Regression using Matrices

We can now present linear regression in matrix terms. Begin with the linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad \text{where } i = 1, \ldots, n$$

This implies

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

We can now arrange these into appropriate vectors and matrices.

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \qquad \underset{2\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \underset{n\times 1}{\boldsymbol{\epsilon}} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

We can write the model using matrix terms very compactly as

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

Dimension
of product

Next we turn to the assumptions of the errors. Since $\mathbf{E}\{\varepsilon_i\} = 0$,

$$\underset{n\times 1}{\mathbf{E}\{\boldsymbol{\epsilon}\}} = \underset{n\times 1}{\mathbf{0}}$$

We also make the assumption that the errors have constant variance σ^2 and that they are independent of one another $\sigma(\varepsilon_i, \varepsilon_j)=0$ for $i \neq j$. The **variance-covariance matrix of the error terms** can therefore be expressed as

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\epsilon}\}} = \begin{bmatrix} \sigma_\epsilon^2 & 0 & \cdots & 0 \\ 0 & \sigma_\epsilon^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_\epsilon^2 \end{bmatrix}$$

This scalar matrix can also be expressed as

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\epsilon}\}} = \sigma^2 \underset{n\times n}{\mathbf{I}}$$

Thus, the **regression model** can be completely expressed in matrix terms as

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

where $\boldsymbol{\varepsilon}$ is a vector of independent random variables with $\mathbf{E}\{\boldsymbol{\varepsilon}\}=0$ and $\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}= \sigma^2\mathbf{I}$

Solving for β

We can re-express the **error vector** as

$$\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$$

The *method of least squares* minimizes the **sum of squared errors** which is expressed as

$$\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

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

Note that each term is a 1x1 matrix, which means that each term is equal to its transpose. We will re-write the third term **Y'X β** as its transpose **β'X'Y**

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

Combining the two middle terms we get

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

To find the values for the elements in $\boldsymbol{\beta}$ that minimize the equation, we differentiate with respect to β_0 and β_1 ($\boldsymbol{\beta}$)

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\underset{1\times n}{\mathbf{Y}'}\underset{n\times 1}{\mathbf{Y}} - 2\underset{1\times 2}{\boldsymbol{\beta}'}\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} + \underset{1\times 2}{\boldsymbol{\beta}'}\underset{2\times n}{\mathbf{X}'}\underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}})$$

Differentiating, we get...

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

Solving this equation for zero...

$$\begin{aligned}\underset{2\times 1}{\mathbf{0}} &= -2\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} + 2\underset{2\times n}{\mathbf{X}'}\underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}} \\ &= -\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} + \underset{2\times n}{\mathbf{X}'}\underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}}\end{aligned}$$

Adding **X'Y** to both sides

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

Now we pre-multiply both sides of the equation by $(\mathbf{X'X})^{-1}$

$$\underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} = \underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\underset{2\times n}{\mathbf{X}'}\underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}}$$

$$\underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} = \underset{2\times 2}{\mathbf{I}}\underset{2\times 1}{\boldsymbol{\beta}}$$

This means

$$\underset{2\times 1}{\boldsymbol{\beta}} = \underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}}$$

The vector of regression coefficients can be obtained directly through manipulation of the design matrix and the vector of outcomes.

$$\underset{2\times 1}{\boldsymbol{\beta}} = \underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}} \underset{2\times n}{\mathbf{X}'} \underset{n\times 1}{\mathbf{Y}}$$

```
> lm.1 = lm(wage ~ 1 + age, data = myData)

> Y = myData$wage
> X = matrix(c(rep(1, 5), myData$age), ncol = 2)

> b = solve(t(X) %*% X) %*% t(X) %*% Y
> b

                  [,1]
(Intercept) -3.8982759
age          0.4993678

> summary(lm.1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.8983    19.3658  -0.201    0.853
age           0.4994     0.6185   0.807    0.479
```

$$\mathbf{b} = \begin{bmatrix} -3.8983 \\ 0.4994 \end{bmatrix}$$

When we estimate the $\boldsymbol{\beta}$ matrix from sample data we refer to it as $\mathbf{b}$

```
> summary(lm.1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.8983    19.3658  -0.201    0.853
age           0.4994     0.6185   0.807    0.479
```

The SEs for the coefficients represent the uncertainty in the estimates.

The uncertainty is also correlated between the coefficients (unless you have a balanced design.)

We represent the uncertainty and the correlation in the estimated variance–covariance matrix of the coefficients,

$$\mathbf{V}_\beta \hat{\sigma}^2$$

where

$$\mathbf{V}_\beta = (\mathbf{X}^t\mathbf{X})^{-1}$$

and

$$\hat{\sigma}^2 = \mathrm{MSE}$$

```
> solve(t(X) %*% X)

            (Intercept)         age
(Intercept)  28.1724138 -0.89655172
age          -0.8965517  0.02873563
```

$$\mathbf{V}_{\beta} = \begin{bmatrix} 28.17 & -0.90 \\ -0.90 & 0.03 \end{bmatrix}$$

$$\hat{\sigma}^2 = 13.31$$

$$\begin{bmatrix} 28.17 & -0.90 \\ -0.90 & 0.03 \end{bmatrix} (13.31) = \boxed{\begin{bmatrix} 375.0 & -11.9 \\ -11.9 & 0.04 \end{bmatrix}}$$

The diagonal elements are the *variances* of the regression coefficients and the off-diagonal elements are the *covariances*.

```
> vcov(lm.1)

            (Intercept)         age
(Intercept)   375.03555 -11.9350358
age           -11.93504   0.3825332
```

We can also compute the estimated SE for each coefficient and the correlation between the coefficients.

$$\mathbf{V}_{\beta} = \begin{bmatrix} 28.17 & -0.90 \\ -0.90 & 0.03 \end{bmatrix}$$

$$\mathrm{SE}_{\beta_0} = (\sqrt{\mathbf{V}_{11}})(\hat{\sigma}) \qquad \mathrm{SE}_{\beta_1} = (\sqrt{\mathbf{V}_{22}})(\hat{\sigma}) \qquad r_{\beta_1,\beta_2} = \frac{\mathbf{V}_{12}}{\sqrt{\mathbf{V}_{11}\mathbf{V}_{22}}}$$

$$\mathrm{SE}_{\beta_0} = (\sqrt{28.17})(3.65) = 19.37$$

$$\mathrm{SE}_{\hat{\beta}_1} = (\sqrt{0.03})(3.65) = 0.62$$

$$r_{\hat{\beta}_1,\hat{\beta}_2} = \frac{-0.90}{\sqrt{(28.17)(0.03)}} = -0.996$$