STATISTICS WORKSHEET-1

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

Ans - a) 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 - a) Central Limit Theorem

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

Ans - b) Modeling bounded count data

4. Point out the correct statement.

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

5. _____ random variables are used to model rates.

Ans - c) Poisson

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

Ans - a) True

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

Ans - b) Hypothesis

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

Ans - a) 0

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

Ans - c) Outliers cannot conform to the regression relationship

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

Ans - A normal distribution is a probability distribution that is symmetrical and bell-shaped. It is characterized by its mean (average) and standard deviation.

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

Ans - Deletion: Removing rows or columns with missing values

Mean/Median Imputation: Replacing missing values with the mean or median of the respective column.

Mode Imputation: Replacing missing values with the most frequent value in the column

K-Nearest Neighbors (KNN): Imputing missing values based on values from similar data points.

Multiple Imputation: Creating multiple imputed datasets to account for uncertainty in the imputation process.

12. What is A/B testing?

Ans - A/B testing is a statistical method used to compare two versions of a webpage, app, or other product to determine which performs better. It involves randomly assigning users to different versions and measuring the impact on a specific metric

13. Is mean imputation of missing data acceptable practice?

Ans - Mean imputation is generally not recommended as it can underestimate variability and lead to biased results.

14. What is linear regression in statistics?

Ans - Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the variables and aims to find the best-fitting line that minimizes the sum of squared errors.

15. What are the various branches of statistics?

Ans - Descriptive Statistics: Summarizes and describes data using measures like mean, median, mode, standard deviation, and visualizations.

Inferential Statistics: Draws conclusions about a population based on a sample of data. It involves hypothesis testing, confidence intervals, and regression analysis.