STATISTICS WORKSHEET

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?
   Answer :- 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
   Answer :- d) All of the mentioned
5. _____ random variables are used to model rates.
   Answer :- c) Poisson
6. Usually replacing the standard error by its estimated value does change the CLT
   Answer :- b) False
7. Which of the following testing is concerned with making decisions using data?
   Answer :- a) Probability
8. 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 :- A significant class of statistical distributions with numerous uses is the normal distribution. The majority of machine learning algorithms use this distribution. The Gaussian or Gauss distribution is another name for the Normal distribution. The mean and standard deviation are the two key components of a normal distribution. We can choose the distribution's shape and probabilities in relation to our problem statement with the aid of these factors. The distribution's form changes as the parameter value changes.
11. How do you handle missing data? What imputation techniques do you recommend?
    Answer :- In order to handle missing data there are three ways :- Deletion to eliminate data, regression analysis to eliminate data and imputation of data. To deal with missing data, we employ two data imputation methods: average imputation and common-point imputation. When a value is absent, average imputation fills it in using the average of the replies from the other data entries. The dataset's variability may be artificially reduced by employing this method, therefore use with caution. On the other hand, when data scientists use common-point imputation, they choose the midpoint or most frequently selected value. I'd suggest using the Average imputation.
12. What is A/B testing?
    Answer :- An elementary randomised control experiment is A/B testing. It is a method for contrasting two variations of a variable to see which performs better in a regulated setting. A/B testing is a type of experiment in which you divide your website traffic or user base into two groups and provide them with two different iterations of the same web page, app, email, etc. The objective is to compare the results and identify the version that performs better.
13. Is mean imputation of missing data acceptable practice?
    Answer :- Yes, Mean imputation of data is acceptable especially when we are dealing which

large segment of numeric data where data s heterogenous but still proves to have relation within.

14. What is linear regression in statistics?|
    Answer :- A fundamental and widely used form of predictive analysis is linear regression. Regression analysis' main goal is to look at two things: (1) Is it possible to accurately forecast an outcome (dependent) variable using a set of predictor variables? (2) Which individual variables—as shown by the size and sign of the beta estimates—are highly important predictors of the outcome variable, and how do they affect the outcome variable. These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.
15. What are the various branches of statistics?
    Answer :- There are three branches of statistics:
    data collection,
    descriptive statistics and
    inferential statistics