Statistics Assignment 2

1. How can we figure out what the interquartile range is?
    - Interquartile range is difference between third quartile and first quartile.
    - First quartile is nothing but 25^{th} percentile value.
    - Third quartile is nothing but 75^{th} percentile value.
    - IQR= Q3-Q1

2. What exactly is the value of the 5-number theory?
    - We find 5-number theory concept in case of boxplot.
    - 5-number theory is:
        - Minimum value
        - First Quartile
        - Median
        - Third Quartile
        - Maximum Value
    - In boxplot, there are 5 horizontal line indicates 5-number theory
    - First quartile is nothing but 25^{th} percentile value.
    - Third quartile is nothing but 75^{th} percentile value.
    - Median means middle value or 50^{th} percentile value.
    - Minimum value is lowest value of that column and minimum value should be: min=Q1-1.5*IQR
    - Maximum value is highest value of that column and maximum value should be: Max=Q3+1.5*IQR

3. What is the relationship between standard deviation and variance?
    - Both standard deviation and variance explains the measure of spread of data.
    - Standard deviation is just square root of variance.
    - Population variance is summation of square of difference between random variable and mean of data, divided by number of datapoints.

4. What does the difference between variance and standard deviation mean?
    - Standard deviation is just square root of variance.
    - Population variance is summation of square of difference between random variable and mean of data, divided by number of datapoints.
    - Standard deviation talks about how far the data point is away from mean.
    - Variance talks about measure of spread of distribution and value of variance usually in squares

5. When is it appropriate to refer to a skewed data distribution?
    - Skewed distribution means abnormal distribution. There are two types of skewed data distribution: 1. Left skew distribution and 2. Right skewed distribution.
    - In left skewed distribution, left part of graph is slightly elongated and in right skewed distribution right part of graph is elongated.
    - Whenever there is skewed distribution, it is preferred to use median value instead of mean value.
    - As distribution is not normal mean do not perform well and there is large difference between mean and median and also not show middle part of data correctly.
    - Example of skewed distribution is annual of people; very few people have income less than 1 lakh, most of people have income between range 5lakh-15lakh and very few people have income more than 50 lakh.