

## STATISTICS

- 1. A
- 2. A
- 3. B
- 4. C
- 5. C
- 6. B
- 7. B
- 8. A
- 9. C

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

Answer-The normal distribution is a statistical distribution that is also known as Gaussian distribution, and is bell-shaped, it's a type of continuous probability distribution in which most data points cluster toward the middle of the range which indicates that values near the mean occur more frequently than the values that are farther away from the mean.

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

Answer-Handling missing data usually depends on the type and nature of data, the amount of data missing and the reason for missing data, in case of few missed data, we just put N/A Le. null in its place, however with large number of missing data can affect the prediction model so many imputation techniques are used to challenge that

a) Mean, median and mode imputation-Missing values are replaced with the mean, median

or mode of the observed values for that variable. Outliers presence makes this technique not suitable. b) Linear regression imputation - Using a regression model to predict missing values based

on other variables. c) K-Nearest Neighbors (KNN) imputation Data is filled by the values of their k-nearest neighbours

12. What is A/B testing?

Answer- A/B testing also known as split testing, is a statistical method for connecting two variables, A & B against each other and statistically analysed to determine which variation performs according to the goal.

13. Is mean imputation of missing data acceptable practice?

Answer-Mean imputation means the missing values are replaced with the mean of observed values. for the e varia variable. Mean imputation is a sample and asry to implement sechnique which is actually an effective way to substitute mining values, given that there are no mich outliers present

14. What is linear regression in statistics?

## STATISTICS

regression is a statistical method used to model the relationship between Answer Linear regression dependent variable and independent variables. Its form is in the simple linear regression equation  $(Y = \beta_0 + \beta_1 X + e)$ . Where  $Y$  is a dependent variable,  $\beta_1$  is slope or coefficient,  $X$  is the independent variable and  $e$  is the error.

15. What are the various branches of statistics?

Answer - Statistics has mainly 2 branches, Descriptive statistics and Inferential statistics. Descriptive statistics has two branches (a) Central tendency which has mean, median, mode and (b) Dispersion of data which has range, variance, standard deviation, percentile, skew etc.

Inferential statistics has hypothesis testing, z score testing, t test, regression test, chi-square test etc.