STATISTICS WORKSHEET -1

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

1.a) true

2.a) Central Limit Theorem.

3.b) Modelling bounded count data

4.d) All of the mentioned

5.c) Poisson

6.b) False

7.b) Hypothesis

8.a) 0

9.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, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

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

1. Use deletion methods to eliminate missing data. The deletion methods only work for certain datasets where participants have missing fields.
2. Use regression analysis to systematically eliminate data.
3. Data scientists can use data 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.

Imputation of Arbitrary value, Frequent category 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 to find out which performs better in a controlled environment.

For instance, let’s say you own a company and want to increase the sales of your product. Here, either you can use random experiments, or you can apply scientific and statistical methods. A/B testing is one of the most prominent and widely used statistical tools.

13. Is mean imputation of missing data acceptable practice?

It is 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. Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. What is linear regression in statistics?

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. A linear regression line has an equation of the form Y = a + b X, where X is the independent 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.