Political Science 239 - Hints for Problem Set 2

Rocio Titiunik

Fall 2006

1 OLS in Matrix Form

The multiple linear regression model for unit i can be written as:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i$$

and since we have n observations, we have the system:

$$\begin{aligned} y_1 &= \beta_1 x_{11} + \beta_2 x_{12} + ... + \beta_k x_{1k} + \varepsilon_1 \\ y_i &= \beta_1 x_{21} + \beta_2 x_{22} + ... + \beta_k x_{2k} + \varepsilon_2 \\ & ... \\ y_i &= \beta_1 x_{n1} + \beta_2 x_{n2} + ... + \beta_k x_{nk} + \varepsilon_i \end{aligned}$$

This system can be expressed in matrix form as

$$\begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1k} \\ x_{21} & x_{22} & ... & x_{2k} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & ... & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ ... \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ ... \\ \varepsilon_k \end{pmatrix}$$

where $\mathbf{y}$ is $(n \times 1)$, X is $(n \times k)$, $\boldsymbol{\beta}$ is $(k \times 1)$ and $\boldsymbol{\varepsilon}$ is $(n \times 1)$.

So the multiple regression model can be written as

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

In this notation, the sum of squared residuals is

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \sum_{i=1}^{n} \varepsilon_i^2$$

and the vector $\widehat{\boldsymbol{\beta}}$ that minimizes $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ is

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \arg\min \left\{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right\} \\
&= \arg\min \left\{(\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta})\right\} \\
&= \arg\min \left\{\mathbf{y}'\mathbf{y} - \mathbf{2}X^T\mathbf{y}\boldsymbol{\beta} + \boldsymbol{\beta}X^T X\boldsymbol{\beta}\right\}
\end{aligned}$$

The first order conditions are:

$$-2X^T\mathbf{y} + X^{\mathbf{T}}X\widehat{\boldsymbol{\beta}} = \mathbf{0}$$

and if the matrix $X^T X$ is invertible, we have

$$\widehat{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T\mathbf{y}$$

Also, remember that

$$Var\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\sigma}^2 \left(X^T X\right)^{-1}$$

where $\boldsymbol{\sigma}^2$ is the variance of the residuals. This parameter $\boldsymbol{\sigma}^2$ is unknown, so we can estimate it using

$$\widehat{\boldsymbol{\sigma}^2} = \frac{1}{n-k}\sum_{i=1}^{n} \widehat{\varepsilon}_i^2 = \frac{\widehat{\boldsymbol{\varepsilon}}^T\widehat{\boldsymbol{\varepsilon}}}{n-k}$$

where n is the number of observations (rows in X) and k is the number of covariates (columns in X).

So when you write your function in R, just define your covariate matrix X and your outcome vector $\mathbf{y}$, and perform the matrix multiplication $\left(X^T X\right)^{-1} X^T \mathbf{y}$ to get your estimated betas. Then, use these estimated betas to obtain the residuals by doing

$$\widehat{\boldsymbol{\varepsilon}}= \mathbf{y}-X\widehat{\boldsymbol{\beta}}$$

and then obtain the estimated $\widehat{\boldsymbol{\sigma}^2}$ by doing

$$\widehat{\boldsymbol{\sigma}^2}=\frac{1}{n-k}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2 = \frac{\widehat{\boldsymbol{\varepsilon}}^T\widehat{\boldsymbol{\varepsilon}}}{n-k}$$

Calculate the variance-covariance matrix of the betas by doing

$$Var\left(\widehat{\boldsymbol{\beta}}\right)=\frac{\widehat{\boldsymbol{\varepsilon}}^T\widehat{\boldsymbol{\varepsilon}}}{n-k}\left(X^T X\right)^{-1}$$

and then observe that the variance of each beta is in the diagonal of this matrix. If you want the standard errors, just take the square root of elements of the diagonal. In R, if you have a matrix A and you want to take the elements in the diagonal of A you do "diag(A)". And if you want the square root of these elements you do "sqrt(diag(A))".