STATISTICS WORKSHEET-1

1) a) True
2) b) Central Mean Theorem
3) a) Modeling event/time data
4) b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
5) a) Empirical
6) b) False
7) b) Hypothesis
8) a) 0
9) c) Outliers cannot conform to the regression relationship

10) Normal distribution represents the probability distribution of a continuous variable. It is symmetric about the mean. It represents data near the mean has maximum probability.

11) How do you handle missing data? What imputation techniques do you recommend?
There are two types of variables:
   (a) continuous and numerical- For this type of variable missing data can be replaced by mean/median of the data
   (b) categorical/discrete – For categorical variable/discrete variable missing data can be replaced by mode of the variable.

12) A/B testing is a method to compare two different machine learning models based on the statistical significance using hypothesis testing.

13) Yes, mean imputation of continuous variable for missing data is an acceptable practice, if the number of missing values is not too much.

14) Linear regression is the method of fitting a best fit line between a dependent and independent variables in a data set. The approach used to fit this line is called least square error method.

15) There are two types of statistics, namely inferential and descriptive.