

Q 01. Bernoulli random variables take (only) the values 1 and 0.

A 01. A) True

Q 02. 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 02. A) Central Limit Theorem

Q 03. Which of the following is incorrect with respect to use of Poisson distribution?

A 03. C) Modeling contingency tables

Q 04. Point out the correct statement.

A 04. D) All of the mentioned

Q 05. _____ random variables are used to model rates.

A 05. C) Poisson

Q 06. Usually replacing the standard error by its estimated value does change the CLT.

A 06. B) False

Q 07. Which of the following testing is concerned with making decisions using data?

A 07. B) Hypothesis

Q 08. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

A 08. A) 0

Q 09. Which of the following statement is incorrect with respect to outliers?

A 09. C) Outliers cannot conform to the regression relationship

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

A 10. Normal distribution (also known as the Gaussian distribution), is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3.

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

A 11. There are 3 major categories of missing values - Missing Completely at Random (MCAR), Missing at Random (MAR) and Missing Not at Random (MNAR).

Missing data can be handled using following techniques -

1. Deletion - Deletion technique deletes the missing values from a dataset. It can be done via - List wise deletion, Pairwise Deletion or Dropping complete columns.
2. Imputation techniques - Imputation technique replaces missing values with substituted values. It can be done in any of the following manner - Imputation with constant value, Imputation using Statistics, Advanced Imputation Technique or K_Nearest Neighbour Imputation.

Q12. What is A/B testing?

A 12. A/B testing is a basic randomised control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

It is a hypothetical testing methodology for making decisions that estimate population parameters based on sample statistics.

Q 13. Is mean imputation of missing data acceptable practice?

A 13. Mean imputation is the replacement of a missing observation with the mean of the non-missing observations for that variable. Imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. Therefore mean imputation of missing data is an acceptable practice.

Q 14. What is linear regression in statistics?

A 14. Linear regression is a kind of statistical analysis that attempts to show a relationship between two variables. Linear regression looks at various data points and plots a trend line. Linear regression can create a predictive model on apparently random data, showing trends in data.

Q 15. What are the various branches of statistics?

A 15. There are three main branches of statistics- data collection, descriptive statistics and inferential statistics.

- Data Collection - Data collection is all about how the actual data is collected.
- Descriptive Statistics - Descriptive statistics is the part of statistics that deals with presenting the data we have.
- Inferential Statistics - Inferential statistics is the aspect that deals with making conclusions about the data.