STATISTICS ASSIGNMENT

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

Ans.) 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?

Ans.) a.) Central limit Theorem

Q3. Which of the following is incorrect with respect to use of Poisson distribution?

Ans.) b) Modeling bounded count data

Q4. Point out the correct statement.

Ans) d) All of the mentioned

5. _____ random variables are used to model rates.

Ans.) c) Poisson

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

Ans.) b) False

Q7. Which of the following testing is concerned with making decisions using data?
Ans.) b) Hypothesis

Q8.Normalized data are centered at_____and have units equal to standard deviations of the original data.

Ans) a) 0

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

Ans) c) Outliers cannot conform to the regression relationship

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

Ans.) The Normal Distribution is a dataset layout in which the majority of values cluster in the middle of the range and the rest taper off symmetrically on both sides. The mean, median, and mode of a normal distribution are all the same. The graphical representation of this is known as a bell curve because of its flared appearance. It is symmetric because the majority of the observations cluster around the curve's central peak.

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

Ans.)
1. Remove rows with incomplete data;
2. Imputation of Mean/Median/Mode
3. Assigning a unique value
4. Predicting the values that are lacking
5. Using a missing-values-supporting algorithm, such as random forests.

12. What is A/B testing?

Ans. A/B testing is also known as split testing or bucket testing. It is a way to compare two versions of a single variable, by testing someone's response to variant A against variant B and determine which is more effective.

13. Is mean imputation of missing data acceptable practice?

Mean imputation is generally considered poor practise since it ignores feature correlation. Consider the following scenario: we have a table with age and

fitness scores, and an eighty-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

14. What is linear regression in statistics?

Ans. Linear regression in statistics is a linear approach to modelling the relationship between a dependent variable and independent variable.

15. What are the various branches of statistics?

Ans.
Statistics have two main branches, namely:

a. Descriptive Statistics: This usually summarizes the data from the sample by making use of an index like mean or standard deviation. The methods which are used in the descriptive statistics are displaying, organizing, and describing the data.

b. Inferential Statistics: These conclude from data which are subject to random variations like observation mistakes and other sample variation.