Discriminant Analysis

Discriminant analysis has several interrelated objectives that include:

- identification of how one or more groups of observations differ as described by one or more, usually interrelated variables;
- assignment or classification of new observations to one or another of the groups based on the values of the variables.

In the general case, there are m groups, of sizes $n_1, n_2, ..., n_m$, and p variables $X_1, X_2, ..., X_p$ that describe the observations. The objective of determining *how* the groups differ can be met by deriving a new variable (a linear combination of the existing variables) that, as in principal components analysis, has certain desirable properties. One desirable property of the new variable is that when plotted along it, the group centroids should show maximum separation, and it is possible to show that this is achieved when the between-groups variance is maximized relative to the within-groups variance.

Let the $p \times p$ symmetric matrix $\mathbf{T}$ be a matrix of the sums-of-squares and cross-products of the variables calculated over all observations, and let $\mathbf{W}_i$ be a similar matrix for just the observations in group i, and let

$$\mathbf{W} = \mathbf{W}_1 + \mathbf{W}_2 + ... + \mathbf{W}_m.$$

$\mathbf{W}$ is therefore composed of the "pooled within-groups" sums-of-squares and cross-products. The "between-groups" sums-of-squares and cross-products matrix is then given by

$$\mathbf{B} = \mathbf{T} - \mathbf{W}.$$

A linear combination, referred to as a *canonical discriminant function*, Z_1 can be defined as

$$Z_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p$$

such that it would produce the maximum possible F-ratio in a one-way analysis of variance of Z_1, that is, the maximum possible ratio of the "between" to the "within"

groups sum-of-squares; these are given by $\mathbf{a'Ba}$, and $\mathbf{a'Wa}$ respectively.

The optimization problem in discriminant analysis is therefore to

$$\max(\lambda_1) = (\mathbf{a'Ba})/(\mathbf{a'Wa}).$$

To maximize λ_1, the derivative $\partial\lambda_1/\partial\mathbf{a'}$ is found and set equal to zero. After some manipulation, this gives

$$(\mathbf{B} - \lambda_1\mathbf{W})\mathbf{a} = 0$$

and λ_1 can be recognized as the first eigenvalue of the matrix $\mathbf{W^{-1}B}$ and $\mathbf{a}$ as the corresponding eigenvector.

When the number of groups is greater than 2, additional canonical dicriminant functions can be defined in much the same way.

An overall test, (Wilk's test, or Wilk's lambda) of the significance of the difference among group centroids (i.e. multivariate analysis of variance, or MANOVA) can be expressed in terms of the λ's:

$$\Lambda = \frac{|\mathbf{B}|}{|\mathbf{W}+\mathbf{B}|} = \frac{1}{1+\lambda_1} \times \frac{1}{1+\lambda_2} \times ... \times \frac{1}{1+\lambda_p}$$

and in turn a statistic based on Λ can be compared with the F distribution.