

Table of Contents

Word Solutions for chapter 2	2
Case Solutions for chapter 2	20
Excel Solutions for chapter 2	41
1 2.3Soln	41
1.1 CondFormat	41
1.2 Sheet2	42
2 2.12Soln	43
2.1 Data	43
3 2.17Soln	45
3.1 Sheet1	45
4 2.19Soln	47
4.1 Sheet1	47
5 2.24Soln	49
5.1 xl5galry.xls	49
6 2.25Soln	52
6.1 Data	52
7 2.26Soln	54
7.1 Data	54

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

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%
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
		Previous Year	Current Year	
		On-time	On-time	Change in On-
1	Carrier	Percentage	Percentage	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
		Previous Year	Current Year	
		On-time	On-time	Change in On-
1	Carrier	Percentage	Percentage	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	% 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.

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

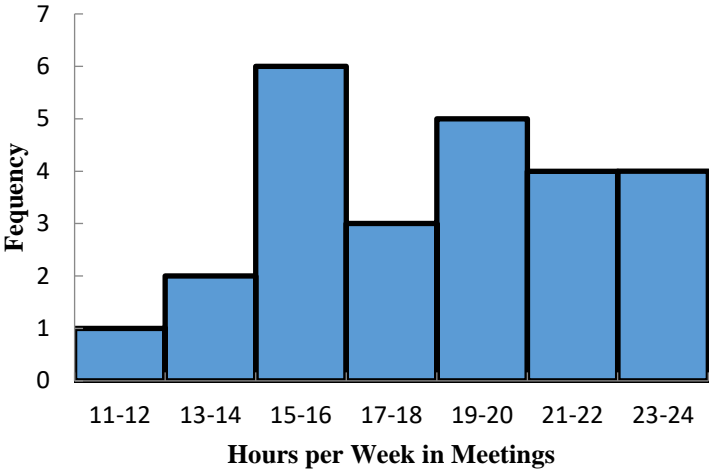
c. The most frequent most-visited-website is google.com (GOOG); second is yahoo.com (YAH).

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%
	25	100%

c.



The distribution is slightly skewed to the left.

7. a.

Industry	Frequency	% Frequency
Bank	26	13%
Cable	44	22%
Car	42	21%
Cell	60	30%
Collection	28	14%
Total	200	100%

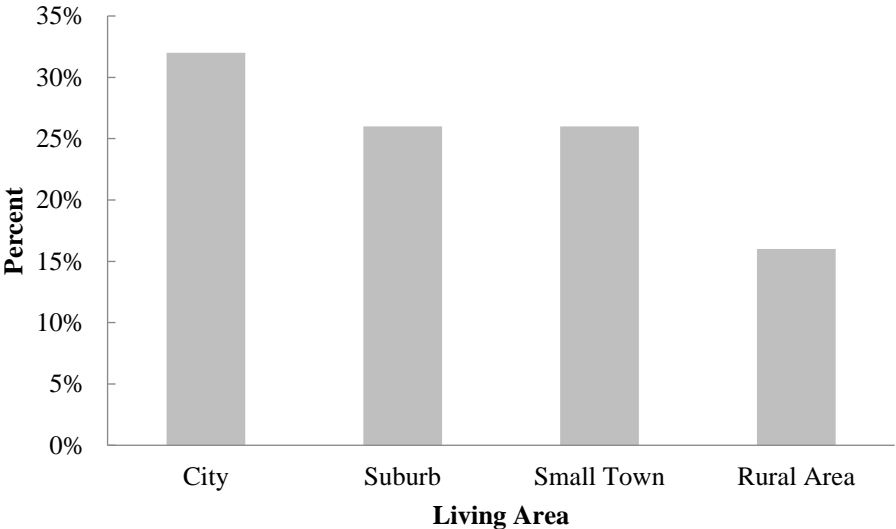
- b. The cellular phone providers had the highest number of complaints.
- c. The percentage frequency distribution shows that the two financial industries (banks and collection agencies) had about the same number of complaints. Also, new car dealers and cable and satellite television companies also had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

8. Percent Frequency Distribution:

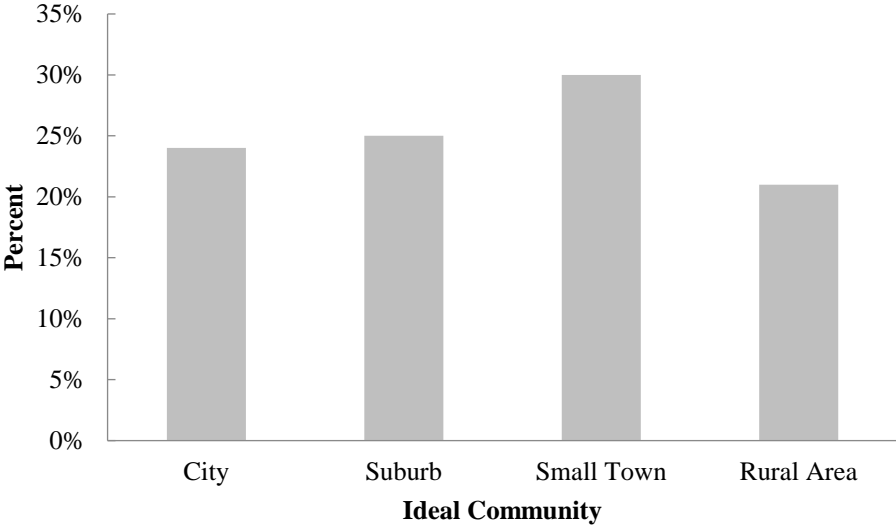
Living Area	Live Now	Ideal Community
City	32/100=32%	24/100=24%
Suburb	26/100=26%	25/100=25%
Small Town	26/100=26%	30/100=30%
Rural Area	16/100=16%	21/100=21%
Total	100%	100%

Histograms:

Where do you live now?



What do you consider the ideal community?



- a. Most adults are now living in a city (32%).
- b. Most adults consider the ideal community a small town (30%).
- c. Changes in percentages by living area: City -8%, Suburb -1%, Small Town +4%, and Rural Area +5%.
Suburb living is steady, but the trend would be that living in the city would decline while living in small towns and rural areas would increase.

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

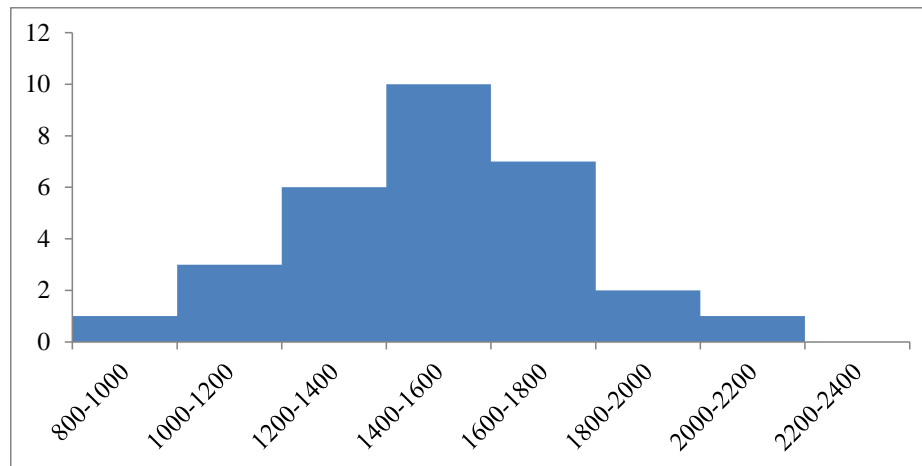
11. a – d.

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.

12. a.

Class	Frequency
800-1000	1
1000-1200	3
1200-1400	6
1400-1600	10
1600-1800	7
1800-2000	2
2000-2200	1
2200-2400	0



- b. The distribution is slightly skewed to the right.
- c. The most common score for students is between 1400 and 1600. No student scored above 2200, and only 3 students scored above 1800. Only 4 students scored below 1200.

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 p th percentile is given by the formula $L_p = \frac{p}{100}(n + 1)$

For our data 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

$$15 + 0.80 \cdot (20 - 15) = 19.$$

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

$$3500=5000 \left[(x_1)(x_2) \cdots (x_9) \right], \text{ so } \left[(x_1)(x_2) \cdots (x_9) \right] = 0.700$$

where x_1, x_2, \dots 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,

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

where x_1, x_2, \dots are the growth factors for years, 1, 2, etc. through year 8.

Next, we calculate $\bar{x}_g = \sqrt[n]{(x_1)(x_2) \cdots (x_n)} = \sqrt[8]{1.80} = 1.07624$

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

For the Trippi mutual fund we have:

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

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

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.

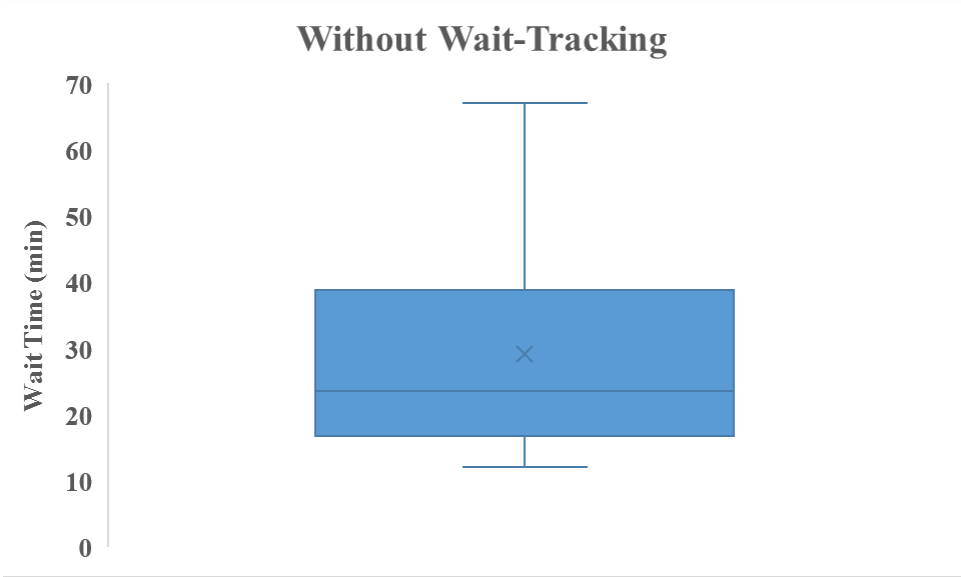
Alternatively, we can use Excel and the function GEOMEAN as shown below:

	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				

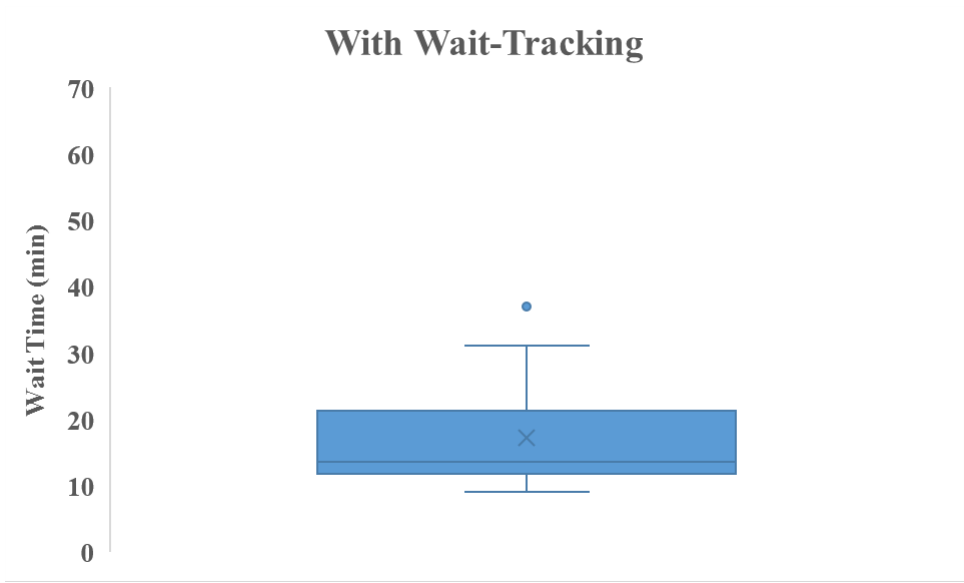
	A	B	C	D	E
1		Stivers		Trippi	
2	Year	End of Year Value	Growth Factor	End of Year Value	Growth Factor
3	0	10000		5000	
4	1	11000	=B4/B3	5600	=D4/D3
5	2	12000	=B5/B4	6300	=D5/D4
6	3	13000	=B6/B5	6900	=D6/D5
7	4	14000	=B7/B6	7600	=D7/D6
8	5	15000	=B8/B7	8500	=D8/D7
9	6	16000	=B9/B8	9200	=D9/D8
10	7	17000	=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. $\text{Mean} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1291.5}{48} = 26.906$
- 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.

c.

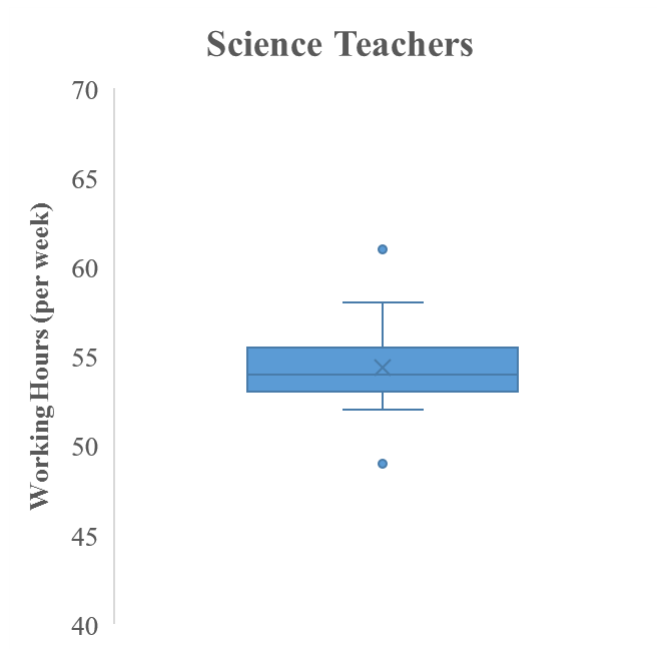


d.

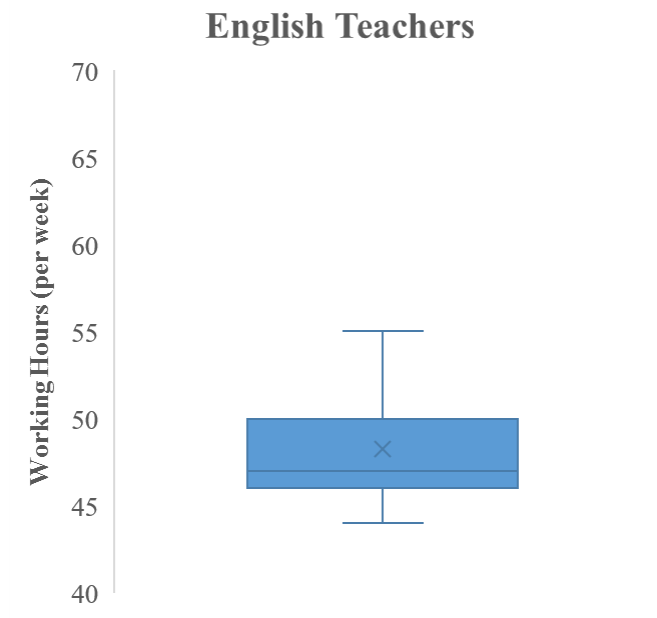


e. Wait times for patients with the wait-tracking system are substantially shorter than those for patients without the wait-tracking system. However, some patients with the wait-tracking system still experience long waits.

- 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.



d.



- e. The box plots show that science teachers spend more hours working per week than English teachers. The box plot 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, $z = \frac{37-29.1}{16.6} = 0.48$.
- 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, $z = \frac{37-17.2}{9.28} = 2.13$.
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

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.

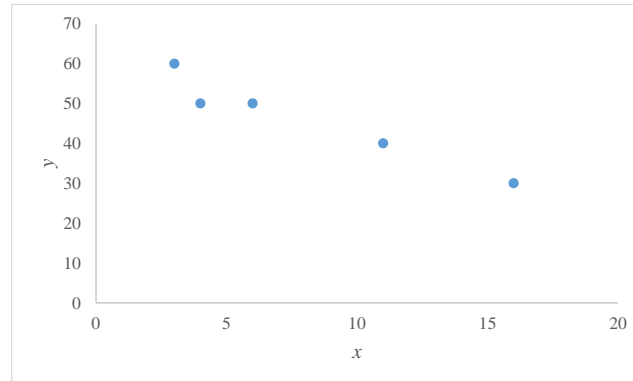
Without Wait-Tracking System		With Wait-Tracking System	
Wait Time	z -Score	Wait Time	z -Score
24	-0.31	31	1.49
67	2.28	11	-0.67
17	-0.73	14	-0.34
20	-0.55	18	0.09
31	0.11	12	-0.56
44	0.90	37	2.13
12	-1.03	9	-0.88
23	-0.37	13	-0.45
16	-0.79	12	-0.56
37	0.48	15	-0.24

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. 615 is one standard deviation above the mean. The empirical rule states that 68% of data values will be within one standard deviation of the mean. Because a bell-shaped distribution is symmetric half of the remaining values will be greater than the (mean + 1 standard deviation) and half will be below (mean - 1 standard deviation). In other words, we expect that $0.5*(1 - 68\%) = 16\%$ of the data values will be greater than (mean + 1 standard deviation) = 615.
- b. 715 is two standard deviations above the mean. The empirical rule states that 95% of data values will be within two standard deviations of the mean, and we expect that $0.5*(1 - 95\%) = 2.5\%$ of data values will be above two standard deviations above the mean.
- c. 415 is one standard deviation below the mean. The empirical rule states that 68% of data values will be within one standard deviation of the mean, and we expect that $0.5*(1 - 68\%) = 16\%$ of data values will be below one standard deviation below the mean. 515 is the mean, so we expect that 50% of the data values will be below the mean. Therefore, we expect $50\% - 16\% = 36\%$ of the data values will be between the mean and one standard deviation below the mean (between 414 and 515).
- d. $z = \frac{620-515}{100} = 1.05$

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.

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	-2083.97	-26636.85	4342921.55	709521692.00	55510332.79
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	<u>5570696.31</u>	<u>266269017.70</u>	<u>38513683.74</u>
			Total	368589209.4	62647162947	3954149359

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

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

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

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

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
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

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

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

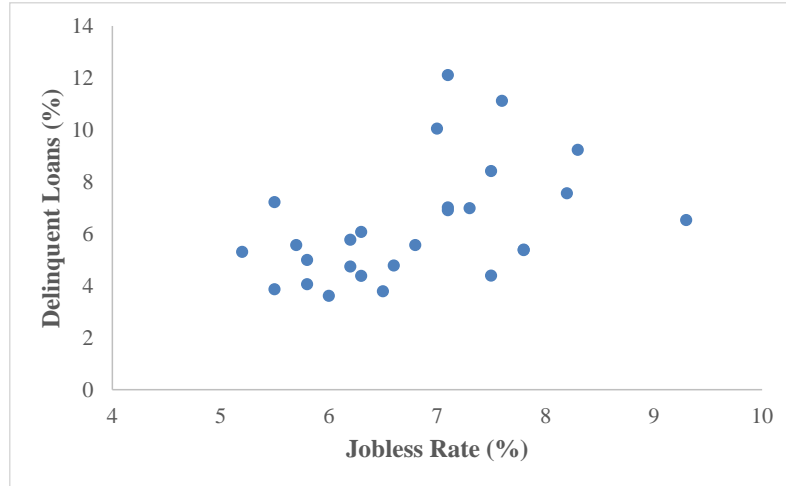
$$s_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n - 1}} = \sqrt{\frac{130.0594}{26}} = 2.2366$$

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

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.

# 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 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.

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.

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.

CH 01 - Introduction

Multiple Choice

1. The decisions concerning an organization's goals and future plans are called

- a. financial decisions.
- b. tactical decisions.
- c. strategic decisions.
- d. operational decisions.

ANSWER:

c

RATIONALE:

Strategic decisions involve higher-level issues concerned with the overall direction of the organization.

POINTS:

1

DIFFICULTY:

Easy

REFERENCES:

DECISION MAKING

QUESTION TYPE:

Multiple Choice

HAS VARIABLES:

False

NATIONAL STANDARDS:

United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS:

Bloom's: Remember

DATE CREATED:

1/23/2018 10:15 AM

DATE MODIFIED:

3/7/2018 10:56 AM

2. Tactical decisions are concerned with

- a. the day-to-day activities of the organization.
- b. the goals and plans of the organization.
- c. the domain of operations managers, who are close to the customer.
- d. how the organization should achieve the goals and objectives set by its strategy.

ANSWER:

d

RATIONALE:

Tactical decisions concern how the organization should achieve the goals and objectives set by its strategy.

POINTS:

1

DIFFICULTY:

Moderate

REFERENCES:

DECISION MAKING

QUESTION TYPE:

Multiple Choice

HAS VARIABLES:

False

NATIONAL STANDARDS:

United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS:

Bloom's: Remember

DATE CREATED:

1/23/2018 10:15 AM

DATE MODIFIED:

3/7/2018 10:59 AM

3. Picks and Axes Inc. is an Internet-based retail seller of hiking boots and mountaineering gear. The company decides to open retail stores across the major areas of the city to help complement its Internet-based strategy. This activity would be categorized as a(n)

- a. tactical decision.
- b. operational decision.
- c. strategic decision.
- d. financial decision.

ANSWER:

c

RATIONALE:

Strategic decisions involve higher-level issues concerned with the overall direction of the organization. These decisions define the organization's overall goals and aspirations for the

CH 01 - Introduction

future. Strategic decisions are usually the domain of higher-level executives and have a time horizon of three to five years.

POINTS: 1
DIFFICULTY: Moderate
REFERENCES: DECISION MAKING
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Apply
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 10:56 AM

4. _____ is the most critical step of the decision-making process.
- a. Choosing an alternative
 - b. Identifying and defining the problem
 - c. Evaluating the alternatives
 - d. Determining the set of alternatives

ANSWER: b
RATIONALE: Step 1 of decision making, identifying and defining the problem, is the most critical. Only if the problem is well-defined, with clear metrics of success or failure (step 2), can a proper approach for solving the problem (steps 3 and 4) be devised. Decision making concludes with the choice of an alternative (step 5).

POINTS: 1
DIFFICULTY: Moderate
REFERENCES: DECISION MAKING
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Understand
DATE CREATED: 3/7/2018 11:03 AM
DATE MODIFIED: 3/7/2018 1:40 PM

5. Which of the following is not an approach to making decisions?
- a. Tradition
 - b. Rules of thumb
 - c. Intuition
 - d. Guess and check

ANSWER: d
RATIONALE: There are a number of approaches to making decisions: tradition ("We've always done it this way"), intuition ("gut feeling"), and rules of thumb ("As the restaurant owner, I schedule twice the number of waiters and cooks on holidays").

POINTS: 1
DIFFICULTY: Moderate
REFERENCES: BUSINESS ANALYTICS DEFINED
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False

CH 01 - Introduction

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Understand

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/28/2018 2:22 PM

6. Data-driven decision making tends to decrease a firm's

- a. market value.
- b. productivity.
- c. risk.
- d. profit.

ANSWER: c

RATIONALE: Firms guided by data-driven decision making have higher productivity and market value and increased output and profitability.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: BUSINESS ANALYTICS DEFINED

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Understand

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 10:56 AM

7. Data dashboards are a type of _____ analytics.

- a. predictive
- b. descriptive
- c. prescriptive
- d. decision

ANSWER: b

RATIONALE: Descriptive analytics encompass the set of techniques that describes what has happened in the past.

POINTS: 1

DIFFICULTY: Easy

REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Remember

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 11:08 AM

8. The extraction of information on the number of shipments, how much was included in each shipment, the date each shipment was sent, and so on from the manufacturing plant's database exemplifies

- a. spreadsheet models.
- b. data dashboards.
- c. data mining.
- d. data queries.

ANSWER: d

CH 01 - Introduction

RATIONALE: A data query is a request for information with certain characteristics from a database.
POINTS: 1
DIFFICULTY: Moderate
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 11:10 AM

9. Corporate-level managers use _____ to summarize sales by region, current inventory levels, and other company-wide metrics all in a single screen.

- a. simulations
- b. crosstabulation
- c. data dashboards
- d. tables

ANSWER: c
RATIONALE: For corporate-level managers, daily data dashboards might summarize sales by region, current inventory levels, and other company-wide metrics.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 10:56 AM

10. A forecast that helps direct police officers to areas where crimes are likely to occur based on past data is an example of

- a. predictive analytics.
- b. decision analysis.
- c. prescriptive analytics.
- d. descriptive analytics.

ANSWER: a
RATIONALE: Predictive analytics consists of techniques that use models constructed from past data to predict the future.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

CH 01 - Introduction

KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 11:12 AM

11. Which one of the following is used in predictive analytics?

- a. Data dashboard
- b. Linear regression
- c. Data visualization
- d. Optimization model

ANSWER: b

RATIONALE: Linear regression, time series analysis, some data-mining techniques, and simulation, often referred to as risk analysis, all fall under the banner of predictive analytics.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Understand

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/28/2018 2:26 PM

12. A retail store owner offers a discount on product A and predicts that the customers would purchase products B and C in addition to product A. Identify the technique used to make such a prediction.

- a. Data query
- b. Simulation
- c. Data mining
- d. Data dashboards

ANSWER: c

RATIONALE: Data mining is a technique used to find patterns or relationships among elements of the data in a large database.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Apply

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 10:56 AM

13. _____ are used in the pharmaceutical industry to assess the risk of introducing a new drug.

- a. Data dashboards
- b. Charts
- c. Spreadsheet models
- d. Simulations

ANSWER: d

RATIONALE: Simulation involves the use of probability and statistics to construct a computer model to study the impact of uncertainty on a decision.

CH 01 - Introduction

POINTS: 1
DIFFICULTY: Moderate
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Apply
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/28/2018 2:27 PM

14. Which of the following analytical techniques helps us arrive at the best decision?
- a. Predictive analytics
 - b. Data mining
 - c. Prescriptive analytics
 - d. Descriptive analytics

ANSWER: c
RATIONALE: Prescriptive analytics indicate a best course of action to take; that is, the output of a prescriptive model is a best decision.

POINTS: 1
DIFFICULTY: Moderate
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Understand
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 10:56 AM

15. Simulation optimization helps
- a. in identifying the constraints of the situation.
 - b. to find good decisions in highly complex and highly uncertain settings.
 - c. in assigning values to outcomes.
 - d. to model certainty using optimization techniques.

ANSWER: b
RATIONALE: Simulation optimization combines the use of probability and statistics to model uncertainty with optimization techniques to find good decisions in highly complex and highly uncertain settings.

POINTS: 1
DIFFICULTY: Moderate
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember

CH 01 - Introduction

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 10:56 AM

16. When a decision maker is faced with several alternatives and an uncertain set of future events, s/he uses _____ to develop an optimal strategy.

- a. utility theory
- b. predictive analytics
- c. data mining
- d. decision analysis

ANSWER: d

RATIONALE: The techniques of decision analysis can be used to develop an optimal strategy when a decision maker is faced with several decision alternatives and an uncertain set of future events.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Understand

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/28/2018 2:27 PM

17. _____ assigns values to outcomes based on the decision maker's attitude toward risk, loss, and other factors.

- a. Simulation optimization
- b. Utility theory
- c. Optimization model
- d. Data dashboard

ANSWER: b

RATIONALE: Decision analysis employs utility theory, which assigns values to outcomes based on the decision maker's attitude toward risk, loss, and other factors.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Remember

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 12:24 PM

18. Which of the following best exemplifies big data?

- a. Five hundred Facebook users upload one thousand pictures per day.
- b. Cellphone owners around the world generate vast amounts of data by calling, texting, tweeting, and browsing the Web on a daily basis.
- c. A local grocery store collects data from those that scan their loyalty card.
- d. A pharmacy keeps track of customer purchases to send its customers coupons.

CH 01 - Introduction

ANSWER: b
RATIONALE: Big data is simply a set of data that cannot be managed, processed, or analyzed with commonly available software in a reasonable amount of time.
POINTS: 1
DIFFICULTY: Moderate
REFERENCES: BIG DATA
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Apply
DATE CREATED: 3/7/2018 12:25 PM
DATE MODIFIED: 3/28/2018 2:27 PM

19. Which of the following sources of big data is not publicly available?
- a. Twitter
 - b. Weather data
 - c. Medical records
 - d. Sports records

ANSWER: c
RATIONALE: While some sources of big data are publicly available (Twitter, weather data, etc.), other sources contain private information. Medical records, bank account information, and credit card transactions, for example, are all highly confidential and must be protected from computer hackers.
POINTS: 1
DIFFICULTY: Moderate
REFERENCES: BIG DATA
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Apply
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 12:28 PM

20. Advanced analytics generally refers to
- a. descriptive and prescriptive analytics.
 - b. simulation.
 - c. predictive and prescriptive analytics.
 - d. decision analysis.

ANSWER: c
RATIONALE: Predictive and prescriptive analytics are sometimes referred to as advanced analytics.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: BUSINESS ANALYTICS IN PRACTICE
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False

CH 01 - Introduction

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Remember

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/28/2018 2:28 PM

21. In the financial sector, _____ are used to construct financial instruments such as derivatives.

- a. descriptive and prescriptive models
- b. predictive models
- c. descriptive models
- d. prescriptive models

ANSWER: b

RATIONALE: Predictive models are used to forecast future financial performance and to construct financial instruments such as derivatives.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: BUSINESS ANALYTICS IN PRACTICE

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Understand

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 12:31 PM

22. Optimization models can be used to

- a. assess the risk of investment portfolios.
- b. forecast future financial performance.
- c. successfully manage commercial real estate risk.
- d. decide on how to invest cash received from insurance policies.

ANSWER: d

RATIONALE: GE Asset Management uses optimization models to decide how to invest its own cash received from insurance policies and other financial products.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: BUSINESS ANALYTICS IN PRACTICE

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Understand

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 12:32 PM

23. Utility theory is the study of the _____ or relative desirability of a particular outcome that reflects the decision maker's attitude toward a collection of factors, such as profit, loss, and risk.

- a. total worth

CH 01 - Introduction

- b. total cost
- c. feasibility
- d. financial wellness

ANSWER: a

RATIONALE: Utility theory is the study of the total worth or relative desirability of a particular outcome that reflects the decision maker's attitude toward a collection of factors, such as profit, loss, and risk.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Remember

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 12:33 PM

24. _____ refers to the technology that allows data, collected from sensors in all types of machines, to be sent over the Internet to repositories where it can be stored and analyzed.

- a. Internet of Things (IoT)
- b. MapReduce
- c. Hadoop
- d. Advanced analytics

ANSWER: a

RATIONALE: Internet of Things (IoT) refers to the technology that allows data, collected from sensors in all types of machines, to be sent over the Internet to repositories where it can be stored and analyzed.

POINTS: 1

DIFFICULTY: Easy

REFERENCES: BIG DATA

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Remember

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 12:37 PM

25. _____ refers to a programming model used within Hadoop that performs the two major steps for which it is named: the map step and the reduce step.

- a. MapReduce
- b. Internet of Things (IoT)
- c. Advanced analytics
- d. Optimization model

CH 01 - Introduction

ANSWER: a
RATIONALE: MapReduce refers to a programming model used within Hadoop that performs the two major steps for which it is named: the map step and the reduce step.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: BIG DATA
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 12:39 PM

26. _____ is an open-source programming environment that supports big data processing through distributed storage and distributed processing on clusters of computers.

- Hadoop
- Excel
- Java
- MapReduce

ANSWER: a
RATIONALE: Hadoop is an open-source programming environment that supports big data processing through distributed storage and distributed processing on clusters of computers.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: BIG DATA
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 12:40 PM

27. _____ analytics are techniques that use models, constructed from past data, to predict the future or to ascertain the impact of one variable on another.

- Predictive
- Descriptive
- Simulation
- Prescriptive

ANSWER: a
RATIONALE: Predictive analytics are techniques that use models, constructed from past data, to predict the future or to ascertain the impact of one variable on another.
POINTS: 1
DIFFICULTY: Moderate

CH 01 - Introduction

REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Remember

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 12:43 PM

28. A _____ decision involves higher-level issues and is concerned with the overall direction of the organization, defining the overarching goals and aspirations for the organization's future.

- a. strategic
- b. tactical
- c. intuitive
- d. operational

ANSWER: a

RATIONALE: A strategic decision involves higher-level issues and is concerned with the overall direction of the organization, defining the overarching goals and aspirations for the organization's future.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: DECISION MAKING

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Remember

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 12:46 PM

29. A _____ decision is concerned with how the organization should achieve the goals and objectives set by its strategy.

- a. tactical
- b. strategic
- c. intuitive
- d. operational

ANSWER: a

RATIONALE: A tactical decision is concerned with how the organization should achieve the goals and objectives set by its strategy.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: DECISION MAKING

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

CH 01 - Introduction

KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 12:47 PM

30. _____ analytics use techniques that take input data and yield a best course of action.
- Prescriptive
 - Simulation
 - Strategic
 - Operational

ANSWER: a
RATIONALE: Predictive analytics uses techniques that take input data and yield a best course of action.
POINTS: 1
DIFFICULTY: Moderate
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 12:49 PM

31. In the spectrum of business analytics, which is the most complex?
- Descriptive
 - Predictive
 - Prescriptive
 - Operational

ANSWER: c
RATIONALE: Prescriptive analytics is the most complex in the spectrum of business analytics.
POINTS: 1
DIFFICULTY: Moderate
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Multiple Choice
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Understand
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 12:50 PM

32. In order to manage an organization's human resource activities, such as hiring employees, tracking, and influencing employee retention, HR personnel use
- descriptive and predictive analytics.
 - descriptive and prescriptive analytics.
 - predictive and prescriptive analytics.
 - predictive analytics.

ANSWER: a

CH 01 - Introduction

RATIONALE: The HR analytics team uses descriptive and predictive analytics to support employee hiring and to track and influence retention.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Understand

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 10:56 AM

33. A better understanding of consumer behavior through analytics directly leads to
- a. more profits.
 - b. better pricing strategies.
 - c. reduced advertising costs.
 - d. reduced risk.

ANSWER: b

RATIONALE: A better understanding of consumer behavior through analytics leads to the better use of advertising budgets, more effective pricing strategies, improved forecasting of demand, improved product line management, and increased customer satisfaction and loyalty.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: BIG DATA

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Understand

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 10:56 AM

34. A light bulb manufacturer uses descriptive analytics
- a. to present supply chain to managers visually.
 - b. to achieve efficiency in delivery of goods.
 - c. to schedule staff and vehicle for delivery.
 - d. to plan capacity utilization by incorporating the inherent uncertainty in commodities pricing.

ANSWER: a

RATIONALE: The light bulb manufacturer has successfully used descriptive analytics to present the status of its supply chain to managers visually.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: BUSINESS ANALYTICS IN PRACTICE

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

CH 01 - Introduction

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Apply

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 1:01 PM

35. The U.S. Internal Revenue Service uses _____ to identify patterns that distinguish questionable annual personal income tax filings.

- a. utility theory
- b. prescriptive analytics
- c. data mining
- d. decision analysis

ANSWER: c

RATIONALE: The U.S. Internal Revenue Service uses data mining to identify patterns that distinguish questionable annual personal income tax filings.

POINTS: 1

DIFFICULTY: Easy

REFERENCES: BUSINESS ANALYTICS IN PRACTICE

QUESTION TYPE: Multiple Choice

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Remember

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 1:03 PM

Subjective Short Answer

36. _____ may be used to develop an optimal strategy when a decision maker is faced with several decision alternatives and an uncertain set of future events.

ANSWER: Decision analysis

RATIONALE: Decision analysis is a technique used to develop an optimal strategy when a decision maker is faced with several decision alternatives and an uncertain set of future events.

POINTS: 1

DIFFICULTY: Moderate

REFERENCES: BUSINESS ANALYTICS IN PRACTICE

QUESTION TYPE: Subjective Short Answer

HAS VARIABLES: False

NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics

KEYWORDS: Bloom's: Understand

DATE CREATED: 1/23/2018 10:15 AM

DATE MODIFIED: 3/7/2018 1:05 PM

37. An increase in data _____ would help to protect stored data from destructive forces or unauthorized users.

ANSWER: security

RATIONALE: Data security helps to protect stored data from destructive forces or unauthorized users.

POINTS: 1

CH 01 - Introduction

DIFFICULTY: Moderate
REFERENCES: BUSINESS ANALYTICS IN PRACTICE
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:07 PM

38. _____ are analytical tools that describe what has happened.

ANSWER: Descriptive analytics
RATIONALE: Descriptive analytics are analytical tools that describe what has happened.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:11 PM

39. The use of analytical techniques for better understanding patterns and relationships that exist in large data sets is _____.

ANSWER: data mining
RATIONALE: Data mining is a technique used to find patterns or relationships among elements of the data in a large database.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:14 PM

40. A dashboard is a collection of tables, charts, and maps to help management _____ selected aspects of the company's performance.

ANSWER: monitor
RATIONALE: A dashboard is a collection of tables, charts, and maps to help management monitor selected aspects of the company's performance.

CH 01 - Introduction

POINTS: 1
DIFFICULTY: Moderate
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:16 PM

41. A decision concerned with how the organization is run from day to day is known as a(n) _____.

ANSWER: operational decision

RATIONALE: An operational decision is a decision concerned with how the organization is run from day to day.

POINTS: 1
DIFFICULTY: Easy
REFERENCES: DECISION MAKING
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:17 PM

42. A mathematical model that gives the best decision, subject to the situation's constraints, is an a(n) _____.

ANSWER: optimization model

RATIONALE: An optimization model is a mathematical model that gives the best decision, subject to the situation's constraints.

POINTS: 1
DIFFICULTY: Easy
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:18 PM

43. A data _____ is a request to obtain information with certain characteristics from a database.

ANSWER: query

RATIONALE: A data query is a request to obtain information with certain characteristics from a database.

POINTS: 1

CH 01 - Introduction

DIFFICULTY: Easy
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:22 PM

44. Business analytics is the _____ process of transforming data into insight for making better decisions.

ANSWER: scientific
RATIONALE: Business analytics is the scientific process of transforming data into insight for making better decisions.
POINTS: 1
DIFFICULTY: Moderate
REFERENCES: BUSINESS ANALYTICS DEFINED
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:24 PM

45. A data _____ is trained in both computer science and statistics and knows how to effectively process and analyze large amounts of data.

ANSWER: scientist
RATIONALE: A data scientist is trained in both computer science and statistics and knows how to effectively process and analyze large amounts of data.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: BIG DATA
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:27 PM

46. The use of probability and statistics to construct a computer model to study the impact of uncertainty on the decision at hand is called _____.

ANSWER: simulation
RATIONALE: Simulation is the use of probability and statistics to construct a computer model to study the impact of uncertainty on the decision at hand.

CH 01 - Introduction

POINTS: 1
DIFFICULTY: Easy
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:28 PM

47. Predictive and prescriptive analytics can also be referred to as _____.

ANSWER: advanced analytics
RATIONALE: Advanced analytics generally refers to predictive and prescriptive analytics.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:30 PM

48. _____ analytics is the analysis of online activity, such as visits to websites or social media.

ANSWER: Web
RATIONALE: Web analytics is the analysis of online activity, such as visits to websites or social media.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: BUSINESS ANALYTICS IN PRACTICE
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:33 PM

49. One of the 4 Vs of big data that refers to uncertainty due to data inconsistency and incompleteness, ambiguities, latency, deception, and model approximations is _____.

ANSWER: veracity
RATIONALE: Veracity refers to uncertainty due to data inconsistency and incompleteness, ambiguities, latency, deception, and model approximations.
POINTS: 1

CH 01 - Introduction

DIFFICULTY: Moderate
REFERENCES: BIG DATA
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:34 PM

50. Data that are too large or too complex to be handled by standard data-processing techniques and typical desktop software are called _____.

ANSWER: big data
RATIONALE: When data is too large or too complex to be handled by standard data-processing techniques and typical desktop software, this is referred to as big data.
POINTS: 1
DIFFICULTY: Easy
REFERENCES: BIG DATA
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/7/2018 1:36 PM

51. Veracity has to do with how much _____ is in the data.

ANSWER: uncertainty
RATIONALE: Veracity has to do with how much uncertainty is in the data.
POINTS: 1
DIFFICULTY: Moderate
REFERENCES: BIG DATA
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 1/23/2018 10:15 AM
DATE MODIFIED: 3/28/2018 2:30 PM










52. What are the four V's of big data?

ANSWER: Volume, Velocity, Variety, Veracity
RATIONALE: The four V's of big data are volume, velocity, variety, and veracity.
POINTS: 1
DIFFICULTY: Moderate

CH 01 - Introduction

REFERENCES: BIG DATA
QUESTION TYPE: Subjective Short Answer
HAS VARIABLES: False
NATIONAL STANDARDS: United States - BUSPROG: Analytic
United States - DISC: - Descriptive Statistics
KEYWORDS: Bloom's: Remember
DATE CREATED: 3/7/2018 1:38 PM
DATE MODIFIED: 3/7/2018 1:39 PM

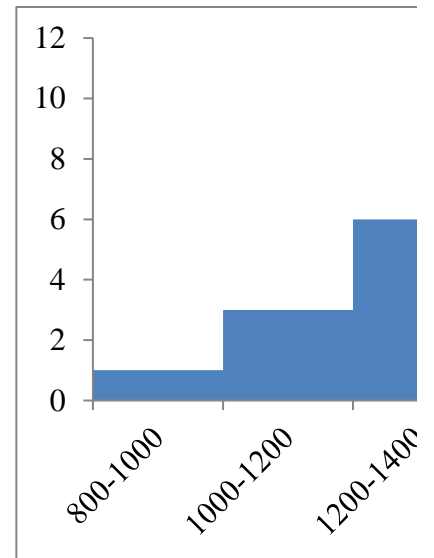
Carrier	Previous Year On-time Percentage	Current Year On- time Percentage	Change in On- time Percentage
Blue Box Shipping	88.4%	94.8%	6.4%
Cheetah LLC	89.3%	91.8%	2.5%
Smith Logistics	84.3%	88.7%	4.4%
Granite State Carriers	81.8%	87.6%	5.8%
Super Freight	92.1%	86.8%	-5.3%
Minuteman Company	91.0%	84.2%	-6.8%
Jones Brothers	68.9%	82.8%	13.9%
Honsin Limited	74.2%	80.1%	5.9%
Rapid Response	78.8%	70.9%	-7.9%

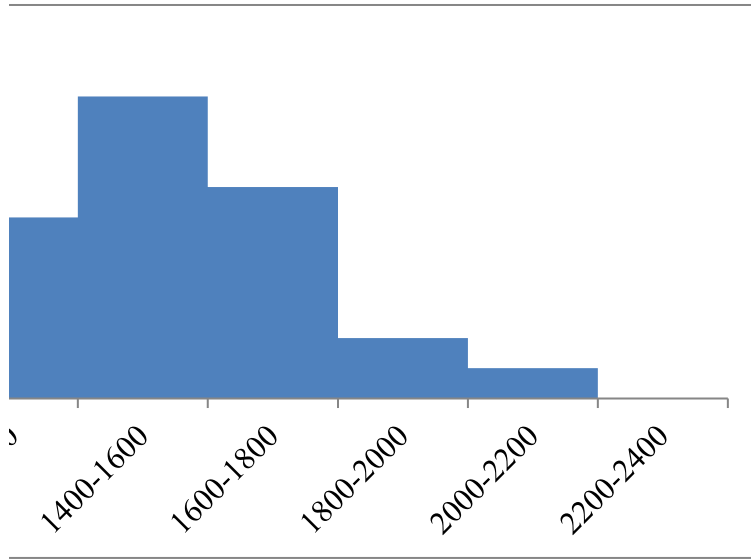
Carrier	Previous Year On-time Percentage	Current Year On-time Percentage	Change in On- time Percentage
Blue Box Shipping	88.4%	94.8%	 6.4%
Cheetah LLC	89.3%	91.8%	 2.5%
Smith Logistics	84.3%	88.7%	 4.4%
Granite State Carriers	81.8%	87.6%	 5.8%
Super Freight	92.1%	86.8%	 -5.3%
Minuteman Company	91.0%	84.2%	 -6.8%
Jones Brothers	68.9%	82.8%	 13.9%
Honsin Limited	74.2%	80.1%	 5.9%
Rapid Response	78.8%	70.9%	 -7.9%

SAT Scores

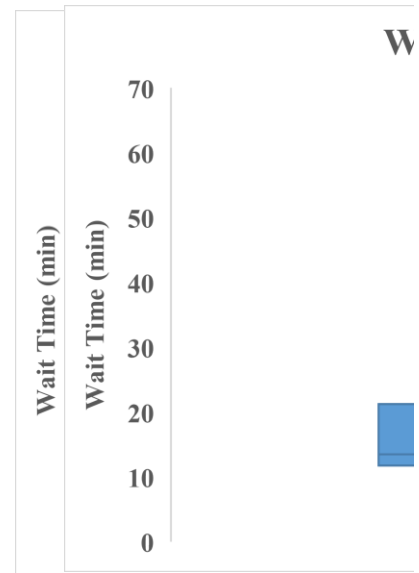
1665		
1275	1000	1
1650	1200	3
1590	1400	6
1475	1600	10
1490	1800	7
1525	2000	2
2135	2200	1
1560	2400	0
1880		
1680		
1560		
1355		
1280		
1150		
1420		
1440		
940		
1645		
1060		
1485		
1755		
1260		
1390		
1780		
1585		
1990		
1375		
1730		
1175		

<u>Class</u>	<u>Frequency</u>
800-1000	1
1000-1200	3
1200-1400	6
1400-1600	10
1600-1800	7
1800-2000	2
2000-2200	1
2200-2400	0





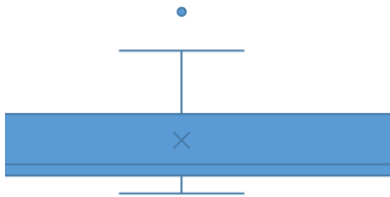
Without Wait-Tracking		With Wait-Tracking System	
Wait Time	z -Score	Wait Time	z -Score
24	-0.31	31	1.49
67	2.28	11	-0.67
17	-0.73	14	-0.34
20	-0.55	18	0.09
31	0.11	12	-0.56
44	0.90	37	2.13
12	-1.03	9	-0.88
23	-0.37	13	-0.45
16	-0.79	12	-0.56
37	0.48	15	-0.24
Mean:	29.1	Mean:	17.2
Median:	23.5	Median:	13.5
Standard Dev.:	16.60	Standard Dev.:	9.28
Variance:	275.66	Variance:	86.18



Without T_i With Tracking

24	31
67	11
17	14
20	18
31	12

With Wait-Tracking



Without Wait-Tracking		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
Mean:	29.1	Mean:	17.2
Median:	23.5	Median:	13.5
Standard Dev.:	16.60	Standard Dev.:	9.28
Variance:	275.66	Variance:	86.18

Without T1 With Tracking

24	31
67	11
17	14
20	18
31	12
44	37
12	9
23	13
16	12
37	15

x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
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

$$\bar{x} = 8$$

$$\bar{y} = 46$$

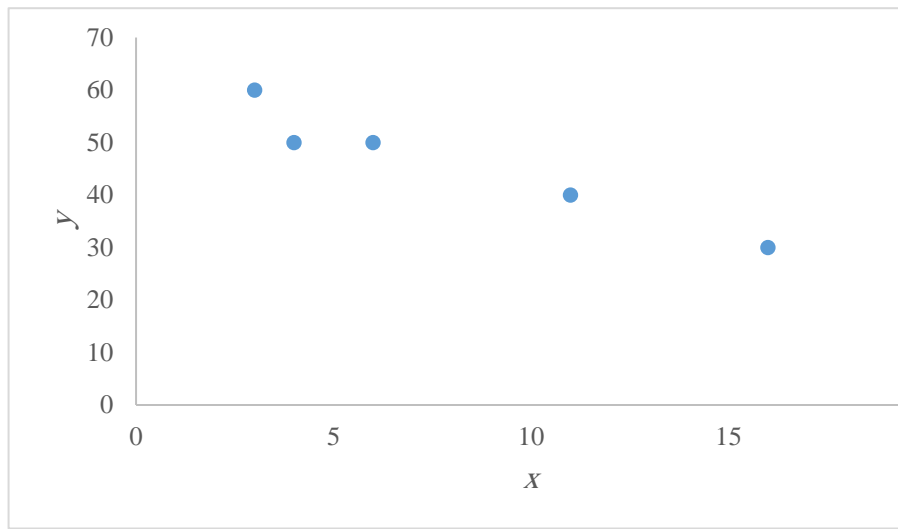
$$s_{xy} = -60$$

$$s_x = 5.43139$$

$$s_y = 11.40175$$

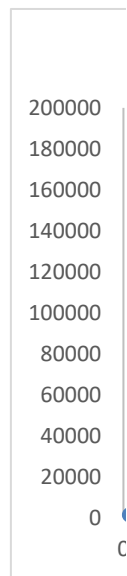
Correlation: -0.96888

CORREL: -0.96888



—
20
—

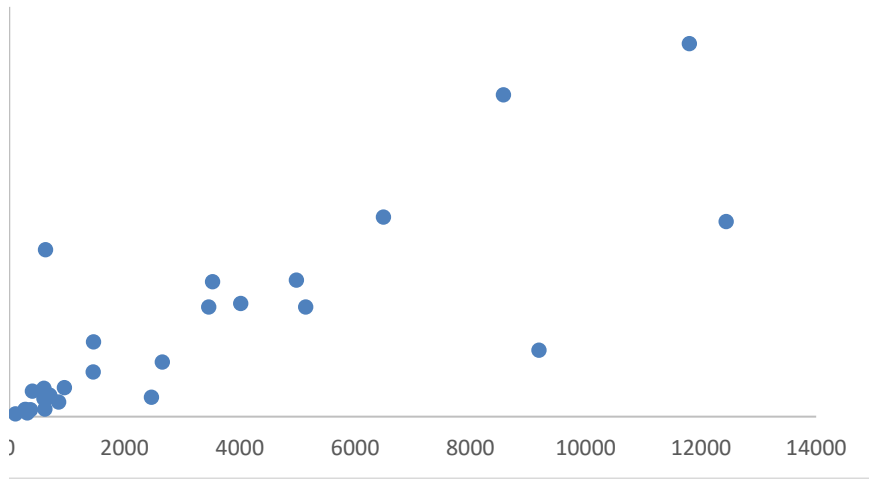
Company	Profits (\$ millions)	Market Capitalization (\$ millions)
Alliant Techsystems	313.2	1891.9
Amazon.com	631	81458.6
AmerisourceBergen	706.6	10087.6
Avis Budget Group	-29	1175.8
Boeing	4,018.00	55188.8
Cardinal Health	959	14115.2
Cisco Systems	6,490.00	97376.2
Coca-Cola	8,572.00	157130.5
ConocoPhillips	12,436.00	95251.9
Costco Wholesale	1,462.00	36461.2
CVS Caremark	3,461.00	53575.7
Delta Air Lines	854	7082.1
Fidelity National Financial	369.5	3461.4
FMC Technologies	399.8	12520.3
Foot Locker	278	3547.6
General Motors	9,190.00	32382.4
Harley-Davidson	599.1	8925.3
HCA Holdings	2,465.00	9550.2
Kraft Foods	3,527.00	65917.4
Kroger	602	13819.5
Lockheed Martin	2,655.00	26651.1
Medco Health Solutions	1,455.70	21865.9
Owens Corning	276	3417.8
Pitney Bowes	617.5	3681.2
Procter & Gamble	11,797.00	182109.9
Ralph Lauren	567.6	12522.8
Rockwell Automation	697.8	10514.8
Rockwell Collins	634	8560.5
United Stationers	109	1381.6
United Technologies	4,979.00	66606.5
UnitedHealth Group	5,142.00	53469.4



Covariance: 131804978.638

Correlation: 0.823

Market Capitalization (\$ millions)



Metro Area	Jobless Rate (%)	Delinquent Loans (%)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
Atlanta	7.1	7.02	0.285185	0.69	0.20
Boston	5.2	5.31	-1.61481	-1.02	1.65
Charlotte	7.8	5.38	0.985185	-0.95	-0.94
Chicago	7.8	5.40	0.985185	-0.93	-0.92
Dallas	5.8	5.00	-1.01481	-1.33	1.35
Denver	5.8	4.07	-1.01481	-2.26	2.29
Detroit	9.3	6.53	2.485185	0.20	0.50
Houston	5.7	5.57	-1.11481	-0.76	0.85
Jacksonville	7.3	6.99	0.485185	0.66	0.32
Las Vegas	7.6	11.12	0.785185	4.79	3.76
Los Angeles	8.2	7.56	1.385185	1.23	1.70
Miami	7.1	12.11	0.285185	5.78	1.65
Minneapolis	6.3	4.39	-0.51481	-1.94	1.00
Nashville	6.6	4.78	-0.21481	-1.55	0.33
New York	6.2	5.78	-0.61481	-0.55	0.34
Orange County	6.3	6.08	-0.51481	-0.25	0.13
Orlando	7.0	10.05	0.185185	3.72	0.69
Philadelphia	6.2	4.75	-0.61481	-1.58	0.97
Phoenix	5.5	7.22	-1.31481	0.89	-1.17
Portland	6.5	3.79	-0.31481	-2.54	0.80
Raleigh	6.0	3.62	-0.81481	-2.71	2.21
Sacramento	8.3	9.24	1.485185	2.91	4.32
St. Louis	7.5	4.40	0.685185	-1.93	-1.32
San Diego	7.1	6.91	0.285185	0.58	0.17
San Francisco	6.8	5.57	-0.01481	-0.76	0.01
Seattle	5.5	3.87	-1.31481	-2.46	3.24
Tampa	7.5	8.42	0.685185	2.09	1.431529492
					<u>25.5517037</u>
Mean	6.814814815	6.330740741			
Std Dev	0.995645791	2.236578647			

$s_{xy} = 0.982757835$
 $s_x = 0.995645791$
 $s_y = 2.236578647$

$r_{xy} = 0.441323931$

CORREL: 0.441323931

$$(x_i - \bar{x})(y_i - \bar{y})^2$$

0.081331	0.475078
2.607627	1.041912
0.97059	0.903908
0.97059	0.866278
1.029849	1.770871
1.029849	5.110949
6.176145	0.039704
1.242812	0.578726
0.235405	0.434623
0.616516	22.937
1.918738	1.511078
0.081331	33.39984
0.265034	3.766475
0.046145	2.404797
0.377997	0.303315
0.265034	0.062871
0.034294	13.83289
0.377997	2.498741
1.728738	0.790782
0.099108	6.455364
0.663923	7.348115
2.205775	8.463789
0.469479	3.72776
0.081331	0.335541
0.000219	0.578726
1.728738	6.055245
0.469479	4.365004
25.77407	130.0594

