***Assignment no 2***

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.

**Ans:** *Dataset:* {29, 45, 49, 56, 57, 60, 62, 65}

**Mean** = {Sum of all Numbers in a dataset}/ Total numbers

**Median (for odd nos.)** = Middle no. of the dataset after arranging the nos. in the ascending order

**Median (for even nos.)** = Average of the two middlemost nos. after arranging the nos. in the ascending order.

|  |  |  |
| --- | --- | --- |
| Method | With Outlier | Without Outlier |
| 1. MEAN | {29+45+49+56+57+60+62+65}/8 = 52.875 | {45+49+56+57+60+62+65}/7 = 56.285 |
| 1. MEDIAN | {56+57}/2 = 56.5 | 57 |

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

**Ans:** From the above observation,the presence of the outlier affects the mean more than the median because the mean is the average of all the numbers in a dataset and if a single number is removed, it affects the average. Whereas, median only focuses on the middlemost number and wouldn’t affect much if a number is removed. Hence, the presence of outlier affects the mean more after the outlier is removed.

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

**Ans:** Consider a dataset = {2, 3, 4, 4, 5, 6, 9, 15}

Total numbers in dataset are 8

Q2 = (4+5)/2 = 4.5

Two halves are (2, 3, 4, 4) and (5, 6, 9, 15)

Q1 = (3+4)/2 = 3.5

Q3 = (6+9)/2 = 7.5

**To find outliers:** 1.) IQR = Q3-Q1 = 7.5 – 3.5 = 4

2.) 1.5\* IQR = 6.0

3.) Outliers will be any points below Q1 - 1.5 \*IQR = 3.5 – 6.0 = -2.5 and above

Q3 + 1.5\*IQR = 7.5 + 6 .0 = 13.5

4.) Hence outlier in the dataset is **15**.