<u>STATISTICS WORKSHEET-1</u>

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
   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?
   a) Central Limit Theorem
   b) Central Mean Theorem
   c) Centroid Limit Theorem
   d) All of the mentioned
   ANS: A) Central Limit Theorem

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
   ANS: B) Modeling bounded count data

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
   ANS: C

5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
   ANS: C) POISSON

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
   ANS: 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
   ANS: Hypothesis

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
ANS: A) 0

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

ANS: C) Outliers cannot conform to the regression relationship

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

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

    ANS: Normal Distribution also known as Gaussian Distribution or bell curve. It is continuous probability distribution that is symmetric around its mean, showing that data near the mean more frequent occurrence than data far from mean.

11. How do you handle missing data? What imputation techniques do you recommend?
 ANS: Presence of missing data can impact the performance of machine learning model. The choice of imputation technique depends on nature of the data and the reason for missingness. Here are some techniques to handle missing data: 1) list wise detection and column wise detection. 2) Mean, median mode imputation, Linear regression imputation, K-Nearest Neighbors, Multiple Imputation 3) Matrix Factorization technique and deep learning imputation

12. What is A/B testing?
    ANS: A/B testing also known as split testing or controlled experimentation .This technique is used to compare to more machine Learning model or algorithms. This testing in machine learning essential for making informed decision about model selection, Hyperparameter tuning and feature engineering.

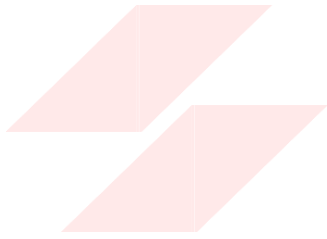13. Is mean imputation of missing data acceptable practice?
    Mean imputation where missing value replaced with observed value in a variable. It is commonly used method for missing data but its acceptability depend on the nature of the data. There are some advantages and disadvantages of using this imputation

14. What is linear regression in statistics?
    ANS: Linea regression is a statistical method used to model relationship between dependent variable and one or more independent variable. The relationship modeled as a linear equation where the coefficient represent weight assigned to each independent variable.

15. What are the various branches of statistics?
    ANS: There are two main branches of statistics 1) Descriptive statistics and 2) Inferential Statistics