STATISTICS WORKSHEET-1

1. a) True
2. a) Central Limit Theorem
3. b) Modelling bounded count data
4. d) All of the above
5. c) Poisson
6. b) False
7. b) Hypothesis
8. a) 0
9. c) Outliers cannot conform to the regression relationship
10. Normal Distribution in statistics is basically a probability distribution function where the graph is plotted with symmetric of both sides of the mean. The right side of the mean is the mirror image of the left side. The graph under normal distribution represents the probability and total area under the curve is equal to one.
11. Missing data in statistics can be dealt with two primary methods – Imputation or removal of data.
    The imputation technique is used when the percentage of missing data is too low.
    Imputation using Mean-Median values and imputation using most frequent values.
    In Mean-mode-median imputation, we replace all the occurrences of missing values with the value of mean or median of that values
    The most frequent imputation works with the categorical type of data and the missing values are replaced with the most frequent values occurring in each column.
12. A/B Testing is a randomized control testing mechanism. The mechanism compares the two variable to find out which variable performs better in a control environment. They are widely used in industry to make product launch decisions.
13. No, it is not good practise to use mean imputation for missing data values. Replacing all the missing values with the mean of the data leads to an underestimate of the standard deviation. It distorts the relationship between the variables by taking the correlation estimates to zero.
14. Linear regression is the most common and basic algorithm used for predictive analysis. With the help of linear regression, we identify whether set of predictor variable able to predict the outcome variable. Secondly, what particular types of variables are used to identify the outcome variable? This regression algorithm is used to explain the relationship between one dependent variables with one or more independent variables.
15. The two branches of statistics are – Descriptive Statistics and Inferential Statistics. Descriptive statistics basically deals with the presentation and collection of data, while inferential statistics deals with the hypothesis and drawing conclusion from the data.