STATISTICS WORKSHEET-1

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

Ans. 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?

Ans. Central Limit Theorem

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

Ans. Modeling bounded count data

4. Point out the correct statement.

Ans. All of the mentioned

5. random variables are used to model rates.

Ans. Poisson

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

Ans. False

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

Ans. Hypothesis

8. 4. Normalized data are centered at and have units equal to standard deviations of the original data.

Ans. 0

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

Ans. Outliers cannot conform to the regression relationship

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

Ans. The Normal distribution is a symmetric probability distribution centered on the mean, indicating that data around the mean occur more frequently than data far from it. The normal distribution will show as a bell curve on the graph. The mean of normal distribution is 0 and the standard deviation is 1. It has a kurtosis of 3 and zero skew. Although all symmetrical distributions are normal, but not all normal distributions are symmetric.

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

12. What is A/B testing?

Ans. A/B testing is a method of comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

13. Is mean imputation of missing data acceptable practice?

14. What is linear regression in statistics?

Ans. Linear regression is the simplest and most extensively used statistical technique for predictive modelling analysis. It is a way to explain the relationship between a dependent variable (target) and one or more explanatory variables(predictors) using a straight line.

15. What are the various branches of statistics?

Ans. There are two main branches of statistics that are Descriptive Statistics and Inferential Statistics. Descriptive statistics deals with the presentation and collection of data. Inference statistics are statistical techniques that allow to utilize data from a sample to conclude, predict the behavior of a given population, and make judgments or decisions.