STATISTICS WORKSHEET-1

1. True

2. Central Limit Theorem

3. Modelling contingency tables

4. The square of a standard normal random variable follows what is called chi-squared distribution

5. Poisson

6. False

7. Hypothesis

8. 0

9. Outliers cannot conform to the regression relationship

10.

Normal Distribution is a probability distribution that is symmetrical and bell-shaped, which means that the distribution of values in a data set follows a specific pattern. In a normal distribution, the majority of the data is cantered around the mean, and the further the values are from the mean, the less frequently they occur.

11.

Handling missing data is an important part of data analysis, and there are several techniques to impute missing values in a dataset, including mean imputation, mode imputation, median imputation, and hot-deck imputation. The choice of imputation technique depends on the type of data and the nature of the missingness. For example, if the data is continuous, mean imputation can be used, while if the data is categorical, mode imputation can be used.

12.

A/B testing is a statistical method used to compare the performance of two versions of a webpage or an app to determine which one performs better. In A/B testing, a random sample of users is split into two groups, and each group is shown a different version of the webpage or app. The performance of each version is then measured, and statistical tests are used to determine whether there is a significant difference in performance between the two versions.

13.

Mean imputation is a commonly used technique to impute missing data, but it has its limitations. Mean imputation assumes that the missing values are missing completely at random and that the missingness is not related to any other variable in the dataset. However, this assumption may not hold in all cases, and mean imputation can introduce bias and affect the accuracy of the analysis.

14.

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. In linear regression, the goal is to find the best-fit line that can explain the relationship between the variables. The line is represented by an equation of the form y = mx + b, where y is the dependent variable, x is the independent variable, m is the slope of the line, and b is the y-intercept.

15.

Statistics is a broad field that has several branches, including descriptive statistics, inferential statistics, probability theory, regression analysis, time series analysis, and Bayesian statistics. Descriptive statistics deals with summarizing and presenting data, while inferential statistics deals with making inferences and predictions based on the data. Probability theory deals with the study of random events, and regression analysis deals with modelling the relationship between variables. Time series analysis deals with modelling data that varies over time, while Bayesian statistics deals with updating beliefs based on new evidence.