

Flip Robo Statistics Assignment-1

Answers

[Q No. 10] What do you understand by the term Normal Distribution?

Ans-

- Normal distribution is a continuous probability distribution which is symmetric about mean and shows that data which are near to the mean are more frequent in occurrence than data far from the mean.
- It is also known as **Gaussian Distribution**
- The Normal distribution has two Parameters 1. **Mean** 2. **Standard Deviation**.

[Q No. 11] How do you handle missing data? What imputation techniques do you recommend?

Ans-

- Missing data are the data which are not stored or the data are absent for some variables in given data sets.
- After getting the data first we have to categorise the missing values and have to check whether the data is **missing completely at random (MCAR)**, or **Data is missing at Random (MAR)** or **Missing not at random (MNAR)**.
- After detecting the missing values we have to find the relationship among the missing data by using **matrix plot** or by using **heat map**.
- Missing values can be treated with two techniques 1. **Deletion technique** such as pair wise deletion, List wise deletion or Dropping rows and Dropping complete columns 2. **Imputation technique** such as **Imputation with constant Value, Imputation using Statistics and K-nearest neighbour Imputation**.
- Imputation techniques should be recommended based on variables such as **Numerical Variables, Categorical Variable or by Both Variables**.
- For **Numerical Variables** the imputation techniques should be **Mean/Median Imputation, Arbitrary Value Imputation and Mode Imputation**.
- For **Categorical Variable** the imputation techniques should be **Frequent Category Imputation, Adding a "Missing" category**.
- For Both the cases the imputation techniques should be **complete case analysis, Adding a "Missing" Indicator and Random Sample Imputation**.

[Q No. 12] What is A/B testing?

Ans-

- A/B testing is a statistical hypothesis testing where a hypothesis is considered between two Data sets and those data sets are then again compared against each other to determine that is there any statistical relationship exists between the two data sets or not.

[Q No. 13] Is mean imputation of missing data acceptable practice?

Ans-

- No, mean imputation of missing data is not an acceptable practice.

Flip Robo Statistics Assignment-1

Answers

[Q No. 14] What is linear regression in statistics?

Ans-

- Linear regression is a type of statistical analysis which is used to predict value of a variable (commonly known as Dependent Variable) by using the values of another variable (commonly known as independent variable).

[Q No. 15] What are the various branches of statistics?

Ans-

- Statistics is broadly classified into two branches 1. Descriptive statistics
2. Inferential Statistics.

Submitted By- Abhishek Behera

Internship Batch -28