Jared Yu

Module 3 Discussion

A set of subjects give data on a response variable and two regressors, and . The data are fitted to two different models

Model 1:

Model 2:

where is the simple average of values from this set of data, ; and are random errors with mean zero.

Discuss with mathematical arguments whether the ordinary least-squares estimator of under Model 1 is equal to the ordinary least-squares estimator of under Model 2, for . State the assumptions in your discussion. Discuss whether or not the regressors are random or non-random change your discussion.

Ans:

Let be the design matrix for , the data matrix with dimensions . It has dimensions , where , since it includes the column of 1’s in the first position. Let be the vector of the observations. An assumption for OLS is that , , , and that the errors are uncorrelated. The same applies to in place of .

Using the least-squares normal equations, the OLS estimates of for , are as follows,

An assumption for this is that exists, which is possible if the regressors (i.e., and ) are linearly independent. The next step is to find . Let

then

*Note: To simplify notation, allow to be shortened to .*

Furthermore, let . The following steps will attempt to find the inverse of . The row operations will be abbreviated with R1, R2, and R3.

Divide R1 by n:

R2 – R1\*:

R3 – R1\*

Divide R2 by

R3 – R2 \*

Divide R3 by

R2 – R3 \*

R1 – R3\*:

R1 – R2\*

The following has been shown:

Where

Then, from equation (7):

From this result it follows that for **:**

To double-check the results, it can be shown that , as stated in equation (6):

Now look towards Model 2:

Let , denote the centered terms,

Then the following also holds:

The following can also be shown:

which is showing that . The following can also be shown:

So, it follows that:

It has been shown that and . However, the following can be shown:

In conclusion it can be said that in general,

In the process of finding the inverse of , an assumption is that is required is that it is doable. That is, there is no perfect collinearity issue. In other words, exists. So,

That means we need , or . If the regressors are random, then , the sample correlation, is also a random variable. It cannot be guaranteed since it is random. This must also apply to the centered versions of the data, , . So, for certain sets of sample data, if the regressors are random, it is possible that the above statements do not hold. Another unlikely situation is if . In such a situation, it would hold that , which would make .