

1. Bernoulli random variables take (only) the values 1 and 0.

Sol. 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?

Sol. c) Centroid Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?

Sol. b) Modeling bounded count data

4. Point out the correct statement.

Sol. c) The square of a standard normal random variable follows what is called chi-squared distribution

5. _____ random variables are used to model rates.

Sol. c) Poisson

6. Usually replacing the standard error by its estimated value does change the CLT.

Sol. b) False

7. Which of the following testing is concerned with making decisions using data?

Sol. b) Hypothesis

8. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

Sol. a) 0

9. Which of the following statement is incorrect with respect to outliers?

Sol. c) Outliers cannot conform to the regression relationship

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

Sol. The normal distribution is a probability distribution that is symmetric and bell-shaped. It is characterized by its mean (average) and standard deviation, and many natural phenomena in the world tend to follow this distribution pattern. In a normal distribution, the majority of the data points cluster around the mean, and the probability of observing values decreases as you move away from the mean in either direction

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

Sol. Missing data can be handled through various imputation techniques. Some recommended methods include Mean Imputation i.e. replacing missing values with the mean of the observed data, Median Imputation i.e. using the median, and Regression Imputation i.e. predicting missing values using regression models.

12. What is A/B testing?

Sol. A/B testing, also known as split testing, is a statistical method used to compare two versions of a data to determine which one performs better in terms of a desired outcome. It involves randomly assigning users or subjects to different versions and measuring their

responses. This helps to assess the impact of changes and make data-driven decisions about which version is more effective.

13. Is mean imputation of missing data acceptable practice?

Sol. Mean imputation is a simple method to handle missing data, it has limitations as well. It can introduce bias and underestimate the variability in the data since it ignores relationships between variables. Mean imputation assumes that missing values are missing completely at random, which might not hold true in many real-world scenarios. More advanced imputation techniques, such as regression imputation or multiple imputation, are often recommended.

14. What is linear regression in statistics?

Sol. 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 the observed data. The goal is to find the best-fitting line (regression line) that minimizes the sum of squared differences between the observed data points and the predicted values from the linear equation. Linear regression is commonly used for prediction, understanding relationships between variables, and assessing the strength and direction of associations.

15. What are the various branches of statistics?

Sol. Below are the branches of statistics:

Descriptive Statistics: Involves summarizing and describing data using measures such as mean, median, mode, variance, and standard deviation.

Inferential Statistics: Focuses on making predictions or inferences about populations based on sample data.

Probability Theory: Deals with the study of random events and uncertainty, providing a foundation for statistical analysis.

Biostatistics: Applies statistical methods to analyse data in fields related to biology and health sciences

Econometrics: Applies statistical techniques to economic data to understand and model economic relationships.

Social Statistics: Applies statistical methods to study social phenomena, such as demographics and social trends

Multivariate Statistics: Deals with the analysis of multiple variables simultaneously to understand complex relationships

Nonparametric Statistics: Involves methods that do not assume specific distributions or parameters in the data