One-way ANOVA – an Agricultural Example

Arunsoumya Basu
Roll No. bs2029

We now turn our attention to an application of linear models in agricultural statistics. We consider a practical situation, where we have a few varieties of a certain crop. Suppose we want to identify the varieties which produce good yield, and also the ones which do not. So how would we perform the task?

The first step would be to take some plots of land. Ideally, these plots should be identical to one another so that any possible differences in yields occur only due to the varieties being sown on them. But realistically speaking, this is too much to expect, so we should try to take plots with similar intrinsic properties and any differences present would be attributed to the random error. So for example, we might have three distinct varieties, denoted by $\mathbf{1}, \mathbf{2}$ and $\mathbf{3}$, and 8 (almost) identical plots of land (Fig. 1). This is a typical scenario in India, where different varieties of a crop are grown on small, say, $2m \times 2m$ plots to measure their yields and compare the varieties.

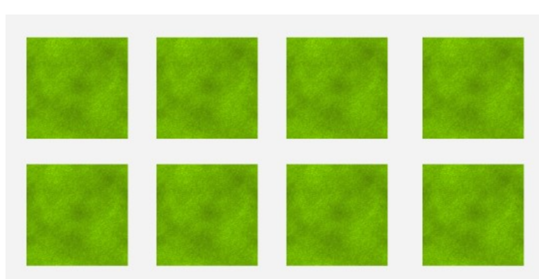

Fig. 1

Now, we would like to assign the varieties to the plots of land as equally as possible (so in this case, each variety would be assigned to 2 or 3 plots of land), while maintaining random assignment. Random assignment is essential because even if some plot of land has slightly different characteristics compared to the others, each variety has an equal probability of getting assigned to it. So, we pick 3 plots at random (via SRSWOR, **S**imple **R**andom **S**ampling **W**ith**O**ut **R**eplacement) and assign variety **1** to them. The scenario now is somewhat like this: (Fig. 2)

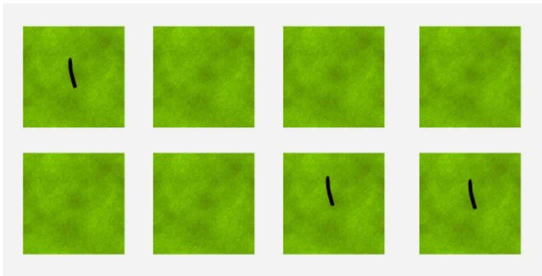

Fig. 2

Next out of the remaining 5 plots, we select 3 plots at random again, via SRSWOR, and assign variety **2** to them. (Fig. 3) Finally, variety **3** is assigned to the remaining 2 plots. (Fig. 4).

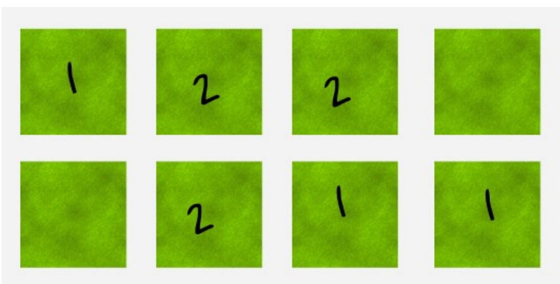

Fig. 3

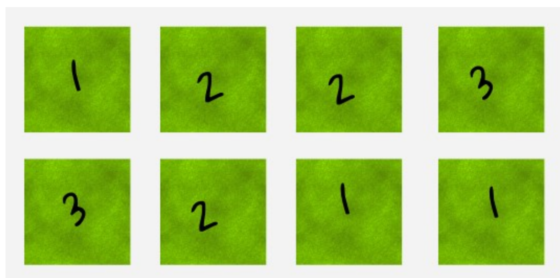

Fig. 4

The assigned varieties are now sown on the corresponding plots, and at the end of the season, the yields are measured. Our blackbox diagram will thus look similar to the one in Fig. 5:

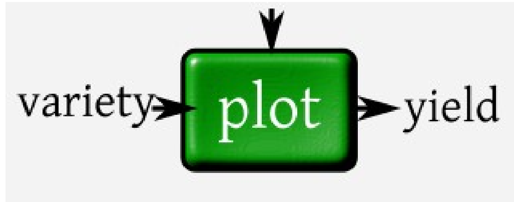

Fig. 5

To elaborate, the blackbox 'plot' takes the input 'variety' and gives an output 'yield'. The vertical arrow is due to the random error, as usual. The blackbox diagram indicates that the tabular form of the data will have two columns, one each for variety and yield, and 8 rows corresponding to the varieties sown in the 8 plots. The tabular form of the collected data is somewhat like the one below (yields are measured in some standard unit):

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

In the upcoming sections, we look forward to mathematical modelling of the data and subsequent analysis through R.