**ASSIGNMENT 2**

**STATISTICS**

**1. What is the difference between Descriptive statistics and inferential statistics?**

**Descriptive statistics**- describes data according to its characteristics such as

Measure of central tendency

Measures of dispersion

Skewness

Kurtosis

**Inferential statistics**-  allows you to make predictions (“inferences”) from that data . using test such as

T-test

Chi-square test

Normal test

F-test and soon

2. **What is the difference between population and sample in inferential statistics?**

A population is the entire group that you want to draw conclusions about. A sample is the specific group that you will collect data from.

Most common characteristics used in descriptive statistics

Measure of central tendency- mean, median, mode

Measures of dispersion- mean deviation, standard deviation, variance

Skewness

Kurtosis

3. **How to calculate range and inter quartile range?**

Range is the difference between maximum and minimum value in the data set

Range = max - min

The interquartile range formula is the first quartile subtracted from the third quartile:

Inter quartile range = Q3 – Q1.

4. **How is statistical significance of an inference assessed ?**

**Statistical significance can be accessed using hypothesis testing:**  
– Stating a null hypothesis which is usually the opposite of what we wish to test (classifiers A and B perform equivalently, Treatment A is equal of treatment B)  
– Then, we choose a suitable statistical test and statistics used to reject the null hypothesis  
– Also, we choose a critical region for the statistics to lie in that is extreme enough for the null hypothesis to be rejected (p-value)  
– We calculate the observed test statistics from the data and check whether it lies in the critical region  
**Common tests:**  
– One sample Z test  
– Two-sample Z test  
– One sample t-test  
– paired t-test  
– Two sample pooled equal variances t-test  
– Two sample unpooled unequal variances t-test and unequal sample sizes  
– Chi-squared test for variances  
– Chi-squared test for goodness of fit  
– Anova   
– Regression F-test