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.
   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?
   a) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?
   b) Modeling bounded count data

4. Point out the correct statement.
   d) All of the mentioned

5. _____random variables are used to model rates.
   c) Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   b) False

7. 1. Which of the following testing is concerned with making decisions using data?
   b) Hypothesis

8. 4. Normalized data are centered at___and have units equal to standard deviations of the original data.
   a) 0

9. Which of the following statement is incorrect with respect to outliers?
   d) None of the mentioned

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

Normal 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.

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

There are several methods for handling missing data. Some common techniques include mean imputation, regression imputation, and multiple imputations. The choice of method depends on the nature of the data and the analysis being performed.

12. What is A/B testing?

A/B testing is a statistical method used to compare the performance of two versions of a statistical model. The performance of each version is then measured and compared using statistical analysis to determine which version has a more effective prediction.

13. Is mean imputation of missing data acceptable practice?

Mean imputation involves replacing missing values with the mean value of the variable. While it is a simple method, it is generally not recommended as it can lead to biased estimates and reduce the variability of the data.

14. What is linear regression in statistics?

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data.

15. What are the various branches of statistics?

Statistics has two main branches: descriptive statistics and inferential statistics. Descriptive statistics involves organizing, summarizing, and presenting data in an informative way. Inferential statistics involves using sample data to make estimates or decisions about a larger population.