PMDS504L: Regression Analysis and Predictive Models

Introduction to Simple Linear Regression

Dr. Jisha Francis

Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology
Vellore Campus, Vellore - 632 014
India

Learning Goals

1. Deterministic Relationship
2. Non-Deterministic Relationship
3. Simple Linear Regression Model

Regression

- A statistical technique for modeling and analyzing numerical data.
- Exploits the relationship between two or more variables.
- Helps gain information about one variable through known values of others.

Deterministic vs. Non-Deterministic Relationships

Deterministic Relationships:

- Occurs when one variable's value completely determines the other's value.

Example:

- A cab service charges a fixed base fare of Rs. 450, plus Rs. 15 per kilometer.

- Total fare $y = 450 + 15x$.

- For 10 kilometers, $y = 450 + 15 \times 10 = 600$.

Linear Deterministic Relationship

- The simplest deterministic mathematical relationship is linear: $y = b_0 + b_1 x$.
- b_0: Y-intercept (value of y when $x = 0$).
- b_1: Slope (change in y for a one-unit change in x).
- Forms a straight line on a graph.

Non-Deterministic Relationships

- More complex, where the relationship is not exact but shows a discernible pattern.

- Example: High school GPA X and college GPA Y.

- Predicts trends but not exact values.

Non-Deterministic Relationships: Examples

- Child's age and vocabulary size.

- Engine size and fuel efficiency in automobiles.

- Generalize deterministic linear relationships into a probabilistic model.

Non-Deterministic Relationships: Error Term ε

- Represents random variation or noise.
- Accounts for deviations from the deterministic path.
- $Y = \beta_0 + \beta_1 X + \varepsilon$.

Key Variables

Independent Variable (X):

- Predictor or explanatory variable.

Dependent Variable (Y):

- Outcome or response variable.

The Simple Linear Regression Model

- Relates Y and X through the model equation:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

Where:

- Y: Dependent variable (response).

- X: Independent variable (predictor).

- β_0: Y-intercept.

- β_1: Slope.

- ε: Random error term.

Error Term (ε)

- Randomly distributed with mean 0 and variance σ^2.
- Accounts for deviations from the true regression line.
- Points scatter around the true regression line.

Regression Models

- Simple Linear Regression Model.
- Multiple Linear Regression Model.

Summary

- Bivariate data.

- Deterministic linear relationships.

- Non-deterministic linear relationships.

- Simple linear regression model.

- Example.

Abstract on Next Session

- Simple linear regression model.
- Estimation of regression coefficients.

Thank You!

Thank you for your attention!