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**
   b) False
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**
   b) Central Mean Theorem
   c) Centroid Limit Theorem
   d) All of the mentioned
3. Which of the following is incorrect with respect to use of Poisson distribution?
   a) Modeling event/time data
   b) **Modeling bounded count data**
   c) Modeling contingency tables
   d) All of the mentioned
4. Point out the correct statement.
   a) The exponent of a normally distributed random variables follows what is called the log-normal distribution
   b) **Sums of normally distributed random variables are again normally distributed even if the variables are dependent**
   c) The square of a standard normal random variable follows what is called chi-squared distribution
   d) All of the mentioned
5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) **Poisson**
   d) All of the mentioned
6. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) **False**
7. Which of the following testing is concerned with making decisions using data?
   a) Probability
   b) **Hypothesis**
   c) Causal
   d) None of the mentioned
8. Normalized data are centered at_____and have units equal to standard deviations of the original data.
   a) **0**
   b) 5
   c) 1
   d) 10
9. Which of the following statement is incorrect with respect to outliers?
   a) Outliers can have varying degrees of influence
   b) Outliers can be the result of spurious or real processes
   c) **Outliers cannot conform to the regression relationship**
   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 around the mean, where most of the observations cluster around the mean and decrease symmetrically as you move away from the mean. It is characterized by its mean and standard deviation, and it forms a bell-shaped curve when plotted.

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

Missing data can be handled through various imputation techniques such as mean imputation, median imputation, mode imputation, regression imputation, or using sophisticated methods like k-nearest neighbors (KNN) imputation or multiple imputation. The choice of imputation technique depends on the nature of the data and the underlying assumptions.

12. What is A/B testing?

A/B testing is a statistical hypothesis testing method used to compare two or more versions of a product or service to determine which one performs better. It involves randomly assigning users or subjects to different variants and then comparing their outcomes to determine the most effective variant.

13. Is mean imputation of missing data acceptable practice?

Mean imputation of missing data can introduce bias and distort the distribution of the data, especially if the missing data are not missing completely at random (MCAR). While it's a simple method, it's generally not recommended as a standalone technique for handling missing data, particularly when there are better imputation methods available.

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. It is used for predicting the value of the dependent variable based on the values of the independent variables.

15. What are the various branches of statistics?

Branches of statistics include descriptive statistics, which involve summarizing and organizing data, and inferential statistics, which involve making inferences and predictions about populations based on sample data. Other branches include probability theory, hypothesis testing, regression analysis, and multivariate analysis.