STATISTICS WORKSHEET-1

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

Ans: **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?

Ans: a) Central Limit Theorem

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

b) Modeling bounded count data

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

Ans: d) All of the mentioned

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

Ans: c) Poisson

d) All of the mentioned

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

a) True

Ans: 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.

Ans: 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

Ans: c) Outliers cannot conform to the regression relationship

d) None of the mentioned WORKSHEET

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. Data is normal when it is symmetric around the mean. It looks like Bell Curve.

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

Ans: Missing data can be found through regression. It can be used to predict the null values using other information from the dataset.

We can find out the mean of that particular variables and use mean as a nan value.

In case of any character values are missing like locations, names etc we can use maximum occurred data to fill the missing value.

12. What is A/B testing?

Ans: Like any type of scientific testing, A/B testing is basically statistical hypothesis testing, or, in other words, statistical inference. It is **an analytical method for making decisions that estimates population parameters based on sample statistics**

13. Is mean imputation of missing data acceptable practice?

Ans: Yes, it is accepted. It is the process of replacing the null/nan values in dataset with the mean of that particular variable.

Sometime mean decrease the variance of the data which leads to increase in the bias. Due to this model' accuracy might reduce. 14. What is linear regression in statistics?

15. What are the various branches of statistics

Ans: 1) Descriptive Statistics

2) Inferential Statistics