Analysis of Variance

In the following, analysis of variance (ANOVA or AOV) is illustrated for a case where there are k groups or regions, and n_i, $i = 1, \dots k$, observations in the i-th group or region. The model that underlies analysis of variance assumes that each observation has several components

$$X_{ij} = \overline{X}_T + (\overline{X}_i - \overline{X}_T) + (X_{ij} - \overline{X}_i), \text{ where}$$

X_{ij} = the j-th observation in the i-th group or region,

$\overline{X}_T$ = the grand mean (or all observations in all groups),

$(\overline{X}_i - \overline{X}_T)$ = is an "among (or between) groups" deviation (the difference between the mean of the i-th group and the grand mean, and

$(X_{ij} - \overline{X}_i)$ = is a "within groups" deviation (the difference between the j-th observation in the i-th, and the mean of the i-th group).

The means are defined as follows

$$\overline{X}_T = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij}, \text{ and}$$

$$\overline{X}_i = \frac{1}{n_i} \sum_{i=1}^{n_i} X_{ij}, \; i = 2, \dots k, \text{ where}$$

$$N = \sum_{\mathrm{i}=1}^{k} n_i, \text{ is the total number of observations.}$$

With this model for an individual observation in mind, the total variability of the data can be decomposed into two components:

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_T)^2 = \sum_{i=1}^{k} (\overline{X}_i - \overline{X}_T)^2 + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2$$

$$TotalSS = SS_A + SS_W$$

where $TotalSS$ is the “total sum of squares if the data (the sum of the squared deviation of each observation in each group about the grand mean), SS_A is the “among (or between) groups sum of squares” (the sum of the squared deviations of each group mean about the grand mean), and SS_W is the “within-groups sum of squares” (the sum of the deviatins within each group about the group mean).

The among-groups and within-groups sum of squares are adjusted for the appropriate degrees of freedom to produce the following variance-like quantities:

$$MS_A = \frac{SS_A}{df_A} = \frac{SS_A}{k-1}, \text{ and}$$

$$MS_W = \frac{SS_W}{df_W} = \frac{SS_W}{N-k}.$$

The test statistic, F , provides a measure of the size of the among-groups variability relative to the within-groups variability

$$F = \frac{MS_A}{MS_W}.$$

F may be compared to the F distribution with $k=1,$ and $N-k$ degrees of freedom