Jared Yu

Module 9 Discussion

Suppose that we fit the full model to statistically independent observations and obtain the least-squares estimator of .

and are fixed (that is, they are not random variables).

Then we fit the subset model to the same observations and obtain another least-squares estimator of .

Discuss which least-squares estimator of is more biased for the parameter and which has smaller variance. State assumptions underlying your discussion.

Ans: (*Note: This largely references pp. 329-330 from the Textbook.*)

A first assumption that we can add is that , where here in the full model and in the subset model. This is should not be an issue, since the requirement here is only that or which is quite minimal.

From previous chapters, we have learned that the (full) model can be written in matrix notation,

where the only difference here is that and are used in place of and respectively. This just follows the traditional notation, rather than the notation of the problem. Equation (1) can be rewritten as follows,

where corresponds with (and which are all 1’s), , and , while corresponds with and .

For the full model we have the following estimates for and ,

and

The subset model, , has the corresponding estimates,

and

Regarding , the expected valued of is

where is also known as the alias matrix. Equation (7) indicates then that is biased in the case of the subset model, since . The exception would be in the case that or and are orthogonal, in which case would be zero. On the other hand, we know from before that the multiple linear regression (MLR) model is unbiased. So, given Equation (3), we can see that within the full model, it is unbiased for .

Regarding the variance of , we have that for and that

and

From this, the matrix is positive semidefinite. This implies that the variance of the estimates for the full model is greater than or equal to the variance of the estimate in the subset model.

For simplicity, we can denote and . The variance of then can be found by looking at and . Since the resulting matrix is positive semidefinite, that implies that the variance for will also be the same if not larger.

*Note: I am somewhat confused about one issue regarding my discussion at the end. When I discuss the resulting matrix that is calculated from , I am directly sourcing this from the textbook. However, since these are matrices based on the inverse of the covariance matrix between the corresponding and , my thinking is that their dimensions will not match, given that the former is larger than the latter. I think I am missing something either obvious or not stated explicitly, but in my current thinking it’d require these matrices to have matching dimensions before the difference can be taken. The idea however is simply that the variance of , found by looking at the diagonal is either equal or larger in the full model.*