Statistic 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, becomes that of a standard normal as the sample size increases?
   Answer- Central Mean 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- It is the distribution where most of the result lies near the mean. Half of the value will be less than mean or average and half of them will be higher. Most results will be lay close to the mean and tail off symmetrical in both of the directions.
11. How do you handle missing data? What imputation techniques do you recommend?
    Answer- We handle the missing datas or NAN with the help of Imputers.
    I would recommend of using Simple Imputer and KNN Imputer.
12. What is A/B testing?
    Answer- 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?
    Answer- No. Mean imputation is typically considered terrible practice since it ignores feature correlation. For example in a table age and weight is given and a 2year old child is missing weight record. If we average the weight of people between the ages of 10 and 70, the 2year old child will appear to have a significantly more weight than he actually does.
    Mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14. What is linear regression in statistics?

    Answer- It is a statistical method which provides the relationship between two continuous variable. 1st variable is x which is known as predictor or independent variable. 2nd one is y which is response or outcome. Formula- y=mx+c

15. What are the various branches of statistics

    Answer- There are 2 various branches of statistic.

    1) Descriptive which describes the characteristic of data using the measures with central tendency like mean, median, mode and measure of dispersion like range, standard deviation etc.
    2) Inferential which It determines the probability of the characteristics of the sample using probability theory. The most common methodologies used are hypothesis tests, Analysis of variance etc.