Matrix Approach to Simple Linear Regression: An Overview

- Aspects of matrices that you should know:
  - Definition of a matrix
  - Addition/subtraction/multiplication of matrices
  - Symmetric/diagonal/identity matrix
  - Transpose/rank/inverse of a matrix
  - Determinants
- Random vectors and matrices
- Simple linear regression model in matrix terms
- Least squares estimation of regression parameters
- Fitted values and residuals
- Inferences in regression analysis
- ANOVA results

Random Vector and Matrix

- A random vector or a random matrix contains elements that are random variables.
- Simple linear regression: The response variables $Y_1, \ldots, Y_n$ can be written in the form of a random vector

$$\boldsymbol{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

- Alternative notation: $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ or $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$.

Expectation of Random Vector/Matrix

- The expectation of an $n \times 1$ random vector $\boldsymbol{Y}$ is

$$\mathbb{E}(\boldsymbol{Y})_{n\times 1} = [\mathbb{E}(Y_i) : i = 1, \ldots, n] = \begin{bmatrix} \mathbb{E}(Y_1) \\ \vdots \\ \mathbb{E}(Y_n) \end{bmatrix}$$

- Since $E(Y_i) = \beta_0 + \beta_1 X_i$ for $i = 1, \ldots, n$,

$$\mathbb{E}(\boldsymbol{Y}) = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix}$$

- In general, the expectation of an $n \times p$ random matrix $\boldsymbol{Y}$ is

$$\mathbb{E}(\boldsymbol{Y})_{n\times p} = [\mathbb{E}(Y_{ij}) : i = 1, \ldots, n; j = 1, \ldots, p].$$

Variance-Covariance Matrix of Random Vector

- The variance-covariance of an $n \times 1$ random vector $\boldsymbol{Y}$ is $\boldsymbol{\Sigma}$ (or $\sigma^2\{\boldsymbol{Y}\}$)

$$
\begin{aligned}
\boldsymbol{\Sigma} &= Var(\boldsymbol{Y}) = \mathbb{E}\left[(\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y}))(\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y}))^T\right] \\
&= \mathbb{E}\begin{bmatrix} (Y_1 - \mathbb{E}(Y_1))^2 & \cdots & \cdots & (Y_1 - \mathbb{E}(Y_1))(Y_n - \mathbb{E}(Y_n)) \\ \cdots & \cdots & \cdots & (Y_2 - \mathbb{E}(Y_2))(Y_n - \mathbb{E}(Y_n)) \\ \vdots & \vdots & & \vdots \\ \cdots & \cdots & \cdots & (Y_n - \mathbb{E}(Y_n))^2 \end{bmatrix} \\
&= \begin{bmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_1, Y_n) \\ \cdots & Var(Y_2) & \cdots & Cov(Y_2, Y_n) \\ \vdots & \vdots & & \vdots \\ \cdots & \cdots & \cdots & Var(Y_n) \end{bmatrix}
\end{aligned}
$$

- Since $Cov(Y_i, Y_j) = Cov(Y_j, Y_i)$, $\boldsymbol{\Sigma}$ is symmetric.

Variance-Covariance Matrix of Random Vector

- The independent random errors $\epsilon_1, \ldots, \epsilon_n$ can be written in the form of a random vector

$$\boldsymbol{\epsilon}_{n\times 1} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- The expectation of $\boldsymbol{\epsilon}$ is

$$\mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}_{n\times 1}.$$

- The variance-covariance matrix of $\boldsymbol{\epsilon}$ is

$$\sigma^2\{\boldsymbol{\epsilon}\} = \boldsymbol{\Sigma}_\epsilon =$$

Some Basic Results

- For $\boldsymbol{Y}$ ($n \times 1$ random vector), $\boldsymbol{A}$ ($n \times n$ non-random matrix), and $\boldsymbol{b}$ ($n \times 1$ non-random vector), we have

$$\begin{aligned} \mathbb{E}(\boldsymbol{AY} + \boldsymbol{b}) &= \boldsymbol{A}\mathbb{E}(\boldsymbol{Y}) + \boldsymbol{b} \\ \text{Var}(\boldsymbol{AY} + \boldsymbol{b}) &= \boldsymbol{A}\text{Var}(\boldsymbol{Y})\boldsymbol{A}^T \end{aligned}$$

- Also for $\boldsymbol{a}_{n\times 1} = (a_1, \ldots, a_n)^T$ ($n \times 1$ non-random vector),

$$\boldsymbol{a}^T\boldsymbol{a} = \sum_{i=1}^{n} a_i^2 := ||\boldsymbol{a}||_2^2$$

and for $\boldsymbol{J} = [1]_{n\times n}$ ($n \times n$ matrix of 1's),

$$\boldsymbol{a}^T\boldsymbol{J}\boldsymbol{a} =$$

Multivariate Normal Distribution

- Let $\boldsymbol{Y}_{m\times 1} = (Y_1, \ldots, Y_m)^T$ follow a multivariate normal distribution with mean

$$\boldsymbol{\mu}_{m\times 1} = (\mu_1, \ldots, \mu_m)^T$$

and variance

$$\boldsymbol{\Sigma}_{m\times m} = (\sigma_{ij}^2)_{i=1,\ldots,m;j=1,\ldots,m}.$$

- We denote this by

$$\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

- The probability density function is

$$f(\boldsymbol{Y}) = \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu})\right],$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$.

Using the Notation with Simple Linear Regression

- Let $\boldsymbol{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ denote the $n \times 1$ vector of response variables.
- Let $\boldsymbol{X}_{n\times 2} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$ denote the $n \times 2$ design matrix of predictor variables (2 columns for simple LR)
- Let $\boldsymbol{\epsilon}_{n\times 1} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$ denote the $n \times 1$ vector of random errors.
- Let $\boldsymbol{\beta}_{2\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ denote the 2×1 vector of regression coefficients.

Simple Linear Regression in Matrix Terms

- The simple linear regression model in matrix terms is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}).$$

- Recall that $E(\boldsymbol{\epsilon}) = \mathbf{0}_{n\times 1}$ and $\Sigma_\epsilon = \sigma^2 \boldsymbol{I}_{n\times n}$.
- Thus, we have

$$\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}).$$

since

$$\begin{aligned}
\mathbb{E}(\boldsymbol{Y}) &= \mathbb{E}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = \boldsymbol{X}\boldsymbol{\beta} + \mathbb{E}(\boldsymbol{\epsilon}) = \boldsymbol{X}\boldsymbol{\beta} \\
\boldsymbol{\Sigma} &= \text{Var}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = \text{Var}(\boldsymbol{\epsilon}) = \sigma^2\{\boldsymbol{\epsilon}\} = \sigma^2 \boldsymbol{I}.
\end{aligned}$$

Least Squares Estimation

- Let $\hat{\boldsymbol{\beta}}_{2\times 1} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$ denote the least squares estimate of $\boldsymbol{\beta}$.
- As we will show,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}.$$

- Recall that the least squares method minimizes

$$Q = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$$

- In matrix terms,

Normal Equations

- Let
$$\frac{\partial Q}{\partial \boldsymbol{\beta}}_{2\times 1} = \left(\frac{\partial Q}{\partial \beta_0}, \frac{\partial Q}{\partial \beta_1} \right)'.$$

- Differentiate Q with respect to $\boldsymbol{\beta}$ to obtain:

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} =$$

- Set the equation above to $\mathbf{0}_{2\times 1}$ and obtain a set of normal equations:

- Thus the least squares estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} =$$

assuming that the 2×2 matrix $\boldsymbol{X}^T\boldsymbol{X}$ is nonsingular and thus invertible.

Fitted Values

- Let $\hat{\boldsymbol{Y}}_{n\times 1} = \begin{bmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$ denote the $n \times 1$ vector of fitted vales.
- Since $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$, we have

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}.$$

- Note that

$$\begin{bmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 + \hat{\beta}_1 X_1 \\ \vdots \\ \hat{\beta}_0 + \hat{\beta}_1 X_n \end{bmatrix}$$

Hat Matrix

- Rewrite the vector of fitted values as

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} := \boldsymbol{H}\boldsymbol{Y}$$

  where the $n \times n$ hat matrix is defined as

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T.$$

- The hat matrix $\boldsymbol{H}$ is symmetric and idempotent, as

$$\boldsymbol{H}^T =$$

$$\boldsymbol{H}\boldsymbol{H} =$$

Residuals

- Let $\boldsymbol{e}_{n\times 1} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$ denote the $n \times 1$ vector of residuals.
- Since $e_i = Y_i - \hat{Y}_i$, we have

  $$\boldsymbol{e} =$$

- $(\boldsymbol{I} - \boldsymbol{H})$ is symmetric and idempotent.

  Proof:

Variance-Covariance Matrix of Residuals

- The variance-covariance matrix of residuals is

- The estimated variance-covariance matrix of residuals is

Fitted Values

- The variance-covariance matrix of fitted values is

- The estimate of $\sigma^2\{\hat{\boldsymbol{Y}}\}$ is

Regression Coefficients

- The variance-covariance matrix of least squares estimates is

- The estimate of $\sigma^2\{\hat{\boldsymbol{\beta}}\}$ is

Estimation of Mean Response and Prediction of New Observation

- Define $\boldsymbol{X}_h = [\ 1 \quad X_h\ ]$.
- NOTE!! Book uses column vector for $\boldsymbol{X}_h$, not row. So many transposes are opposite from book's presentation!
- The variance of $\hat{Y}_h$ for estimating $E(Y_h)$ is
- The estimate of $\sigma^2\{\hat{Y}_h\}$ is
- Similarly, the variance of $\hat{Y}_h$ for predicting $Y_{h(\text{new})}$ is
- The estimate of $\sigma^2\{\text{pred}\}$ is

Sums of Squares

- Total sum of squares:

$$SSTO = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 =$$

- Error sum of squares:

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \boldsymbol{e}^T\boldsymbol{e} =$$

- Regression sum of squares:

$$SSR = SSTO - SSE \quad =$$

Sums of Squares as Quadratic Forms

- For a non-random symmetric $n \times n$ matrix $\boldsymbol{A}$, a quadratic form is defined as

$$Q(\boldsymbol{Y}) = \boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} Y_i Y_j, \quad a_{ij} = a_{ji}.$$

- Total sum of squares:

  $SSTO =$

- Error sum of squares:

  $SSE =$

- Regression sum of squares:

  $SSR =$