STATISTICS WORKSHEET-1

1. a) true

2. a) The Central Limit Theorem

3. b) Modelling bounded count data

4. d) All of the mentioned

5. c) poisson

6. b) false

7. b) Hypothesis

8. a) 0

9. c) Outliners cannot conform to the regression relationship

10. Normal distribution is a kind of probability distribution that shows data near the mean are more frequent than data far from mean, i.e. a probability distribution symmetric about the mean.

11. In case of missing data we can use Deletion techniques like listwise deletion, pairwise deletion or dropping columns if more than 80% is missing value and the column is not important.

One can use imputation techniques such as: i) Imputation with constant value where any constant value or zero replaces the missing value. ii) imputation using statistics where the mean or most frequent value of a column replaces the missing value. iii) KNN imputation: this algorithm helps replace missing data by analyzing the closest neighbours by using Euclidean distance metric in observation to the non-missing values in the neighbours.

12. A/B testing is a user experience analyzing methodology consisting of random tests with two variants A and B. It uses application of statistical hypothesis testing or two-sample hypothesis testing.

13. No, mean imputation of missing data is not considered acceptable practice.

It is not an acceptable practice because i) it does not preserve the relationship between variables and ii) it leads to an underestimate of standard errors.

14. Linear Regression analysis is a method used for predicting the value of another dependent variable. It attempts to set a relationship between the explanatory and the dependent variable by fitting a linear equation.

15. There are three branches of statistics:

i) data collection: it is the process of observing, gathering and measuring data to help finding solution to research problems.

ii) descriptive statistics: in this type the data is summarized through the given observations. This summarization or organization is of samples of population using parameters such as the mean or standard deviation.

iii) inferential statistics: This type of statistics is used to interpret the meaning of Descriptive statistics. When the data has been collected, summarized and analyzed it is used to describe the data. It uses the information from the data and predicts a conclusion or infers a solution statement that was not in the raw data.