**What is Discriminant Function Analysis?**

Discriminant Function Analysis is used to determine which variables discriminate between two or more naturally occurring groups. The model is composed of a discriminant based on linear combinations of the predictor variables that provide the best discrimination between the groups. Discriminant Analysis is the appropriate statistical technique when the dependent variable is categorical and the independent variables are quantitative.

In other words, the basic purpose of discriminant analysis is to estimate the relationship between a single categorical dependent variable and a set of quantitative independent variables in establishing procedures for classifying individuals into groups on the basis of their scores on a set of independent variables.

To put it simply, the basic idea underlying discriminant function analysis is to determine whether groups differ with regard to the mean of a variable, and then to use that variable to predict group membership. If calculating the mean* of a variable is not possible then the procedure will not result in accurate groupings.

This approach to grouping statistical units has important assumptions.

These assumptions are:

1. It is assumed that the data (for the variables) represent a sample from a multivariate normal distribution. However, note that violations of the normality assumption are usually not "fatal," meaning, that the resultant significance tests etc. are still "trustworthy."

2. It is assumed that the variance/covariance matrices of variables are homogeneous across groups. Again, minor deviations are not that important.

3. Invalidity of significance tests occurs when the means for variables across groups are correlated with the variances or standard deviations. This pattern may occur if one group in the study contains a few extreme outliers.

4. Another assumption of discriminant function analysis is that the variables that are used to discriminate between groups are not completely redundant. As part of the computations involved in discriminant analysis, one inverts the variance/covariance matrix of the variables in the model. If any one of the variables is completely redundant with the other variables then the matrix is said to be ill-conditioned, and it cannot be inverted. For example, if a variable is the sum of three other variables that are also in the model, then the matrix is ill-conditioned. This is an important assumption of the procedure and deviations from it result in invalid conclusions (Debra Wetcher-Hendricks, 2011).
As Discriminant Function Analysis is used to determine which continuous variables discriminate between two or more naturally occurring groups, Logistic Regression answers the same questions as Discriminant Analysis and it is often preferred to discriminate analysis as it is more flexible in its assumptions and types of data that can be analysed.

Logistic regression can handle both categorical and continuous variables, and the predictors do not have to be normally distributed, linearly related, or of equal variance within each group (Tabachnick and Fidell 1996).

*The values in a categorical variable exist on a nominal scale: they each represent a logically separate concept, cannot necessarily be meaningfully ordered, and cannot be otherwise manipulated as numbers could be. Instead, valid operations are equivalence, set membership, and other set-related operations. As a result, the central tendency of a set of categorical variables is given by its mode; neither the mean nor the median can be defined (see e.g. Perriere and Thioulouse, 2003 and Abdi, 2007).