STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.
   a) True
   b) False
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
   b) Central Mean Theorem
   c) Centroid Limit Theorem
   d) All of the mentioned
3. Which of the following is incorrect with respect to use of Poisson distribution?
   a) Modeling event/time data
   b) Modeling bounded count data
   c) Modeling contingency tables
   d) All of the mentioned
4. Point out the correct statement.
   a) The exponent of a normally distributed random variables follows what is called the log- normal distribution
   b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
   c) The square of a standard normal random variable follows what is called chi-squared distribution
   d) All of the mentioned
5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
7. 1. Which of the following testing is concerned with making decisions using data?
   a) Probability
   b) Hypothesis
   c) Causal
   d) None of the mentioned
8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
   a) 0
   b) 5
   c) 1
   d) 10
9. Which of the following statement is incorrect with respect to outliers?
   a) Outliers can have varying degrees of influence
   b) Outliers can be the result of spurious or real processes
   c) Outliers cannot conform to the regression relationship
   d) None of the mentioned

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

    Normal distribution is the arrangement of data in which data is distributed equally to the both side of curve. If the curve is normal then the data is suppose to the normal distributed and free from outliers.

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

    Dealing with missing data is one of the most important stage in the process of building data model. In accurate model will lead to less accurate model. There are multiple imputation technique available in the SKlearn like – SimpleImputer, HotcodeEncoder, LabelEncoder and many more.

12. What is A/B testing?

    The A/B testing is the process in which we compare one model to another ideal model to understand the difference of the accuracy in the model. We can also say this is proceed to test the performance of model with respect to another model.

13. Is mean imputation of missing data acceptable practice?

    According to me imputation of missing data is acceptable because in the imputation technique the data will fill by the same data by taking the mean, median and mode. So, the dataset lost the original identity and it will lead to disturbed the accuracy of the model.

14. What is linear regression in statistics?

    Linear regression is the technique to explain the relationship between the features and labels. With the help of relationship we will predict the model.

15. What are the various branches of statistics?

Generally there are two branches of statistics Descriptive and inferential statistic. In descriptive statistic we collect data representation of the data and in inferential statistic we making conclusion based on the result of the descriptive statistic.