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.

Answer:-(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?

Answe:- (a) Central Limit Theorem

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

Answer:- (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

Answer;-(d) All of the mentioned

5. \_\_\_\_\_\_ random variables are used to model rates.

Answer:-(c) Poisson

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

Answer:- (b) False

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

Answer;- (b) Hypothesis

8. 4. Normalized data are centered at\_\_\_\_\_\_and have units equal to standard deviations of the

original data.

Answer;- ( a) 0

9. Which of the following statement is incorrect with respect to outliers?

Answer:- (c) Outliers cannot conform to the regression relationship

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

Answer:- The mean, median, and mode are all equal. Half of the data is less than the mean and half is greater than the mean.

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

Answer;- When dealing with missing data, we can use two primary methods to solve the error: imputation or the removal of data. I will recommend Multiple imputation, Multiple imputation is considered a good approach for data sets with a large amount of missing data. Instead of substituting a single value for each missing data point, the missing values are exchanged for values that encompass the natural variability and uncertainty of the right values. Using the imputed data, the process is repeated to make multiple imputed data sets. Each set is then analysed using the standard analytical procedures, and the multiple analysis results are combined to produce an overall result.

The various imputations incorporate natural variability into the missing values, which creates a valid statistical inference. Multiple imputations can produce statistically valid results even when there is a small sample size or a large amount of missing data.

12. What is A/B testing?

Answer;- To determine which technique of two different versions performs better in a statistical way we do A/B testing.

13. Is mean imputation of missing data acceptable practice?

Answer:- Imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. What is linear regression in statistics?

Answer:- Linear regression quantifies the relationship between one or more predictor variable(s) and one outcome variable. Linear regression is commonly used for predictive analysis and modelling. Linear regression is also known as multiple regression, multivariate regression, ordinary least squares (OLS), and regression.

15. What are the various branches of statistics?

Answer:- There are two branches of statistics <1> Descriptive Statistics And <2> inferential Statistics. Descriptive statistics are procedure used to summarize, organize and make sense of a set source or observation, in descriptive statistics we work on central tendency & spread of data . Inferential statistics are procedure used that all researches to infer or generalized observations made with sample to the larger data.