STATISTICS

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

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?

Normal distribution is the proper term for probability bell curve which is an arrangement of dataset in which most values, cluster in middle of range and rest taper off symmetrically towards either extreme. The standard normal deviation has two parameters the mean and standard deviation where mean=0 and standard deviation=1.

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

The two main methods used for handling the missing data are imputation and deletion. Imputation techniques help in predicting the missing value based on the estimate and analyzing if the imputed values were actual observed values.

Each imputation technique can be applicable for different datasets. when mean imputation is easier it also has few drawbacks. The imputation method is based on the missing pattern of dataset and also the impact of the method when data is modified.

12. What is A/B testing?

A/B testing in other words statistically hypothesis testing/statistical inference is an analytical method where a hyposthesis is made on relationship betweentwo datasets and compared with each other to know the statistical significance of each other. It helps in predicting the unknown outcomes from the known information with reasonable level of accuracy.

13. Is mean imputation of missing data acceptable practice?

Even though mean imputation is the easy one, there are many drawbacks in this imputation. The data does not maintain feature correlation with other data and gives a negative effect on accuracy and mean also reduces the variance of data. It creates a bias in the multivariate estimates.

14. What is linear regression in statistics?

It is basic type of predictive analysis to model the relationship between two variables by fitting a linear equation to observed data. It is to predict the value of a variable based on value of another variable. The variable for which we predict the value is called dependent variable and the variable which we used to predict the variable is called independent variable

15. What are the various branches of statistics?

(1) Descriptive Statistics

Descriptive statistics deals with the collection and summarizing of data, its presentation in various forms, such as tables, graphs and diagrams and finding averages and other measures which would describe the data.The statiscian need to be aware of designing experiments,choose right focus group and avoid future biases.

(2) Inferential Statistics

Inferential statistics deals with techniques used for the analysis of data, making estimates and drawing conclusions from limited information obtained through sampling and testing the reliability of the estimates.