Inferential statistics & Descriptive statistics

Inferential statistics	Descriptive statistics
Concerned with describing the target population	Make inference from the sample and generalize them to the population
Organize, analyze, and present the data in a meaningful manner	Compare, test and predicts future outcomes
Final results are shown in graphs, charts and tables	Final result is the probability scores
Describes the data which is already known	Tries to make conclusion about the population that is beyond the data available
Measure of central tendency Mean median, mode	Measure of variability

Population

The population is the complete set.
Measurable quality is called a parameter.

Sample

The sample is a subset of population.
Measurable quality is called a statistic.

Hypothesis

It is the statistical assumption about the population parameter. The assumption made by researcher may or may not be true, the job of the researcher is to test that assumption be true or false.

Null hypothesis

A null hypothesis is a type of statistical hypothesis that proposes that no statistical significance exists in a set of given observations. Hypothesis testing is used to assess the credibility of a hypothesis by using sample data. It is represented as H0

Alternative hypothesis

An alternative hypothesis is an opposing theory to the null hypothesis. It is represented as H1 or Ha

Central limit theorem

Central limit theorem is a statistical theory which states that when the large sample size has a finite variance, the samples will be normally distributed and the mean of samples will be approximately equal to the mean of the whole population .

Type 1 Error

A type I error is a kind of fault that occurs during the hypothesis testing process when a null hypothesis is rejected, even though it is accurate and should not be rejected.

Type 2 Error

In statistical hypothesis testing, a type II error is a situation wherein a hypothesis test fails to reject the null hypothesis that is false.

Linear regression

Linear regression models the relationships between at least one explanatory variable and an outcome variable. These variables are known as the independent and dependent variables, respectively. When there is one independent variable, the procedure is known as simple linear regression.

Assumptions required for linear regression

- Linearity: The relationship between X and the mean of Y is linear.
- Homoscedasticity: The variance of residual is the same for any value of X.
- Independence: Observations are independent of each other.
- Normality: For any fixed value of X, Y is normally distributed.