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: b) False

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: Normal Distribution, also known as Gaussian distribution, is a fundamental concept in statistics and probability. It represents a symmetric bell-shaped curve that describes the distribution of continuous data in many natural phenomena. In a normal distribution, the mean, median, and mode are all the same, and the data clusters around this central value. The distribution is fully characterized by its mean and standard deviation, which determine its shape and spread. Many real-world phenomena, like measurements and errors, often follow this distribution due to the central limit theorem.

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

    Ans: Handle missing data by imputing estimated values. Recommended techniques include mean/median imputation, KNN imputation, regression imputation, and multiple imputation. Choose based on data type and context.

12. What is A/B testing?

    Ans: A/B testing is a method to compare two versions of a webpage or app against each other to determine which one performs better. It involves randomly assigning users to either version A or B and analyzing their responses to identify which version yields higher engagement, conversions, or other desired outcomes.

13. Is mean imputation of missing data acceptable practice?

Ans: Mean imputation of missing data is a simple technique, but it has limitations. While it can work well for small amounts of missing data, it may introduce bias and underestimate variability. It assumes missing values are missing completely at random, which might not always hold. More advanced imputation methods might be preferred in many cases.

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 and aims to find the best-fitting line (or hyperplane in higher dimensions) that minimizes the difference between predicted and observed values. It's widely used for prediction and understanding the impact of variables on an outcome.

15. What are the various branches of statistics?

Ans: Various branches of statistics are as follows:

- Descriptive Statistics
- Inferential Statistics
- Social Statistics
- Time Series Analysis
- Bayesian Statistics
- Survival Analysis

These branches specialize in different aspects of data inference.