ASSIGNMENT-2

Poorvi Raut- 20009560

Chapter 2, Page 40, Problem 5

Make up a data set consisting of eight scores on an exam in which one of the scores is an outlier.

a.) Find the mean score and the median score, with and without the outlier.

Student ID	Marks
01	40
02	73
03	70
04	65
05	75
06	72
07	98
08	75

Outliers: 40 and 98

With Outliers:

Mean: (40+73+70+65+75+72+98+75)/8 = **71**

Median: Arrange in ascending order and then calculate the value

40,65,70,72,73,75,75,98= 4th+5th/2= (72+73)/2= **72.5**

Without Outliers:

Mean: (65+70+72+73+75+75)/6 =**71.66**

Median: 65,70,72,73,75,75= 3rd+4th/2= (72+73)/2= **72.5**

b.) State which measure, the mean or the median, the presence of the outlier affects more, and why.

Answer:

In the given dataset the mean score and median score values are calculated by two methods i.e with and without outliers present.

1. In With outliers method, mean is calculated by summation of all the marks divided by the total number of student id where n=8. In   Without outliers method since we are excluding the extreme values i.e. 40 and 98 and the average of marks has number of student Id's as n=6. Therefore, there is a difference between the mean score values.

2. For median calculation, we can see that the median value remains constant in with and without outlier's methods because the median is the middle value in the dataset and the presence or absence of outliers do not affect the median score value.

c.) Verify that the outlier is indeed an outlier, using the IQR method.

The given marks are: 40,73,70,65,75,72,98,75.

Arranging in ascending order: 40,65,70,72,73,75,75,98

Lower Quartile: Q1:(65+70)/2= **67.5**

Median: Q2: **72.5**

Upper Quartile: Q3: (75+75)/2= **75**

IQR= Q3-Q1= 75-67.5= **7.5**

Q1-(1.5*IQR) =67.5 - (1.5*7.5) =**56.25** which is in lower fence

Q3+(1.5*IQR) = 75 + (1.5*7.5) = **86.25** which is in upper fence

Conclusion: We can observe that the lower value 40 is lower than 56.25 which is the lower fence and value 98 is greater than the upper fence value that is 86.25 hence we can conclude that 40 & 98 are outlier proved by IQR METHOD.