



SNS COLLEGE OF ENGINEERING



Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

Course Name – 23ADT202 Fundamental of Data
science and Analytics

II Year / IV Semester

Unit 2 – Descriptive Analytics

Correlation and scatter plots





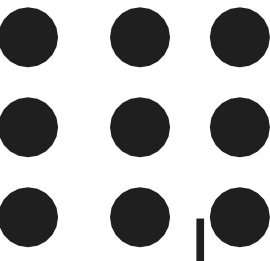
What is Correlation?



Correlation measures the relationship between two variables. It quantifies how changes in one variable are associated with changes in another. A correlation coefficient (r) is the numerical value that represents this relationship, ranging from -1 to 1.



Types of Correlation



1. **Positive Correlation:** As one variable increases, the other increases
2. **Negative Correlation:** As one variable increases, the other decreases.
3. **No Correlation:** No noticeable relationship between the variables.



Types of Correlation



1. **Positive Correlation:** As one variable increases, the other increases
2. **Negative Correlation:** As one variable increases, the other decreases.
3. **No Correlation:** No noticeable relationship between the variables.



Correlation Coefficient (r)



The correlation coefficient (r) is a value between -1 and 1 that indicates the strength and direction of the relationship.

$r = 1$: Perfect positive correlation

$r = -1$: Perfect negative correlation

$r = 0$: No correlation

Calculating Correlation

To calculate the correlation coefficient (r), use the Pearson's formula.

Formula:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

Where X and Y are data points and their means are \bar{X} and \bar{Y} .



Example of Correlation Calculation



Consider the dataset:

$$X = [10, 20, 30, 40]$$

$$Y = [15, 25, 35, 50]$$

Steps:

Compute the means of X and Y.

Apply the Pearson formula to find the correlation coefficient.



Scatter Plots



A scatter plot is a graph using Cartesian coordinates to represent two variables. Each point represents an observation. Scatter plots help visualize the relationship between variables.



Interpreting Scatter Plots



Positive Slope: Indicates a positive correlation.

Negative Slope: Indicates a negative correlation.

No Pattern: Suggests no correlation between variables.



Example: Scatter Plot

Create a scatter plot for the dataset:

$$X = [1, 2, 3, 4, 5], Y = [2, 4, 6, 8, 10]$$

The plot should show a straight line with a positive slope, indicating a perfect positive correlation.



Applications of Correlation

Business: Examine how advertising spend impacts sales.

Healthcare: Study the relationship between medication dosage and recovery time.

Education: Analyze how attendance affects academic performance.



THANK YOU