Descriptive Business Analytics Using Excel: Exercise 1

After attending Business Analytics evening classes, Carlos Gonzales got an attractive job offer as a business data scientist from a large bank located in New York City (NYC). After accepting the job offer, Carlos started to commute to NYC by bus. Unfortunately, the bus often arrives later than scheduled. The tardiness of arrivals varies substantially. Carlos believes that most influential factors for the amount of delays are the bus driver and the day of week. To examine the issue, each time Carlos commutes to NYC he records the lateness of arrival (in minutes), along with the first name of the bus driver, day of the week, and the time of scheduled bus departure.

Open the **BUS-to-NYC.txt** file located at the Business Analytics website: www.small-big-data.com.

1. Based on the data, how many observations for the departure delays did Carlos collect?
A. 30
B. 36
C. 38
D. 39
E. None of the above

2. Based on the data, what is the mean for the departure delays?
A. 7.5
B. 8.5
C. 13.3
D. 14.7
E. None of the above

3. Based on the data, what is the median for the departure delays?
A. 7.5
B. 8.5
C. 13.3
D. 14.7
E. None of the above

4. Based on the data, what is the standard deviation for the departure delays?
A. 7.5
B. 8.5
C. 13.3
D. 14.7
E. None of the above

5. Compute the range (largest value less the smallest value) for all the departure delays. Based on the data, what is the range?
A) 61
B) 62
C) 63
D) 64
E) None of the above

6. Compute the skewness for all the departure delays. Based on the data, what is the skewness?
A) 1.0
B) 1.5
C) 2.0
D) 2.5
E) None of the above

7. Based on the data, what is the shape of the distribution of data values for the departure delays?
A) Negative, or left-skewed
B) Symmetric, or zero skewness
C) Positive, or left-skewed
D) Positive, or right-skewed
E) None of the above

Solution

The objective of this exercise is to practice how to conduct a basic data analysis using an application software such as MS Excel.

MS Excel formulas to be used include:
=COUNT()
=AVERAGE()
=MEDIAN()
=STDEV.S()
=MIN()
=MAX()
=SKEW()

First, you need to import the **BUS-to-NYC.txt** file located at the Business Analytics website: www.small-big-data.com (depicted in Figure 1) into Excel. The file is comma delimited. The procedure of importing text files into Excel is described at:
http://www.small-big-data.com/baexcelworkshop.htm

www.small-big-data.com/Bus-to-NYC.txt

```
TripID,Late,Bus Driver,Weekday,Time
001,12,John,Wednesday,8:45 AM
003,1,Paulo,Tuesday,9:25 AM
005,3,Mark,Tuesday,9:25 AM
007,45,Raymond,Thursday,9:25 AM
009,61,Mark,Friday,9:25 AM
011,2,Monica,Tuesday,8:45 AM
013,5,John,Tuesday,8:45 AM
121,10,John,Wednesday,7:15 AM
127,-2,Mark,Monday,9:25 AM
016,14,John,Wednesday,7:15 AM
098,0,Raymond,Friday,9:25 AM
103,-1,Paulo,Friday,9:25 AM
108,3,Mark,Friday,8:00 AM
113,45,John,Monday,9:25 AM
118,7,John,Monday,8:45 AM
757,17,Paulo,Wednesday,9:25 AM
477,8,John,Tuesday,9:25 AM
133,19,Raymond,Tuesday,8:00AM
138,22,John,Tuesday,7:15 AM
128,0,John,Monday,7:15 AM
```

Figure 1

After the text file is imported into Excel as depicted in Figure 2, basic data analysis can be conducted.

	A	B	C	D	E	F	G
1	TripID	Late	Bus Driver	Weekday	Time		
2	001	12	John	Wednesday	8:45 AM		
3	003	1	Paulo	Tuesday	9:25 AM		
4	005	3	Mark	Tuesday	9:25 AM		
5	007	45	Raymond	Thursday	9:25 AM		
6	009	61	Mark	Friday	9:25 AM		
7	011	2	Monica	Tuesday	8:45 AM		
8	013	5	John	Tuesday	8:45 AM		
9	121	10	John	Wednesday	7:15 AM		
10	127	-2	Mark	Monday	9:25 AM		
11	016	14	John	Wednesday	7:15 AM		
12	098	0	Raymond	Friday	9:25 AM		
13	103	-1	Paulo	Friday	9:25 AM		
14	108	3	Mark	Friday	8:00 AM		
15	113	45	John	Monday	9:25 AM		
16	118	7	John	Monday	8:45 AM		

Figure 2

1. The COUNT function was utilized to count the number of observations. In order to get the right answer, it is necessary to select only the cells that contain data used in following calculations. Thus, in this case we chose the whole “Late” column, which is the range of cells B to B. In the picture below, it is shown that we selected those cells for the COUNTA function. The right answer is letter C, which is 38 observations.

   Formula: **=COUNT(B:B)**
   Result: **38**

	A	B	C	D	E
1	TripID	Late	Bus Driver	Weekday	Time
2	001	12	John	Wednesday	8:45 AM
3	003	1	Paulo	Tuesday	9:25 AM
4	005	3	Mark	Tuesday	9:25 AM
5	007	45	Raymond	Thursday	9:25 AM
6	009	61	Mark	Friday	9:25 AM
7	011	2	Monica	Tuesday	8:45 AM
8	013	5	John	Tuesday	8:45 AM
9	121	10	John	Wednesday	7:15 AM
10	127	-2	Mark	Monday	9:25 AM
11	016	14	John	Wednesday	7:15 AM
12	098	0	Raymond	Friday	9:25 AM
13	103	-1	Paulo	Friday	9:25 AM
14	108	3	Mark	Friday	8:00 AM
15	113	45	John	Monday	9:25 AM
17	118	7	John	Monday	8:45 AM
19	757	17	Paulo	Wednesday	9:25 AM
20	477	8	John	Tuesday	9:25 AM
21	133	19	Raymond	Tuesday	8:00AM
22	138	22	John	Tuesday	7:15 AM
23	128	0	John	Monday	7:15 AM
24	019	2	John	Monday	8:45 AM
25	021	7	Paulo	Thursday	9:25 AM
26	023	-1	Paulo	Monday	8:45 AM
27	025	32	Mark	Thursday	8:00AM

Figure 3

2. The function used to get the Mean departure delays is AVERAGE. For this question, we again need to select the values in the departure delay columns first, which is the range of cells B to B (as shown in Figure 4 below) and use that data for our calculation.
Formula: **=AVERAGE(B:B)**
Result: **12.32**

The answer is E, none of the above.

	A	B	C	D	E
1	TripID	Late	Bus Driver	Weekday	Time
2	001	12	John	Wednesday	8:45 AM
3	003	1	Paulo	Tuesday	9:25 AM
4	005	3	Mark	Tuesday	9:25 AM
5	007	45	Raymond	Thursday	9:25 AM
6	009	61	Mark	Friday	9:25 AM
7	011	2	Monica	Tuesday	8:45 AM
8	013	5	John	Tuesday	8:45 AM
9	121	10	John	Wednesday	7:15 AM
10	127	-2	Mark	Monday	9:25 AM
11	016	14	John	Wednesday	7:15 AM
12	098	0	Raymond	Friday	9:25 AM
13	103	-1	Paulo	Friday	9:25 AM
14	108	3	Mark	Friday	8:00 AM
15	113	45	John	Monday	9:25 AM
17	118	7	John	Monday	8:45 AM
19	757	17	Paulo	Wednesday	9:25 AM
20	477	8	John	Tuesday	9:25 AM
21	133	19	Raymond	Tuesday	8:00AM
22	138	22	John	Tuesday	7:15 AM
23	128	0	John	Monday	7:15 AM
24	019	2	John	Monday	8:45 AM
25	021	7	Paulo	Thursday	9:25 AM
26	023	-1	Paulo	Monday	8:45 AM
27	025	32	Mark	Thursday	8:00AM

Figure 4

3. To obtain the Median departure delays, the "Late" column needs to be selected again and the MEDIAN function is utilized. The answer is A, which is 7.5.
   Formula: **=MEDIAN(B:B)**
   Result: **7.5**

4. This question also deals with the departure delays, therefore, we are again using the "Late" column. The standard deviation function could be either STDEV.P or STDEV.S. However, it is important to read the question well to see if we are being asked the population or the sample standard deviation. In this case, we are using STDEV.S. since we have a sample.

Formula: **=STDEV.S(B:B)**
Result: **14.7**

The answer is D, which is 14.7.

5. To calculate the range, we need to subtract the lowest value from the highest value. In order to do this, we start with the MAX function for the "Late" column (B:B). Then, we use the minus sign to indicate subtraction and utilize the MIN function for the same set of values (B:B).

   Formula: **=MAX(B:B)-MIN(B:B)**
   Result: **63**

   The answer is C, which is 63.

6. The skewness of the departure delays is obtained by using the function SKEW and selecting the range of values in that column.

   Formula: **=SKEW(B:B)**
   Result: **1.63**

   After we round down to 1.5, the closest answer B.

7. We built on the previous question.
   In essence, if the SKEW formula yields a positive result the data is positive or right skewed.
   If the SKEW formula yields a negative result the data is negative or left skewed.

   In our case, the result of the SKEW formula is 1.63 so the data is positive or right skewed.

   Therefore, the answer is D.

Answers

Exercise
1. C
2. E
3. A
4. D
5. C
6. B
7. D