**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

WORKSHEET

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: A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

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

ANS: We handle missing data using Mean, Median and Mode and imputation techniques we use is multiple imputation

12. What is A/B testing?

ANS: a/b testing is a method of comparing two versions of a webpage or app against each other to determine which one performs better.

13. Is mean imputation of missing data acceptable practice?

ANS:

* Bad practice in general
* If just estimating means: mean imputation preserves the mean of the observed data
* Leads to an underestimate of the standard deviation
* Distorts relationships between variables by “pulling” estimates of the correlation toward zero

14. What is linear regression in statistics?

ANS: Linear regression establishes the linear relationship between two variables based on a line of best fit. Linear regression is thus graphically depicted using a straight line with the slope defining how the change in one variable impacts a change in the other.

15. What are the various branches of statistics?

ANS: There are two main branches of statistics

- Inferential Statistic.

- Descriptive Statistic.