

Name- Nikhil Lalwani

Batch- DS2307

Statistics - STATISTICS WORKSHEET-1

Q1) Bernoulli random variables take (only) the values 1 and 0.

a) True

b) False

A1) a) True

Q2) 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

A2) a) Central Limit Theorem

Q3) 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

A3) b) Modeling bounded count data

Q4) 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

A4) d) All of the mentioned

Q5) _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

A5) c) Poisson

Q6) Usually replacing the standard error by its estimated value does change the CLT.

a) True

b) False

A6) b) False

Q7) Which of the following testing is concerned with making decisions using data?

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

A7) b) Hypothesis

Q8) Normalized data are centered at _____ and have units equal to standard deviations of the original data.

a) 0

b) 5

c) 1

d) 10

A8) a) 0

Q9) 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

A9) c) Outliers cannot conform to the regression relationship

Q10) What do you understand by the term Normal Distribution?

A10) Normal distribution, also known as Gaussian distribution, is a fundamental concept in statistics and probability theory. It is a continuous probability distribution that is characterized by its bell-shaped curve when plotted on a graph. The shape of the curve is symmetric around its mean, which is also the median and mode of the distribution.

Q11) How do you handle missing data? What imputation techniques do you recommend?

A11) Handling missing data is an important step in data preprocessing and analysis to ensure the accuracy and reliability of your results. There are several strategies you can employ to deal with missing data:

- Remove Rows with Missing Data
- Fill with Default Values
- Forward Fill or Backward Fill
- Interpolation
- Predictive Modeling/Impute Missing Data
- Impute Missing Data

Q12) What is A/B testing?

A12) A/B testing, also known as split testing or bucket testing, is a controlled experimentation method used to compare two versions of a webpage, app, email, or other elements of digital marketing or user experience. The goal of A/B testing is to determine which version performs better in terms of user engagement, conversion rates, or other desired outcomes.

Q13) Is mean imputation of missing data acceptable practice?

A13) Mean imputation, which involves replacing missing values with the mean of the non-missing values in the same column, is a simple and quick method for handling missing data. However, while it might seem appealing due to its simplicity, mean imputation has several limitations and potential drawbacks that should be carefully considered.

Q14) What is linear regression in statistics?

A14) Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. The primary goal of linear regression is to find the best-fitting linear relationship that explains the variation in the dependent variable based on the values of the independent variables.

Q15) What are the various branches of statistics?

A15) Statistics is a broad field that encompasses various branches, each focusing on different aspects of data analysis, interpretation, and inference. Some of the major branches of statistics include:

- Descriptive Statistics
- Inferential Statistics
- Probability Theory
- Statistical Inference
- Regression Analysis

-----END-----