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. 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. 4. 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

Q10 and 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 also called as Bell Shaped Curve because it is a symmetrical curve that looks like a bell, showing that data near the mean are more frequent in occurrence than data far from the mean. In a normal distribution the mean is zero and the standard deviation is 1.

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

Missing data are defined as not available values, and that could be meaningful if observed. Missing data can be anything from files missing, information incomplete, data entry error etc. Most datasets in the real world contain missing data. We can transform the missing data manually through analysis and coding strategy, also there are machine learning algorithms and packages that can automatically detect and deal with missing data. Mean, Median and Mode is one of the most common and recommended techniques of imputing values when dealing with missing data.

12. What is A/B testing?

A/B testing is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

13. Is mean imputation of missing data acceptable practice?

Mean imputation is a popular solution to missing data, but it's not an acceptable practise, also, by imputing the mean, we will able to keep our sample size up to the full sample size and the data may be skewed

14. What is linear regression in statistics?

Linear regression allows us to model mathematically the relationship between two or more variables by fitting a linear equation to observed data

15. What are the various branches of statistics?

Statistics is the main branch of mathematics. Used to perform different operations, i.e., Data collection, organization, analysis, and so on. The two branches of statistics, Descriptive and Inferential Statistics. Descriptive statistics is the first part of statistics that deals with the collection of data. Whereas Inference statistics are techniques that enable statisticians to use the information collected from the sample to conclude, bring decisions, or predict a defined population.