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

Regression Line & Least square regression line

What is a Regression Line?

• A regression line represents the average relationship between two variables, typically denoted as Y (dependent) and X (independent). It is a straight line that aims to summarize the data points in a scatter plot. This line shows how changes in X affect Y, helping in predictions.

Regression Line Equation

- The regression line can be expressed by the equation:

- $$Y = \beta_0 + \beta_1 X$$

Where:

- β_0 is the intercept (value of Y when X = 0).

- β_1 is the slope (rate of change of Y with respect to X).

The regression line fits the data as closely as possible.

What is Least Squares Regression?

- The Least Squares method finds the best fitting line by minimizing the sum of squared errors between observed and predicted values. This approach ensures that the line best represents the overall trend of the data.

Least Squares Formula

- To calculate the regression line parameters, we use the following formulas:
- β_1 (slope) $= [\Sigma(X_i - \bar{X})(Y_i - \bar{Y})] / \Sigma(X_i - \bar{X})^2$
- **β_0 (intercept)** $= \bar{Y} - \beta_1\bar{X}$

    Where:

- X_i and Y_i are individual data points.
- $\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively.

Steps to Calculate Least Squares Regression Line

- Compute the means of X and Y.
- Calculate the slope (β_1) using the formula.
- Calculate the intercept (β_0) using the formula.
- Write the regression equation: $Y = \beta_0 + \beta_1 X$.

Example 1 – Least Squares Calculation

Given X = [10, 20, 30, 40] and Y = [15, 30, 25, 40]:

- Step 1: Compute means of X and Y.
- Step 2: Calculate the slope β_1.
- Step 3: Calculate the intercept β_0.
- Step 4: Write the regression equation.

Complete calculation is demonstrated in the next slides.

Example 1 – Detailed Calculation

Mean of X: $\bar{X}$ = (10 + 20 + 30 + 40) / 4 = 25

Mean of Y: $\bar{Y}$ = (15 + 30 + 25 + 40) / 4 = 27.5

Slope β_1 = [Σ(Xi - $\bar{X}$)(Yi - $\bar{Y}$)] / Σ(Xi - $\bar{X}$)^2 = 0.625

Intercept β_0 = $\bar{Y}$ - $\beta_1\bar{X}$ = 27.5 - (0.625 * 25) = 11.25

Thus, the regression equation is: **Y = 11.25 + 0.625X**

Example 2 – Plotting the Regression Line

Given X = [1, 2, 3, 4, 5] and Y = [3, 6, 9, 12, 15]:

This is a perfect linear relationship, where the slope is 3 and intercept is 0.

Regression equation: Y = 3X

We will now plot this line along with the data points on a graph.

Applications of Regression Lines

Business: Predict sales based on advertising spend.

Healthcare: Model the relationship between dosage and recovery time.

Education: Forecast student performance based on study hours.

Regression lines are essential tools for decision-making in various industries.

THANK YOU