Chapter 5: Matrix Approach to Simple Linear Regression Analysis

Some basic notations

In regression analysis, one basic matrix is the vector $\mathbf{Y}$, consisting of the n observations on the response variable:

Note that the transpose $\mathbf{Y}'$ is the row vector:

Another basic matrix in regression analysis is the $\mathbf{X}$ matrix, which is defined as follows for simple linear regression analysis:

The matrix $\mathbf{X}$ consists of a column of Is and a column containing the n observations on the predictor variable X. Note that the transpose of $\mathbf{X}'$ is:

The $\mathbf{X}$ matrix is often referred to as the ____________.

Simple Linear Regression Model in Matrix Terms

We are now ready to develop simple linear regression in matrix terms. We begin with the **normal error** regression model;

This implies:

Now, we put above equations in the matrix form:

With respect to the error terms, regression model assumes that ___________, ___________, and that the ϵ_i are independent normal random variables. The condition ___________ in matrix terms is:

The condition that the error terms have constant variance a σ^2 and that all covariances ___________ for $i \neq j$ are zero (since the ϵ_i are independent) is expressed in matrix terms through the variance-covariance matrix of the error terms:

Since this is a scalar matrix, it can be expressed in the following simple fashion:

Thus, the normal error regression model in matrix terms is:

where: ϵ is a vector of independent normal random variables with $E\epsilon = 0$ and $Var(\epsilon) = \sigma^2\mathbf{I}$

1.1 Least Squares Estimation of Regression Parameters

Recall Normal equations:

Normal equations in matrix terms are:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

where **b** is the vector of the least squares regression coefficients:

$$\mathbf{b} = \begin{bmatrix} b_0 \\ \\ b_1 \end{bmatrix}$$

Estimated Regression Coefficients

To obtain the estimated regression coefficients from the normal equations by matrix methods, We multiply both sides by the ______________________.

Example 1. We shall USe matrix methods to obtain the estimated regression coefficients for the Toluca Company example. The data On the Y and X variables Were given in http://users.stat.ufl.edu/ rrandles/sta4210/Rclassnotes/data/textdatasets/KutnerData/Chapter%20%201%20Data%20Sets/CH01TA01.txt.

1. Using these data, define the Y observations vector

2. Using these data, define the X matrix

3. Find $X'X$

4. Find $X'Y$

5. Find $(X'X)^{-1}$

6. Find the **b**, vector of regression coefficients

1.2 Fitted Values and Residuals

Fitted Values

Let the vector of the fitted values $\hat{Y}_i$ be denoted by $\hat{\mathbf{Y}}$:

In matrix notation, we then have:

because,

Example 2. For the Toluca Company example, obtain the vector of fitted values using the matrices in R.

Hat Matrix: We can express the matrix result for $\hat{\mathbf{Y}}$ as follows by using only $\mathbf{X}$ and $\mathbf{Y}$ matrices (no need to find the matrix $\mathbf{b}$)

Note 1. The matrix $\mathbf{H}$ is symmetric and has the special property (called idempotency):

Residuals

Let the vector of the residuals $e_i = Y_i - \hat{Y}_i$ be denoted by $\mathbf{e}$:

In matrix notation, we then have:

Example 3. For the Toluca Company example, obtain the vector of the residuals by using these results and R

Variance-Covariance Matrix of Residuals: The residuals e_i, like the fitted values Y_i, can be expressed as linear combinations of the response variable observations Y_i

We thus have the important result:

Then, the variance-covariance matrix of the vector of residuals:

and is estimated by:

Note 2. The matrix $\mathbf{I} - \mathbf{H}$, like the matrix $\mathbf{H}$, is symmetric and idempotent.

Short notes

Short notes

Short notes