Unit II

1. Descriptive Multivariate Analysis:

Descriptive Multivariate Analysis in data analytics involves analyzing and understanding the relationships and interactions between multiple variables in a dataset simultaneously. Unlike univariate analysis that focuses on a single variable, multivariate analysis explores how different variables relate to each other and how they collectively contribute to patterns and insights.

Here are some common techniques used in descriptive multivariate analysis:

1. Scatter Plots: Scatter plots are used to visualize the relationship between two continuous variables. Each data point represents an observation, and the position of the point on the graph is determined by the values of the two variables being compared. Scatter plots can reveal patterns like correlations or clusters of data points.
2. Correlation Analysis: Correlation analysis measures the strength and direction of the linear relationship between two or more continuous variables. It quantifies the degree to which changes in one variable are associated with changes in another. The correlation coefficient ranges from -1 to +1, where -1 indicates a perfect negative correlation, +1 indicates a perfect positive correlation, and 0 indicates no correlation.
3. Covariance: Covariance is a measure of the joint variability of two continuous variables. It indicates the extent to which two variables change together. Positive covariance indicates that the variables tend to increase or decrease together, while negative covariance indicates they move in opposite directions.
4. Heatmaps: Heatmaps are graphical representations of data where values in a matrix are represented as colors. They are useful for visualizing the relationships between multiple variables in a dataset. Heatmaps are commonly used in correlation matrices to show the strength and direction of correlations between all pairs of variables.
5. Cluster Analysis: Cluster analysis is a technique used to group similar observations together based on their characteristics. It helps identify natural clusters or patterns within the data. This can be valuable in segmenting customers, grouping products, or understanding similarities in any dataset.
6. Principal Component Analysis (PCA): PCA is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space while preserving the most important patterns and relationships. It helps identify the main components that explain the most variance in the data.
7. Factor Analysis: Factor analysis is used to identify underlying factors that explain the common variance among a set of observed variables. It helps in understanding the underlying structure of the data and reducing the number of variables to a smaller set of factors.

Advantages of multivariate data analysis:

The following are the advantages of multivariate data analysis:

1. As multivariate data analysis deals with multiple variables, all the variables can either be independent or dependent on each other. This helps the analysis to search for factors that can help in drawing accurate conclusions.
2. Since the analysis is tested, the drawn conclusions are closer to real-life situations.

Disadvantages of multivariate data analysis:

The following are the disadvantages of multivariate data analysis:

1. Multivariate data analysis includes many complex computations and hence can be laborious.
2. The analysis necessitates the collection and tabulation of a large number of observations for various variables. This process of observation takes a long time.