Multiple Linear Regression Model

Expressing the Multiple Linear Regression Model Using Matrix Notation

The multiple linear regression model shown on the previous page:

$Y = \beta_0 = \beta_1 X_1 + ... + \beta_k X_k + \varepsilon$

can be expressed in terms of four matrices:

- Y: the n x 1 column vector of values of the dependent variable Y
- X: the n x p matrix consisting of a column of ones, followed by k column vectors for the independent variables. Each column of X contains the values for a particular independent variable
- β: the p x 1 vector of parameters to be estimated, where $p = k + 1$
- ε: the n x 1 vector of random errors

The linear model can now be written in matrix notation as the following:

$Y = \beta_0 = X\beta + \varepsilon$

where

$$\underset{(n\times 1)}{\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}, \quad \underset{(n\times p)}{\boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & & X_{nk} \end{bmatrix}}, \quad \underset{(p\times 1)}{\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}, \quad \underset{(n\times 1)}{\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}$$

In matrix notation, the normal equations are written as the following:

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

The normal equations have a solution, given as the following:

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

If $X'X$ has an inverse, then the normal equations have a unique solution.

Testing Model Significance Using the *F* Statistic

The significance of the model can be tested using the F statistic shown in the ANOVA table below.

Source of Variation	Degrees of Freedom (df)	Sum of Squares (SS)	Mean Squares (MS=SS/df)	*F* value	*p*-value
Due to regression	k	$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$	MSM=SSM/dfR	MSM/MSE	If < α (predefined, for example, 0.05), then significant model
Error	$n - k - 1$		MSE=SSE/dfE		

		$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$			
Total, corrected for mean $\bar{Y}$	*n - 1*	$\sum_{i=1}^{n}(Y_i - \bar{Y})^2$			

Close