

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.

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. d) All of the mentioned

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

Q10 and 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 the Gaussian distribution or bell curve, is a statistical distribution that is symmetric and bell-shaped. It is a continuous probability distribution characterized by its probability density function.

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

Ans. Handling missing data is a crucial aspect of data preprocessing in any data analysis or machine learning task. There are various techniques to deal with missing data, and the choice of method depends on the nature of the data and the specific context of the analysis. Here are some common techniques for handling missing data-

Imputation:

1. Mean, Median, or Mode Imputation
2. Forward Fill or Backward Fill
3. Linear Regression Imputation
4. K-Nearest Neighbors (KNN) Imputation

12. What is A/B testing?

Ans. A/B testing is also known as split testing, it is a method used in marketing and product development to compare two versions of a webpage, application feature, or other elements to determine which one performs better. The goal is to assess changes and improvements objectively

13. Is mean imputation of missing data acceptable practice?

Ans. Mean imputation is a straightforward and commonly used method for handling missing data by replacing missing values with the mean of the observed values for that variable. While it is simple and easy to implement

14. What is linear regression in statistics?

Ans. Linear regression is a statistical method used for modeling the relationship between a dependent variable and one or more independent variables. The goal of linear regression is to find the best-fitting linear relationship that describes the average relationship between the variables

15. What are the various branches of statistics?

- Ans.
1. Descriptive Statistics
 2. Inferential Statistics
 3. Bayesian Statistics
 4. Statistical Learning Theory
 5. Resampling Methods
 6. Time Series Analysis

7.Spatial Statistics