STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0. - 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? - Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution? - Modeling bounded count data

4. Point out the correct statement. - All of the mentioned

5. _____ random variables are used to model rates - Poisson

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

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

8. Normalized data are centered at___and have units equal to standard deviations of the original data. -0

9. Which of the following statement is incorrect with respect to outliers? - Outliers cannot conform to the regression relationship

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

Normal distribution refers to the cluster of most values in the data set is arranged in the mid range and rest of the data tapers symmetrically at the extreme.

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

Missing data can be handled by "Imputing the missing values". There are several imputing techniques available. Depending on the missing value category we can make decision which one to use.
Some of the techniques are: Forward fill, Backward fill, Univariate approach, Multivariate approach, Interpolate approach.

12. What is A/B testing?

It is a simple randomized controlled experiment, in which two samples(A and B) of single vector variables compared.

13. Is mean imputation of missing data acceptable practice?

No, Mean imputation ignores feature correlation. It does not preserve relationships between variables. It shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval.

14. What is linear regression in statistics?

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data.
Also, model that estimates the relationship between one independent variable and one dependent variable using a straight line.

15. What are the various branches of statistics?

Three branches of statistics.

1) Data collection
2) Descriptive statistics
3) inferential statistics