STATISTICS

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.

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?

a) Central Limit Theorem

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

b) Modeling bounded count data

4. Point out the correct statement.

d) All of the mentioned

5. _____ random variables are used to model rates.

c) Poisson

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

b) False

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

b) Hypothesis

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.

a) 0

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

c) Outliers cannot conform to the regression relationship

10. What do you understand by the term Normal Distribution?
Ans.: Normal distribution is the graphical representation of the data which have shape of bell curve. It is symmetric about the mean of the set of data which shows that the data which are near to mean are more frequent in occurrence than data far from mean. It is known as gaussian distribution as well.

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

To handle missing data we can use the following techniques:

1. Delete the rows containing missing data;
2. Assign the median values to the missing data
3. Assign the mean value to the missing data

Simple imputation can effectively used to handle the missing data.

12. What is A/B testing?
Ans.: A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment. Suppose you have developed one website or application. If we compare the two different versions of the same website or application and present it to the randomized users. Based on the feedback you can decide which performs better.

13. Is mean imputation of missing data acceptable practice?
Ans.: Mean imputation can be used for the data containing numerical values, as it replaces the missing values with the mean value. In some cases, it may create a false data, hence generally avoided.
For ex. Salary of the person is calculated based on the mean of the other data and based on the experience and age of the person the salary shall be less than the predicted values based on mean.

14. What is linear regression in statistics?
Ans.: Linear regression is the technique used to predict the value of the variable with the help of the other variables. The variable of which value need to predict is called the dependent variable and the variables which are used to predict another variable value are called the independent variables.

15. What are the various branches of statistics?
Ans.: There are majorly two branches of the statistics.

1. Descriptive statistics
2. Inferential statistics