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

a) True

b) False

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

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

8. 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?

* It is type of probability distribution which is symmetric about the mean, represented in a ways that data near the mean are more frequent in occurrence compared to the data far from the mean
* The graphical representation of normal distribution will appear as a bell curve
* In ideal scenario, mean is 0 and the standard deviation is 1

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

The missing data can be handled in following ways:

* Use Delete/Remove method for eliminating the missing values

12. What is A/B testing?

* It is method to create a hypothesis and test it.
* There are two possible hypothesis – null hypothesis and alternate hypothesis.
* I.e. we reject null hypothesis because we have got the strong evidence while, in other scenario we cannot reject null hypothesis since, we didn’t got any interpretation.

13. Is mean imputation of missing data acceptable practice?

* Ideally, No, because in the case of dataset containing the outliers it many mislead us to wrong outcomes. However, mean can be helpful, in scenarios where data is missing completely at random.

14. What is linear regression in statistics?

* It attempts to model the relationship between two sets of variables
* In this, one of the variables is predictor, explanatory or independent variable. While other variable called as response, outcome or dependent variable
* The relationship is mathematically given by straight line equation

15. What are the various branches of statistics?

* Broadly, there are two types of branches under statistics:
  + Descriptive Statistics: branch of statistics which focuses on collecting, summarizing and presenting a set of data
  + Inference Statistics: branch of statistics that analyses the sample dataset to draw conclusion from the population