Multiple Linear Regression and Least Squares estimators

Jhonatan Felix

1 Remembering the Linear Model

The linear regression model classic can be wrote in the following way:

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p} + E_i$$

where:

- Y_i is the random variable that models the ith observation
- $x_{i,k}$ is the value of the kth variable for the ith observation
- E_i is the error with a distribution i.i.d. $\mathcal{N}(0, \sigma^2)$

We can write under the matricial form like the following:

$$Y = X\beta + E \tag{1}$$

where:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ 1 & x_{2,1} & \cdots & x_{2,p} \\ \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}, E = \begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{pmatrix}$$

With these notations, **Y** and **E** are column vector of size $n \times 1$, **X** is a matrix with n lines and $p+1$ columns, and β is a column vector having $p+1$ lines.

2 Estimation of β_i

The parameters β_i can be estimated using the maximum likelihood method, minimizing the criterion of the sum of the squared vertical deviations, *i.e.*

$$(\hat{\beta}_0, ..., \hat{\beta}_p) = \underset{(\beta_0, ..., \beta_p)}{Argmin} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_{i,1} - ... - \beta_p x_{i,p})^2$$

And this criterion can be rewrote in the following sentence:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^{n+1}}{Argmin} ||Y - X\beta||^2 = \underset{\beta \in \mathbb{R}^{n+1}}{Argmin} (Y - X\beta)'(Y - X\beta) \tag{2}$$

Proposition 1. *The estimator of β in (1) defined by (2) satifies:*

$$(\boldsymbol{X'X})\hat{\beta} = \boldsymbol{X'Y}$$

and under the hypothesis that $\boldsymbol{X'X} isinvertible$ we have

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

Furthermore, the covariance matrix of $\hat{\beta}$ defined by $Cov(\hat{\beta}) = \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$ is such that:

$$Cov(\hat{\beta}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

Proof. Note that:

$$\begin{aligned} (Y - X\beta)'(Y - X\beta) &= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta \\ &= Y'Y - 2\beta'X'Y + \beta'X'X\beta \end{aligned}$$

Now, we can calculate:

□