FLIP ROBO

STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

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

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

   b)Modeling bounded count data

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

5. _____random variables are used to model rates.

   c)Poisson

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

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

   b)Hypothesis

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

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

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

   Normal distribution is a probability distribution which is symmetric about the mean, such that data near the mean occurs more frequently than data far from the mean. In graphical form, the normal distribution appears as a "bell curve".

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

   There is no single optimum method to handle missing values. Before applying any methods, it is necessary to understand the type of missing values, then check the datatype and skewness of the missing column, and then decide which method is best for a particular problem.

   The following Imputation techniques can be used -

   a) Basic Imputation Techniques
   • Imputation with a constant value
   • Imputation using the statistics (mean, median, mode)
   b)
   • K-Nearest Neighbor Imputation

12. What is A/B testing?

   A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable, say A & B, to find out which performs better based on a given metric in a controlled environment. It's a common method used in marketing, web design, product development, and user experience design to improve campaigns.

13. Is mean imputation of missing data acceptable practice?
   No, it is considered as a terrible practice.
   If the data missing are at random, then imputing mean is not a good idea since it does not take into account correlations among different features. Also to note that mean reduces the variance of the data.

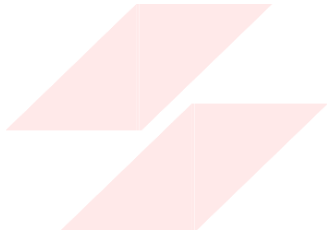14. What is linear regression in statistics?

   Linear regression is a method to model the relationships between one or more feature variable and an outcome variable. These variables are known as the independent and dependent variables, respectively. When there is one independent variable, the procedure is known as simple linear regression. When there are more it is known as multiple regression.
   A linear regression equation allows to predict the mean value of the dependent variable given values of the independent variables

15. What are the various branches of statistics?
   There are three real branches of statistics:
      1. data collection
      2. descriptive statistics
      3. inferential statistics