STATISTICS WORKSHEET-1

Answers of Q1 to Q9

1. **(a) True**
2. **(a) Central Limit Theorem**
3. **(b) Modeling 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**

Answers of Q10 to Q15

10. **Normal Distribution which is also known as 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. Normal Distribution will appear as a Bell Curve in graph.**

11. **As there are different types of data can be missing in a particular dataset. So there are different techniques to recognize the nature of missing data and to handle the missing data according to nature of data missing in the dataset.**

The commonly used Imputation Techniques to handle missing data are-

(a) Mean or Median Imputation
(b) Multivariate Imputation by Chained Equations
(c) Random Forest

12. **A/B testing is a popular way to test your products and is gaining steam in the data science field. 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. A/B Testing is a widely used concept in most industries nowadays, and data scientists are at the forefront of implementing it. A data scientist collects and studies the data available to help optimize the website for a better consumer experience.**

13. Mean imputation of missing data is acceptable up to some extent. 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.

    Outliers' data points will have a significant impact on the mean and hence in such cases it is not recommended to use the mean for replacing the missing values.

14. In Statistics, Linear Regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression, for more than one, the process is called multiple linear regression.

15. Statistics is concerned with developing and studying different methods for collecting, analysing and presenting the empirical data. There are two various branch of statistics-

    (a) Descriptive Statistics

    (b) Inferential Statistics

Submitted By
Raushan Kumar
(Intern Data Science)