PMDS504L: Regression Analysis and Predictive Models

Introduction to Multiple Linear Regression

Dr. Jisha Francis

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

Introduction

- Regression models help describe the relationship between a dependent variable and one or more independent variables.
- A multiple regression model involves more than one regressor variable.
- These models are extensions of simple linear regression.

Introduction

- Regression models help describe the relationship between a dependent variable and one or more independent variables.
- A multiple regression model involves more than one regressor variable.
- These models are extensions of simple linear regression.

Key Idea

A multiple regression model provides a way to predict or explain the response variable using multiple predictors.

Multiple Regression Model

- The general form of a multiple linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$$

- Y: Response variable
- $X_1, X_2, \ldots, X_k$: Predictor variables
- $\beta_0, \beta_1, \ldots, \beta_k$: Regression coefficients
- ϵ: Random error term

Multiple Regression Model

- The general form of a multiple linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$$

- Y: Response variable
- $X_1, X_2, \ldots, X_k$: Predictor variables
- $\beta_0, \beta_1, \ldots, \beta_k$: Regression coefficients
- ϵ: Random error term

Interpretation of Coefficients

β_j represents the expected change in y for a one-unit change in x_j, keeping other predictors constant.

Example: Chemical Process Yield

- Consider a chemical process where yield (Y) depends on:
  - X_1: Temperature
  - X_2: Catalyst concentration
- The model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

Interaction Effects

- Interaction models consider combined effects of predictors:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \epsilon$$

- $X_1 X_2$: Interaction term
- Example model:

$$Y = 50 + 10X_1 + 7X_2 + 5X_1 X_2$$

Second-Order Models with Interaction

- A second-order model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \epsilon$$

- Example:

$$E(Y) = 800 + 10X_1 + 7X_2 - 8.5X_1^2 - 5X_2^2 + 4X_1 x_2$$

Summary

- Multiple linear regression extends simple regression to multiple predictors.
- Interaction terms allow for more complex relationships.
- Second-order models introduce curvature and flexibility.

Summary

- Multiple linear regression extends simple regression to multiple predictors.
- Interaction terms allow for more complex relationships.
- Second-order models introduce curvature and flexibility.

Applications

Used in various fields such as chemistry, economics, biology, and engineering for predictive and explanatory modeling.

Data Table and Model Assumptions

Suppose that $n > k$ observations are available. Let y_i denote the i-th observed response, and x_{ij} denote the i-th observation or level of regressor x_j. The data can be summarized as:

Observation (i)	x_1	x_2	$\cdots$	x_k	Response, y
1	x_{11}	x_{12}	$\cdots$	x_{1k}	y_1
2	x_{21}	x_{22}	$\cdots$	x_{2k}	y_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
n	x_{n1}	x_{n2}	$\cdots$	x_{nk}	y_n

Model Assumptions

Assumptions on the Error Term (ϵ):

- $\mathbb{E}(\epsilon) = 0$
- $\text{Var}(\epsilon) = \sigma^2$
- Errors are uncorrelated.

Model Equation Recap

The regression model is:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i$$

where:

- y_i: Observed response for the i-th observation.
- x_{ij}: Value of the j-th regressor for the i-th observation.
- β_j: Regression coefficients to be estimated.
- ϵ_i: Random error term.

Least-Squares Function

The least-squares function is:

$$S(\beta_0, \beta_1, \ldots, \beta_k) = \sum_{i=1}^{n} \epsilon_i^2$$

Substituting $\epsilon_i = y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij}$:

$$S(\beta) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2$$

- Goal: Minimize $S(\beta)$ to estimate $\beta_0, \beta_1, \ldots, \beta_k$.

Derivation of Normal Equations

To minimize $S(\beta)$, compute partial derivatives:

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)$$

$$\frac{\partial S}{\partial \beta_j} = -2 \sum_{i=1}^{n} x_{ij} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right) \quad (j = 1, 2, \ldots, k)$$

Derivation of Normal Equations

Setting derivatives to zero:

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_{i1} + \cdots + \beta_k \sum_{i=1}^{n} x_{ik}$$

$$\sum_{i=1}^{n} x_{ij} y_i = \beta_0 \sum_{i=1}^{n} x_{ij} + \beta_1 \sum_{i=1}^{n} x_{i1} x_{ij} + \cdots + \beta_k \sum_{i=1}^{n} x_{ij}^2$$

Regression Model in Matrix Form

The regression model can be represented in matrix form as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Explanation of Terms

- $\boldsymbol{y}$: Vector of observed responses ($n \times 1$).
- $\boldsymbol{X}$: Design matrix of regressors ($n \times (k+1)$), including the intercept term.
- $\boldsymbol{\beta}$: Vector of regression coefficients ($(k+1) \times 1$).
- $\boldsymbol{\epsilon}$: Vector of random error terms ($n \times 1$).

Key Assumptions:

- $\mathbb{E}(\boldsymbol{\epsilon}) = 0$.
- $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.
- Errors are uncorrelated.

Least-Squares Objective

The sum of squared residuals is:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \epsilon_i^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

Expanding $S(\boldsymbol{\beta})$:

$$S(\boldsymbol{\beta}) = \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$$

Key Point: $\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y}$ is scalar, so $(\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y})' = \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta}$.

Normal Equations

To minimize $S(\boldsymbol{\beta})$, set the derivative with respect to $\boldsymbol{\beta}$ to zero:

$$\frac{\partial S}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = 0$$

This simplifies to the **normal equations**:

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{y}$$

Solving for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

Condition: $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ exists if $\boldsymbol{X}$ has full column rank (linearly independent regressors).

Matrix Form of Normal Equations

Writing out the normal equations in detail:

$$\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \cdots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \cdots & \sum x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{ik} & \sum x_{i1}x_{ik} & \sum x_{i2}x_{ik} & \cdots & \sum x_{ik}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \vdots \\ \sum x_{ik}y_i \end{bmatrix}.$$

Fitted Values and Residuals

The vector of fitted values is:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\beta = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{H}\boldsymbol{y}$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is the **hat matrix**.

The residuals are:

$$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}.$$

Properties of the Hat Matrix

The hat matrix $\boldsymbol{H}$ has the following properties:

- Symmetric: $\boldsymbol{H}' = \boldsymbol{H}$.
- Idempotent: $\boldsymbol{H}^2 = \boldsymbol{H}$.
- Maps observed values to fitted values: $\hat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$.

References

- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis, Fifth Edition*. Wiley.

Thank You!

Thank you for your attention!