Variance/Covariance Matrices

For a vector of random variables, $(Y_1, Y_2, \ldots, Y_n)$, we can write a matrix containing their variances and their covariances. Let σ_i^2 be the variance of Y_i and let cov_{ij} be the covariance between Y_i and Y_j, $i < j$. Then the variance/covariance matrix for $(Y_1, Y_2, \ldots, Y_n)$ is

$$\begin{bmatrix} \sigma_1^2 & cov_{12} & \cdots & cov_{1n} \\ cov_{12} & \sigma_2^2 & \cdots & cov_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ cov_{1n} & cov_{2n} & \cdots & \sigma_n^2 \end{bmatrix}.$$

This can be standardized to give the correlation matrix

$$\begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n} & \rho_{2n} & \cdots & 1 \end{bmatrix}$$

where ρ_{ij} is the correlation of Y_i and Y_j, $i < j$.
Note that both of these matrices are symmetric. Furthermore, the terms on the diagonal of the variance/covariance matrix must be positive and the terms off the diagonal of the correlation matrix are bounded between -1 and 1.

Multiple Regression in Matrix Notation

We have a response variable Y and a set of independent variables $X_1, \ldots, X_p$. We can write our data for n observations as

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}.$$

The multiple regression model can then be written as

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

where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

and

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

We assume $\boldsymbol{\epsilon}$ has a multivariate normal distribution:

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I_n}).$$

Least squares minimizes the function

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

with respect to $\boldsymbol{\beta}$. Note that this is the same as

$$\begin{aligned}
&(\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta}) \\
&= \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \ldots + \beta_p x_{1p} \\ \beta_0 + \beta_1 x_{21} + \ldots + \beta_p x_{2p} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \ldots + \beta_p x_{np} \end{bmatrix} \right)' \\
&\qquad \times \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \ldots + \beta_p x_{1p} \\ \beta_0 + \beta_1 x_{21} + \ldots + \beta_p x_{2p} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \ldots + \beta_p x_{np} \end{bmatrix} \right) \\
&= \begin{bmatrix} y_1 - (\beta_0 + \beta_1 x_{11} + \ldots + \beta_p x_{1p}) \\ y_2 - (\beta_0 + \beta_1 x_{21} + \ldots + \beta_p x_{2p}) \\ \vdots \\ y_n - (\beta_0 + \beta_1 x_{n1} + \ldots + \beta_p x_{np}) \end{bmatrix}' \begin{bmatrix} y_1 - (\beta_0 + \beta_1 x_{11} + \ldots + \beta_p x_{1p}) \\ y_2 - (\beta_0 + \beta_1 x_{21} + \ldots + \beta_p x_{2p}) \\ \vdots \\ y_n - (\beta_0 + \beta_1 x_{n1} + \ldots + \beta_p x_{np}) \end{bmatrix} \\
&= \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}) \right]^2 .
\end{aligned}$$

Recall that the least squares estimate is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ and that $\hat{\boldsymbol{\beta}}$ is an unbiased estimate of $\boldsymbol{\beta}$. We can show this as follows

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\right] \\
&= E\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon})\right] \\
&= E\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\epsilon}\right] \\
&= E\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}\right]+E\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\epsilon}\right] \\
&= E(\boldsymbol{\beta})+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{\epsilon}) \\
&= E(\boldsymbol{\beta})+\mathbf{0} \\
&= E(\boldsymbol{\beta})
\end{aligned}$$