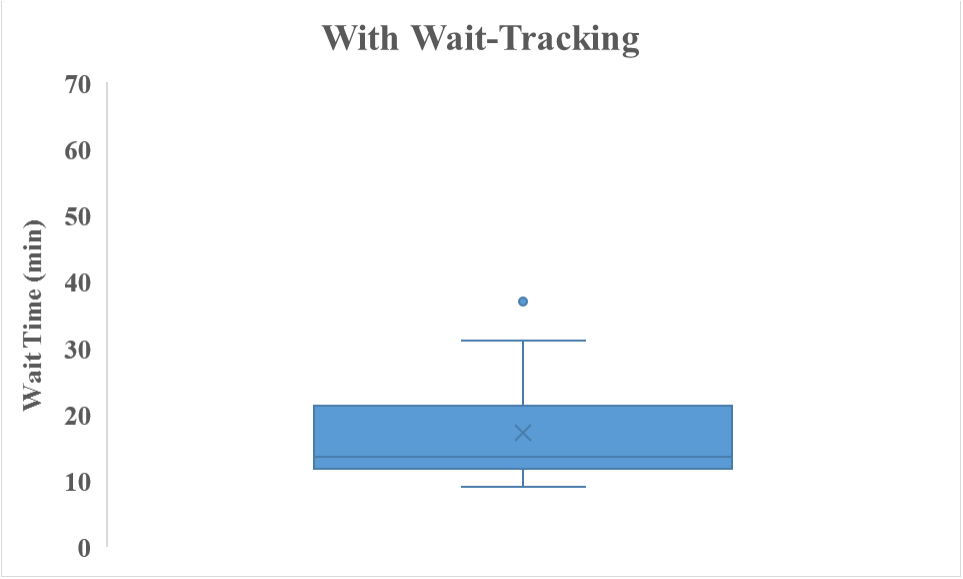
b. The standard deviation of waiting time for patients with the wait-tracking system is 9.28 and the variance is 86.18. The standard deviation of waiting time for patients without the wait-tracking system is 16.60 and the variance is 275.66.

Business Analytics 4th edition Camm Solutions Manual

c.



d.



e. Wait times for patients with the wait-tracking system are substantially shorter than those for patients without the wait-tracking system. However, some patients with the wait-tracking system still experience long waits.

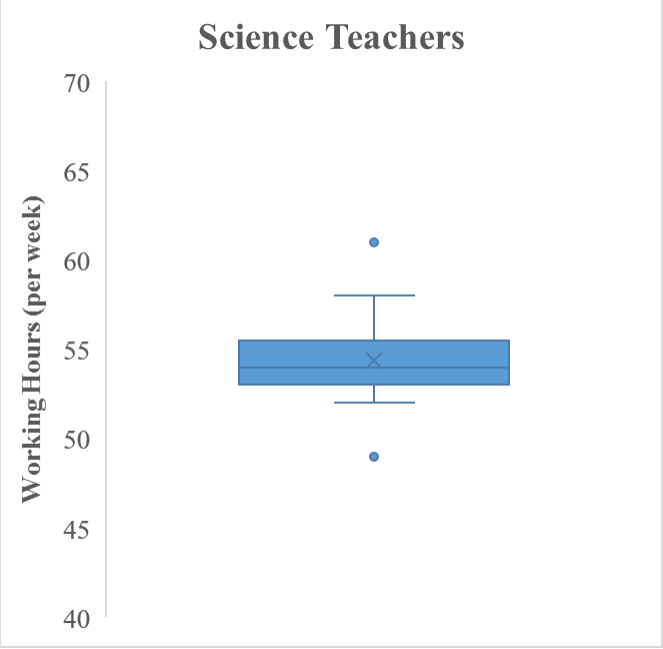
20. a. The median number of hours worked for science teachers is 54.

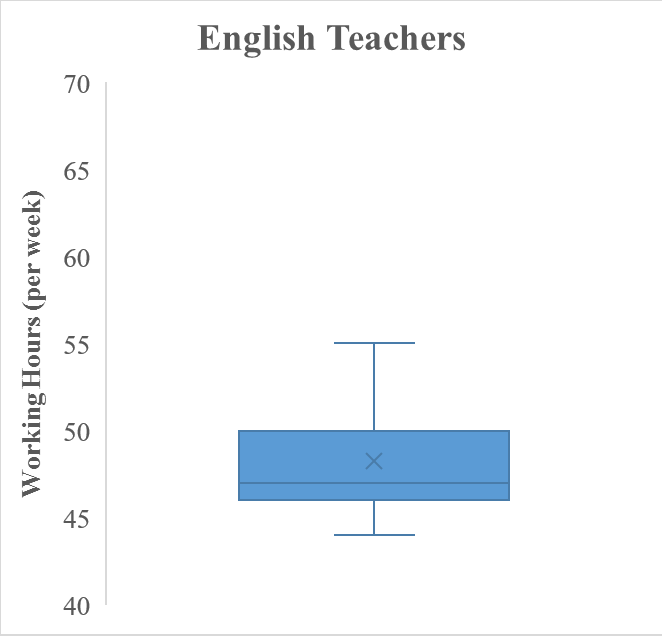
b. The median number of hours worked for English teachers is 47.

c.

Business Analytics 4th edition Camm Solutions Manual

d.





e. The boxplots show that science teachers spend more hours working per week than English teachers. The boxplot for science teachers also shows that most science teachers work about the same amount of hours; in other words, there is less variability in the number of hours worked for science teachers.

21. a. Recall that the mean patient wait time without wait-time tracking is 29.1 and the standard deviation of wait times is 16.6. Then the z-score is calculated as follows: $z = \frac{37-29.1}{16.6} = 0.48$.

b. Recall that the mean patient wait time with wait-time tracking is 17.2 and the standard deviation of wait times is 9.28. Then the z-score is calculated as follows: $z = \frac{37-17.2}{9.28} = 2.13$.
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

Business Analytics 4th edition Camm Solutions Manual

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.

c. To calculate the z-score for each patient waiting time, we can use the formula $z = \frac{x_i - \bar{x}}{s}$ or we can use the Excel function STANDARDIZE. The z-scores for all patients follow.

Without Wait-Tracking System		With Wait-Tracking System	
Wait Time	**z-Score**	**Wait Time**	**z-Score**
24	–0.31	31	1.49
67	2.28	11	–0.67
17	–0.73	14	–0.34
20	–0.55	18	0.09
31	0.11	12	–0.56
44	0.90	37	2.13
12	–1.03	9	–0.88
23	–0.37	13	–0.45
16	–0.79	12	–0.56
37	0.48	15	–0.24

No z-score is less than –3.0 or above +3.0; therefore, the z-scores do not indicate the existence of any outliers in either sample.

22. a. According to the empirical rule, approximately 95% of data values will be within two standard deviations of the mean. 4.5 is two standard deviation less than the mean and 9.3 is two standard deviations greater than the mean. Therefore, approximately 95% of individuals sleep between 4.5 and 9.3 hours per night.

b. $z = \frac{8-6.9}{1.2} = 0.9167$

c. $z = \frac{6-6.9}{1.2} = -0.75$

23.

a. The value 647 is one standard deviation above the mean. Approximately 68% of the scores are between 447 and 647 with half of 68%, or 34%, of the scores between the mean of 547 and 647. Also, because the distribution is symmetric, 50% of the scores are above the mean of 547. With 50% of the scores above 547 and with 34% of the scores between 547 and 647, 50% – 34% = 16% of the scores are 647 or higher.

b. The value 747 is two standard deviations above the mean. Approximately 95% of the scores are between 347 and 747 with half of 95%, or 47.5%, of the scores between the mean of 547 and 747. Also, because the distribution is symmetric, 50% of the scores are above the mean of 547. With 50% of the scores above 547 and 47.5% of the scores between 547 and 747, 50% – 47.5% = 2.5% of the scores are

Business Analytics 4th edition Camm Solutions Manual

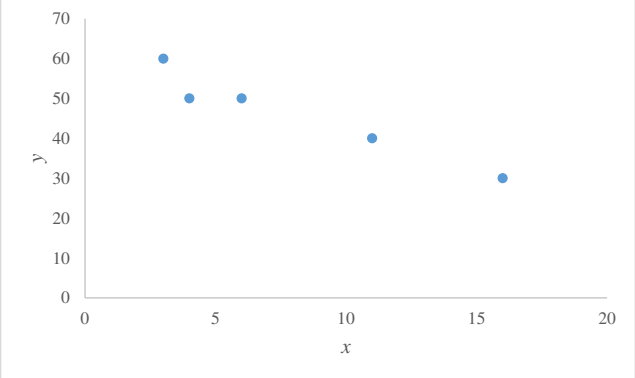
747 or higher.

c. Approximately 68% of the scores are between 447 and 647 with half of 68%, or 34%, of the scores are between 447 and the mean of 547.

d. Approximately 95% of the scores are between 347 and 747 with half of 95%, or 47.5%, of the scores between 347 and the mean of 547. Approximately 68% of the scores are between 447 and 647 with half of 68%, or 34%, of the scores between the mean of 547 and 647. Thus, 47.5% + 34% = 81.5% of the scores are between 347 and 647.

e. $z = \frac{405-515}{100} = -1.10$

24. a.



b. There appears to be a negative linear relationship between the *x* and *y* variables.

c. Without Excel, we can use the calculations shown below to calculate the covariance:

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
4	50	–4	4	–16
6	50	–2	4	–8
11	40	3	–6	–18
3	60	–5	14	–70
16	30	8	–16	–128

$\bar{x} =$ 8

$\bar{y} =$ 46

$s_{xy} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n-1} = \frac{-16-8-18-70-128}{4} = -60$

Or, using Excel, we can use the COVARIANCE.S function.

The negative covariance confirms that there is a negative linear relationship between the *x* and *y* variables in this data set.

Business Analytics 4th edition Camm Solutions Manual

d. To calculate the correlation coefficient without Excel, we need the standard deviation for *x* and *y*: $s_x = 5.43, s_y = 11.40$. Then the correlation coefficient is calculated as:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-60}{(5.43)(11.40)} = -0.97$$

Or, we can use the Excel function CORREL.
The correlation coefficient indicates a strong negative linear association between the *x* and *y* variables in this data set.

25. a. The scatter chart indicates that there may be a positive linear relationship between profits and market capitalization.

b. Without Excel, we can use the calculations below to find the covariance and correlation coefficient:

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
313.2	1891.9	–2468.57	–35259.75	6093826.70	1243249856.32	87041077.46
631	81458.6	–2150.77	44306.95	4625801.88	1963105961.23	–95293962.27
706.6	10087.6	–2075.17	–27064.05	4306321.16	732462715.10	56162440.18
–29	1175.8	–2810.77	–35975.85	7900415.30	1294261667.17	101119754.14
4,018.00	55188.8	1236.23	18037.15	1528270.20	325338838.31	22298108.67
959	14115.2	–1822.77	–23036.45	3322482.24	530677954.29	41990095.01
6,490.00	97376.2	3708.23	60224.55	13750986.48	3626996616.98	223326625.02
8,572.00	157130.5	5790.23	119978.85	33526789.60	14394924834.35	694705416.89
12,436.00	95251.9	9654.23	58100.25	93204200.49	3375639237.48	560913323.32
1,462.00	36461.2	–1319.77	–690.45	1741786.89	476718.98	911231.51
3,461.00	53575.7	679.23	16424.05	461356.46	269749471.38	11155745.66
854	7082.1	–1927.77	–30069.55	3716288.47	904177740.20	57967105.40
369.5	3461.4	–2412.27	–33690.25	5819035.66	1135032836.38	81269899.40
399.8	12520.3	–2381.97	–24631.35	5673770.32	606703323.37	58671077.30
278	3547.6	–2503.77	–33604.05	6268852.91	1129232068.00	84136732.35
9,190.00	32382.4	6408.23	–4769.25	41065440.67	22745730.18	–30562451.36
599.1	8925.3	–2182.67	–28226.35	4764038.47	796726743.27	61608740.10
2,465.00	9550.2	–316.77	–27601.45	100341.80	761839953.07	8743248.48
3,527.00	65917.4	745.23	28765.75	555371.12	827468465.86	21437166.03
602	13819.5	–2179.77	–23332.15	4751387.41	544389148.36	50858664.40
2,655.00	26651.1	–126.77	–10500.55	16070.06	110261516.43	1331130.81
1,455.70	21865.9	–1326.07	–15285.75	1758455.66	233654103.75	20269937.85
276	3417.8	–2505.77	–33733.85	6278871.98	1137972527.00	84529189.10
617.5	3681.2	–2164.27	–33470.45	4684054.86	1120270915.23	72439011.75
11,797.00	182109.9	9015.23	144958.25	81274412.67	21012894710.67	1306832306.01
567.6	12522.8	–2214.17	–24628.85	4902538.79	606580172.87	54532401.62
697.8	10514.8	–	–26636.85	4342921.55	709521692.00	55510332.79

Business Analytics 4th edition Camm Solutions Manual

		2083.97				
634	8560.5	2147.77	–28591.15	4612906.27	817453766.09	61407146.21
109	1381.6	2672.77	–35770.05	7143687.40	1279496361.62	95605031.46
4,979.00	66606.5	2197.23	29454.85	4827829.60	867588283.54	64719150.12
5,142.00	53469.4	2360.23	16317.75	5570696.31	266269017.70	38513683.74
			Total	368589209.4	62647162947	3954149359

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{3954149359}{30} = 131804978.6$$

$$s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{368589209.4}{30}} = 3505.18$$

$$s_y = \sqrt{\frac{\sum(y - \bar{y})^2}{n-1}} = \sqrt{\frac{62647162947}{30}} = 45697.25$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{131804978.6}{(3505.18)(45697.25)} = 0.8229$$

Or using Excel, we use the formula = COVARIANCE.S(B2:B32,C2:C32) to calculate the covariance, which is 131804978.638. This indicates that there is a positive relationship between profits and market capitalization.

c. In the Excel file, we use the formula =CORREL(B2:B32,C2:C32) to calculate the correlation coefficient, which is 0.8229. This indicates that there is a strong linear relationship between profits and market capitalization.

26. a. Without Excel, we can use the calculations below to find the correlation coefficient:

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
7.1	7.02	0.2852	0.6893	0.0813	0.4751	0.1966
5.2	5.31	–1.6148	–1.0207	2.6076	1.0419	1.6483
7.8	5.38	0.9852	–0.9507	0.9706	0.9039	–0.9367
7.8	5.40	0.9852	–0.9307	0.9706	0.8663	–0.9170
5.8	5.00	–1.0148	–1.3307	1.0298	1.7709	1.3505
5.8	4.07	–1.0148	–2.2607	1.0298	5.1109	2.2942
9.3	6.53	2.4852	0.1993	6.1761	0.0397	0.4952
5.7	5.57	–1.1148	–0.7607	1.2428	0.5787	0.8481
7.3	6.99	0.4852	0.6593	0.2354	0.4346	0.3199
7.6	11.12	0.7852	4.7893	0.6165	22.9370	3.7605
8.2	7.56	1.3852	1.2293	1.9187	1.5111	1.7028
7.1	12.11	0.2852	5.7793	0.0813	33.3998	1.6482
6.3	4.39	–0.5148	–1.9407	0.2650	3.7665	0.9991
6.6	4.78	–0.2148	–1.5507	0.0461	2.4048	0.3331
6.2	5.78	–0.6148	–0.5507	0.3780	0.3033	0.3386
6.3	6.08	–0.5148	–0.2507	0.2650	0.0629	0.1291
7.0	10.05	0.1852	3.7193	0.0343	13.8329	0.6888
6.2	4.75	–0.6148	–1.5807	0.3780	2.4987	0.9719

Business Analytics 4th edition Camm Solutions Manual

5.5	7.22	–1.3148	0.8893	1.7287	0.7908	–1.1692
6.5	3.79	–0.3148	–2.5407	0.0991	6.4554	0.7999
6.0	3.62	–0.8148	–2.7107	0.6639	7.3481	2.2088
8.3	9.24	1.4852	2.9093	2.2058	8.4638	4.3208
7.5	4.40	0.6852	–1.9307	0.4695	3.7278	–1.3229
7.1	6.91	0.2852	0.5793	0.0813	0.3355	0.1652
6.8	5.57	–0.0148	–0.7607	0.0002	0.5787	0.0113
5.5	3.87	–1.3148	–2.4607	1.7287	6.0552	3.2354
7.5	8.42	0.6852	2.0893	0.4695	4.3650	1.4315
			Total	25.77407	130.0594	25.5517

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{25.5517}{26} = 0.9828$$

$$s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{25.77407}{26}} = 0.9956$$

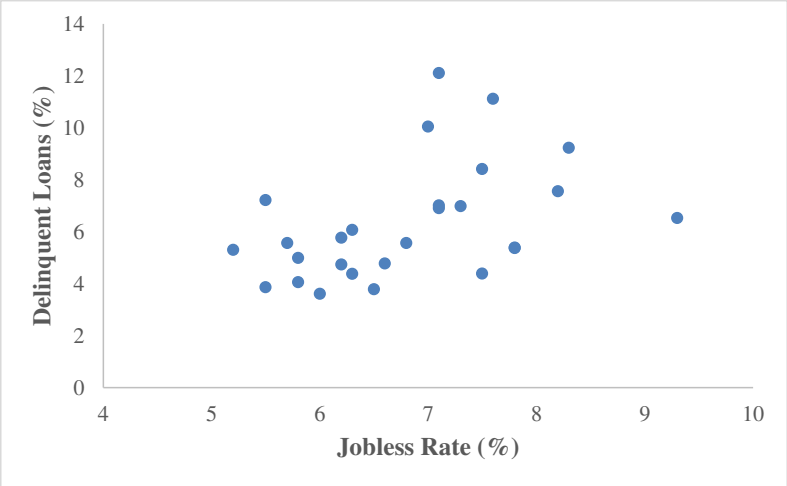
$$s_y = \sqrt{\frac{\sum(y - \bar{y})^2}{n-1}} = \sqrt{\frac{130.0594}{26}} = 2.2366$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{0.9828}{(0.9956)(2.2366)} = 0.44$$

Or we can use the Excel function CORREL.

The correlation coefficient indicates that there is a moderate positive linear relationship between jobless rate and delinquent loans. If the jobless rate were to increase, it is likely that an increase in the percentage of delinquent housing loans would also occur.

b.



27. a. Using the Excel function COUNTBLANK we find that there is one blank response in column C (Texture) and one blank response in column F (Depth of Chocolate Flavor of the Cup). With further investigation we find that the value of texture for respondent 157 is missing, and the value of Depth of the Chocolate Flavor of the Cup for respondent 199 is missing.

b. To help us identify erroneous values, we calculate the Average, Standard Deviation, Minimum and Maximum values for each variable.

Business Analytics 4th edition Camm Solutions Manual

# Missing Values:	0	1	0	0	1
Average:	72.29	105.94	77.11	78.51	77.36
Standard Deviation:	20.21	447.89	19.14	66.15	17.69
Minimum:	8.40	13.00	19.00	0.67	11.00
Maximum:	100.00	6666.00	120.00	997.00	100.00

We can immediately spot some surprising values in Column C (Texture) and column E (Sweetness). Further examination identifies the following: the value of Texture for respondent 68 is 6666, which is outside the range for this variable; the value of sweetness for respondent 72 is 997, which is outside the range for this variable; the value of Sweetness for respondent 85 is 0.67, which is outside the range for this variable and is not an integer. Additional examination of the other responses shows that the value of Taste for respondent 90 is 8.4, which is not an integer, and the value of Creaminess of filling for respondent 197 is 120, which is outside the range for this variable.

Business Analytics 4th edition Camm Solutions Manual

28. a. Using the Excel function COUNTBLANK we find that one observation is missing in column B and one observation missing in column C. Additional investigation shows that the missing value in column B is for year 2016 for the Phillies and the missing value in column C is for year 2016 for the Marlins. Review of major league baseball attendance data that are available from a reliable source shows that the Phillies' attendance in 2016 was 1,915,144, which is consistent with the value of the Phillies' attendance for the observation with the missing value of season. This supports our suspicion that the value of season for this observation is 2016. Review of major league baseball attendance data that are available from a reliable source shows that the Marlins' 2016 attendance was 1,712,417. (Note that we use the reference ESPN.com, at http://www.espn.com/mlb/attendance as of May 14, 2017, as our reliable data source for comparison.)

b. To help us identify erroneous values, we calculate the average, standard deviation, minimum and maximum values for Season and Attendance. (Note that we use the reference ESPN.com, at http://www.espn.com/mlb/attendance as of May 14, 2017, as our reliable data source for comparison.)

# Missing Values:	1	1
Average:	2,004	2,651,944.64
Standard Deviation:	124	2,333,151.88
Minimum:	214	-3,365,256.00
Maximum:	2,016	26,426,820.00

We immediately identify that there is an erroneous value for Season as all values should be between 2014 and 2016, but the minimum value is 214. We also identify that the minimum and maximum values for Attendance appear questionable. Review of major league baseball attendance data that are available from a reliable source shows that the Cubs' attendance in 2014 was 2,652,113, which is consistent with the value of the Cubs' attendance for the observation with the missing value of season. This supports our suspicion that the value of season for this observation is 2014.

The value for attendance for the Giants in 2016 is –3,365,256, which is unrealistic. Review of major league baseball attendance data that are available from a reliable source shows that the Giants' 2016 attendance was 3,365,256.

The value for attendance for the Cubs in 2013 is 26,426,820, which is unusually large. Review of major league baseball attendance data that are available from a reliable source shows that the Cubs' 2013 attendance was 2,642,682.

We can also sort the data in Excel by Team Name to help us identify attendance values that seem outside the norm for that team. Additional analysis of individual attendance values shows the following.

The value for attendance for the Royals in 2011 is 172,445, which is unusually small. Review of major league baseball attendance data that are available from a reliable source shows that the Royals' 2011 attendance was 1,724,450.

The value for attendance for the Marlins in 2014 is 9,732,283, which is unusually large compared to other season attendance values for the Marlins. Review of major league baseball attendance data that are available from a reliable source shows that the Marlins' 2014 attendance was 1,732,283.

The value for attendance for the Marlins in 2015 is 752,235, which is unusually small compared to other season attendance values for the Marlins. Review of major league baseball attendance data that are available from a reliable source shows that the Marlins' 2015 attendance was 1,752,235.

Business Analytics 4th edition Camm Solutions Manual

The value for attendance for the Orioles in 2014 is 22,464,473, which is unusually large. Review of major league baseball attendance data that are available from a reliable source shows that the Orioles' 2014 attendance was 2,464,473.

Business Analytics 4th edition Camm Solutions Manual

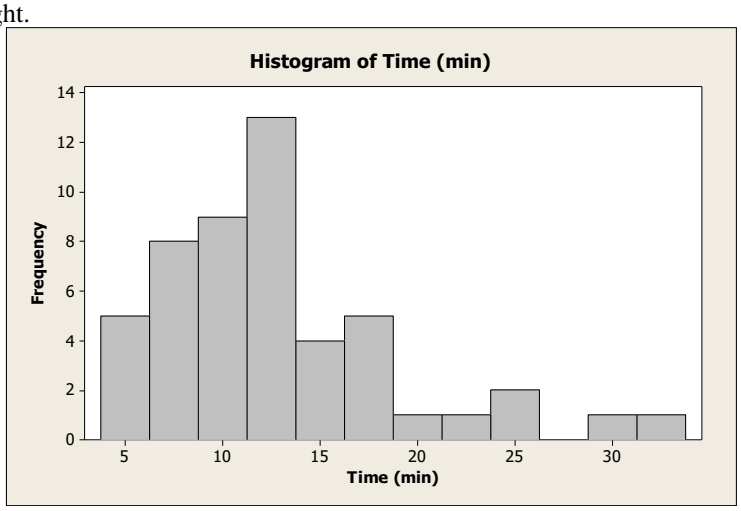
Chapter 2
Descriptive Statistics

Case Problem 1: Heavenly Chocolates Web Site Transactions

1. Descriptive statistics for the time spent on the web site, number of pages viewed, and amount spent are shown below.

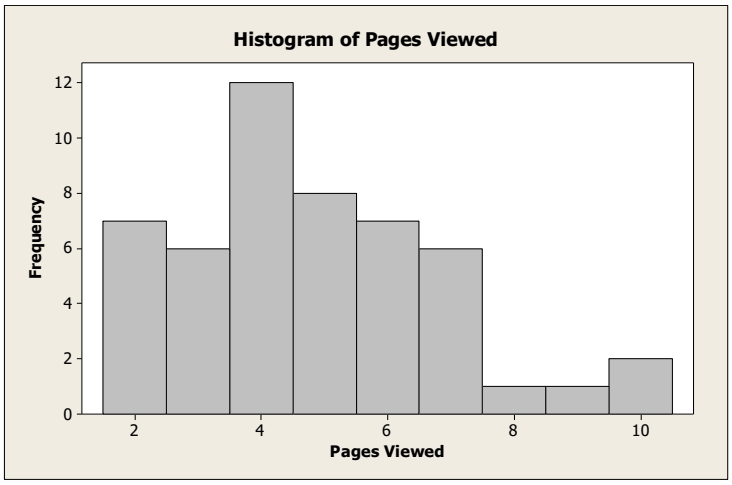
	Time (min)	Pages Viewed	Amount Spent ($)
Mean	12.8	4.8	68.13
Median	11.4	4.5	62.15
Standard Deviation	6.06	2.04	32.34
Range	28.6	8	140.67
Minimum	4.3	2	17.84
Maximum	32.9	10	158.51
Sum	640.5	241	3406.41

The mean time a shopper is on the Heavenly Chocolates web site is 12.8 minutes, with a minimum time of 4.3 minutes and a maximum time of 32.9 minutes. The following histogram demonstrates that the data are skewed to the right.

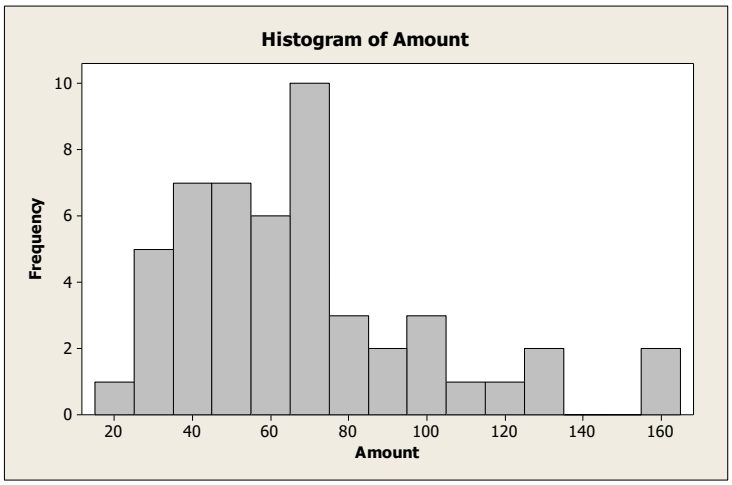


The mean number of pages viewed during a visit is 4.8 pages with a minimun of 2 pages and a maximum of 10 pages. A histogram of the number of pages viewed indicates that the data are slightly skewed to the right.

Business Analytics 4th edition Camm Solutions Manual



The mean amount spent for an online shopper is $68.13 with a minimum amount spent of $17.84 and a maximum amount spent of $158.51. The following histogram indicates that the data are skewed to the right.



2. Summary by Day of Week

Day of Week	Frequency	Total Amount Spent ($)	Average Amount Spent ($)
Sunday	5	218.15	43.63
Monday	9	813.38	90.38
Tuesday	7	414.86	59.27
Wednesday	6	341.82	56.97
Thursday	5	294.03	58.81
Friday	11	945.43	85.95
Saturday	7	378.74	54.11
Total	50	3406.41	

Business Analytics 4th edition Camm Solutions Manual

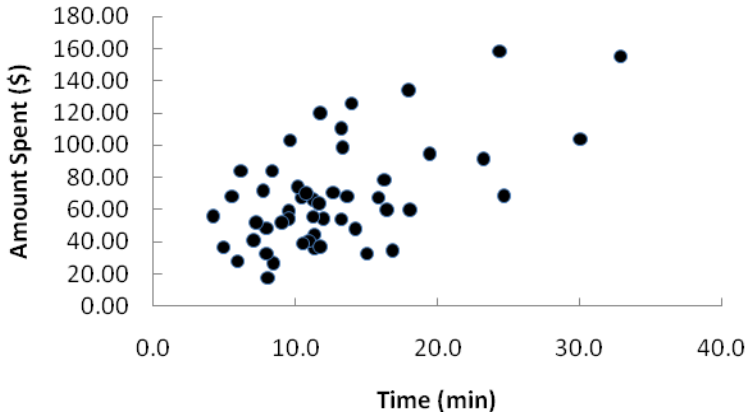
The above summary shows that Monday and Friday are the best days in terms of both the total amount spent and the averge amount spent per transaction. Friday had the most purchases (11) and the highest value for total amount spent ($945.43). Monday, with nine transactions, had the highest average amount spent per transaction ($90.38). Sunday was the worst sales day of the week in terms of number of transactions (5), total amount spent ($218.15), and average amount spent per transaction ($43.63). However, the sample size for each day of the week are very small, with only Friday having more than 10 transactions. We would suggest a larger sample size be taken before recommending any specific stratgegy based on the day of week statistics.

3. Summary by Type of Browser

Browser	Frequency	Total Amount Spent ($)	Average Amount Spent ($)
Firefox	16	1228.21	76.76
Chrome	27	1656.81	61.36
Other	7	521.39	74.48

Chrome was used by 27 of the 50 shoppers (54%). But, the average amount spent by customers who used Chrome ($61.36) is less than the average amount spent by customers who used Firefox ($76.76) or some other type of browser ($74.48). This result would suggest targeting special promotion offers to Firefox users or users of other types of browsers. But, before recommending any specific strategies based upon the type of browser, we would suggest taking a larger sample size.

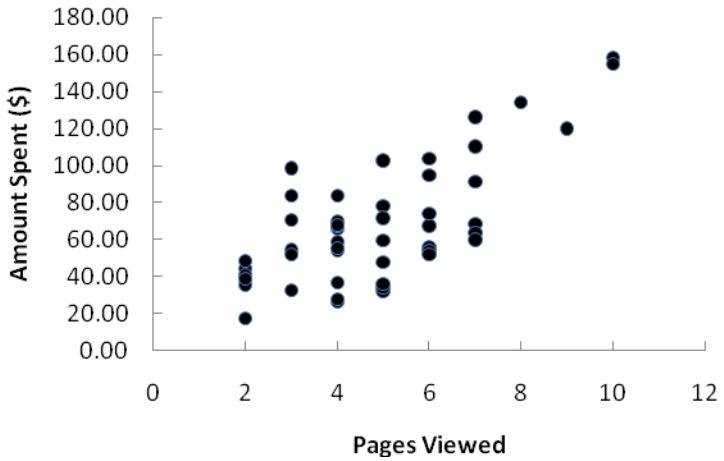
4. A scatter diagram showing the relationship between time spent on the web site and the amount spent follows:



The sample correlation coefficient between these two variables is .580. The scatter diagram and the sample correlation coefficient indicate a postive relationship between time spent on the web site and the total amount spent. Thus, the sample data support the conclusion that customers who spend more time on the web site spend more.

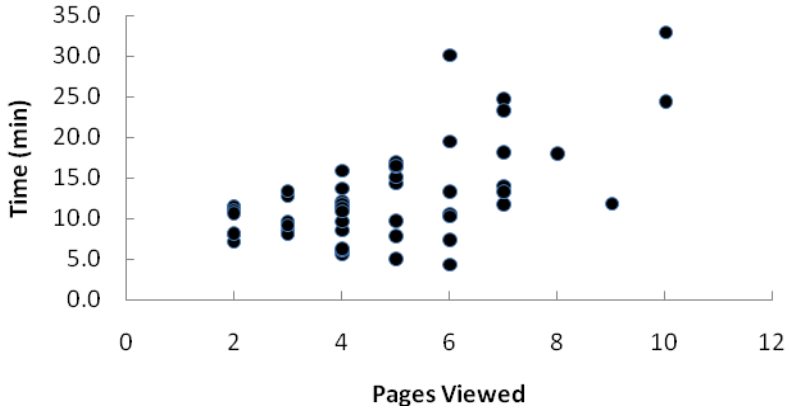
5. A scatter diagram showing the relationship between the number of pages viewed and the amount spent follows:

Business Analytics 4th edition Camm Solutions Manual



The sample correlation coefficient between these two variables is .724. The scatter diagram and the sample correlation coefficient indicate a postive relationship between time spent on the web site and the number of pages viewed. Thus, the sample data support the conclusion that customers who view more web site pages spend more.

6. A scatter diagram showing the relationship between the number of pages viewed and the time spent on the web site follows:



The sample correlation coefficient between these two variables is .596. The scatter diagram and the sample correlation coefficient indicate a postive relationship between the number of pages viewed and the time spent on the web site.

Summary: The analysis indicates that online shoppers who spend more time on the company's web site and/or view more web site pages spend more money during their visit to the web site. If Heavenly Chocolates can develop an attractive web site such that online shoppers are willing to spend more time on the web site and/or view more pages, there is a good possiblity that the company will experience greater sales. And, consideration should also be given to developing marketing strategies based upon possible differences in sales associated with the day of the week as well as differences in sales associated with the type of browser used by the customer.

Business Analytics 4th edition Camm Solutions Manual

Case Problem 2: African Elephant Populations

This case provides the student with the opportunity to use the geometric mean in conjunction with a graph (such as the boxplot) to analyze changes over time in the populations of elephants in several African nations.

1. Let's calculate the proportional change for each country over the 10-year period, 1979–1989. We'll begin by considering the Central African Republic. We have:

   19,000 = 63,000$\left[(x_1)(x_2)\cdots(x_{10})\right]$, so $\left[(x_1)(x_2)\cdots(x_{10})\right]$ = 0.301587 and

   $$\bar{x}_g = \sqrt[n]{(x_1)(x_2)\cdots(x_{10})} = \sqrt[10]{0.301587} = 0.887036$$

   So the mean annual change in the elephant population for the Central African Republic during this period is (0.887036 – 1)100% = –11.3%. From 1979 to 1989, the elephant population in the Central African Republic declined at an annual rate of 11.3%.

   Repeating these calculations for each nation yields the values in the following table.

Country	$(x_1)(x_2)\cdots(x_{10})$	$\bar{x}_g$	Mean Annual Change
Angola	1.0000	1.0000	0.0000
Botswana	2.5500	1.0981	0.0981
Cameroon	1.3086	1.0273	0.0273
Cen African Rep	0.3016	0.8870	–0.1130
Chad	0.2067	0.8541	–0.1459
Congo	6.4815	1.2055	0.2055
Dem Rep of Congo	0.2250	0.8614	–0.1386
Gabon	5.6716	1.1895	0.1895
Kenya	0.2923	0.8843	–0.1157
Mozambique	0.3394	0.8976	–0.1024
Somalia	0.2469	0.8695	–0.1305
Tanzania	0.2529	0.8716	–0.1284
Zambia	0.2733	0.8784	–0.1216
Zimbabwe	1.4333	1.0367	0.0367

   The elephant populations in several nations (Central African Republic, Chad, Democratic Republic of the Congo, Kenya, Mozambique, Somalia, Tanzania, and Zambia) declined at an annual rate of 10% or more from 1979 to 1989. During the same period a few nations (Botswana, Congo, and Gabon) experienced growth in their elephant populations.

2. Now let's calculate the proportional change for each country over the 18-year period, 1989–2007. We'll again begin by considering the Central African Republic. We have:

   3,334 = 19,000$\left[(x_1)(x_2)\cdots(x_{18})\right]$, so $\left[(x_1)(x_2)\cdots(x_{18})\right]$ = 0.175474 and

   $$\bar{x}_g = \sqrt[n]{(x_1)(x_2)\cdots(x_{18})} = \sqrt[18]{0.175474} = 0.907845$$

Business Analytics 4th edition Camm Solutions Manual

So the mean annual change in the elephant population for the Central African Republic during this period is (0.907845 – 1)100% = –9.2155%. From 1979 to 1989, the elephant population in the Central African Republic declined at an annual rate of 9.2%.

Repeating these calculations for each nation yields the values in the following table.

Country	$(x_1)(x_2)\cdots(x_{18})$	$\bar{x}_g$	**Mean Annual Change**
Angola	0.2040	0.9155	–0.0845
Botswana	3.4409	1.0711	0.0711
Cameroon	0.7258	0.9824	–0.0176
Cen African Rep	0.1755	0.9078	–0.0922
Chad	2.0758	1.0414	0.0414
Congo	0.3157	0.9380	–0.0620
Dem Rep of Congo	0.2790	0.9315	–0.0685
Gabon	0.9294	0.9959	–0.0041
Kenya	1.6651	1.0287	0.0287
Mozambique	1.4026	1.0190	0.0190
Somalia	0.0117	0.7809	–0.2191
Tanzania	2.0875	1.0417	0.0417
Zambia	0.7130	0.9814	–0.0186
Zimbabwe	2.3048	1.0475	0.0475

Only one country (Somalia) continues to experience average annual declines of 10% or more in its elephant population from 1989 to 2007, while most other nations had relatively small mean annual changes in their elephant populations.

3. Now let's calculate the proportional change for each country over the five-year period, 2007–2012. We'll again begin by considering the Central African Republic. We have:

2,285 = 3,334$\left[(x_1)(x_2)\cdots(x_5)\right]$, so $\left[(x_1)(x_2)\cdots(x_5)\right]$ = 0.685363 and

$$\bar{x}_g = \sqrt[n]{(x_1)(x_2)\cdots(x_5)} = \sqrt[5]{0.685363} = 0.927223$$

So the mean annual change in the elephant population for the Central African Republic during this period is (0.927223 – 1)100 = –7.2777%. From 2007 to 2012, the elephant population in the Central African Republic declined at an annual rate of 7.3%.

Repeating these calculations for each nation yields the values in the following table.

Country	$(x_1)(x_2)\cdots(x_5)$	$\bar{x}_g$	**Mean Annual Change**
Angola	1.0000	1.0000	0.0000
Botswana	0.9998	1.0000	0.0000
Cameroon	0.9130	0.9820	–0.0180
Cen African Rep	0.6854	0.9272	–0.0728

Business Analytics 4th edition Camm Solutions Manual

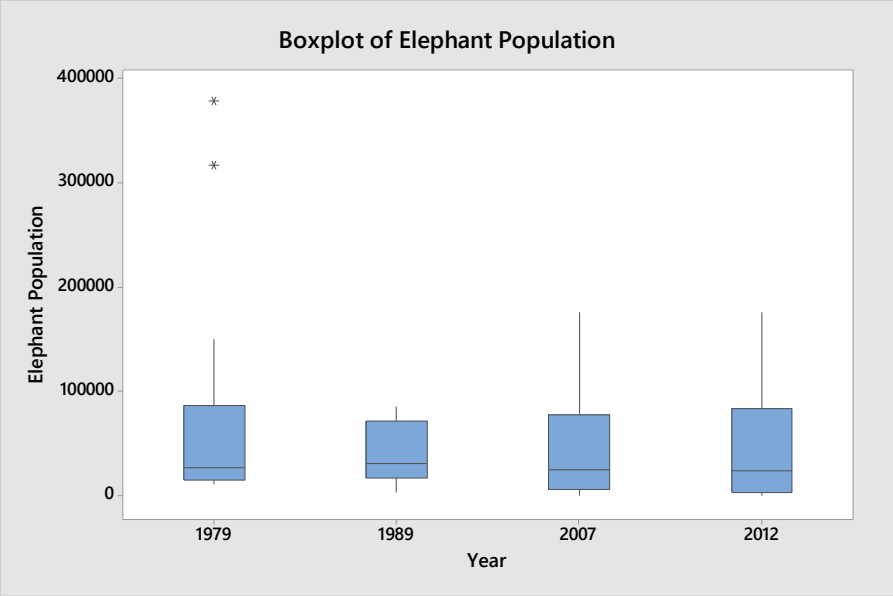
Chad	0.4668	0.8587	–0.1413
Congo	2.2298	1.1740	0.1740
Dem Rep of Congo	0.5766	0.8957	–0.1043
Gabon	1.0936	1.0181	0.0181
Kenya	1.1462	1.0277	0.0277
Mozambique	1.0163	1.0032	0.0032
Somalia	1.0000	1.0000	0.0000
Tanzania	0.7033	0.9320	–0.0680
Zambia	0.7386	0.9412	–0.0588
Zimbabwe	1.0119	1.0024	0.0024

Two countries (Chad and Democratic Republic of the Congo) experienced average annual declines in their elephant populations of 10% or more from 2007 to 2012, while most other nations had relatively small mean annual changes in their elephant populations.

4. Now we compare the results of our three analyses and draw conclusions.

Country	Mean Annual Change 1979–1989	Mean Annual Change 1989–2007	Mean Annual Change 2007–2012
Chad	–0.1459	0.0414	–0.1413
Dem Rep of Congo	–0.1386	–0.0685	–0.1043
Somalia	–0.1305	–0.2191	0.0000
Tanzania	–0.1284	0.0417	–0.0680
Zambia	–0.1216	–0.0186	–0.0588
Kenya	–0.1157	0.0287	0.0277
Cen African Rep	–0.1130	–0.0922	–0.0728
Mozambique	–0.1024	0.0190	0.0032
Angola	0.0000	-0.0845	0.0000
Cameroon	0.0273	–0.0176	–0.0180
Zimbabwe	0.0367	0.0475	0.0024
Botswana	0.0981	0.0711	0.0000
Gabon	0.1895	–0.0041	0.0181
Congo	0.2055	–0.0620	0.1740

Business Analytics 4th edition Camm Solutions Manual



From these boxplots we can see that the population of elephants declined dramatically from 1979 to 1989, generally started to come back between 1989 and 2007, and stabilized between 2007 and 2012. We can also see that the declining trend that was established between 1979 and 1989 continues for the elephant populations in some African nations.

Several nations appear to have reversed the declines in elephant populations they experienced from 1979 to 1989, but the growth rates are still generally low (and in some countries still negative). At 2007–2012 rates of change, it will take many decades for the elephant populations to recover to their 1979 levels.

Business Analytics 4th edition Camm Solutions Manual