STATISTICS WORKSHEET-1

1. A
2. A
3. C
4. B
5. C
6. B
7. B
8. A
9. C

NORMAL DISTRIBUTION(Q 10)

Normal distribution was discovered by Carl Friedrich Gauss. So it is also called **Gaussian Distribution.** While performing data analysis, we first explore the data and aim to find its probability distribution and most common type of distribution is normal distribution.

It is also known as bell shaped curves. When we plot histrograms and connect the histrogram then bell shaped curved is formed, this curved is known as Normal Distribution.

Many natural datasets basically forms normal distribution. For ex iq level, height of gender in particular area.

MISSING VALUES AND IMPUTATION TECHNIQUE

In data science many times we get the data in which many values are missing. This may deviate our model when it is used for any other set of data.so for this there are many method for handling missing values.for example

1 deleting the row or column from the data set

2 mean, median mode replacement method

3 Algorithms that Support Missing Values k nearest neighour

4 fillna method in pandas

I recommend the algorithm that support the missing values data that is knn.

A/B TESTING

A/B testing is an experiment on the two variants of the same data to see which perform better. It compare the performatance of two variants in the controlled enviroments.

Mean imputation of missing data acceptable practice

we can the mean of imputation for missing data in data sets.but mean imputation does not always works.we cannot always consider the practice for replacing the values.For example we have data sets of average height of person but age is unknown and supposingly his age is 9 year, so we cannot take average height of persons of age group 20 to 50.

Draw back of mean imputation is it do not consider the relations of variables

And secondly mean reduces the variance of the data

. What is linear regression in statistics

Linear regression is supervised machine learning model.in this model we have to find out best fitted line. We plot the line which is at minimum distance from data points. It uses least square method.In this model we try to predict the variable from another variable.the variable we try to predict is known as dependent variable and variable from which we predict the any another variable is known as independent variable.

What are the various branches of statistics

Basically there are four branches of statistics

Mathematical or theorical statitics

It helps in forming the experimental and statistical distribution

Statitical methods or funcations

Its helps in the collection,tabulation and interpretation of the data. Its helps in analyzing the data and returns insight from the data.

Descripitive statitics

It helps in summarizing and organizing any data set characteristics. It also helps in the representation of data in both classification and diagrammatic way.

Inferenial statitics

Its helps in finding the conclusion regarding the population after analysis on the sample drawn from it.