Matrix reexpression for simple linear regression:

Consider our simple regression model:

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

This could be written out as:

$$Y_1 = \beta_0 + \beta_1 X_1 + \epsilon_1$$

$$Y_2 = \beta_0 + \beta_1 X_2 + \epsilon_2$$

$$\vdots$$

$$Y_n = \beta_0 + \beta_1 X_n + \epsilon_n$$

Now, define

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}; \boldsymbol{X} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}; \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

My equations could be written as:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

or

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

or as

$$\boldsymbol{Y} = E(\boldsymbol{Y}) + \boldsymbol{\epsilon}$$

because

$$E(\boldsymbol{Y}) = \begin{bmatrix} E(Y_1) \\ \vdots \\ E(Y_n) \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} = \boldsymbol{X}\boldsymbol{\beta}$$

Defining the Variance-Covariance Matrix:

Suppose I have a random vector $\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$

Let $\sigma^2(Y_i)$ be the variance of Y_i;

$\sigma(Y_i, Y_j)$ be the covariance of Y_i and Y_j, i.e.,

$\sigma(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i)E(Y_j) =$

$E[(Y_i - E(Y_i))(Y_j - E(Y_j))]$

Variance-Covariance Matrix:

$$\sigma^2(\boldsymbol{Y}) = \begin{bmatrix} \sigma^2(Y_1) & \sigma(Y_1, Y_2) & \ldots & \sigma(Y_1, Y_n) \\ \sigma(Y_2, Y_1) & \sigma^2(Y_2) & \ldots & \sigma(Y_2, Y_n) \\ \vdots & & & \vdots \\ \sigma(Y_n, Y_1) & \sigma(Y_n, Y_2) & \ldots & \sigma^2(Y_n) \end{bmatrix} =$$

$$E[(\boldsymbol{Y} - E(\boldsymbol{Y}))(\boldsymbol{Y} - E(\boldsymbol{Y}))^{'}]$$

Remember that the expectation operator, $E(\cdot)$, just means to take the expectation of each entry.

Thus, the simple regression model can be restated as:

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

where $\boldsymbol{\epsilon}$ is a vector of independent normal random variables with

$E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\sigma^2(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}$

The regression coefficients can be obtained as the solution to the normal equation:

$$(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{Y}$$

where $\boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$

or

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

This minimizes

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})$$

Fitted values:

$$\hat{\boldsymbol{Y}} = \begin{bmatrix} \widehat{Y}_1 \\ \vdots \\ \widehat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_1 \\ \vdots \\ b_0 + b_1 X_n \end{bmatrix} = \boldsymbol{X}\boldsymbol{b}$$

So,

$$\boldsymbol{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} Y_1 - \widehat{Y}_1 \\ \vdots \\ Y_n - \widehat{Y}_n \end{bmatrix} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}$$

Computations:

$\boldsymbol{Y}'\boldsymbol{Y} = \sum Y_i^2$

$$\boldsymbol{X}'\boldsymbol{X} = \left[\begin{array}{cc} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{array} \right]$$

$$\boldsymbol{X}'\boldsymbol{Y} = \left[\begin{array}{c} \sum Y_i \\ \sum X_i Y_i \end{array} \right]$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{1}{n\sum X_i^2 - (\sum X_i)^2} \left[\begin{array}{cc} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{array} \right]$$

So,

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \frac{1}{n^2 S_X^2} \left[\begin{array}{c} \sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i \\ -\sum X_i \sum X_i Y_i + n \sum X_i Y_i \end{array} \right] =$$

$$\begin{bmatrix} \bar{Y} - r_{XY}\frac{S_Y}{S_X}\bar{X} \\ r_{XY}\frac{S_Y}{S_X} \end{bmatrix}$$

Residuals:

$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$: original model

$\boldsymbol{Y} = \boldsymbol{Xb} + \boldsymbol{e}$: fitted model

So, $\boldsymbol{Y} - \boldsymbol{Xb} = \boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} =$

$(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$

$\boldsymbol{H} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is called the "hat" matrix;

$\boldsymbol{HH} = \boldsymbol{H}$ and is thus referred to as "idempotent"