INDIAN STATISTICAL INSTITUTE
Saptarshi Sinha(bs2031)

Acknowledgement

I want to thank Prof. Arnab Chakraborty for providing detailed intuitive explanations about the topic and neat visuals in his web page.

1 Two factor model: Additive model

In this section, we will mathematically model the interaction chart(from previous example) properly. Let Y_{ijk} denotes the yield on a particular plot of land. The yield depends on the variety of crop, tilt of the land and some unknown factors(denoted by the random error), as depicted in the **"blackbox diagram"**. Here the i^{th} index denotes the crop variety, j^{th} index denotes tilt and k^{th} index is the randomness.

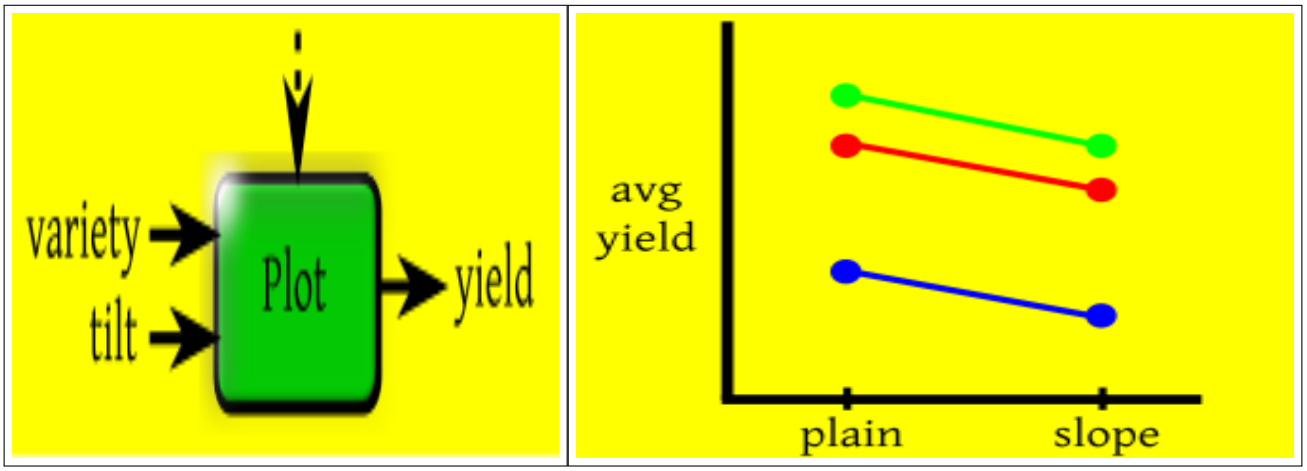

Figure 1: (a)Blackbox diagram (b)Interaction chart

The three different lines are the profiles of the three varieties. One of the striking features of the chart is that the difference in average yield between two varieties on the same type of land(same slope) is independent of the type of land(plane/slope). This feature is reflected by the profiles being parallel(vertical distances between any two parallel lines are fixed). Similarly the difference in average yield between two types of land on the same variety profile does not depend on the variety.[1] The best, perhaps the only mathematical formula which captures this idea is:

$$Y_{ijk} \;=\; \alpha_i + \beta_j + \epsilon_{ijk} \;, \qquad \vec{\epsilon} \sim \; \left(\vec{0}, \sigma^2 I_n\right), \quad n = ijk$$

where:
α_i := average yield of the i^{th} crop variety — **(factor 1)**
β_j := average yield of crop in j^{th} type of land — **(factor 2)**
ϵ_{ijk} := randomness in yield of i^{th} crop type for a particular plot of j^{th} type

It is easy to see that under this model, if we find the difference in yield between any two crop type "a" and "b" over a particular type of land "j", it will give

$$Y_{ajk} - Y_{bjk} \;=\; (\alpha_a - \alpha_b) + (\epsilon_{ajk} - \epsilon_{bjk})$$

which is independent of the land dependent yield β_j. Similarly, the yield difference between land type "c" and "d" with same crop type "i" is:

$$Y_{ick} - Y_{idk} \;=\; (\beta_c - \beta_d) + (\epsilon_{ick} - \epsilon_{idk})$$

This does not depend on α_i the average yield for i^{th} type of crop.
The given mathematical expression can be used to model any two-factor input system with this additive property.

[1] We wont be getting this kind of privilege everytime. We will be using additive models only when we see this kind of properties.