A NOTE ON MULTIVARIATE PARALLEL REGRESSION

by

Bernhard N. Flury*
Purdue University and University of Berne

Technical Report #83-42

Department of Statistics
Purdue University

September 1983

*This work was done under contract no. 82.008.0.82 of the Swiss National Science Foundation.

Abstract: We analyze the multivariate linear regression model $\underset{\sim}{Y} = \underset{\sim}{Z}\,\underset{\sim}{B} + \underset{\sim}{U}$, where one column $\underset{\sim}{w}$ of $\underset{\sim}{Z}$ is an indicator variable. Relations between this model, two-group linear discriminant analysis, and a multiple linear regression model with $\underset{\sim}{w}$ as response variable are investigated. Some remarks on the case of several indicator variables are given.

Keywords: multivariate regression; multiple regression; discriminant analysis; Mahalanobis distance; Wilk's Λ; conditional mean difference.

AMS-classification: primary 62J05, 62H30;
secondary 62H15.

1. A multivariate parallel regression model

We consider the model defined by

$$\underset{\sim}{Y} = \underset{\sim}{Z}\,\underset{\sim}{B} + \underset{\sim}{U} \tag{1.1}$$

where

$$\underset{\sim}{Y}(\overline{n_1+n_2} \times \overline{p-q}) = \begin{pmatrix} Y_{q+1,1} & \cdots & Y_{p,1} \\ \vdots & & \vdots \\ Y_{q+1,n_1} & \cdots & Y_{p,n_1} \\ Y_{q+1,n_1+1} & \cdots & Y_{p,n_1+1} \\ \vdots & & \vdots \\ Y_{q+1,n_1+n_2} & \cdots & Y_{p,n_1+n_2} \end{pmatrix} \quad (q<p) \tag{1.2}$$

is an observed matrix of p - q response variables on each of $n_1 + n_2$ individuals,

$$\underset{\sim}{Z}(\overline{n_1+n_2} \times \overline{q+2}) = \begin{pmatrix} 1 & X_{1,1} & \cdots & X_{q,1} & W_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & X_{1,n_1} & \cdots & X_{q,n_1} & W_{n_1} \\ 1 & X_{1,n_1+1} & \cdots & X_{q,n_1+1} & W_{n_1+1} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & X_{1,n_1+n_2} & & X_{q,n_1+n_2} & W_{n_1+n_2} \end{pmatrix} \tag{1.3}$$

$$= (\underset{\sim}{1} : \underset{\sim}{X} : \underset{\sim}{W})$$

is a known matrix of rank q + 2,

$$\underset{\sim}{B}(\overline{q+2} \times \overline{p-q}) = \begin{pmatrix} \beta_{0,q+1} & \cdots & \beta_{0,p} \\ \beta_{1,q+1} & \cdots & \beta_{1,p} \\ \vdots & & \vdots \\ \beta_{q,q+1} & \cdots & \beta_{q,p} \\ \beta_{q+1,q+1} & \cdots & \beta_{q+1,p} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\beta}_0 \ (1 \times \overline{p-q}) \\ \underset{\sim}{\beta} \times (q \times \overline{p-q}) \\ \underset{\sim}{\beta}_w \ (1 \times \overline{p-q}) \end{pmatrix} \qquad (1.4)$$

is a matrix of unknown regression parameters, and $\underset{\sim}{U}(\overline{n_1+n_2} \times \overline{p-q})$ is a matrix of (unobserved) random disturbances, whose rows are uncorrelated, each with mean $\underset{\sim}{0}$ and common covariance matrix $\underset{\sim}{\Sigma}$. Without loss of generality we can assume that the columns of $\underset{\sim}{X}$ are all centered, i.e. $\underset{\sim}{X}'\underset{\sim}{1} = \underset{\sim}{0}$. The elements of $\underset{\sim}{w}$ are assumed to take only two values $c_1 = n_2/(n_1+n_2)$ and $c_2 = c_1 - 1 = -n_1/(n_1+n_2)$. Furthermore, we assume that the rows of the matrix $\underset{\sim}{Z}$ are ordered such that

$$w_i = \begin{cases} c_1 & \text{if} \quad 1 \le i \le n_1 \\ c_2 & \text{if} \quad n_1+1 \le i \le n_1+n_2 \end{cases} \qquad (1.5)$$

w is called a dummy-or indicator-variable, and the special choice (1.5) of its values simplifies some calculations.

The model (1.1) corresponds to $p - q$ simultaneous multiple regressions on $q + 1$ variables. Alternatively, we can look at it as a regression model for two samples of size n_1 and n_2 respectively, thus representing $2(p-q)$ multiple regressions on q variables, the two regression hyperplanes associated with each Y-variable being parallel. We call (1.1) therefore a multivariate parallel regression model. Note that for any fixed point $(x_1,\ldots,x_q) \in \mathbb{R}^q$ the values of the two regression functions of variable Y_j differ by $\beta_{q+1,j}$.

If we assume multivariate normality for the rows of $\underset{\sim}{U}$, it is well known that the maximum likelihood estimates of $\underset{\sim}{B}$ and $\underset{\sim}{\Sigma}$ are (Johnson and Wichern 1982, p. 324)

$$\hat{\underset{\sim}{B}} = (\underset{\sim}{Z}'\underset{\sim}{Z})^{-1}\underset{\sim}{Z}'\underset{\sim}{Y} = \begin{pmatrix} \hat{\underset{\sim}{\beta}}_0 \\ \hat{\underset{\sim}{\beta}}_x \\ \hat{\underset{\sim}{\beta}}_w \end{pmatrix} \tag{1.6}$$

and

$$\hat{\underset{\sim}{\Sigma}} = (n_1+n_2)^{-1}\, \hat{\underset{\sim}{U}}'\hat{\underset{\sim}{U}} \tag{1.7}$$

where

$$\hat{\underset{\sim}{U}}\ (\overline{n_1+n_2} \times \overline{p-q}) = \underset{\sim}{Y} - \underset{\sim}{Z}\,\hat{\underset{\sim}{B}} \tag{1.8}$$

is the matrix of estimated residuals. $\hat{\underset{\sim}{U}}'\hat{\underset{\sim}{U}}$ is often called the residual sum of squares and products (SSP) - matrix. Its distribution is Wishart, and it is stochastically independent of the parameter estimate $\hat{\underset{\sim}{B}}$ which has a multivariate normal distribution.

Let us now consider the hypothesis H_0: $\underset{\sim}{\beta}_w = \underset{\sim}{0}$. Under this hypothesis, each of the p - q pairs of parallel regression hyperplanes coincide to one hyperplane. To test H_0, we can use Wilk's Λ-statistic, which is the same as the likelihood-ratio statistic (Johnson and Wichern 1982, p. 327). Alternatively, a test can be constructed using the facts that

$$\underset{\sim}{\beta}_w \sim N_{p-q} (\underset{\sim}{\beta}_w, k\underset{\sim}{\Sigma}) \tag{1.9}$$

where k is the element in the last row and last column of $(Z'Z)^{-1}$, and

$$\hat{\underset{\sim}{U}}'\hat{\underset{\sim}{U}} \sim W_{p-q} (\underset{\sim}{\Sigma}, n_1+n_2-q-2) \tag{1.10}$$

(Mardia et al 1979, p. 160). Under H_0, the statistic

$$F = \frac{n_1+n_2-p-1}{p-q} \; \hat{\underset{\sim}{\beta}}_w (\hat{\underset{\sim}{U}}'\hat{\underset{\sim}{U}})^{-1} \hat{\underset{\sim}{\beta}}_w'/k \tag{1.11}$$

has the central F-distribution with $p - q$ and n_1+n_2-p-1 degrees of freedom (Mardia et al 1979, theorems 3.5.1. and 3.5.2.). This test statistic will later be seen to be the same as Wilk's Λ.

Let us now give some more details about the parameter estimates. Let $r = n_1 n_2/(n_1+n_2)$,

$$d_j = n_1^{-1} \sum_{i=1}^{n_1} X_{ji} - n_2^{-1} \sum_{i=n_1+1}^{n_1+n_2} X_{ji} \quad (j = 1,\dots,q). \tag{1.12}$$

Then $\underset{\sim}{d}_x = (d_1,\dots,d_q)'$ is the vector of mean differences for the x-variables, and

$$\underset{\sim}{Z}'\underset{\sim}{Z} = \begin{pmatrix} n_1+n_2 & \underset{\sim}{0} & 0 \\ \underset{\sim}{0} & \underset{\sim}{X}'\underset{\sim}{X} & r\underset{\sim}{d}_x \\ 0 & r\underset{\sim}{d}_x' & r \end{pmatrix} . \tag{1.13}$$

Let $(\underset{\sim}{Z}'\underset{\sim}{Z})^{-1} = \underset{\sim}{G} = \begin{pmatrix} g_0 & \underset{\sim}{0} & 0 \\ \underset{\sim}{0} & \underset{\sim}{G}_{11} & \underset{\sim}{G}_{12} \\ 0 & \underset{\sim}{G}_{21} & g_{22} \end{pmatrix}$; (1.14)

$$\overline{Y}_j^{(1)} = n_1^{-1} \sum_{i=1}^{n_1} Y_{ji} \,, \quad \overline{Y}_j^{(2)} = n_2^{-1} \sum_{i=n_1+1}^{n_1+n_2} Y_{ji} \; (j=q+1,\ldots,p);$$

$$\underset{\sim}{\overline{Y}}^{(1)} = (\overline{Y}_{q+1}^{(1)},\ldots,\overline{Y}_p^{(1)})' \; ; \quad \underset{\sim}{\overline{Y}}^{(2)} = (\overline{Y}_{q+1}^{(2)},\ldots,\overline{Y}_p^{(2)})' \;,$$

$$\underset{\sim}{d}_y = \underset{\sim}{\overline{Y}}^{(1)} - \underset{\sim}{\overline{Y}}^{(2)} \text{ and } \underset{\sim}{\overline{Y}} = (n_1+n_2)^{-1}(n_1\underset{\sim}{\overline{Y}}^{(1)}+n_2\underset{\sim}{\overline{Y}}^{(2)}).$$

Then we have

$$\underset{\sim}{Z}'\underset{\sim}{Y} = \begin{pmatrix} (n_1+n_2)\underset{\sim}{\overline{Y}}' \\ \underset{\sim}{X}'\underset{\sim}{Y} \\ r\underset{\sim}{d}'_y \end{pmatrix} \qquad (1.15)$$

and it follows easily that

$$\underset{\sim}{\hat{\beta}}_0 = \underset{\sim}{\overline{Y}}', \qquad (1.16)$$

the vector of overall means of the y-variables.

Using formula (A.2.4g) from Mardia et al (1979) we get

$$\begin{aligned} \underset{\sim}{\hat{\beta}}_x &= \underset{\sim}{G}_{11}(\underset{\sim}{X}'\underset{\sim}{Y} - r\underset{\sim}{d}_x\underset{\sim}{d}'_y) \\ &= (\underset{\sim}{X}'\underset{\sim}{X} - r\underset{\sim}{d}_x\underset{\sim}{d}'_x)^{-1}(\underset{\sim}{X}'\underset{\sim}{Y} - r\underset{\sim}{d}_x\underset{\sim}{d}'_y) \\ &= \underset{\sim}{A}_{xx}^{-1}\, \underset{\sim}{A}_{xy} \end{aligned} \qquad (1.17)$$

where

$$\underset{\sim}{A} = \begin{pmatrix} \underset{\sim}{A}_{xx} & \underset{\sim}{A}_{xy} \\ \underset{\sim}{A}_{yx} & \underset{\sim}{A}_{yy} \end{pmatrix} \tag{1.18}$$

is the sum of squares and products-matrix of the combined data matrix $(\underset{\sim}{X} : \underset{\sim}{Y})$, pooled over both groups. The last line of (1.17) follows easily by writing the typical element of $\underset{\sim}{A}_{xx}$ as the pooled covariance of two variables, and analogously for $\underset{\sim}{A}_{xy}$.

Using again (A.2.4g) from Mardia et al (1979) we get

$$\hat{\underset{\sim}{\beta}}_w = (1-r\underset{\sim}{d}_x' (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{d}_x)^{-1}\underset{\sim}{d}_y' - \underset{\sim}{d}_x' (\underset{\sim}{X}'\underset{\sim}{X}-r\underset{\sim}{d}_x\underset{\sim}{d}_x')^{-1}\underset{\sim}{X}'\underset{\sim}{Y} \; . \tag{1.19}$$

Noting that

$$(1-r\underset{\sim}{d}_x'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{d}_x)^{-1} = 1 + r\underset{\sim}{d}_x' (\underset{\sim}{X}'\underset{\sim}{X} - r\underset{\sim}{d}_x\underset{\sim}{d}_x')^{-1}\underset{\sim}{d}_x \tag{1.20}$$

it follows that

$$\begin{aligned} \hat{\underset{\sim}{\beta}}_w &= (1+r\underset{\sim}{d}_x'\underset{\sim}{A}_{xx}^{-1}\underset{\sim}{d}_x)\, \underset{\sim}{d}_y' - \underset{\sim}{d}_x'\underset{\sim}{A}_{xx}^{-1}\underset{\sim}{X}'\underset{\sim}{Y} \\ &= \underset{\sim}{d}_y' - \underset{\sim}{d}_x'\underset{\sim}{A}_{xx}^{-1} (\underset{\sim}{X}'\underset{\sim}{Y} - r\underset{\sim}{d}_x\underset{\sim}{d}_y') \\ &= \underset{\sim}{d}_y' - \underset{\sim}{d}_x'\underset{\sim}{A}_{xx}^{-1}\underset{\sim}{A}_{xy} = \underset{\sim}{d}_{y.x}' \end{aligned} \tag{1.21}$$

which is the conditional (sample) mean difference of $\underset{\sim}{Y}$, given $\underset{\sim}{X}$.

To analyze the SSP-matrix $\hat{\underset{\sim}{U}}'\hat{\underset{\sim}{U}}$, let $\hat{\underset{\sim}{u}}_i$, $\underset{\sim}{y}_i$ and $\underset{\sim}{x}_i$ $(1 \leq i \leq n_1+n_2)$ denote the i-th row of $\hat{\underset{\sim}{U}}$, $\underset{\sim}{Y}$ and $\underset{\sim}{X}$, respectively. Then

$$\hat{\underset{\sim}{u}}_i = \underset{\sim}{y}_i - \bar{\underset{\sim}{y}}' - \underset{\sim}{x}_i \underset{\sim}{A}_{xx}^{-1}\underset{\sim}{A}_{xy} - w_i d'_{y.x}$$

$$= \begin{cases} \underset{\sim}{y}_i - \bar{\underset{\sim}{y}}^{(1)'} - (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}}^{(1)'})\underset{\sim}{A}_{xx}^{-1}\underset{\sim}{A}_{xy} & \text{for } 1 \leq i \leq n_1 \\ \underset{\sim}{y}_i - \bar{\underset{\sim}{y}}^{(2)'} - (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}}^{(2)'})\underset{\sim}{A}_{xx}^{-1}\underset{\sim}{A}_{xy} & \text{for } n_1+1 \leq i \leq n_1+n_2 \end{cases} \tag{1.22}$$

and from $\hat{\underset{\sim}{U}}'\hat{\underset{\sim}{U}} = \underset{\sim}{Y}'\hat{U}$ it follows that the residual SSP-matrix is

$$\hat{U}'\hat{U} = \sum_{i=1}^{n_1+n_2} \underset{\sim}{y}_i'\hat{\underset{\sim}{u}}_i$$

$$= \sum_{i=1}^{n_1} \underset{\sim}{y}_i'(\underset{\sim}{y}_i-\bar{\underset{\sim}{y}}^{(1)'} - (\underset{\sim}{x}_i-\bar{\underset{\sim}{x}}^{(1)'})\underset{\sim}{A}_{xx}^{-1}\underset{\sim}{A}_{xy})$$

$$+ \sum_{i=n_1+1}^{n_1+n_2} \underset{\sim}{y}_i'(\underset{\sim}{y}_i-\bar{\underset{\sim}{y}}^{(2)'} - (\underset{\sim}{x}_i-\bar{\underset{\sim}{x}}^{(2)'})\ \underset{\sim}{A}_{xx}^{-1}\underset{\sim}{A}_{xy})$$

$$= \underset{\sim}{A}_{yy} - \underset{\sim}{A}_{yx}\underset{\sim}{A}_{xx}^{-1}\underset{\sim}{A}_{xy} = \underset{\sim}{A}_{yy.x}\ . \tag{1.23}$$

In order to write the F-statistic (1.11) in a different form, we apply formula A.2.4f of Mardia et al (1979) to g_{22}, the element in the last row and last column of (1.14), and get

$$g_{22} = r^{-1}(1 - r\underset{\sim}{d}_x'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{d}_x)^{-1}$$

$$= r^{-1} + \underset{\sim}{d}_x'\underset{\sim}{A}_{xx}^{-1}\underset{\sim}{d}_x \ . \qquad (1.24)$$

Thus the F-statistic (1.11) can be written as

$$F = \frac{n_1+n_2-p-1}{p-q} \cdot \frac{\underset{\sim}{d}_{y.x}'\underset{\sim}{A}_{yy.x}^{-1}\underset{\sim}{d}_{y.x}}{r^{-1} + \underset{\sim}{d}_x'\underset{\sim}{A}_{xx}^{-1}\underset{\sim}{d}_x} \qquad (1.25)$$

If we write $d' = (d_x' : d_y')$, then (1.25) becomes (see [4], formula 3.6.7)

$$F = \frac{n_1+n_2-p-1}{p-q} \cdot \frac{\underset{\sim}{d}'\underset{\sim}{A}^{-1}\underset{\sim}{d} - \underset{\sim}{d}_x\underset{\sim}{A}_{xx}^{-1}\underset{\sim}{d}_x}{r^{-1} + \underset{\sim}{d}_x'\underset{\sim}{A}_{xx}^{-1}d_x} \ . \qquad (1.26)$$

Let us now look at the Λ-statistic for the same problem. The residual SSP-matrix under H_0 is

$$\underset{\sim}{B}_{yy.x} = \underset{\sim}{B}_{yy} - \underset{\sim}{B}_{yx}\underset{\sim}{B}_{xx}^{-1}\underset{\sim}{B}_{xy} \qquad (1.27)$$

where

$$\underset{\sim}{B} = \begin{pmatrix} \underset{\sim}{B}_{xx} & \underset{\sim}{B}_{xy} \\ \underset{\sim}{B}_{yx} & \underset{\sim}{B}_{yy} \end{pmatrix} = \underset{\sim}{A} + r \begin{pmatrix} \underset{\sim}{d}_x\underset{\sim}{d}_x' & \underset{\sim}{d}_x\underset{\sim}{d}_y' \\ \underset{\sim}{d}_y\underset{\sim}{d}_x' & \underset{\sim}{d}_y\underset{\sim}{d}_y' \end{pmatrix} \qquad (1.28)$$

is the combined SSP-matrix of $(\underset{\sim}{X} : \underset{\sim}{Y})$, ignoring the group structure. Wilk's Λ-statistic is

$$\Lambda = \frac{|\underset{\sim}{A}_{yy.x}|}{|\underset{\sim}{B}_{yy.x}|} = \frac{|\underset{\sim}{A}|\ |\underset{\sim}{B}_{xx}|}{|\underset{\sim}{A}_{xx}|\ |\underset{\sim}{B}|}$$

$$= \frac{|\underset{\sim}{I}_q + \underset{\sim}{A}_{xx}^{-1} r \underset{\sim}{d}_x \underset{\sim}{d}_x'|}{|\underset{\sim}{I}_p + \underset{\sim}{A}^{-1} r \underset{\sim}{d}\underset{\sim}{d}'|}$$

$$= \frac{1 + r\underset{\sim}{d}_x' \underset{\sim}{A}_{xx}^{-1} \underset{\sim}{d}_x}{1 + r\underset{\sim}{d}' \underset{\sim}{A}^{-1} \underset{\sim}{d}} \tag{1.29}$$

by (A.2.3n) of Mardia et al (1979). The equivalence with the F-statistic follows now from

$$F = \frac{n_1 + n_2 - p - 1}{p-q} \cdot \frac{1 - \Lambda}{\Lambda} \ . \tag{1.30}$$

Note that, for q = 0, (1.11) is the same as the usual T^2 statistic for testing the equality of the mean vectors of two multivariate normal populations with identical covariance matrices.

2. The linear discriminant analysis model

The coefficients of the linear discriminant function between two normal populations are defined as $\underset{\sim}{\alpha} = \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\delta}$, where $\underset{\sim}{\delta} = \underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)}$ is the difference of mean vectors, and $\underset{\sim}{\Sigma}$ is the common covariance matrix. If $\underset{\sim}{\alpha}$ is partitioned in q and p-q components as $\underset{\sim}{\alpha}' = (\underset{\sim}{\alpha}_1' : \underset{\sim}{\alpha}_2')$, then the hypothesis

$$H_0: \ \underset{\sim}{\alpha}_2 = \underset{\sim}{0} \tag{2.1}$$

says that the discriminant function does not depend on the last p-q variables.

Partitioning $\underset{\sim}{\delta}' = (\underset{\sim}{\delta}_1' : \underset{\sim}{\delta}_2')$ and $\underset{\sim}{\Sigma} = \begin{pmatrix} \underset{\sim}{\Sigma}_{11} & \underset{\sim}{\Sigma}_{12} \\ \underset{\sim}{\Sigma}_{21} & \underset{\sim}{\Sigma}_{22} \end{pmatrix}$ anologously, we can define

$\Delta_p^2 = \underset{\sim}{\delta}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\delta}$ and $\Delta_q^2 = \underset{\sim}{\delta}_1'\underset{\sim}{\Sigma}_{11}^{-1}\underset{\sim}{\delta}_1$ as the Mahalanobis - distance between the two populations, based on p and q variables, respectively. Let $\underset{\sim}{\delta}_{2.1}$ denote the conditional mean difference of the last p-q variables, given the first q ones. Rao (1970) has shown that

$$H_0': \quad \Delta_p^2 = \Delta_q^2 \tag{2.2}$$

and

$$H_0'': \quad \underset{\sim}{\Delta}_{2.1} = \underset{\sim}{0} \tag{2.3}$$

are both equivalent to H_0.

Let us denote by $\underset{\sim}{d}' = (\underset{\sim}{d}_1' : \underset{\sim}{d}_2')$ the sample mean difference, by

$\underset{\sim}{S} = \begin{pmatrix} \underset{\sim}{S}_{11} & \underset{\sim}{S}_{12} \\ \underset{\sim}{S}_{21} & \underset{\sim}{S}_{22} \end{pmatrix}$ the usual (pooled) unbiased estimate of $\underset{\sim}{\Sigma}$ and by

$D_p^2 = \underset{\sim}{d}'\ \underset{\sim}{S}^{-1}\ \underset{\sim}{d}$, $D_q^2 = \underset{\sim}{d}_1'\ \underset{\sim}{S}_{11}^{-1}\ \underset{\sim}{d}_1$ the sample counterparts of Δ_p^2 and Δ_q^2, all based on samples of size n_1 and n_2, respectively. To test H_0, a statistic based on D_p^2 and D_q^2 can be used:

$$F = \frac{n_1+n_2-p-1}{p-q} \quad \frac{D_p^2 - D_q^2}{\dfrac{(n_1+n_2-2)(n_1+n_2)}{n_1n_2} + D_q^2} \tag{2.4}$$

which has the central F distribution with p-q and n_1+n_2-p-1 degrees of freedom under H_0. Mardia et al (1979) give a proof for this which is based on distribution theory for Wishart matrices. Of course, (2.4) is the same as (1.26). Rao (1970) has derived (2.4) using the equivalent condition (2.3). He uses essentially the same approach as our model (1.1), writing it in the form

$$E(\underset{\sim}{Y}|\underset{\sim}{X}) = \underset{\sim}{\alpha}_1 + \underset{\sim}{X}\underset{\sim}{B} \qquad \text{in population 1}$$

$$E(\underset{\sim}{Y}|\underset{\sim}{X}) = \underset{\sim}{\alpha}_2 + \underset{\sim}{X}\underset{\sim}{B} \qquad \text{in population 2,} \tag{2.5}$$

thus avoiding the indicator variable w. Testing for $\underset{\sim}{\alpha}_1 = \underset{\sim}{\alpha}_2$ in Rao's model is the same as testing for $\underset{\sim}{\beta}_w = \underset{\sim}{0}$ in (1.1). Using Wilk's Λ criterion as a test statistic, Rao's proof turns out to be algebraically simpler than the one given in section 1 of this paper. However, Rao's approach does not provide us with the interesting relations to a multiple linear regression model which are to be discussed in the following sections.

Rao's approach is based on the fact that if a random vector $\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix}$ is p-variate normal, the conditional mean of $\underset{\sim}{X}_1$, given $\underset{\sim}{X}_2$, is a linear function of $\underset{\sim}{X}_2$. Multivariate normality is a necessary assumption in the linear discriminant analysis approach. However, if we start with our model (1.1), nothing is said about the distribution of the X-variables, on which the regression approach actually conditions. All we need is the (p-q)-dimensional normality of the random disturbances. In this sense, the multivariate regressions approach is more general than the linear discriminant analysis approach.

3. The multiple regression approach

A formal analogy between linear discriminant analysis and a multiple regression model has been known since Fisher's(1936) first publication on the linear discriminant function. This analogy is often considered as a lucky algebraic coincidence. In this section we are going to show that it occurs in a natural way as a relation between a formal multiple regression model and the multivariate regression model of section 1.

The multiple regression model is

$$\underset{\sim}{w} = (\underset{\sim}{X} : \underset{\sim}{Y}) \begin{pmatrix} \underset{\sim}{\gamma}_x \\ \underset{\sim}{\gamma}_y \end{pmatrix} + \underset{\sim}{e} , \tag{3.1}$$

where $\underset{\sim}{w}$, $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are as in section 1. $\underset{\sim}{\gamma}_x$ and γ_y are q- and (p-q) - vectors of regression coefficients, and $\underset{\sim}{e}$ is formally the (n_1+n_2)-vector of residuals. We assume for simplicity that $\underset{\sim}{Y}'\underset{\sim}{1} = \underset{\sim}{0}$ as well as $\underset{\sim}{X}'\underset{\sim}{1} = \underset{\sim}{0}$. (Alternatively, we can assume that $\underset{\sim}{X}$ is of dimension $\overline{n_1+n_2} \times \overline{q+1}$ with a first row $\underset{\sim}{1}$, and $\underset{\sim}{\gamma}_x$ has dimension q+1, thus allowing for an intercept). We wish to test for $\underset{\sim}{\gamma}_y = \underset{\sim}{0}$, that is, we wish to compare model (3.1) with the restricted model

$$\underset{\sim}{w} = \underset{\sim}{X}\, \underset{\sim}{\gamma}^* + \underset{\sim}{e} . \tag{3.2}$$

Applying least squares estimation to (3.1) and (3.2) yields

$$\hat{\underset{\sim}{\gamma}} = \begin{pmatrix} \underset{\sim}{\gamma}_x \\ \underset{\sim}{\gamma}_y \end{pmatrix} = \left[\begin{pmatrix} \underset{\sim}{X}' \\ \underset{\sim}{Y}' \end{pmatrix} (\underset{\sim}{X} : \underset{\sim}{Y}) \right]^{-1} \begin{pmatrix} \underset{\sim}{X}' \\ \underset{\sim}{Y}' \end{pmatrix} \underset{\sim}{w} \tag{3.3}$$

for the model with p variables, and

$$\hat{\gamma}^* = (X'X)^{-1}X'w \tag{3.4}$$

for the restricted model. The residual sum of squares is

$$SS(p) = w'w - w'(X : Y)\,\hat{\gamma}$$

$$= w'w - w'(X : Y) \begin{pmatrix} X'X & X'Y \\ Y'X & Y'Y \end{pmatrix}^{-1} \begin{pmatrix} X' \\ Y' \end{pmatrix} w \tag{3.5}$$

and

$$SS(q) = w'w - w'X\,(X'X)^{-1}\,X'w \tag{3.6}$$

for models (3.1) and (3.2), respectively. The ratio of the two sums of squares is

$$\Lambda^* = \frac{SS(q)}{SS(p)} = \frac{[SS(p)]^{-1}}{[SS(q)]^{-1}} . \tag{3.7}$$

Noting that $w'w = r$, $X'w = rd_x$, and applying formula A.2.4f of Mardia et al (1979), we get

$$[SS(q)]^{-1} = r^{-1} + d_x' \;(X'X - rd_x d_x')^{-1}\, dx$$

$$= r^{-1} + d_x'\, A_{xx}^{-1} d_x \tag{3.8}$$

and analogously

$$[SS(p)]^{-1} = r^{-1} + d'A^{-1}d . \tag{3.9}$$

Assuming that we allow for an intercept in models (3.1) and (3.2), the F-statistic for the hypothesis $\underset{\sim}{\gamma}_y = \underset{\sim}{0}$ is

$$F = \frac{n_1+n_2-p-1}{p-q} \left(\frac{SS(q)}{SS(p)} - 1\right)$$

$$= \frac{n_1+n_2-p-1}{p-q} \cdot \frac{\underset{\sim}{d}'A\underset{\sim}{d} - \underset{\sim}{d}_x'A_{xx}^{-1}\underset{\sim}{d}_x}{r^{-1}+\underset{\sim}{d}_x'A_{xx}^{-1}\underset{\sim}{d}_x} \quad . \tag{3.10}$$

This is the same as (1.26). Note that, to establish (3.10), we did not use the proportionality between the regression estimates $\hat{\gamma}$ and the sample discriminant function, as shown, e.g., by Lachenbruch (1975, p. 17).

The relation between the parameter estimates $\hat{\underset{\sim}{\gamma}}$ of the multiple regression model and the estimates $\hat{\underset{\sim}{\beta}}_w' = \underset{\sim}{d}_{y.x}$ of the multivariate model (1.1) can also be established easily. First note that

$$(\underset{\sim}{X}'\underset{\sim}{X})^{-1} = (A_{xx} + r\underset{\sim}{d}_x\underset{\sim}{d}_x')^{-1}$$

$$= A_{xx}^{-1} - A_{xx}^{-1}\, r\underset{\sim}{d}_x\, (1 + r\underset{\sim}{d}_x'A_{xx}^{-1}\underset{\sim}{d}_x)^{-1}\underset{\sim}{d}_x'A_{xx}^{-1} \quad . \tag{3.11}$$

This yields

$$\hat{\underset{\sim}{\gamma}}^* = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'w = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}r\underset{\sim}{d}_x$$

$$= A_{xx}^{-1}r\underset{\sim}{d}_x \left(1 - \frac{r\underset{\sim}{d}_x'A_{xx}^{-1}\underset{\sim}{d}_x}{1 + r\underset{\sim}{d}_x'A_{xx}^{-1}\underset{\sim}{d}_x}\right)$$

$$= \frac{1}{1+r\underset{\sim}{d}_x'A_{xx}^{-1}\underset{\sim}{d}_x}\; A_{xx}^{-1}r\underset{\sim}{d}_x . \tag{3.12}$$

For the model with p variables we get anologously

$$\hat{\gamma} = \begin{pmatrix} \hat{\gamma}_x \\ \hat{\gamma}_y \end{pmatrix} = \frac{r}{1+rd'A^{-1}d} \; A^{-1}d \tag{3.13}$$

and therefore

$$\hat{\gamma}_y = A^{-1}_{yy.x} \; (d_y - A_{yx}A^{-1}_{xx}d_x) \; \cdot \; \frac{1}{r^{-1}+d'A^{-1}d}$$

$$= \frac{1}{r^{-1}+d'A^{-1}d} \; A^{-1}_{yy.x}\hat{\beta}'_w \; . \tag{3.14}$$

It is worth noting that $\hat{\gamma}_y$ can be written (from 3.3) as

$$\hat{\gamma}_y = (Y'Y - Y'X\,(X'X)^{-1}X'Y)^{-1}(Y'w - Y'X(X'X)^{-1}X'w). \tag{3.15}$$

Since $\hat{B}_0 = (X'X)^{-1}X'Y$ is the matrix of regression coefficients of Y on X, ignoring group structure, the first term in (3.15) is the inverse of the residual SSP-matrix of model (1.1) under H_0: $\beta_w = 0$. Similarly, the second term is $r(d_y - \hat{B}_0'd_x)$. This shows again the strong relation between models (1.1) and (3.1). However, the proof of (3.14), using (3.15), seems more difficult than the one given above.

Another indication of the relation between the two models can be seen in the fact that the estimate (3.4) $\hat{\gamma}^* = r(X'X)^{-1}dx$ of the restricted model (3.2) appears in (1.19) as well as in (1.24). Interestingly, all these relations are somehow "cross-wise": quantities encountered in the analysis of the unrestricted model in this section are related to quantities occurring in the restricted model of section 1 (see 3.15), while the restricted model (3.2) is is rather related to the unrestricted model of section 1. See also (3.7) for a similar phenomenon involving the residual sums of squares.

4. The k-sample case

The multivariate parallel regression model of section 1 has a straightforward generalization to the case of $k > 2$ samples by using $k - 1$ indicator variables (column vectors) $\underset{\sim}{w}_1, \underset{\sim}{w}_2, \ldots, \underset{\sim}{w}_{k-1}$, $\underset{\sim}{w}_j$ $(j=1,\ldots,k-1)$ being defined by

$$w_{ji} = \begin{cases} \dfrac{\sum_{\ell \neq j} n_\ell}{\sum_{\ell=1}^{k} n_\ell} & \text{if the i-th data vector belongs to sample j} \\[2ex] \dfrac{-n_j}{\sum_{\ell=1}^{k} n_\ell} & \text{otherwise.} \end{cases} \tag{4.1}$$

The code (4.1) has the advantage that the two values differ by unity, and that $\sum_{i=1}^{n_1+\ldots+n_k} w_{ji} = 0$ $(j=1,\ldots,k-1)$, which simplifies the algebraic calculations. In practical situations, however, any two different code values (e.g. 0 and 1) do the job as well.

If we assume $\underset{\sim}{X}'\underset{\sim}{1} = \underset{\sim}{0}$ and $\underset{\sim}{Y}'\underset{\sim}{1} = \underset{\sim}{0}$ in order to avoid an intercept term, we can write the model analogous to (1.1) as

$$\underset{\sim}{Y} = (\underset{\sim}{X} : \underset{\sim}{w}) \begin{pmatrix} \underset{\sim}{\beta}_x \\ \underset{\sim}{\beta}_w \end{pmatrix} + \underset{\sim}{U} \tag{4.2}$$

where $\underset{\sim}{Y}$, $\underset{\sim}{X}$, $\underset{\sim}{\beta}_x$ and $\underset{\sim}{U}$ are as in section 1, (with n_1+n_2 replaced by $\sum_{j=1}^{k} n_j$), $\underset{\sim}{\beta}_w$ is a $\overline{k-1} \times \overline{p-q}$ matrix of unknown parameters, and $\underset{\sim}{w} = (\underset{\sim}{w}_1 : \ldots : \underset{\sim}{w}_{k-1})$ is a $n \times \overline{k-1}$ matrix of dummy variables. Wilk's Λ-statistic for testing $\underset{\sim}{\beta}_w = \underset{\sim}{0}$ is readily available as

$$\Lambda = \frac{\left| \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{Y}'(\underset{\sim}{X}:\underset{\sim}{w}) \begin{pmatrix} \underset{\sim}{X}'\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{w} \\ \underset{\sim}{w}'\underset{\sim}{X} & \underset{\sim}{w}'\underset{\sim}{w} \end{pmatrix}^{-1} \begin{pmatrix} \underset{\sim}{X}' \\ \underset{\sim}{w}' \end{pmatrix} \underset{\sim}{Y} \right|}{\left| \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{Y}'\underset{\sim}{X} (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}'\underset{\sim}{Y} \right|} . \qquad (4.3)$$

The formal multiple regression analog (3.1) is this time itself a multivariate model and can be written as

$$\underset{\sim}{w} = (\underset{\sim}{X} : \underset{\sim}{Y}) \begin{pmatrix} \underset{\sim}{\gamma}_x \\ \underset{\sim}{\gamma}_y \end{pmatrix} + \underset{\sim}{E} \qquad (4.4)$$

where $\underset{\sim}{w}$, $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are as above, $\gamma = \begin{pmatrix} \underset{\sim}{\gamma}_x \\ \underset{\sim}{\gamma}_y \end{pmatrix}$ is a $p \times \overline{k-1}$ matrix of parameters, and $\underset{\sim}{E}$ is formally the matrix of disturbances. The analog to the ratio of the two residual sums of squares (3.7) for testing $\underset{\sim}{\gamma}_y = \underset{\sim}{0}$ is

$$\Lambda^* = \frac{\left| \underset{\sim}{w}'\underset{\sim}{w} - \underset{\sim}{w}'(\underset{\sim}{X}:\underset{\sim}{Y}) \begin{pmatrix} \underset{\sim}{X}'\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{Y} \\ \underset{\sim}{Y}'\underset{\sim}{X} & \underset{\sim}{Y}'\underset{\sim}{Y} \end{pmatrix}^{-1} \begin{pmatrix} \underset{\sim}{X}' \\ \underset{\sim}{Y}' \end{pmatrix} \underset{\sim}{w} \right|}{\left| \underset{\sim}{w}'\underset{\sim}{w} - \underset{\sim}{w}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{w} \right|} . \qquad (4.5)$$

In contrast to section 3, Λ^* is in general not equal to Λ, and so the analogy ends here. Also the relation (3.14) between $\hat{\underset{\sim}{\beta}}_w$ and $\hat{\underset{\sim}{\gamma}}_y$ becomes much more complicated, since there is a different factor of proportionality for each column of $\hat{\underset{\sim}{\gamma}}_y$.

There is one exception, however: the case q = 0. In this case there are no X-variables, and the models (4.2) and (4.4) reduce to

$$\underset{\sim}{Y} = \underset{\sim}{w}\, \underset{\sim}{\beta}_w + \underset{\sim}{U} \tag{4.6}$$

and

$$\underset{\sim}{w} = \underset{\sim}{Y}\, \underset{\sim}{\gamma}_y + \underset{\sim}{E}\ , \tag{4.7}$$

respectively. The Λ-statistic (4.3) for testing $\underset{\sim}{\beta}_w = \underset{\sim}{0}$ becomes

$$\Lambda = \frac{|\underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{Y}'\underset{\sim}{w}\,(\underset{\sim}{w}'\underset{\sim}{w})^{-1}\underset{\sim}{w}'\underset{\sim}{Y}|}{|\underset{\sim}{Y}'\underset{\sim}{Y}|}\ , \tag{4.8}$$

which is simply a test for equality of the mean vectors of k multivariate normal populations. The formal analog (4.5) becomes

$$\Lambda^* = \frac{|\underset{\sim}{w}'\underset{\sim}{w} - \underset{\sim}{w}'\underset{\sim}{Y}(\underset{\sim}{Y}'\underset{\sim}{Y})^{-1}\underset{\sim}{Y}'\underset{\sim}{w}|}{|\underset{\sim}{w}'\underset{\sim}{w}|}$$

$$= |\underset{\sim}{I}_{k-1} - (\underset{\sim}{w}'\underset{\sim}{w})^{-1}\underset{\sim}{w}'\underset{\sim}{Y}(\underset{\sim}{Y}'\underset{\sim}{Y})^{-1}\underset{\sim}{Y}\underset{\sim}{w}|$$

$$= |\underset{\sim}{I}_p - (\underset{\sim}{Y}'\underset{\sim}{Y})^{-1}\underset{\sim}{Y}'\underset{\sim}{w}(\underset{\sim}{w}'\underset{\sim}{w})^{-1}\underset{\sim}{w}'\underset{\sim}{Y}| = \Lambda \tag{4.9}$$

by (A.2.3n) of Mardia et al (1979). Note that to prove (4.9) we didn't use the special structure of $\underset{\sim}{w}$. It is, in fact, a more general result: if X_1 and X_2 are random vectors, measured on the same objects, then testing their linear relationship by the regression of $\underset{\sim}{X}_1$ on $\underset{\sim}{X}_2$ is the same as by the regression of $\underset{\sim}{X}_2$ on $\underset{\sim}{X}_1$.

5. Final remarks

The relations between the linear two-group discriminant analysis model (section 2) and a multiple regression model, regressing an indicator variable w on the measured variables, (section 3) have been known long ago, and they are fairly easy to establish (see Lachenbruch 1975, p. 17-19). Despite this, these relations seem somehow artificial, and no deeper reason has been given for their existence. However, if we start with the multivariate regression model of section 1, looking at the problem of estimating and testing conditional mean differences, then these relations appear much less as a lucky coincidence, since both models have a strong relationship with the more general multivariate model.

In the two sample case, the relations between the three models are of considerable practical importance, since software for multiple linear regression is easily available. In the k-sample case there is no such practical advantage, since the model (4.4) is itself multivariate. However, we can still learn something from this case: Suppose that we deal with $k > 2$ samples, but we are mainly interested in a contrast between two groups defined by the values of a binary variable, say w_1. To represent the group structure, we need also $k - 2$ binary variables $w_2,\ldots,w_{k-1}$. The multivariate regression setup is therefore given by (4.2). However, since we are only interested in tests concerning w_1, we can write the model in the form (1.1), where $w_2,\ldots,w_{k-1}$ appear as binary variables in the matrix $\underset{\sim}{X}$. Since the multivariate regression model conditions on the data matrix $\underset{\sim}{X}$, this will not affect the correctness of the results, and the multiple regression approach (section 3) can still be used. This can be viewed as a justification for including binary variables in a linear two-group discriminant analysis.

Acknowledgement

This work was done under contract no. 82.008.0.82 of the Swiss National Science Foundation.

References

FISHER, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. Ann. Eugen. 7, 179-188.

JOHNSON, R.A. and WICHERN, D.W. (1982). Applied Multivariate Statistical Analysis. Prentice-Hall, Inc., Englewood Cliffs (NJ).

LACHENBRUCH, P.A. (1975). Discriminant Analysis. Hafner Press, New York.

MARDIA, K.V., KENT, J.T. and BIBBY, J.M. (1979). Multivariate Analysis. Academic Press, London.

RAO, C.R. (1970). Inference on Discriminant Function Coefficients. In Essays in Probability and Statistics (R.C. Bose et al, Eds.), pp. 587-602. The University of North Carolina Press, Chapel Hill.