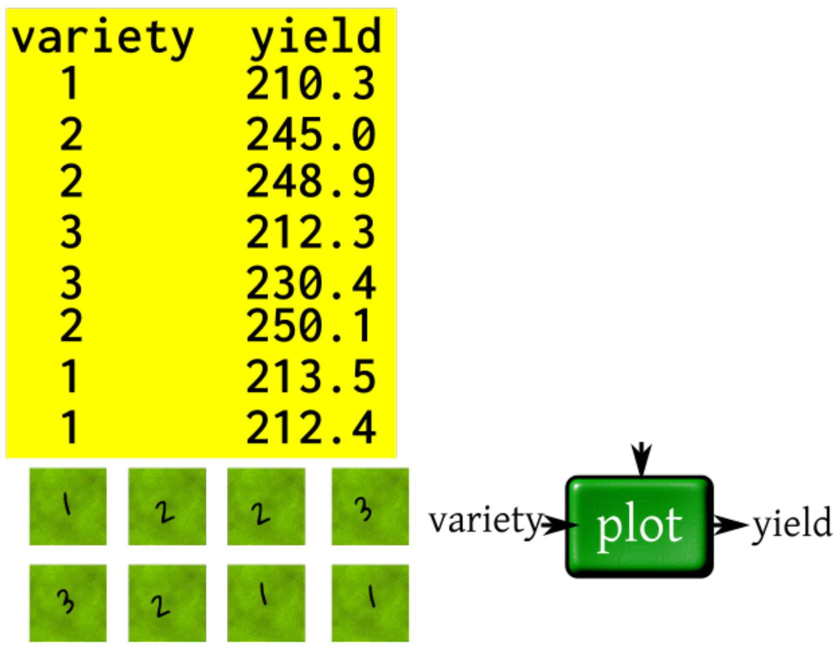
Linear Statistical Models

Video 42 : SECOND PARAMETRISATION

Anushka De
Roll: BS2042



variety	yield
1	210.3
2	245.0
2	248.9
3	212.3
3	230.4
2	250.1
1	213.5
1	212.4

Figure 1: *Figure showing yield for each variety of crop*

Many times, people do not prefer to write the model in the form:

$$y_{ij} = \alpha_i + \varepsilon_{ij}$$

So we instead rewirte the model using a slightly different parametric form like :

$$\boxed{y_{ij} = \mu + \alpha_i + \varepsilon_{ij}}$$

Note that here the number of parameters have increased to four. As a result not of all them are estimable from this data anymore, they are not identifiable.

However there is a particular advantage of this model over the previous model in terms of *interpretability*.

- **Interpretation of μ :** Average yield of this type of crop irrespective of the variety
- **Interpretation of α_1, α_2 and α_3 :** additional effect due to that variety of crop.

Some α_i's may be positive and some may be negative. For instance; $\alpha_1 = 50$ indicates that this variety of crop is actually *bad*. Positive α_i's will be better than the average.

In this case also, the rank of the design matrix will be 3.
NOTE:
The **Design Matrix** for the new model is :

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Observation from the Design Matrix: The second, third and fourth columns are linearly independent and first column is sum of second,third and fourth columns. Hence rank of this matrix is 3.

So we need some additional constraints by which we can say $\alpha_i's$ and μ have the interpretation that we want to imply.