

Multivariate Analysis of Variance (MANOVA)

In *MANOVA*, there are in general g groups of observations, of sizes n_1, n_2, \dots, n_g , and p variables X_1, X_2, \dots, X_p that describe the observations. It is useful to express the variables as deviations, x 's, from the grand mean or centroid (over all groups). The vector of observations of the p variables, for the i th observation in the k th group is \mathbf{x}_{ki} , and these values can be decomposed into two components:

$$\mathbf{x}_{ki} = (\mathbf{m}_k - \mathbf{m}) + (\mathbf{X}_{ki} - \mathbf{m}_k)$$

where $(\mathbf{m}_k - \mathbf{m})$ is the deviation between the centroid of the k th group and the grand centroid, and $(\mathbf{X}_{ki} - \mathbf{m}_k)$ is the deviation between the i th observation in the k th group and the centroid for that group. The first term here could be thought of as analogous to the *systematic* component of some data, while the second term can be thought of as the *irregular* or *unpredictable* component.

As in univariate analysis of variance, the total sum of squares of the dependent variables (the x 's) can be decomposed into two parts:

$$\sum_{k=1}^g \sum_{i=1}^{n_k} \mathbf{x}_{ki} \mathbf{x}_{ki}' = \sum_{k=1}^g \sum_{i=1}^{n_k} (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})' + \sum_{k=1}^g \sum_{i=1}^{n_k} (\mathbf{X}_{ki} - \mathbf{m}_k)(\mathbf{X}_{ki} - \mathbf{m}_k)'$$

Each of the individual terms is a matrix, e.g.:

$$\mathbf{T} = \sum_{k=1}^g \sum_{i=1}^{n_k} \mathbf{x}_{ki} \mathbf{x}_{ki}'$$

is the “total” sums-of-squares matrix,

$$\mathbf{A} = \sum_{k=1}^g \sum_{i=1}^{n_k} (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})'$$

is the “among-groups” sum-of-squares matrix, and

$$\mathbf{W} = \sum_{k=1}^g \sum_{i=1}^{n_k} (\mathbf{X}_{ki} - \mathbf{m}_k)(\mathbf{X}_{ki} - \mathbf{m}_k)'$$

is the “within-groups” sum-of-squares matrix, and so

$$\mathbf{T} = \mathbf{A} + \mathbf{W}$$

A statistic that can be used to test the null hypothesis that the individual group centroids (the \mathbf{m}_k 's) are all equal is Wilk's Lambda,

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|}$$

where $|\mathbf{W}|$ and $|\mathbf{T}|$ are the determinants of the within-group and total sums-of-squares matrices, respectively. As the within-groups sums-of-squares gets smaller relative to the total sums-of-squares, the value of Λ decreases, which in practice also signals a decrease in the P -value of Λ . In other words, as Λ decreases we should be more inclined to reject the null hypothesis that the individual group centroids (the \mathbf{m}_k 's) are all equal.