

EDA-Bivariate Analysis

Bivariate analysis is a statistical technique used to explore the relationship between two quantitative variables. It's an essential step in exploratory data analysis (EDA) that helps identify patterns, correlations, and trends in the data.

Goals of Bivariate Analysis

- 1. Understand the relationship:** Determine if there's a significant association between two variables.
- 2. Identify patterns:** Look for clusters, outliers, or anomalies in the data.
- 3. Visualize relationships:** Use plots to illustrate the connection between the two variables.

Types of Bivariate Analysis

- 1. Scatterplots:** A graphical representation of the relationship between two quantitative variables.
 - Example: Analyzing the relationship between a person's height (inches) and weight (pounds).
- 2. Correlation Coefficient:** Measures the strength and direction of the linear relationship between two quantitative variables.
 - Example: Calculating the correlation coefficient between exam scores and study time for a group of students.
- 3. Contingency Tables:** A table showing the frequency distribution of two categorical variables.
 - Example: Analyzing the relationship between job satisfaction and employee turnover.

Example Walkthrough

Suppose we want to analyze the relationship between **coffee consumption (cups)** and **productivity (points)** among a group of students. We collect data on 50 students, including their coffee consumption habits and productivity scores.

Student ID	Coffee Consumption (cups)	Productivity (points)
1	2	80
2	3	90
...

Step 1: Scatterplot

Create a scatterplot to visualize the relationship between coffee consumption and productivity. Observe any patterns, such as clusters or outliers.

- Interpretation: The scatterplot shows a moderate positive correlation between coffee consumption and productivity. As coffee consumption increases, productivity also tends to increase.

Step 2: Correlation Coefficient

Calculate the correlation coefficient (r) to measure the strength and direction of the linear relationship.

Coffee Consumption	Productivity
-1 (cups)	-20 points
0 (cups)	0 points
1 (cups)	20 points

- Interpretation: The correlation coefficient is approximately 0.6, indicating a moderate positive linear relationship between coffee consumption and productivity.

By performing bivariate analysis, we gain insights into the relationship between two variables, which can inform further exploration and decision-making in the context of our study.