Two Factor Model: Interaction Term

In the two factor additive model, we had the property that the effects due to the two types of factors acted independently of each other, For example: considering the problem of effect of tilt and variety of plant on average yield, the difference in average yield between two varieties, stayed constant whether they were on tilted plot or plane plots.This lead to the interaction chart to have parallel lines. Intuitively, this means that the effect of land tilt on yield, acted without consideration of the variety of plant sowed and vice versa. However, this might not always be true. Sometimes the two factors might act in a dependent manner, For example: the effect of tilt on one variety of the plant maybe different from effect of tilt on another variety of the plant. So, in these cases, the simple additive model seems to be inappropriate. We need to bring in some "*Interaction*" term along with the individual effect terms as well.

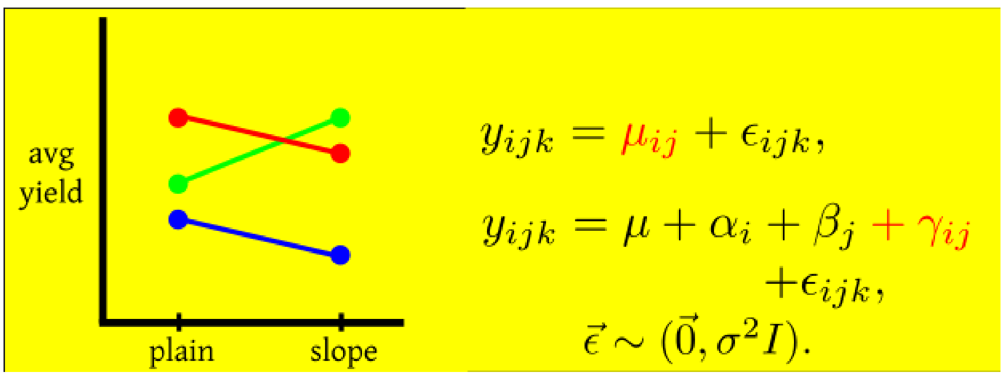

Figure 1: Cell means Model for Interaction

Two Factor Models that are used in cases where the effects of the two type of factors don't act independently, are called "$TwoFactorInteractionModel$". If we plot the interaction chart for cases where there is an interaction effect, we will find, (unlike cases where the effects are independent and additive models are applicable) atleast one pair of lines which are not parallel, this is due to the effect of the interaction . This feature helps us identify, by looking at the data, when we should use the interaction model.

So, what do we do to bring in the effect of this interaction in our model? We simply add a term $\boldsymbol{\gamma_{i,j}}$ which depends both on i and j, to our additive model. This gives us the second equation in Figure 1.

$$\boldsymbol{Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{i,j} + \epsilon_{ijk}, \quad \vec{\epsilon} \sim \left(\overrightarrow{0}, \sigma^2 I_n\right), \quad n = ijk}$$

where:

Y_{ijk} :=Yield in $\mathrm{k^{th}}$ field of $\mathrm{i^{th}}$ crop grown on plot of $\mathrm{j^{th}}$ type

μ := mean effect/ intercept term

α_i := effect on yield due to the crop variety - (factor 1)

β_j := effect on yield due to the type of land - (factor 2)

$\gamma_{i,j}$:= the effect that remains when we strip off the mean effect, effect of tilt and effect of variety

ϵ_{ijk} := effect of randomness in yield of the k^{th} plot with i^{th} crop type and j^{th} plot type

The $\boldsymbol{\gamma_{i,j}}$ term models the interaction between the two factors and is adequately called the "*InteractionTerm*".

A quick glance at our model reveals that it is heavily over-parametrized. This leads to a problem of identifiability! Also, the interpretation of the $\boldsymbol{\gamma_{i,j}}$ term is, the effect that remains when we strip off the mean effect, effect of tilt and effect of variety, quite a hectic interpretation!So what can we do? The simplest thing to do is use a model as follows:

$$\boldsymbol{Y_{ijk} = \mu_{i,j} + \epsilon_{ijk}, \quad \vec{\epsilon} \sim \left(\overrightarrow{0}, \sigma^2 I_n\right), \quad n = ijk}$$

The $\boldsymbol{\mu_{i,j}}$ term encompasses the values of $\boldsymbol{\mu, \alpha_i, \beta_j}$ ***and*** $\boldsymbol{\gamma_{i,j}}$. It has an easy interpretation of being the expected yield for a field having the ith and jth value of factor1 and factor2 respectively. It is also easily estimable by taking the mean of yield for the fields having i^{th} and j^{th} value of factor1 and factor2 respectively. This model is called the *CellMeansModel* since we are using the mean of each cell (cells we get from different combinations of input factors) as our parameter.