Matrix Approach to Simple Linear Regression Analysis

Least squared estimation

The simple linear regression model and be formed in the matrix form as

$$Y = X\beta + \varepsilon, \quad E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2 I,$$

where

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \underset{2\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Now, we apply the least squared method t
o find the estimation of the vector β as follows:

$$Q = \sum_{i=1}^{n} e_i^2 = e'e = (Y - X\beta)'(Y - X\beta)$$

Expanding, we obtain:

$$Q = \mathbf{Y'Y} - \boldsymbol{\beta}'\mathbf{X'Y} - \mathbf{Y'X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X'X}\boldsymbol{\beta}$$

since $(\mathbf{X}\boldsymbol{\beta})' = \boldsymbol{\beta}'\mathbf{X}'$ Note now that $\mathbf{Y'X}\boldsymbol{\beta}$ is 1×1, hence is equal to its transpose, is $\boldsymbol{\beta}'\mathbf{X'Y}$. Thus, we find:

$$Q = \mathbf{Y'Y} - 2\boldsymbol{\beta}'\mathbf{X'Y} + \boldsymbol{\beta}'\mathbf{X'X}\boldsymbol{\beta}$$

To find the value of $\boldsymbol{\beta}$ that minimizes Q, we differentiate with respect to β_0 and β_1. Let:

$$\frac{\partial}{\partial \boldsymbol{\beta}}(Q) = \begin{bmatrix} \dfrac{\partial Q}{\partial \beta_0} \\ \dfrac{\partial Q}{\partial \beta_1} \end{bmatrix}$$

Then it follows that:

$$\frac{\partial}{\partial \boldsymbol{\beta}}(Q) = -2\mathbf{X'Y} + 2\mathbf{X'X}\boldsymbol{\beta}$$

Equating to the zero vector, we get

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \widehat{\beta} = (X'X)^{-1}X'Y$$

where

$$\underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}} = \begin{bmatrix} \frac{\sum X_i^2}{n\sum(X_i - \bar{X})^2} & \frac{-\sum X_i}{n\sum(X_i - \bar{X})^2} \\ \frac{-\sum X_i}{n\sum(X_i - \bar{X})^2} & \frac{n}{n\sum(X_i - \bar{X})^2} \end{bmatrix}$$

$$\underset{2\times 1}{\mathbf{X}'\mathbf{Y}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Example

We shall use matrix methods to obtain the estimated regression coefficients for the Toluca Company example. The data On the Y and X Variables. Using these data, we define the Y observations vector and the X matrix as follows:

$$\mathbf{Y} = \begin{bmatrix} 399 \\ 121 \\ \vdots \\ 323 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 80 \\ 1 & 30 \\ \vdots & \vdots \\ 1 & 70 \end{bmatrix}$$

We now require the following matrix products:

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 80 & 30 & \cdots & 70 \end{bmatrix} \begin{bmatrix} 1 & 80 \\ 1 & 30 \\ \vdots & \vdots \\ 1 & 70 \end{bmatrix} = \begin{bmatrix} 25 & 1{,}750 \\ 1{,}750 & 142{,}300 \end{bmatrix}$$

we find the inverse of $\mathbf{X'X}$:

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix}$$

In subsequent matrix calculations utilizing this inverse matrix and other matrix results, we shall actually utilize more digits for the matrix elements than are shown.

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (\mathbf{X'X})^{-1}\mathbf{X'Y} = \begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix} \begin{bmatrix} 7{,}807 \\ 617{,}180 \end{bmatrix}$$

$$= \begin{bmatrix} 62.37 \\ 3.5702 \end{bmatrix}$$

or $b_0 = 62.37$ and $b_1 = 3.5702$. These results agree with the ones in Chapter 1. Any differences would have been due to rounding effects.

One can use R code direct to get the estimates as:

```
mat <- scan('a.txt')
mat <- matrix(mat, ncol = 2, byrow = TRUE)
mat[,1]
mat[,2]
length(mat[,1])
one=as.vector(rep(1, 25))
x=cbind(one,mat[,1])
y=mat[,2]
b=solve(t(x)%*%x)%*%t(x)%*%y
b
```