Statistics Worksheet-1

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

Answer: True

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, become that of a standard normal as the sample size increases?

Answer: Central Limit Theorem

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

Answer: Modeling bounded count data

4. Point out the correct statement.

Answer: All of the mentioned

5. _____ random variables are used to model rates.

Answer: Poisson

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

Answer: False

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

Answer: Hypothesis

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

Answer: 0

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

Answer: Outliers cannot conform to the regression relationship

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

Answer: Normal Distribution is the probability distribution of the data values, which is more symmetric in nature. The data occurrence is frequently around the mean of the data rather than far away from the mean. The curve of Normal distribution is bell shaped.

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

Answer: In order to handle the missing data i.e. NAN, we can use functions like fillna, backfill, forward fill. Also, techniques like Simple Imputer, KNN Imputer & Iterative Imputer also can be used.

12. What is A/B testing?

Answer: A/B testing is a way to test or compare two variables in order to find out which one performs better. In Python, the A/B testing can be perform using Hypothesis concept i.e. defining Null Hypothesis and Alternate Hypothesis. Further, finding the P values for the Hypothesis made defines whether to accept or reject the Null Hypothesis.

13. Is mean imputation of missing data acceptable practice?

Answer: Mean Imputation is one of the most popular technique as its quite easy and not that complicated. But, it is not advisable to use mean imputation all the time as it can provide biased data sometimes.

14. What is linear regression in statistics?

Answer: Linear Regression is a linear approach of modeling to device a relationship between features and label.

15. What are the various branches of statistics?

Answer: There are 2 branches of Statistics: Descriptive Statistics and Inferential Statistics