STATISTICS WORKSHEET 1

1. Option A- True
2. Option A- Central Limit Theorem
3. Option B - Modeling bounded count data
4. Option D- All of the mentioned
5. Option C- Poisson
6. Option B- False
7. Option B- Hypothesis
8. Option A – 0
9. Option C - Outliers cannot conform to the regression relationship
10. What do you understand by the term Normal Distribution?

Ans-Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, the normal distribution appears as a bell curve.

11. How do you handle missing data? What imputation techniques do you recommend?

Ans- Replacing with Arbitrary Value.

Replacing with Mode.

Replacing with Median.

Replacing with previous value.

Replacing with next value.

Interpolation.

Impute the Most Frequent Value.

Imputation Techniques

- Complete Case Analysis (CCA) - This is a quite straightforward method of handling the Missing Data, which directly removes the rows that have missing data i.e. we consider only those rows where we have complete data i.e. data is not missing. ...
- Arbitrary Value Imputation.
- Frequent Category Imputation.

12. What is A/B testing?

Ans - A/B testing also known as split testing is the process of comparing two versions of a web page, email, or other marketing asset and measuring the difference in performance. You do this giving one version to one group and the other version to another group. Then you can see how each variation performs.

13. Is mean imputation of missing data acceptable practice?

Ans- Mean imputation is typically considered terrible practice since it ignores feature correlation. Mean imputation does not preserve the relationships among variable if the data are missing completely at random, the estimate of the mean remains unbiased.

14. What is linear regression in statistics?

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

15. What are the various branches of statistics?

Ans- There are three real branches of statistics: data collection, descriptive statistics and inferential statistics

1)Data Collection: There are many methods used to collect or obtain data for statistical analysis. Three of the most popular methods are: Direct Observation • Experiments, and • Surveys. A survey solicits information from people

2)Descriptive Statistics: Descriptive statistics describe, show, and summarize the basic features of a dataset found in a given study, presented in a summary that describes the data sample and its measurements. It helps analysts to understand the data better.

3)Inferential Statistics: Inferential statistics is used for comparing the parameters of two or more samples and makes generalizations about the larger population based on these samples.