STATISTICS WORKSHEET – 1

1.a.True

2.a Central Limit Theorem

3.c.Modelling Contingency Tables

4.d All of the mentioned

5.c.Poisson

6.b.False

7.b.Hypothesis

8.a.0

9.c.outliers cannot conform to the regression relationship

10. Normal Distribution is a probability distribution that is symmetric about the mean, showing that the data near the mean are more frequent in occurrence than data far from the mean. In graphical form it as a bell shaped curve. The mean of normal distribution is 0 and S.D is 1.

11. Missing data can be dealt in many ways. One is to ignore it, in such a huge dataset it would hardly matter if we have dropped some of the rows or some not so important columns. But, if there are a larger number of missing values then simply dropping off the rows/columns will impact our analysis and we end up building a wrong model. So, in that case we can go for data imputation and analyse the entire data set as if the imputed values were the true observed values. We can replace the missing values with the mean,median or mode of the variable, or can simply substitute the values depending on some observation made from the data.

12. A/B testing is a scientific method of comparing two versions of a website, app or advertisement to determine which one performs better so that you can optimize future campaigns accordingly. It's also known as split testing. It helps to determine which version will lead to greater sales or leads and lower costs per conversion by showing us where there might be room for improvement before going live with either option.

13. Mean Imputation is one of the simplest and most widely used imputation methods. However, it has some drawbacks, it reduces the variance and standard deviation of the data, ignores data distribution and correlation of the data, and potentially create unrealistic values.

14. Linear Regression analysis is used to predict the value of variable based on the value of another variable . The variable you want to predict is called dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear Regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use 'least squares' method to discover the best fir line for a set of paired data. You then estimate the value of X(dependent variable) from Y(independent variable).

15. There are two main branches of statistics.

a. Descriptive statistics deals with the presentation and collection of data.

b. Inferential statistics as the name suggests involves drawing the right conclusion from the statistical analysis that has been performed using descriptive statistics.