1 Matrix Algebra Approach to Regression

1.1 The Linear Regression Model in Matrix Form

- Suppose we have data

Observation	Predictor Variable X	Response Variable Y
1	X_1	Y_1
2	X_2	Y_2
$\vdots$	$\vdots$	$\vdots$
n	X_n	Y_n

- Recall that the simple linear regression model is

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

for $i = 1, 2, \ldots, n$.

We could write this out as

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \epsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \epsilon_n \end{aligned} \tag{1}$$

- Define the $n \times 1$ ***response vector*** $\mathbf{Y}$ to be

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \tag{2}$$

and the $n \times 1$ ***error vector*** $\boldsymbol{\epsilon}$ to be

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{3}$$

Also, define the 2×1 ***parameter vector*** $\boldsymbol{\beta}$ to be

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \tag{4}$$

and the $n \times 2$ ***design matrix*** $\mathbf{X}$ to be

$$\mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \tag{5}$$

Simple Linear Regression Model (Matrix Approach): The regression model (1) can be written in terms of the vectors and matrix (2), (3), (4) and (5) as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

1.2 Least Squares Estimation in Matrix Form

- Recall that the **least squares estimators** $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$ of β_0 and β_1 are values that minimize

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \tag{6}$$

  as a function of β_0 and β_1.

- Notice that

$$\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} Y_1 - \beta_0 - \beta_1 X_1 \\ Y_2 - \beta_0 - \beta_1 X_2 \\ \vdots \\ Y_n - \beta_0 - \beta_1 X_n \end{bmatrix}$$

  and that Q can be written as

$$\begin{aligned} Q(\beta_0, \beta_1) &= \begin{bmatrix} Y_1 - \beta_0 - \beta_1 X_1, & Y_2 - \beta_0 - \beta_1 X_2, & \cdots, & Y_n - \beta_0 - \beta_1 X_n \end{bmatrix} \begin{bmatrix} Y_1 - \beta_0 - \beta_1 X_1 \\ Y_2 - \beta_0 - \beta_1 X_2 \\ \vdots \\ Y_n - \beta_0 - \beta_1 X_n \end{bmatrix} \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

  where the superscript $\boldsymbol{T}$ denotes ***transpose***.

- The estimates b_0 and b_1 are obtained by setting the partial derivatives of Q (with respect to β_0 and β_1) equal to zero and solving the resulting system of equations:

$$\frac{\partial Q}{\partial \beta_0} = 0 \qquad \text{and} \qquad \frac{\partial Q}{\partial \beta_1} = 0\,,$$

which, it can be shown, are the so-called ***normal equations***:

$$nb_0 + b_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \tag{7}$$

$$b_0 \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i \tag{8}$$

- If we define the 2×1 ***vector of estimated coefficients*** $\boldsymbol{b}$ to be

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

then the **normal equations** can be written as

$$\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

- Noting that

$$\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \\ 1 & X_n \end{bmatrix}$$

$$= \mathbf{X}^T\mathbf{X}\,,$$

we see that the normal equations (7) and (8) can be written as

$$\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{Y} \tag{9}$$

- When the 2×2 matrix $\mathbf{X}^T\mathbf{X}$ is **nonsingular** (i.e. when it's columns are linearly independent), which is usually the case, it's **invertible**. Multiplying both sides of (9) by the inverse $(\mathbf{X}^T\mathbf{X})^{-1}$ of $\mathbf{X}^T\mathbf{X}$ gives the vector of coefficient estimates $\mathbf{b}$:

Least Squares Estimates of β_0 and β_1 (Matrix Approach): The vector of estimated coefficients $\mathbf{b}$ is obtained by:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{10}$$

Although tedious, it can be shown that the right side of (10) is

$$\boldsymbol{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} \bar{Y} - \frac{\sum(Y_i-\bar{Y})(X_i-\bar{X})}{\sum(X_i-\bar{X})^2}\bar{X} \\ \frac{\sum(Y_i-\bar{Y})(X_i-\bar{X})}{\sum(X_i-\bar{X})^2} \end{bmatrix}$$

and so the least squares estimates, obtained using the matrix approach, are

$$b_1 = \frac{\sum(Y_i-\bar{Y})(X_i-\bar{X})}{\sum(X_i-\bar{X})^2} \qquad \text{and} \qquad b_0 = \bar{Y} - b_1\bar{X}\,,$$

the same as were given in Class Notes 1.

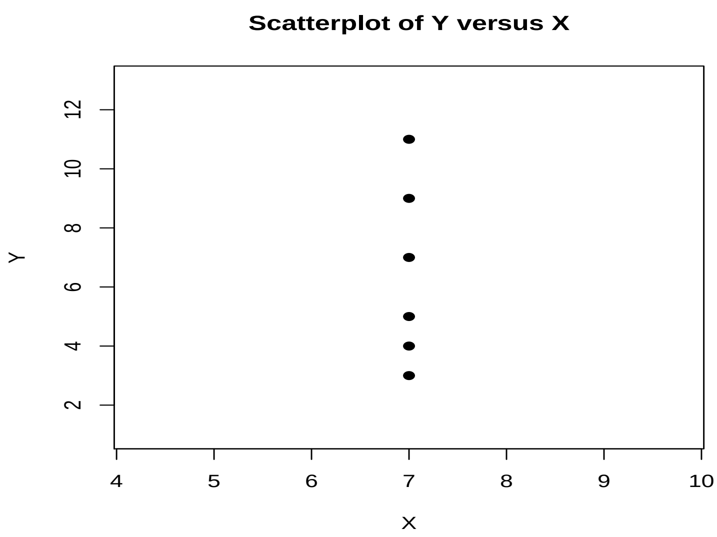
- **Comment**: $\mathbf{X}^T\mathbf{X}$ will be **singular**, and therefore **non-invertible**, if $X_1 = X_2 = \cdots = X_n$ (verify this for yourself!).

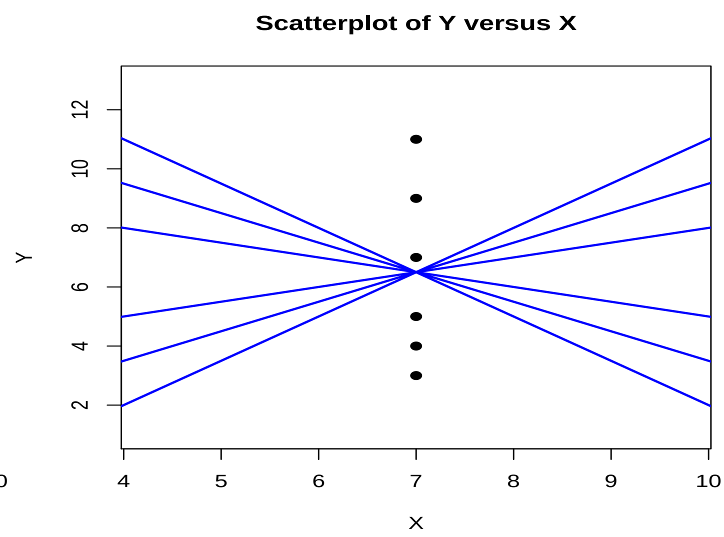
  In this case, there will be **infinitely many solutions** to the normal equations, so $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$ **won't be uniquely determined**.

  For example, consider this data set:

Data

X	Y
7	3
7	5
7	4
7	7
7	9
7	11





1.3 Residuals and Fitted Values in Matrix Form

- Recall that the fitted values are

$$\begin{aligned} \hat{Y}_1 &= b_0 + b_1 X_1 \\ \hat{Y}_2 &= b_0 + b_1 X_2 \\ &\vdots \\ \hat{Y}_n &= b_0 + b_1 X_n \end{aligned}$$

- Define the $n \times 1$ ***vector of fitted values*** $\hat{\mathbf{Y}}$ to be

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} \tag{11}$$

Then the fitted values can be expressed in terms of the design matrix $\mathbf{X}$ and the vector of estimated coefficients $\mathbf{b}$ as:

Fitted Values (Matrix Approach):

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b} \tag{12}$$

- From (10), we can write (12) as

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{H}\mathbf{Y}\,.$$

  where the matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is called the ***hat matrix*** because it "puts a hat" on $\mathbf{Y}$.

- Recall that the **residuals** are

$$\begin{aligned} e_1 &= Y_1 - \hat{Y}_1 \\ e_2 &= Y_2 - \hat{Y}_2 \\ &\vdots \\ e_n &= Y_n - \hat{Y}_n \end{aligned}$$

- Define the $n \times 1$ ***residual vector*** $\mathbf{e}$ to be

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_n - \hat{Y}_n \end{bmatrix} \tag{13}$$

  Then the residuals can be expressed in terms of the response vector $\mathbf{Y}$ and the vector of fitted values $\hat{\mathbf{Y}}$ as:

  Residuals (Matrix Approach):

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\mathbf{b} \tag{14}$$

1.4 Sums of Squares in Matrix Form

- Recall that the error sum of squares is

$$\text{SSE} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

  We can write SSE in terms of the residual vector $\mathbf{e}$ as

  Error Sum of Squares (Matrix Approach):

$$\text{SSE} \;=\; \mathbf{e}^T\mathbf{e} \;=\; (\mathbf{Y}-\hat{\mathbf{Y}})^T(\mathbf{Y}-\hat{\mathbf{Y}}) \;=\; (\mathbf{Y}-\mathbf{X}\mathbf{b})^T(\mathbf{Y}-\mathbf{X}\mathbf{b})$$

- The other sums of squares SSTO and SSR can also be written in matrix form. See the textbook.