

R- Stastistics

R - Statistics

R Statistics concerns data; their collection, analysis, and interpretation.

It has the following two types:

Descriptive statistics: It is about providing a description of the data. It deals with the **quantitative description** of data through numerical representations or graphs.

Example: Normal distribution, Central Tendency, Kurtosis, etc. are some of the statistical techniques in Descriptive Statistics.

Inferential statistics: In inferential statistics, we **draw conclusions or ‘inferences’ from our dataset**. Also, a conclusion is drawn about **the larger population from a data of a much smaller sample**.

Example: Central Limit Theorem, Hypothesis Testing, ANOVA are some of the inferential statistics techniques.

Types of Data in Statistics

Different types of data in Statistics:

- Numerical (discrete and continuous)
- Categorical
- Ordinal

Data is nothing but information that is gathered as a result of a survey.

Data can either be numerical or categorical in nature.

Numerical Data is again of two types –

Discrete

Continuous.

a. **Discrete data** – It represents items that can be counted.

b. **Continuous data** – It represents measurements. Also, their possible values cannot be counted. Although, it can only be described using intervals on the real number line.

2. Categorical Data

Categorical Data is used to **represent characteristics** that are present in the data such as a person's gender, marital status, hometown.

3. Ordinal data

In this form of data, the variables have an ordered category which is natural and the distance between these variables is not known. **Ordinal Data is similar to categorical data with the only difference that the data is ordered.**

For example, Rating a restaurant on a scale of 0 to 4 gives us ordinal data.

ANOVA

ANOVA is used for testing the significance of the differences among more than two sample means.

Assumptions

- ↗ Each sample is randomly drawn from normal population
- ↗ Each of these population have same variance

Analysis of variance is based on comparison of two different estimates of the variance σ^2 , of overall population.

Hypothesis:

- ↗ H_0 : All means are equal
- ↗ H_1 : At least two means are not equal.