# Five number summary and boxplot

Five number summaries

1 Minimum

2 First Quartile (Q1)

3 Median

4 Third Quartile (Q3)

5 Maximum

With help of above 5 factors, we will remove outliers.

Removing the outliers

Data sets - {1,2,2,2,3,3,4,5,5,5,6,6,6,6,6,7,8,8,9,27}

Outlier??

{lower fence < - -> higher fence}

In above data set if there is value -50 than, - 50 is lower fence. And 27 is outlier.

In order define lower and higher fence simple formula =

Interquartile Range (IQR) = Q3-Q1

Q3 = (75%) Percentile

{75 percentile means = 75/100x(19+1) = 15 index = Q3=7}

Q1 = (25%) Percentile

{25 percentile means = 25/100x (19+1) = 5 index = Q1 =3}

Lower Fence =Q1 -1.5(IQR)

Upper fence = Q3 + 1.5 (IQR)

Inter quartile Range (IQR)= Q3-Q1 = 7-3 =4

Lower Fence =Q1 -1.5(IQR) = 3-1.5(4) = 3-6 = -3

Upper fence = Q3 + 1.5 (IQR) = 7+1.5 (4) = 7+6 = 13

Therefore, [Lower Fence < - -> Higher Fence] [ -3 < - - > 13 ]

So, we must remove – observation 27 from the data set because 27 > 13.

Remaining Data sets - {1,2,2,2,3,3,4,5,5,5,6,6,6,6,6,7,8,8,9,~~27~~}

Now from above data set

Minimum = 1

Q1 = 3

Median = 5 5 number Summary with this one can draw box plot.

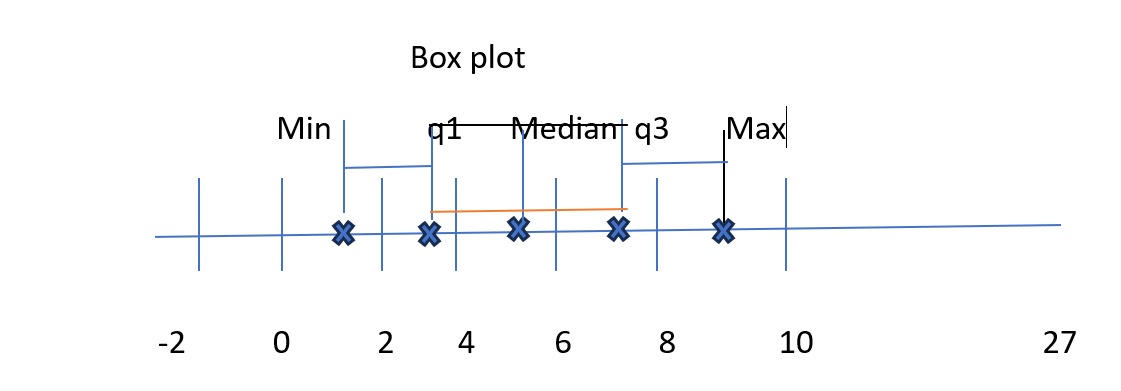
Q3 = 7

Maximum = 9

Box plot

Min q1 Median q3 Max

-2 0 2 4 6 8 10 27



Use for data visualization.

Understand variance – 6.

2/n-1

n-1 = Bessel’s correction or degree of freedom

why Sample Variance is divided by n-1, Please see following link - https://www.youtube.com/watch?v=vGsRwB3TsiE.

What is application use of boxplot?

Box plot can be used to determine outliers, box plot give you visualization way basically to see where the outlier is present.