ANOVA table: Algebraic Identity

Samprit Chakraborty

Roll No.- BS2004

Instructor: Dr. Arnab Chakraborty

Indian Statistical Institute, Kolkata

In the continuing example of fertiliser and plot, there are 3 varieties of fertiliser and n_i many plots under ith variety where i=1,2,3. The outcome shown in a number line:

The black coloured cluster signifies outcome of variety 3, the red coloured cluster signifies outcome of variety 2 and the blue coloured cluster signifies outcome of variety 1.

We start with the output variability. If we call the yield of the j-th plot under the i-th fertiliser by the name y_{ij}, then the output variability may be measured by

Total sum of squares(TSS)=$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$.

where $\bar{y}_{..}$ is the overall mean i.e $\bar{y}_{..}$=$\frac{1}{n} \sum_i \sum_j (y_{ij})$

The error variability is best measured by looking it how much dots of the same colour differ from each other. These are given by (for $i = 1, 2, 3$)

$$\sum_j (y_{ij} - \bar{y}_{i.})^2 .$$

So the total variability due to random error is, this gives an idea about how tight the clusters are

Within sum of squares(WSS)= $\sum_i \sum_j (y_{ij} - \bar{y}_{i}.)^2$.

this gives an idea about how tight the clusters are.

If we want to measure the variability due to fertiliser, then we should first find the average of dots of each colour, and pretend that all the dots of that colour are actually at that average, and then see how much the points differ from each other:

Between sum of squares(BSS)= $\sum_i n_i (\bar{y}_{i.} - \bar{y}_{..})^2$.

where $\bar{y}_{i.}$ is the mean of the ith cluster.

Now, TSS $=\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.} + \bar{y}_{i.}\bar{y}_{..})^2$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_j n_i (\bar{y}_{i.} - \bar{y}_{..})^2 + 2 \sum_i \sum_j (y_{ij} - \bar{y}_{i.}) (\bar{y}_{i.} - \bar{y}_{..}) = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_j n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

So basically, TSS = WSS + BSS.