STATISTICS WORKSHEET 1

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 BOUND COUNT DATA

4. Point out the correct statement.

ANSWER: 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: B] HYPOTHESIS

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

ANSWER: D] 10

9. Which of the following statement is incorrect with respect to outliers?
   ANSWER: C] OUTLIERS CANNOT CONFIRM TO REGRESSION RELATIONSHIP

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

ANSWER: The normal distribution is a probability function that describes how the values of a variable are distributed. It is a symmetric distribution where most of the observations cluster around the central peak and the probabilities for values further away from the mean taper off equally in both directions.

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

ANSWER: generally data features consist of many missing values. The missing values should be filled in the data preprocessing. This missing values can be handled by Imputation methods , they are simple imputer , iterative imputer and knn imputer.
The simple imputer generally used to fill the nan values with the mean or mode values. The knn imputer is used to replace the nearest value comparing the nearest neighbours of the missing value. The iterative imputer takes two factors and makes a relation between them and predicts the values.

12. What is A/B testing?

ANSWER: A/B testing is also known as bucket testing or split-run testing) is a user experience research methodology.A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics. A/B testing is a way to compare two versions of a single variable, typically by testing

a subject's response to variant A against variant B, and determining which of the two variants is more effective.

13. Is mean imputation of missing data acceptable practice?

14. What is linear regression in statistics?
ANSWER: Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable

15. What are the various branches of statistics?

ANSWER: The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics. Descriptive statistics deals with the presentation and collection of data, Inferential statistics, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics