STATISTICS WORKSHEET-1

WORKSHEET

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.

d) All of the mentioned

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

b) False

7. 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

WORKSHEET

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

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

Ans : The Normal Distribution is continuous distribution in nature. In normal distribution the mean , median and mode all line up such that the centre of the distribution is the mean. As a result half of the result fall to either side of the mean. The normal distribution is also referred as bell shaped or bell curve.

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

There are two methods to handle missing data .Imputation (substituting missing data) and Removing data. Imputation techniques- Mean, median and mode imputation and K-nearest neighbours .

12. What is A/B testing?

Ans : Hypothesis testing is a statistical method used to make inferences about a population based on sample data. It involves formulating two competing hypotheses, the null hypothesis (H0) and the alternative hypothesis (Ha), and then collecting data to assess the evidence.

13. Is mean imputation of missing data acceptable practice?

Mean imputation is the replacement of a missing observation with the mean of the non-missing observations for that variable. This is appropriate for data which is MCAR and if you have MAR or MNAR likely to bias in result. Does not reflect un certainity around imputed value..

14. What is linear regression in statistics?

Ans : Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable (y)you want to predict is called the dependent or response variable. The variable(x) you are using to predict the other variable's value is called the independent ,predictor, explanatory variable.

15. What are the various branches of statistics?

Ans: Descriptive statistics and Inferential statistics are branches of the statistics.

In Descriptive statistics, we describe the data using the Mean, Standard deviation, Charts, or Probability distributions.

In Inferential Statistics we interpret the meaning of the descriptive statistics by inferring them to the population.