Simfit

In order to appreciate what is involved in groups and subgroups ANOVA, consider the following data set contained in the SıмFɪT test file `anova4.tf1` and made available after opening the main SıмFɪT menu, followed by choosing [Statistics], [ANOVA], then the groups and subgroups option.

Groups	Subgroups	Observations
1	1	2.1
1	1	2.4
1	1	2.0
1	1	2.0
1	1	2.0
1	2	2.4
1	2	2.1
1	2	2.2
1	3	2.4
1	3	2.2
1	3	2.6
1	4	2.4
1	4	2.4
1	4	2.5
1	5	1.9
1	5	1.7
2	1	2.1
2	1	1.5
2	1	2.0
2	2	1.9
2	2	1.7
2	2	1.9
2	2	1.9
2	2	1.9
2	3	2.0
2	3	2.1
2	3	2.3

The first column is the group number (in nondecreasing order), the second column is the subgroup within the group (in nondecreasing order), and the third column holds the corresponding observations. Note that there are no limits to the number of groups, nor to the sizes of the subgroups, but any data file supplied for groups and subgroups ANOVA must be arranged exactly as above.

These data were obtained as follows.

- The two groups represent data obtained in two consecutive years.

- The subgroups of size five and three are consignments of materials in those two years.

- The observations are replicates for percentages of stretch of the material within the groups and subgroups.

It is wished to test if there are significant differences between the properties of material delivered in the two years, and also between the samples drawn, and analysis then leads to the following results table.

Results for Groups/Subgroups 2-Way ANOVA
Transformation = x (i.e. untransformed data)

Source	SSQ	NDOF	F	p
Between Groups	0.4748	1	16.15	0.0007
Subgroups	0.8162	6	4.626	0.0047
Residual	0.5587	19		
Total	1.850	26		

Group	Subgroup	Mean
1	1	2.100
1	2	2.233
1	3	2.400
1	4	2.433
1	5	1.800
2	1	1.867
2	2	1.860
2	3	2.133

Description	Mean	Sample size
Group 1	2.206	16
Group 2	1.936	11
Grand	2.096	27

These results could be interpreted to suggest a between groups effect ($p = 0.0007$) and also a subgroup effect ($p = 0.0047$) and the requisite theory will be presented next.

Theoretical details

The linear models for ANOVA are easy to manipulate mathematically and trivial to implement in computer programs, and this has lead to a vast number of possible designs for ANOVA procedures. This situation is likely to bewilder users, and may easily mislead the unwary, as it stretches credulity to the limit to believe that experiments, which almost invariably reflect nonlinear non-normal phenomena, can be analyzed in a meaningful way by such elementary models. Nevertheless, ANOVA remains valuable for preliminary data exploration, or in situations like clinical or agricultural trials, where only gross effects are of interest and precise modelling is out of the question. So groups and subgroups ANOVA is a versatile and flexible technique provided by SIMFIT for two-way hierarchical classification with subgroups of possibly unequal size, assuming a fixed effects model.

Suppose, for instance, that there are $k \geq 2$ treatment groups, with group i subdivided into l_i treatment subgroups, where subgroup j contains n_{ij} observations. That is, observation y_{mij} is observation m in subgroup j of group i where

$$1 \leq i \leq k, 1 \leq j \leq l_i, 1 \leq m \leq n_{ij}.$$

The between groups, between subgroups within groups, and residual sums of squares are

$$\text{Group } SSQ = \sum_{i=1}^{k} n_{i.} (\bar{y}_{.i.} - \bar{y}_{...})^2$$

$$\text{Subgroup } SSQ = \sum_{i=1}^{k} \sum_{j=1}^{l_i} n_{ij} (\bar{y}_{.ij} - \bar{y}_{.i.})^2$$

$$\text{Residual } SSQ = \sum_{i=1}^{i} \sum_{j=1}^{l_i} \sum_{m=1}^{n_{ij}} (y_{mij} - \bar{y}_{.ij})^2$$

which, using $l = \sum_{i=1}^{k} l_i$ and $n = \sum_{i=1}^{k} n_i$, and normalizing give the variance ratios

$$F_G = \frac{\text{Group } SSQ/(k-1)}{\text{Residual } SSQ/(n-l)}$$

$$F_S = \frac{\text{Subgroup } SSQ/(l-k)}{\text{Residual } SSQ/(n-l)}$$

to test for between groups and between subgroups effects.

Of course, there are now two p values for significance testing and, also note that, because this technique allows for many designs that cannot be represented by rectangular matrices. However, the data files must have three columns and n rows: column one containing the group numbers, column two containing the subgroup numbers, and column three containing the n observations as a sample vector in the order of groups, and subgroups within groups.

Note that, by defining groups and subgroups correctly in this way, a large number of ANOVA techniques can be performed using this groups and subgroups ANOVA procedure.