Covariance Structures

Creating Data : Rlab

1 Overview

Here we are going to carry out covariance structure design and it's analysis using linear model by creating our own data to get an understanding of how to deal with such cases. We are creating our own data instead of using some pre-defined data so that we can change the data in between just to make concepts more clear.

We are using R for this illustration, though R is not efficient for dealing with covariance structures but this compilation uses R as it's base software for illustrations. There are many other software like SPSS or SAS where all these working regarding to covariance structures are done in completely different way.
For more understanding of literature and it's uses you can refer to the book "*Mixed-Effects Models in S and S-PLUS* " by José C. Pinheiro and Douglas M. Bates. This book uses S and S-Plus which has somewhat similar script as R.

2 Data creation

We are creating a data set where we have three variety of crops and a fertilizer. We need to see the effect of fertilizer on different variety of crops.

The blackbox for this model is :

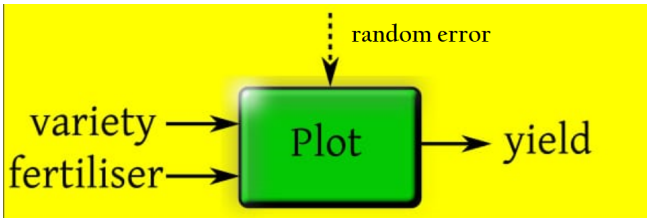

Figure 1

Here we are going to assume random errors as of AR(1) model. So, we are considering all plots which has no fertilizer are correlated with each other and similarly all plots where fertilizers are used has some kind of correlation between them in terms of AR(1) model.
Names of the three variety of crops are 'Jaya', 'Taichung' and 'IRB' while 'Compost' is the name of fertilizer.
We have total six plots, two each for all three varieties where fertilizer is used while we have five plots with no fertilizer in which 'Jaya' variety has one plot and other two varieties have two-two plots. This imbalance of plot distribution among variety-fertilizer is done so that we can understand two correlation matrices easily which will be shown further.

Let's create R code for this data creation in the next page.

3 Rcode

So, here we need to download 'nlme' package i.e., Linear and Nonlinear Mixed Effects Models which let's us specify the variance-covariance structures of the residuals/ random errors.

```
library(nlme)
```

Now lets make vector for distribution vector with respect to variety of crops and fertilizers in farm plots.

```
vnum = c(rep(1,4), rep (2,4), rep(3,4))
fnum = rep(c(rep(1,2), rep (2,2)),3)
```

Note that here right now number of plots in non-fertilizer land is still six. We will delete one data point once we will form data frame.

```
alpha = c(1,2,3)
beta = c(3,3)
mu=100
```

Now let's write the linear model equation.

```
yield = mu + alpha[vnum] +beta[fnum]+ rnorm(12)
```

Now we have used normal distribution for generating random errors in data for our simplicity but we are going to assume it as AR(1) modal and modal it according to that for our predictions.
Now let's name all the varieties of crops and fertilizers.

```
vname=c('Jaya', 'Taichung','IR8')
fname=c('None', 'Compost')
```

It's time to make data frame for the data set.

```
dat = data.frame(yield, variety=vname[vnum], fert=fname[fnum])
dat = dat[-1,]
```

We have removed the first data point as shown in the below figure of data frame to make it imbalance just for easy understanding later. Now it's ready to use.

```
> dat
      yield  variety    fert
2  103.2669     Jaya    None
3  104.2109     Jaya Compost
4  103.0011     Jaya Compost
5  106.0779 Taichung    None
6  103.8010 Taichung    None
7  105.2166 Taichung Compost
8  105.1431 Taichung Compost
9  104.9342      IR8    None
10 105.5714      IR8    None
11 105.3438      IR8 Compost
12 106.9594      IR8 Compost
```