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?
   Ans: b) Modeling bounded count data

4. Point out the correct statement
   Ans : d) All of the mentioned

5. _____ random variables are used to model rates.
   Ans: c) Poisson

6. Usually replacing the standard error by its estimated value does change the CLT.
   Ans : b) False

7. Which of the following testing is concerned with making decisions using data?
   Ans: a) Probability

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.
   Ans: a) 0

9. Which of the following statement is incorrect with respect to outliers?
   Ans: c) Outliers cannot conform to the regression relationship

10. What do you understand by the term Normal Distribution?
    Ans: Normal Distribution also known as Gaussian Distribution. Here 99.7% of data are within the range of (median +/- 3 std). It is represented as a bell curve graphically.

11. How do you handle missing data? What imputation techniques do you recommend?
    Ans: Missing data can be handled in two ways. Either we remove the rows where data is missing or we can replace the missing data with a non-nan data.

    Imputation technique: a) Using pandas with drop function b) using pandas and replacing the missing data with forward fill, backward fill or mean value of the column

12. What is A/B testing?
    Ans: It means creating two or more sample from the dataset and then creating a model from each sample dataset and then testing the accuracy of each to predict the output.

13. Is mean imputation of missing data acceptable practice?
    Ans : NO its not as it doesn't preserve the relationship between the independent variable. Its also leads to underestimating of standard error and low p-values.

14. What is linear regression in statistics?
    Ans: It's a technique to establish linear relationship between independent and dependent variables. Its is represented as y=a+bx (Best Fit Line) where, y is output,a is intercept is coefficient slope and x is the independent variable)

15. What are the various branches of statistics?