ANOVA TABLE

Prisha Reddy

1 Why care ?

The anova table was considered extremely important for a long time, and a constant effort was put into generalizing the concept of ANOVA (Analysis of Variance) to fit any general case with or without the orthogonality condition. However, two very specific cases arose where this was not possible:

1.1 When you have interaction:

In cases where interaction is present, one can make no claim of effect of a particular input independently of all other inputs. The model itself explains that the values are all interdependent. Therefore, the construction of an anova table is not possible, and thus one cannot generate an F statistic to simply claim over the effect of a single input.

1.2 When you have an unbalanced model

To understand the concept of a balanced model,consider our favorite example, namely the plots. The concept is not having equal number of plots in each combination of factor inputs.

Suppose we have a single factor input (for example, variety of seeds). Then, n_i (the no. of plots that consist seeds of variety i) should have a constant value (say k_1) irrespective of i in case of a balanced model.

Similarly, suppose we have 2 factor inputs (say, variety of seeds and the fertilizer type). Then, n_{ij} (the no. of plots that consist seeds of variety i and fertilizer of type j) should have a constant value (say k_2) irrespective of either i or j in case of a balanced model.

Note that, any arrangement varying from this model (where equal no. of plots are assigned to each combination of factor inputs) is no longer a balanced one.

Now again consider the single factor input scenario. In this case, the orthogonality condition is trivially followed (since a single set of vectors, i.e. a single subspace; thus orthogonal to everything else). Therefore, the status of balanced or unbalanced is irrelevant in such a case.

However, the same cannot be said about 2 factor input model. Notice that in any unbalanced model, the orthogonality condition cannot hold strong. Now suppose we had initially planned for an experiment (the plots with 2 factor inputs, say) where we had specifically intended for the orthogonality condition to hold in the model. However, 1 random plot becomes unusable because of some accident for suppose.So basically, even if we had intended to achieve a balanced model, it is extremely rare to achieve it. This model has now become unbalanced, and one can no longer come up with an anova table.

2 Orthogonalization

Thus, due to the above-mentioned reasons, the need for generalizing became eminent. Proceeding mathematically, people started applying the famous 'Gram Schmidt Orthogonalization' on the columns of X. Factor wise columns are considered, and then the effect of one factor is eliminated from the other factor through 'Gram Schmidt Orthogonalization'.

As in the above example, we take the fertilizer effect corrected for variety effect. In detail, consider the columns of the fertilizer parameters, and from that the orthogonal projection of the variety columns is subtracted. Thus, the obtained columns are now orthogonal to the columns of the variety parameter. These columns are interpreted as the effect of fertilizer after correcting for effect of variety.

Due to the above interpretation (not really helpful), and having achieved no alternate way of achieving the orthogonalization, statisticians have eventually stopped considering the ANOVA table of the prime importance it once had. Softwares have also stopped producing the ANOVA table as default output.