STATISTICS WORKSHEET -1

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

   Answer: **True**

2. Which of the following theorem states that the distribution of averages of iid variable, properly normalized, becomes that of a standard normal as the sample size increase?

   Answer: **Central Limit Theorem**

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

   Answer: **Modelling 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 the making decision?

   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 symmetric distribution around the mean. Occurrence is high near the mean and equally distributed in both side of the mean. Occurrence will gradually decrease with increase of distance from the mean. Normal distribution also known as Gaussian distribution; in graph it looks like bell curve.**

11. How do you handle missing data? What imputation techniques do you recommend?
    Answer: **Missing data is an inevitable part of the process. Followings can be done for handling of missing data and accordingly imputation techniques can be adopted:**
    - **Deletion of the particular row / column based on applicability.**
    - **Taking nearest similar value for the missing data.**
    - **Based applicability regression study can be performed to obtain the missing data.**

12. What is A/B testing?
    Answer: **In data science A/B testing is testing with two variable A and B. First A is tried and monitored then B is tried and monitored. Based on better efficacy and results one of the two is selected.**

13. Is mean imputation of missing data acceptable practice?
    Answer: **Imputation of mean preserve the mean of observe data. If data are missing completely at random the estimate of mean is unbiased. It is good as sample size remain full and complete. Only problem with mean imputation is underestimate of standard error.**

14. What is linear regression is statistics?
    Answer: **Linear regression is a model relationship between two variables by establishing a linear equation to observe changes in the dependent variable with changes in the independent variable.**

15. What are the various branches of statistics?
    Answer: **Statistics are mainly divided into two branches. One is descriptive and another is predictive statistics. Again, descriptive statistics is dependent on central tendency and dispersion. On the other side dispersion is can be divided into range, standard deviation, standard error etc. Predictive statistics is depend on different kind of regression studies.**

Name: Humayun Kabir
Batch No.: Internship-25.
Project: Statistics Worksheet – 1.