Mrinal Chaudhari- STAT 614

Assignment 1:

1)(Show Work) For the following data: 2, 5, 6, 7, 12, 3, 11, 9 ,1

1a.Determine the Median

b.Determine Q1 and Q3

c.Find the IQR

d.What is the largest value that would not be considered an outlier? How about the smallest?

Solution:

a: Data- 1,2,3,5,6,7,9,11,12

There are total 9 values which are arrange in ascending order in dataset hence median is the middle value of the data which is 6.

Median- 6

b:

Data- 1,2,3,5,6,7,9,11,12

Q1: left side data- 1,2,3,5

Q1: median of (2,3)

Q1: 2+3/2

Q1:2.5

Q3: Right side data- 7,9,11,12

Q3: median of (7,9,11,12)

Mrinal Chaudhari- STAT 614

Q3:9+11/2

Q3:10

C:

IQR: Q3-Q1

IQR: 10-2.5

IQR: 7.5

D:

Outlier:>Q3+(1.5*IQR) and <Q1-(1.5*IQR)

LF<Q1-1.5*IQR

UF>Q3+1.5*IQR

LF<2.5-1.5*7.5=-8.75

UF>10+1.5*7.5=21.25

Since both points -8.75 and 21.25 are within the interval (1,12) ,none of the points are outliers.

Largest value that would not be considered an outlier is 12

Smallest value that would not be considered an outlier is 1

2) For the data in #1

Mrinal Chaudhari- STAT 614

a.Determine the mean

b.Determine the variance

c.Determine the standard deviation

Solution:

a.

Data- 1,2,3,5,6,7,9,11,12

Obs.	X	$\bar{x}$	$(X- \bar{x})$	$(X- \bar{x})^2$
1	1	6.2	-5.22	27.24
2	2	6.2	-4.22	17.8
3	3	6.2	-3.22	10.36
4	5	6.2	-1.22	1.48
5	6	6.2	0.22	0.04
6	7	6.2	0.78	0.6
7	9	6.2	2.78	7.72

8	11	6.2	4.78	22.84
9	12	6.2	5.78	33.4
Sum	56			121.48
N=9	∑X			$\sum(X- \bar{x})^2$

a : Mean- $\bar{x} = (\sum X)/N = 56/9 = 6.22$

b: Variance- $S^2 = (\sum(X - \bar{x})^2)/(N-1) = 121.48/9\text{-}1 = 15.185$

c: Standard Deviation= Square root of the variance =S

S= $\sqrt{15.25} = 3.89$

3)A study was run on the time difference between how long a surgery should take and the actual time it took for a surgery based on day. The data is in the excel file called Medical Data. It is on the sheet called Surgery Time Study. For this data:

a.Create a histogram of the data. Comment on the shape and if you think that there might be outliers

b.Create a numerical summary. Compare the mean and median. Does the comparison between the two verify your summary of the shape of the distribution? Why or Why not

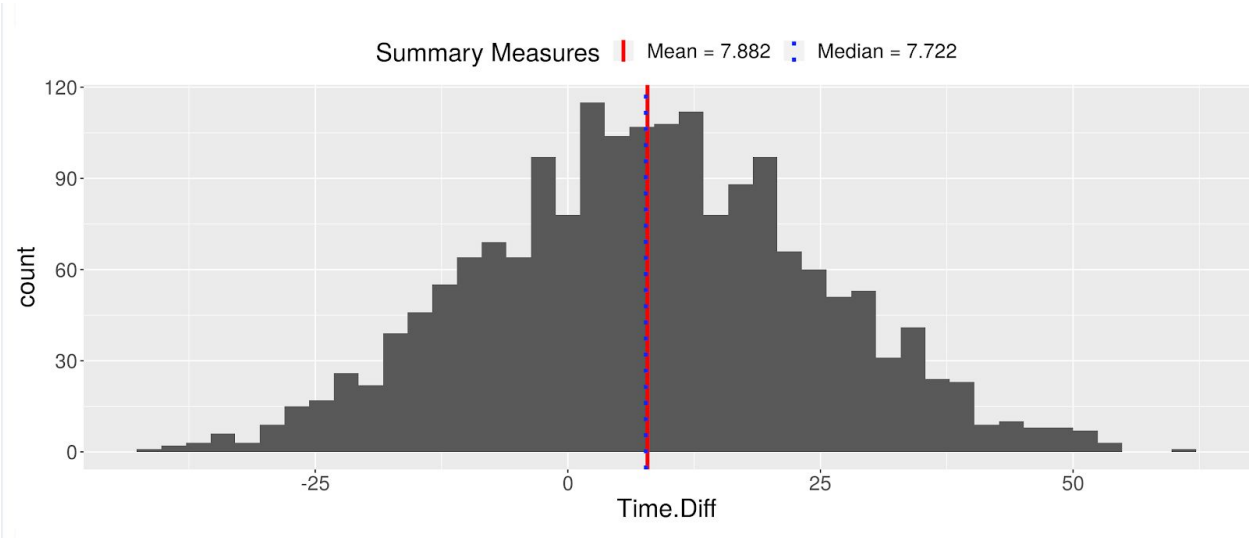
c.Create a boxplot of the data. Based on this plot, does the shape you found in #1 make sense? Explain

d.What numerical summary measure of center and spread do you think should be used in this case? Explain

Mrinal Chaudhari- STAT 614

Solution:

a:



Shape - Symmetrical shape
Outliers are 60.33 and -42.10

b:
Median= 7.722
Mean= 7.88
Min = -42.10
Max = 60.33
Standard deviation = 16.70
Variance = 278.96
Mode = NA
Q1= -3.308549448
Q3= 19.26613289
IQR= Q3-Q1=22.57468234

Outlier:>Q3+(1.5*IQR) and <Q1-(1.5*IQR)

UF = Q3+1.5*IQR= 19.26+1.5*22.57=53.1219
LF =Q1-1.5*IQR=-3.3085-1.5*22.57468= -37.17052

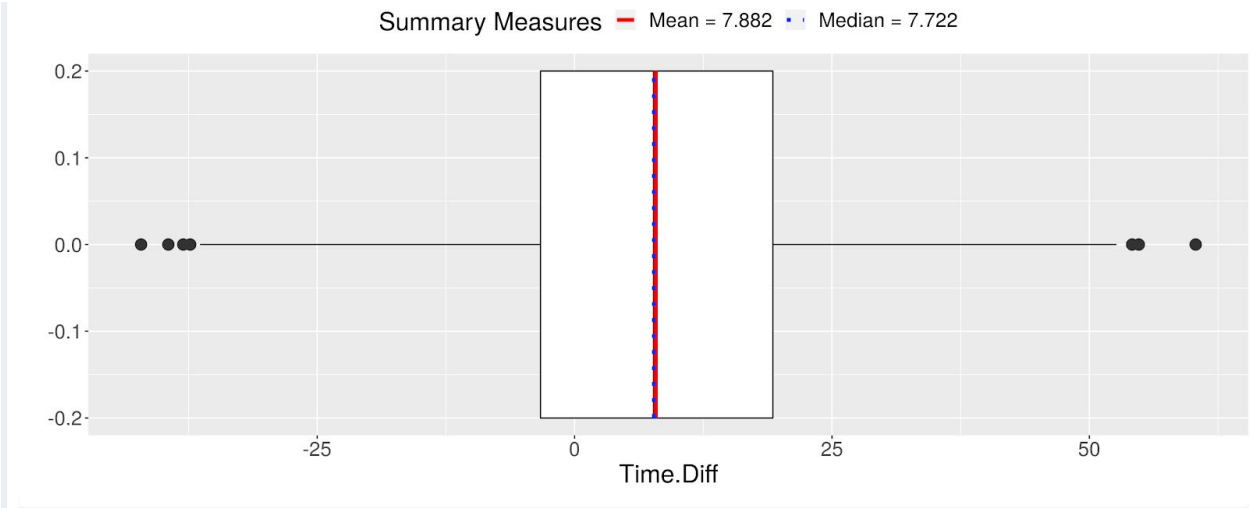
Max value 60.33 falls outside of interval hence it is considered an outlier

Extreme UF = Q3+3*IQR= 19.26+3*22.57= 86.97
Extreme LF=Q1-3*IQR=-3.3085-3*22.57468= -71.03254

Min value -42.10 falls outside of the interval hence it is considered as an outlier.

From above calculation, We have seen both values of mean and median are very close to each other i.e Mean is 7.88 and median is 7.720 therefore shape of distribution is symmetrical.
And from histogram we can observe the symmetry between two parts.

c:



From the above graph the mean and median is quite similar and identical to each other and hence the shape is symmetrical. We conclude that histogram and box plot graphs have similar symmetrical shape.

d:

Since the dataset contains outliers hence we will use median as a measure of central tendency and IQR as measure of spread.

4)A study was run to assess body temperature as taken with an ear thermometer. 10 patients had their temperature taken 3 times in each ear by 3 different nurses. The expectation is that the body temperature is between 35 and 40 degrees C. The data is in the excel file called Medical Data. It is on the sheet called Ear Thermometer Data. Do an analysis of the data. To get full credit you should: provide both a numerical and graphical summary, including checking for outliers. Within the numerical summary, you should provide a measure of center and spread and justify which you chose. Then give the real-world answer.

Solution:

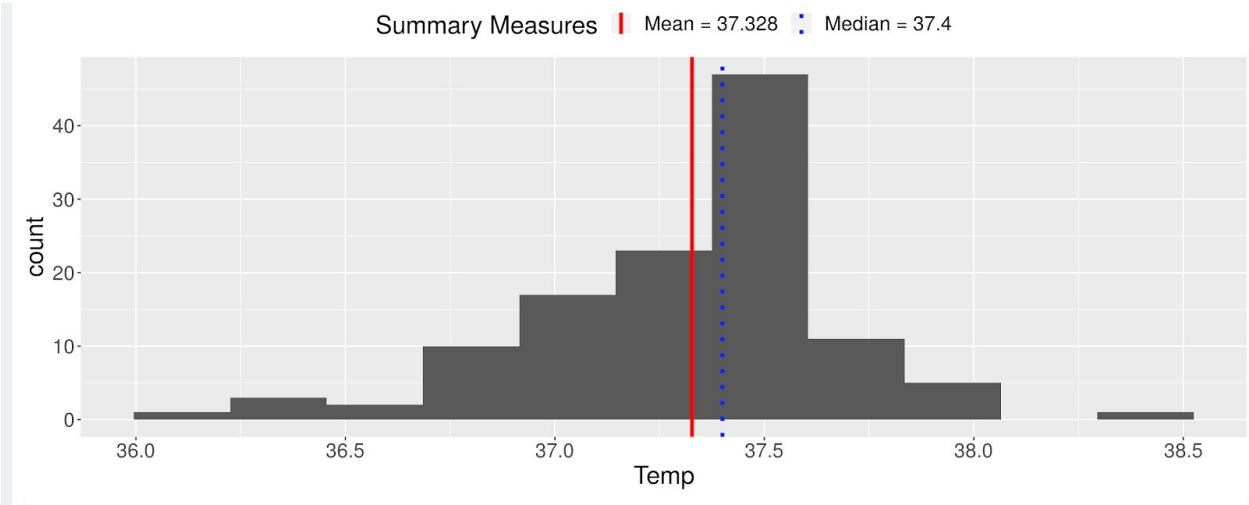
Mrinal Chaudhari- STAT 614

Mean=37.32
Min=36.1
Max=38.4
Median=37.4
Mode=37.5
SD=0.37
Variance=0.136
Q1=37.1
Q3=37.6
IQR=Q3-Q1=0.5

Outlier:>Q3+(1.5*IQR) and <Q1-(1.5*IQR)

LF= Q1-1.5*IQR=37.1-1.5*0.5=36.35
UF=Q3+1.5*IQR=37.6+1.5*0.5=38.35
No outliers

Since both points (36.35, 38.35) are within the interval (36.1,38.4) , none of the points are outliers.
Measure of center is mean and spread is standard deviation
Mean=37.32
Standard Deviation=0.37



The mean value is roughly 37 which is between 35 to 40 degree celsius.
So the expectation of body temperature should be between 35 and 40 degree celsius, none of the data is exceeded far from that range and also we have noticed a very negligible amount of range i.e 0.5 therefore data is not spread.