STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.

Ans- True

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

Ans- Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?

Ans- Modeling bounded count data

4. Point out the correct statement.

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

5. _____ random variables are used to model rates.

Ans- Poisson

6. Usually replacing the standard error by its estimated value does change the CLT.

Ans- False

7. Which of the following testing is concerned with making decisions using data?

Ans- Hypothesis

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.

Ans- 0

9. Which of the following statement is incorrect with respect to outliers?

Ans- Outliers cannot conform to the regression relationship

10. What do you understand by the term Normal Distribution?

Ans- The normal distribution, also known as the Gaussian distribution, is a symmetric probability distribution with a bell-shaped curve. It's characterized by its mean and

standard deviation, and it's widely used in statistics and various fields due to its prevalence in natural phenomena and its mathematical properties.

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

Ans- Handling missing data involves various techniques:

1. Mean/Median/Mode Imputation**: Fill missing values with the mean, median, or mode of the available data in the same column.

2. Forward Fill/Backward Fill**: Use the last known or next known value to fill in missing data, useful for time series.

3. Interpolation**: Estimate missing values based on neighboring data points.

4. Multiple Imputation**: Generate multiple datasets with imputed values and combine results.

5. K-Nearest Neighbors (KNN) Imputation**: Predict missing values based on nearest neighbors.

6. Predictive Modeling**: Use machine learning algorithms to predict missing values based on other features.

12. What is A/B testing?

Answer- A/B testing, also known as split testing, is a method used to compare two versions of a product or service to determine which one performs better. In an A/B test, two versions, A and B, are compared by randomly assigning users to either version A or B and then measuring their responses to determine which version is more effective.

13. Is mean imputation of missing data acceptable practice?

Ans- Mean imputation is a simple method for handling missing data, but it's not always ideal. It can introduce bias and underestimate variability, particularly if data are not missing completely at random.

14. What is linear regression in statistics?

Ans- Linear regression is a statistical method used to model the relationship between one or more independent variables and a dependent variable, assuming a linear relationship.

Linear regression is a statistical method represented by the formula:

$$ Y = \beta_0 + \beta_1X + \varepsilon $$

Where:

- $ Y $ is the dependent variable,

- $ X $ is the independent variable,

- $ \beta_0 $ is the intercept,

- $ \beta_1 $ is the slope,

- $ \varepsilon $ represents the error term.

15. What are the various branches of statistics?

Ans- Statistics encompasses several branches, including:

1. Descriptive Statistics

2. Inferential Statistics

3. Probability Theory

4. Biostatistics

5. Econometrics

6. Bayesian Statistics

7. Multivariate Statistics

8. Spatial Statistics

9. Time Series Analysis

10. Statistical Learning (Machine Learning)