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. 4. Normalized data are centred 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?

The normal distribution also known as the Gaussian distribution. It is the most important probability distribution in statistics for independent, random variables. Its familiar bell-shaped curve is ubiquitous in statistical reports. The graph of the normal distribution is characterized by two parameters: the mean, or average, which is the maximum of the graph and its always symmetric, and the standard deviation, which determines the amount of dispersion away from the mean.

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

Common Methods

1. Mean or Median Imputation. When data is missing at random, we can use list-wise or pair-wise deletion of the missing observations
2. Multivariate Imputation by Chained Equations (MICE) MICE assumes that the missing data are Missing at Random (MAR)
3. Random Forest.

   These are imputation techniques are recommended

Imputation Techniques

- Complete Case Analysis(CCA):- This is a quite straightforward method of handling the Missing Data, which directly removes the rows that have missing data i.e we consider only those rows where we have complete data i.e data is not missing. ...
- Arbitrary Value Imputation. ...
- Frequent Category Imputation.

12. What is A/B testing?

A/B testing is a method of comparing two versions of a webpage or app against each other to determine which one performs better. It is also known as split testing or bucket testing.

Testing takes the guesswork out of website optimization and enables data-informed decisions that shift business conversations from "we think" to "we know." It can ensure that every change produces positive results.

13. Is mean imputation of missing data acceptable practice?

True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased.

14. What is linear regression in statistics?

In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables)

A linear regression line has an equation of the form Y = a + bX, where X is the explanatory variable and Y is the dependent variable.

15. What are the various branches of statistics?

There are three real branches of statistics: data collection, descriptive statistics and inferential statistics

Out of which two major areas of statistics are known as descriptive statistics,
which describes the properties of sample and population data, and inferential statistics, which uses those properties to test hypotheses and draw conclusions.