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. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
7. 1. Which of the following testing is concerned with making decisions using data?
   a) Probability
   b) Hypothesis
   c) Causal
   d) None of the mentioned
8. 4. 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?
11. How do you handle missing data? What imputation techniques do you recommend?
12. What is A/B testing?
13. Is mean imputation of missing data acceptable practice?
14. What is linear regression in statistics?
15. What are the various branches of statistics?

Answers:-

1. True
2. Central Limit Theorem
3. Modeling bounded count data
4. All of the mentioned
5. Poisson
6. False
7. Hypothesis
8. 0
9. None of the mentioned

10. Normal distribution, also known as the Gaussian distribution, is a continuous probability distribution that is symmetrical and characterized by its bell-shaped curve.

11. There are a few ways to handle missing data in dataset which are as follows:-

    A. Deleting Rows:- In deleting rows it removes the rows & Columns which contains the missing values
    B. Imputation Method:- In this method it replaces the missing values with having guess. There are few Imputation Techniques which are listed as (Mean, Median, Mode Imputation; Regression Imputation)

12. A/B Testing is a testing method which is used to compare the 2 variables to determine whose performance is better. This testing is also known as Split Testing. This type of testing is commonly used in Marketing, Web Designing etc.

13. Mean imputation of missing data is acceptable practice only if the percentage of missing data is very low otherwise mean imputation is not considered as an acceptable practice for imputing missing data.

14. Linear regression in statistics is a fundamental statistical technique that provides valuable understanding between variables. There are 2 types of Linear regressions: - Simple Linear Regression & Multiple Linear Regression.

15. Various Branches of Statistics: -

    A. Descriptive Statistics
    B. Biostatistics
    C. Applied Statistics
    D. Mathematical Statistics
    E. Time Series Analysis
    F. Inferential Statistics