Statistics Assignment 2

1. How can we figure out what the Interquartile range is?
Answer:
Interquartile Range(IQR) is the difference between Third Quartile(Q3) and First Quartile(Q1) i.e. **IQR = Q3 – Q1**
Third Quartile(Q3) is 75 percentile i.e. **Q3 = (75/100)*(n+1)**
First Quartile(Q1) is 25 percentile i.e. **Q1 = (25/100)*(n+1)**
Here n is the total number of observation in our dataset.

2. What exactly is the value of the 5-number theory?
Answer:
The 5-number theory is used to find outliers and it is used for plotting Box Plot. The 5-number theory is shown below:-
  I. Minimum – It is the lowest value of the given dataset.
 II. First Quartile(Q1) – It is the 75 percentile value of the given dataset.
III. Median – Calculate the median of the given dataset.
 IV. Third Quartile(Q3) – It is the 25 percentile value of the given dataset.
  V. Maximum – It is the highest value of the given dataset.

3. What is the relationship between standard deviation and variance?
Answer:
Standard Deviation (σ) is the square root of the Variance (σ^2)
i.e. standard deviation = $\sqrt{\text{variance}}$

4. What does the difference between variance and standard deviation mean?
Answer:
Variance tells us about Dispersion of data that is how much our data is spread. More variance mean more spreading of data. Thus, more data around the centre region of dataset whereas Standard Deviation tells us that how much a given point is far from the mean of the data. So, standard deviation is an observation that we get measured through dispersion within a dataset.

5. When is it appropriate to refer to a skewed data distribution?
Answer:
Whenever we see the bell curve is towards on left side and having a tail on right side, it refers to Right skewed (Positive skewed) Data. In this case Mean > Median > Mode. Such type of distribution can also refer as Log Normal Distribution. Example of Right skewed distribution are Wealth Distribution, Number of long comments etc.

And whenever we see the bell curve is towards on right side and having a tail on left side, it refers to Left skewed (Negative skewed) Data. In this case Mean < Median < Mode. Example of Left skewed distribution is Life Span of Human Age.

Whenever, if any one of the above mention condition happens, we will use skewed data distribution.