# 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

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

3. Which of the following is incorrect with respect to use of Poisson distribution?
a) Modeling event/time data

4. Point out the correct statement.
d) All of the mentioned

5. _____ random variables are used to model rates.
c) Poisson

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

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

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0

9. Which of the following statement is incorrect with respect to outliers?
b) Outliers can be the result of spurious or real processes

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

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

Ans: The term normal distribution is associated with data plotting curve where data is plotted using mean as centerline. Normal distribution is plotted up to the third standard deviation. Other than that are considered as outliers.

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

Ans: df.isnull().sum() can identify the null or missing values. The missing values can be replaced by the mean value of that particular column or the data can be dropped if rows are too small.

12. What is A/B testing?

Ans:

13. Is mean imputation of missing data acceptable practice?

Ans:

14. What is linear regression in statistics?

Ans: Linear regression is a type of supervised machine learning. In linear regression the inputs are available and the output is predicted by the machine. It comes in train and test module, where nearly 70% train and 30& test module are selected.

15. What are the various branches of statistics?

Ans: 1. Descriptive 2. Inferential