Topic 11: Matrix Approach to Linear Regression

Outline

- **Linear Regression in Matrix Form**

The Model in Scalar Form

- $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
- The ε_i are independent Normally distributed random variables with mean 0 and variance σ^2
- Consider writing the observations:

$$Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1$$
$$Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2$$
$$\vdots$$
$$Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n$$

The Model in Matrix Form

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

The Model in Matrix Form II

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

The Design Matrix

$$\underset{n\times 2}{\mathbf{X}} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & \vdots \\ 1 & X_n \end{pmatrix}$$

Vector of Parameters

$$\underset{2\times 1}{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Vector of error terms

$$\underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Vector of responses

$$\underset{n\times 1}{\mathbf{Y}} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

Simple Linear Regression in Matrix Form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}$$

Variance-Covariance Matrix

$$\sigma^2(\mathbf{Y}) = \begin{bmatrix} \sigma^2(Y_1) & \sigma(Y_1, Y_2) & \cdots & \sigma(Y_1, Y_n) \\ \sigma(Y_2, Y_1) & \sigma^2(Y_2) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \sigma(Y_{n-1}, Y_n) \\ \sigma(Y_n, Y_1) & \cdots & \sigma(Y_n, Y_{n-1}) & \sigma^2(Y_n) \end{bmatrix}$$

Main diagonal values are the variances and off-diagonal values are the covariances.

<u>Covariance Matrix of ε</u>

$$\underset{n\times n}{\sigma^2(\boldsymbol{\varepsilon})} = \mathrm{Cov}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \sigma^2 \underset{n\times n}{\mathbf{I}}$$

Independent errors means that the covariance of any two residuals is zero. Common variance implies the main diagonal values are equal.

Covariance Matrix of Y

$$\underset{n\times n}{\sigma^2(\mathbf{Y})} = \mathrm{Cov}\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \sigma^2 \underset{n\times n}{\mathbf{I}}$$

Distributional Assumptions in Matrix Form

- $\boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I})$
- $\mathbf{I}$ is an n x n identity matrix
- Ones in the diagonal elements specify that the variance of each ε_i is 1 times σ^2
- Zeros in the off-diagonal elements specify that the covariance between different ε_i is zero
- This implies that the correlations are zero

Least Squares

- We want to minimize $(Y-X\beta)'(Y-X\beta)$
- We take the derivative with respect to the (vector) β
- This is like a quadratic function
- Recall the function we minimized using the scalar form

Least Squares

- **The derivative is 2 times the derivative of $(Y\text{-}X\beta)'$ with respect to β**
- **In other words, $-2X'(Y\text{-}X\beta)$**
- **We set this equal to 0 (a vector of zeros)**
- **So, $-2X'(Y\text{-}X\beta) = 0$**
- **Or, $X'Y = X'X\beta$ (the normal equations)**

Normal Equations

- $X'Y = (X'X)\beta$
- Solving for β gives the least squares solution $b = (b_0, b_1)'$
- $b = (X'X)^{-1}(X'Y)$
- See KNNL p 199 for details
- This same matrix approach works for multiple regression!!!!!!

Fitted Values

$$\hat{\mathbf{Y}} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \begin{pmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \vdots \\ b_0 + b_1 X_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \mathbf{X}\boldsymbol{b}$$

Hat Matrix

$$\hat{\mathbf{Y}} = \mathbf{X}\boldsymbol{b}$$

$$= \underbrace{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{}\mathbf{Y}$$

$$= \mathbf{H}\mathbf{Y}$$

We'll use this H matrix when assessing diagnostics in multiple regression

Estimated Covariance Matrix of b

- This matrix, b, is a linear combination of the elements of Y
- These estimates are therefore Normal if Y is Normal
- These estimates will be approximately Normal in general

A Useful MultivariateTheorem

- U ~ N(μ, Σ), a multivariate Normal vector
- V = c + DU, a linear transformation of U

  c is a vector and D is a matrix
- Then V ~ N(c+Dμ, DΣD′)

Application of theorem

- $b = (X'X)^{-1}X'Y = ((X'X)^{-1}X')Y$
- Since $Y \sim N(X\beta, \sigma^2 I)$ this means the vector b is Normally distributed with mean $(X'X)^{-1}X'X\beta = \beta$ and covariance $\sigma^2 ((X'X)^{-1}X')\, I((X'X)^{-1}X')' = \sigma^2 (X'X)^{-1}$

Background Reading

- We will use this framework to do multiple regression → we have more than one explanatory variable
- Another explanatory variable is comparable to adding another column in the design matrix
- See Chapter 6