Covariance Structures: A Realistic Example

Dr.Arnab Chakraborty * Mrinmoy Banik †

17th October 2022

Abstract

In the previous section we have seen how after dropping the homoscedasticity assumption we end up with the possibility of various covariance structures for the error covariance matrix (they include Scalar,Compound Symmetric,AR1,Any PD, & Block Diagonal).In this section we will explore one such structure in a real life example which we had already discussed in some other previous section.

1 Our Example

So we will be working with our familiar example in which we have an agricultural study with plots of land as our basic unit, with 3 varieties of the same crop and 3 villages in total.Each village with each variety has a some amount of replication too and we are interested in knowing if the varieties have a significant effect on the crop yields.

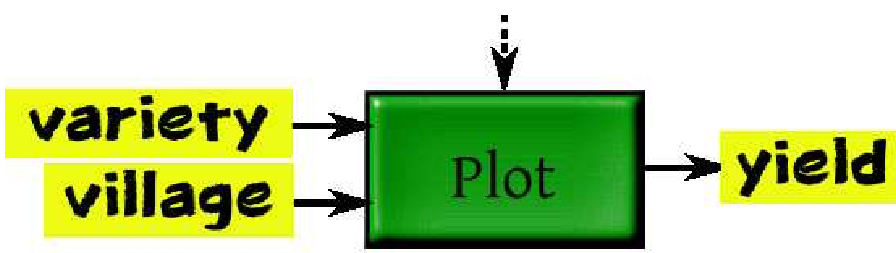

FIGURE 1: Black-Box diagram of our linear model with random error, village and crop varieties as categorical inputs and Yield as continuous output.

*Department of Applied Statistical Division, Indian Statistical Institute,Kolkata Email:http://www.isical.ac.in/ arn-abc/

†This portion of the book is latex-ed and edited by him,Student of B.Stat 3rd yr at ISI,BS2012

But it so happens that each village uses a certain kind of instrument for measuring the yield which is local to each village and hence the village might have an undesirable effect on the yield too. At the same time the different measurements from the same village might be correlated.
Let's see how we incorporate that:- In this case the model is of the form:-

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

where the indicators i,j, & k stand for:-
i:Variety
j:Village
i:Replication Factor

We expect the different entries that share the same j values to be correlated.We can write the entire $\overrightarrow{y_{ijk}}$ vector in the following way:-

$$\therefore \overrightarrow{y_{ijk}} = (y_{111}, y_{112}, ..., y_{115}, y_{121}, y_{122}, ...)$$

2 Covariance Structure

So all the observations are of the following form if we colour each entry sharing the same j index with the same colour as shown in FIGURE 2:-

Now all the things that are sharing the j index are going to be correlated. Then the covariance matrix will have the particular form like the above. All the white entries are actually zero entries.

So when we say block-diagram structures, the actual covariance might not just consist of blocks or the blocks might not be linked together, instead they can be scattered throughout the matrix like in this case. But nonetheless the sub-matrices in each block will be symmetric over the diagonal.

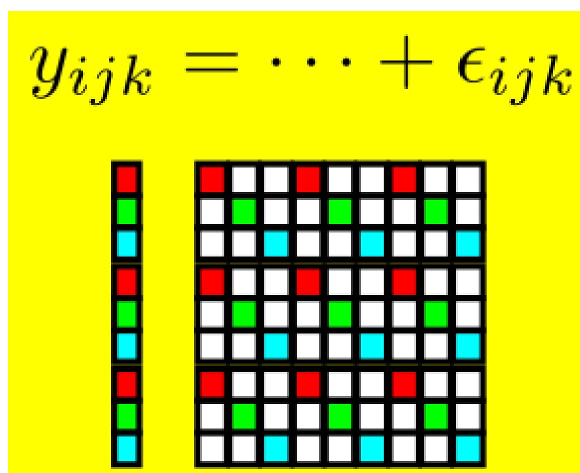

FIGURE 2: All the coloured boxes are of 5x5 dimension with replication $k \in \{1, 2, 3, ...5\}$.All the red boxes are from village 1, so they have identical correlation structure.All the same coloured boxes form a 15x15 submatrix.

3 Parameters

As we deviate from our normal Gauss Markov models we nolonger can say $\epsilon_{ijk} \sim N(0, \sigma^2) \forall i, j, k$ instead we use the following model:-

$$\overrightarrow{Y} = X\overrightarrow{\beta} + \epsilon \quad where \quad \epsilon \sim N(\overrightarrow{0}, \sum(\overrightarrow{\theta}))$$

where $\sum(.)$ is a known function and $\overrightarrow{\theta}$ is the covariance parameter and together they form the covariance structure.
In this example, for the red part representing village 1 you may have compound symmetry, in which case you have two parameters each σ^2 & each ρ^2 (so it's σ_1^2 & ρ_1^2).
Similarly for the green part representing village 2 you may have Autoregressive-1 (AR1) structure, the contributes σ_2^2 & ρ_2^2.
For the third blue part representing village 3 , we may assume they are uncorrelated , so you get a contribution of just σ_3^2.
$\therefore$ the covariance parameters are $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \rho_1, \rho_2)$ with complicated conditions such that the entire covariance matrix **P** remains a Positive Definitive Matrix.(which is a necessity for all covariance matrices)

4 Conclusion

Up untill now we were working with Gauss Markov Models and had assumed homoskedasticity of the errors, i.e the errors all are mutually uncorrelated and independent of the parameters. Once we drop this assumption we have to dean will a wide range of structures of the covariance matrix of our data under our model.Depending on our design matrix these covariance matrix can have various box-diagonal structures with inter-class symmetric covariance sub-matrices scattered throughout. Each of these block martices contribute parameters on which the structure of the overall P matrix depend on. These also have to satisfy complicated conditions to satisfy the Positive Definiteness of P.

[—————————]