STATISTICS WORKSHEET-1

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

   Ans: 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: Central Limit Theorem

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

   Ans: Modeling bounded count data

4. Point out the correct statement.

   A: The square of a standard normal random variable follows what is called chi-squared distribution

5. _____ random variables are used to model rates.

   Ans: Poisson

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

   Ans: False

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

   Ans: Hypothesis

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

   Ans: 0

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

   Ans: Outliers cannot conform to the regression relationship

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

    Ans: Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

    normal distribution has two parameters: the mean and the standard deviation. For a normal distribution, 68% of the observations are within +/- one standard deviation of the mean, 95% are within +/- two standard deviations, and 99.7% are within +- three standard deviations.

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

    Ans: Missing data will give wrong results and it reduces the statistical power of the analysis.

    When dealing with missing data, we use two primary methods to solve the error: imputation or the removal of data.

    The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low.

    The other option is to remove data. When dealing with data that is missing at random, related data can be deleted to reduce bias. Removing data may not be the best option if there are not enough observations to result in a reliable analysis.

12. What is A/B testing?

    Ans: A/B testing basically used to compare two different products based on user input.

    Ecommerce sites, YouTube etc will show recommendations to you based on your input.

13. Is mean imputation of missing data acceptable practice?
    Ans: Yes if missing data is low, mean imputation is the replacement of a missing observation with the mean of the non-missing observations for that variable.
14. What is linear regression in statistics?
    Ans: Linear regression is commonly used type of predictive analysis. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable. This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable.
15. What are the various branches of statistics?
    Ans: Branches of statistics are Descriptive Statistics and Inferential Statistics.
    Descriptive Statistics: focuses on collecting, summarizing, and presenting a set of data.
    Inferential Statistics: analyses sample data to draw conclusions.