Introduction to multivariate analysis of variance (MANOVA)

The multivariate analysis of variance technique is an extension of the standard univariate analysis of variance procedure (ANOVA) to the situation where observations of more than one variable are made for each subject. So, for instance, a simple application of ANOVA would be to assume that multiple observations have been made of a single variable in several groups and, assuming that this variable is distributed normally in each group with the same variance, to test if all the population means are identical. In the corresponding MANOVA case it would be to assume that all the observations are from multivariate normal distributions with the same covariance matrix, and to test for identical mean vectors in the populations.

For example, sometimes a designed experiment is conducted in which more than one response is measured at each treatment, so that there are two possible courses of action.

1. Do a separate ANOVA analysis for each variable.
   The disadvantages of this approach are that it is tedious, and also it relies upon the questionable assumption that each variable is statistically independent of every other variable, with a fixed variance for each variable. The advantages are that the variance ratio tests are intuitive and unambiguous, and also there is no requirement that sample size per group should be greater than the number of variables.
2. Do an overall MANOVA analysis for all variables simultaneously.
   The disadvantages of this technique are that it relies on the assumption of a multivariate normal distribution with identical covariance matrices across groups, it requires a sample size per group greater than the number of variables, and also there is no unique and intuitive best test statistic. Further, the power will tend to be lower than the power of the corresponding ANOVA. The advantages are that analysis is compact, and several useful options are available which simplify situations like the analysis of repeated measurements.

Central to a MANOVA analysis are the assumptions that there are n observations of a random m dimensional vector divided into g groups, each with n_i observations, so that $n = \sum_{i=1}^{g} n_i$ where $n_i \geq m$ for $i = 1, 2, \ldots, g$.

If y_{ij} is the m vector for individual j of group i, then the sample mean $\bar{y}_i$, corrected sum of squares and products matrix C_i, and covariance matrix S_i for group i are

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

$$C_i = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)(y_{ij} - \bar{y}_i)^T$$

$$S_i = \frac{1}{n_i - 1} C_i.$$

For each ANOVA design there will be a corresponding MANOVA design in which corrected sums of squares and product matrices replace the ANOVA sums of squares, but where other test statistics are required in place of the ANOVA F distributed variance ratios. This will be clarified by dealing with typical MANOVA procedures, such as testing for equality of means and equality of covariance matrices across groups.