T^2-TESTS, LINEAR TWO-GROUP DISCRIMINANT ANALYSIS, AND THEIR COMPUTATION BY LINEAR REGRESSION+

by

Bernhard N. Flury*
Purdue University

Hans Riedwyl
University of Berne

Technical Report #82-42

+) revised version of "The Use of Multiple Regression to Perform Discriminant Analysis and T^2 - Tests."

*) Research supported by grant #82.008.0.82 of the Swiss National Science Foundation

Abstract:

Using a general theorem on decomposition of Mahalanobis distance, we give first an integrated view of some T^2-tests. Then we expose the relation between the linear discriminant function and a regression model which was already introduced by Fisher (1936). We show that this relation can be generalized to include the one-sample T^2-test, T^2-tests when only the sample mean is available in one group, and tests for redundancy of variables. Practical and didactical advantages of the regression approach are outlined.

key words: Mahalanobis distance; partial F; multivariate location test; redundant variables; conditional mean difference.

1. INTRODUCTION

Throughout this paper we will use the following abbreviations for books that are to be quoted frequently:

A = Anderson (1958)

JW = Johnson and Wichern (1982)

K = Karson (1982)

L = Lachenbruch (1975)

M = Morrison (1976)

MKB = Mardia, Kent and Bibby (1979)

R = Rao (1973)

SC = Srivastava and Carter (1983)

For instance, (A 375) will stand for (Anderson 1958, p. 375).

Since Fisher (1936) introduced the concept of linear discriminant analysis (DA), it is well known that the linear discriminant function of two samples can be computed by means of multiple linear regression (LR). Fisher's method consists of adding to the measured variables $X_1,\ldots,X_p$ a binary code variable w indicating the group membership. Then the linear regression function of w on $X_1,\ldots,X_p$ is (up to an additive constant) proportional to the linear discriminant function (A 140, K 170, L 17). Moreover, the significance of the discriminant function (L 19) as well as significance of single variables in the discriminant function can be tested by the regression solution.

This paper has two purposes: First, we show that the known relations between LR and DA can easily be generalized to the situation where only the mean is available in one sample. Second, we give some reasons why the stated relations are actually more than "lucky coincidences", and try to outline how they can be used to help practitioners to understand T^2-tests and linear discriminant functions.

To illustrate some parts of this paper which provide new material, we will use data taken on two samples of adult male Tibetans. The variables are:

ST = stature

LH = length of the head

WH = width of the head

WZA = width of zygomatic arch

MFH = morphological facial height.

The first sample, measured by Mullis (1982), consists of $n_1 = 44$ individuals from central Tibet. We will compare this sample with the mean vector of a sample taken by Prince Peter of Greece and Denmark (1966). Unfortunately Prince Peter, unaware of multivariate statisticians' needs, published only mean vectors, and the raw data are no longer available. The sample to be compared with the data of Mullis consists of $n_2 = 51$ individuals of the northeastern race Amdo. Table 1 shows summary statistics for the two samples.

2. SOME MULTIVARIATE TESTS BASED ON CALCULATION OF MAHALANOBIS DISTANCE

2.1. Mahalanobis distance between mean vectors, and the linear discriminant function

Let $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ denote independent p-dimensional random vectors with mean vectors $\underset{\sim}{\mu}$ and $\underset{\sim}{\nu}$, respectively, and with common covariance matrix $\underset{\sim}{\Sigma}$ (assumed to be non-singular). The coefficients of the linear discriminant function between $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are defined as the elements of the vector

$$\underset{\sim}{\alpha} = \underset{\sim}{\Sigma}^{-1}\underset{\sim}{\delta}, \tag{2.1}$$

with

$$\underset{\sim}{\delta} = \underset{\sim}{\mu} - \underset{\sim}{\nu} \tag{2.2}$$

(A 134, JW 464, K 164, L 11, M 235, MKB 303, R 575, SC 232). The Mahalanobis distance between the mean vectors $\underset{\sim}{\mu}$ and $\underset{\sim}{\nu}$ is defined as the quadratic form

$$\Delta_p^2 = \underset{\sim}{\alpha}'\underset{\sim}{\delta} = \underset{\sim}{\delta}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\delta} \tag{2.3}$$

(A 56, JW 467, K 166, L 12, M 235, MKB 31, SC 232). The index p in (2.3) indicates that the calculation of Mahalanobis distance is based on p variables. Using this terminology, the following three forms of the null hypothesis $\underset{\sim}{\mu} = \underset{\sim}{\nu}$ are equivalent:

$$H_0: \underset{\sim}{\delta} = \underset{\sim}{0}; \quad H_0': \underset{\sim}{\alpha} = \underset{\sim}{0}; \quad H_0'': \Delta_p^2 = 0. \tag{2.4}$$

In applied multivariate analysis it is often important not only to reject an overall null hypothesis, but also to identify variables which can be discarded from the analysis without loss of information. In the context of discriminant functions we may wish to know whether some coefficients of the linear discriminant function are zero. To formalize this hypothesis, we introduce the following notation: Partition $\underset{\sim}{\delta}$, $\underset{\sim}{\alpha}$ and $\underset{\sim}{\Sigma}$ as

$$\underset{\sim}{\delta} = \begin{pmatrix} \underset{\sim}{\delta}_1 \\ \underset{\sim}{\delta}_2 \end{pmatrix}; \quad \underset{\sim}{\alpha} = \begin{pmatrix} \underset{\sim}{\alpha}_1 \\ \underset{\sim}{\alpha}_2 \end{pmatrix}; \quad \underset{\sim}{\Sigma} = \begin{pmatrix} \underset{\sim}{\Sigma}_{11} & \underset{\sim}{\Sigma}_{12} \\ \underset{\sim}{\Sigma}_{21} & \underset{\sim}{\Sigma}_{22} \end{pmatrix}, \tag{2.5}$$

where $\underset{\sim}{\delta}_1$ has q components ($0 \leq q < p$) and $\underset{\sim}{\delta}_2$ has p-q components, and $\underset{\sim}{\alpha}$ and $\underset{\sim}{\Sigma}$ are partitioned analogously. The hypothesis of redundancy of the last p-q variables (or sufficiency of the first q variables) can be written as

$$H_q: \underset{\sim}{\alpha}_2 = \underset{\sim}{0}. \tag{2.6}$$

Note that for $q = 0, H_q$ becomes the overall null hypothesis H_0. Assuming multivariate normality for both random vectors, we denote by

$$\underset{\sim}{\delta}_{2.1} = \underset{\sim}{\delta}_2 - \underset{\sim}{\Sigma}_{21}\underset{\sim}{\Sigma}_{11}^{-1}\underset{\sim}{\delta}_1 \tag{2.7}$$

and

$$\underset{\sim}{\Sigma}_{22.1} = \underset{\sim}{\Sigma}_{22} - \underset{\sim}{\Sigma}_{21}\underset{\sim}{\Sigma}_{11}^{-1}\underset{\sim}{\Sigma}_{12} \tag{2.8}$$

the mean difference and covariance matrix of the conditional distribution of the last p - q variables, given the first q ones (A 29, JW 135, K 57, M 92, MKB 63, R 522, SC 30). The fact that $\underset{\sim}{\delta}_{2.1}$ does not depend on the values taken by the first q variables is due to the assumption of identical covariance matrices for both random vectors. Denote by

$$\Delta_q^2 = \underset{\sim}{\delta}_1'\underset{\sim}{\Sigma}_{11}^{-1}\underset{\sim}{\delta}_1 \tag{2.9}$$

the Mahalanobis distance based on q variables. It can be shown that (MKB 78)

$$\Delta_p^2 = \underset{\sim}{\delta}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\delta} = \underset{\sim}{\delta}_1'\underset{\sim}{\Sigma}_{11}^{-1}\underset{\sim}{\delta}_1 + \underset{\sim}{\delta}_{2.1}'\underset{\sim}{\Sigma}_{22.1}^{-1}\underset{\sim}{\delta}_{2.1}$$

$$= \Delta_q^2 + (\Delta_p^2 - \Delta_q^2), \tag{2.10}$$

and therefore the following hypotheses are equivalent to H_q:

$$H_q': \underset{\sim}{\delta}_{2.1} = 0; \quad H_q'': \Delta_p^2 = \Delta_q^2 \tag{2.11}$$

(L 27, MKB 78). This equivalence is also given in Rao (1970), but it antedates Rao's earlier work. For $q = 0$ it is convenient to set $\underset{\sim}{\delta}_{2.1} = \underset{\sim}{\delta}$ and $\Delta_0^2 = 0$ by definition.

2.2. Testing Hypotheses about the Linear Discriminant Function

In almost all practical situations the parameters $\underset{\sim}{\delta}$ and $\underset{\sim}{\Sigma}$ are unknown and have to be estimated. Suppose we have measured a random vector $\underset{\sim}{d}$ (typically a difference of sample mean vectors), which has a multivariate normal distribution with mean vector $\underset{\sim}{\delta}$ and covariance matrix $r\underset{\sim}{\Sigma}$ for some $r > 0$:

$$\underset{\sim}{d} \sim N_p(\underset{\sim}{\delta}, r\underset{\sim}{\Sigma}). \tag{2.12}$$

Suppose furthermore that we have a realization of a positive definite symmetric random matrix $\underset{\sim}{S}$, which is independent of $\underset{\sim}{d}$ and has the Wishart-distribution with m degrees of freedom and parameter matrix $\underset{\sim}{\Sigma}/m$ (typically a sample covariance matrix in the usual unbiased form):

$$\underset{\sim}{S} \sim W_p(m, \frac{1}{m}\underset{\sim}{\Sigma}). \tag{2.13}$$

The reason for this rather general setup is that we can construct one single test for all hypotheses being treated in this paper, and then specify r and m for the various situations. Let us now partition the statistics $\underset{\sim}{d}$ and $\underset{\sim}{S}$ analogously to $\underset{\sim}{\delta}$ and $\underset{\sim}{\Sigma}$ as

$$\underset{\sim}{d} = \begin{pmatrix} \underset{\sim}{d}_1 \\ \underset{\sim}{d}_2 \end{pmatrix}; \quad S = \begin{pmatrix} \underset{\sim}{S}_{11} & \underset{\sim}{S}_{12} \\ \underset{\sim}{S}_{21} & \underset{\sim}{S}_{22} \end{pmatrix} . \tag{2.14}$$

Then we define the sample Mahalanobis distance, based on p and q variables respectively, as

$$D_p^2 = \underset{\sim}{d}' \underset{\sim}{S}^{-1} \underset{\sim}{d} \tag{2.15}$$

and

$$D_q^2 = \underset{\sim}{d}_1' \underset{\sim}{S}_{11}^{-1} \underset{\sim}{d}_1 . \tag{2.16}$$

We can also define the vector of coefficients of the sample linear discriminant function as

$$\underset{\sim}{a} = \begin{pmatrix} \underset{\sim}{a}_1 \\ \underset{\sim}{a}_2 \end{pmatrix} = \underset{\sim}{S}^{-1} \underset{\sim}{d} . \tag{2.17}$$

In order to construct a test for H_q, a theorem can be used which was given by Rao (1970, p. 592; see also L 28) in a slightly less general form.

Theorem (MKB 78): With the above notation and under the hypothesis H_q: $\Delta_p^2 = \Delta_q^2$, the statistic

$$F(H_q) = \frac{m-p+1}{p-q} \cdot \frac{D_p^2 - D_q^2}{mr + D_q^2} \tag{2.18}$$

has the central F distribution with p - q and m - p + 1 degrees of freedom, and F is stochastically independent of D_q^2.

If we put $D_0^2 = 0$ by definition, this theorem holds also for the case $q = 0$.

It is through the statistic (2.18) that the interesting relations between tests in LR and T^2-tests can most easily be established. The theorem shows also the important role of the notion of Mahalanobis distance, which is not sufficiently recognized in most current texts on multivariate statistical analysis (exception: JW).

2.3. Two Sample Discrimination/Two Sample T^2-Test

The most frequent application of the above theory occurs in the two sample case. It is assumed that samples of size n_1 and n_2 are taken from $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$ and $\underset{\sim}{Y} \sim N_p(\underset{\sim}{\nu},\underset{\sim}{\Sigma})$, respectively. Then we have

$$\underset{\sim}{d} = \underset{\sim}{\bar{x}} - \underset{\sim}{\bar{y}} \sim N_p(\underset{\sim}{\delta}, \frac{n_1+n_2}{n_1 n_2} \underset{\sim}{\Sigma}) \tag{2.19}$$

and

$$\underset{\sim}{S} = \frac{1}{n_1+n_2-2}[(n_1-1)\underset{\sim}{S}_1+(n_2-1)\underset{\sim}{S}_2] \sim W_p(n_1+n_2-2, \frac{1}{n_1+n_2-2}\underset{\sim}{\Sigma}), \tag{2.20}$$

where $\underset{\sim}{\bar{x}}$ and $\underset{\sim}{\bar{y}}$ are the sample mean vectors, and $\underset{\sim}{S}_1$ and $\underset{\sim}{S}_2$ are the usual unbiased sample covariance matrices. Note that the assumptions of the theorem hold with $r = (n_1+n_2)/n_1 n_2$ and $m = n_1+n_2-2$, since $\underset{\sim}{S}$ and $\underset{\sim}{d}$ are independent (A 53, JW 148, K 77, M 102, MKB 66, R 537, SC 33). For $q = 0$ the statistic (2.18) reduces to

$$F(H_0) = \frac{(n_1+n_2-p-1)n_1 n_2}{p(n_1+n_2)(n_1+n_2-2)} D_p^2 = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)} T^2, \tag{2.21}$$

where T^2 is the well-known Hotelling T^2-statistic for the two sample case (A 109, JW 239, K 95, M 137, MKB 76, SC 48).

2.4. One Sample T^2-Test

In the one sample case, the mean vector $\underset{\sim}{\nu}$ (but not the covariance matrix $\underset{\sim}{\Sigma}$) of the second model is assumed to be known and is most often denoted by $\underset{\sim}{\mu}_0$ rather than by $\underset{\sim}{\nu}$. The statistic $\underset{\sim}{d}$ is

$$\underset{\sim}{d} = \bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0 \sim N_p(\underset{\sim}{\delta}, \frac{1}{n_1}\underset{\sim}{\Sigma}). \tag{2.22}$$

For $\underset{\sim}{S}$ we take the sample covariance matrix from the X-sample,

$$\underset{\sim}{S} = \underset{\sim}{S}_1 \sim W_p(n_1-1, \frac{1}{n_1-1}\underset{\sim}{\Sigma}). \tag{2.23}$$

Though most authors do not speak of a discriminant function in the one sample case, it is well defined by (2.17). The notions of discriminant function and Mahalanobis distance are very useful here to understand what testing for redundant variables means. To test H_q, (2.18) can be used with $r = 1/n_1$ and $m = n_1-1$. For $q = 0$, we get

$$F(H_0) = \frac{(n_1-p)n_1}{p(n_1-1)} D_p^2 = \frac{n_1-p}{p(n_1-1)} T^2, \tag{2.24}$$

where T^2 is the well known one sample T^2-statistic (A 103, JW 180, K 90, M 131, MKB 125, SC 41). Note that the assumption of equality of covariance matrices is void in the one sample case.

2.5. Two Sample T^2-tests when only the mean is available in one group

Suppose that we have a sample of size n_1 from $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$ and the sample mean vector $\bar{\underset{\sim}{y}}$ of n_2 independent observations from $\underset{\sim}{Y} \sim N_p(\underset{\sim}{\nu},\underset{\sim}{\Sigma})$. In this case, $\underset{\sim}{d}$ is taken as

$$\underset{\sim}{d} = \bar{\underset{\sim}{x}} - \bar{\underset{\sim}{y}} \sim N_p(\underset{\sim}{\delta},\frac{n_1+n_2}{n_1 n_2}\underset{\sim}{\Sigma}) \tag{2.25}$$

as in the two sample case, but all information about variability comes from the first sample:

$$\underset{\sim}{S} = \underset{\sim}{S}_1 \sim W_p(n_1-1,\frac{1}{n_1-1}\underset{\sim}{\Sigma}). \tag{2.26}$$

The test for H_q can therefore be used with $r = (n_1+n_2)/n_1 n_2$ and $m = n_1-1$. The test statistic for the overall hypothesis $\underset{\sim}{\mu} = \underset{\sim}{\nu}$ simplifies to

$$F(H_0) = \frac{n_1 n_2 (n_1-p)}{(n_1-1)(n_1+n_2)p} D_p^2. \tag{2.27}$$

This situation is not as artificial as it might seem at a first glance. It might occur, for instance, when data obtained by a previous researcher are no longer available except for the vector of sample means (as illustrated by the example given in this paper). Moreover, it covers two important special cases:

(i) The case $n_2 = \infty$. In this case, $\bar{\underset{\sim}{y}}$ degenerates to $\underset{\sim}{\nu}$, and we are in the same situation as in section 2.4. The one sample T^2-test can therefore be viewed as a special case of "only the mean available in one group".

(ii) the case $n_2 = 1$, which can also be considered as a special case of the two sample situation (section 2.3.), with one sample consisting of only one observation. It has been described under the name "Identification Analysis" by Riedwyl and Kreuter (1976) and by Flury and Riedwyl (1983a). Note that D_p^2 is asymptotically ($n_1 \to \infty$) distributed as chi square with p degrees of freedom if H_0: $\underset{\sim}{\delta} = \underset{\sim}{0}$ holds. Thus, for large n_1, the Mahalanobis distance can be taken as a convenient test statistic in this case.

Let us illustrate the method given in this section by the example introduced at the end of section 1. Suppose we wish to test whether the variables LH and WH are redundant for discrimination between the two populations of Tibetans, given the variables ST, WZA and MFH. The numerical results are as follows:

a) model with 5 variables:

discriminant function = -.13 ST -.38 LH -1.46 WH + 3.88 WZA + 1.23 MFH

Mahalanobis distance: $D_5^2 = 2.40$

Overall test of significance (using (2.18) with p = 5, q = 0, m = 43, r = 95/2244): $F(H_0) = 10.3$ with 5 and 39 degrees of freedom.

b) model with 3 variables (ST, WZA, MFH):

discriminant function = -.14 ST + 2.56 WZA + 1.11 MFH

Mahalanobis distance: $D_3^2 = 2.15$

Overall test of significance (p = 3, q = 0, m and r as above):

$F(H_0) = 16.2$ with 3 and 41 degrees of freedom.

c) Comparison of solutions (a) and (b): using (2.18) with p = 5, q = 3, m and r as above, gives $F(H_3) = 1.2$ with 2 and 39 degrees of freedom. At any usual level of significance we can conclude that the variables LH and WH are redundant for discrimination (given the other three variables).

It should be noted that in interpreting the above numerical results we tacitly assumed that the covariance matrices are identical in both populations. However, the fact that this assumption is void in the one sample situation suggests that moderate differences between the two covariance matrices should not influence the correctness of the F-statistic, provided that n_2 is large enough.

3. Computation of Mahalanobis Distance and Discriminant Function using Linear Regression

3.1. The Two Sample Case with Full Information in Both Groups

There are different ways to show the proportionality between the sample discriminant function of two groups and the linear regression function of a code variable w on the measured variables. Following Healy (1965), we sketch here a proof which seems memorizable and avoids tricks. See also (A 140, K 170, L 17, Cramer 1967, Kendall 1957, p. 159).

Assume that the code variable w takes the value $c_1 = n_2/(n_1+n_2)$ for the individuals in the first sample, and $c_2 = c_1-1 = -n_1/(n_1+n_2)$ in the second sample. (This is known as Fisher's code.) In order to avoid an intercept b_0 in the regression equation, we use as regressors the centered variables, that is, we minimize the sum

$$SSQ(\underset{\sim}{b}) = \sum_{j=1}^{n_1} (c_1 - \underset{\sim}{b}'(\underset{\sim}{x}_j - \underset{\sim}{g}))^2 + \sum_{h=1}^{n_2} (c_2 - \underset{\sim}{b}'(\underset{\sim}{y}_h - \underset{\sim}{g}))^2 \qquad (3.1)$$

over $\underset{\sim}{b} \in \mathbb{R}^p$, where $\underset{\sim}{g} = (n_1\bar{\underset{\sim}{x}} + n_2\bar{\underset{\sim}{y}})/(n_1+n_2)$ is the vector of "grand means", and $\underset{\sim}{x}_j$, $\underset{\sim}{y}_h$ denote the data vectors of the j-th and h-th individual in the first and second sample, respectively. Thus we wish to find a linear function of the variables which approximates c_1 in sample 1 and c_2 in sample 2 as well as possible in the sense of least squares. The first part of the proof consists of establishing the normal equations:

$$[(n_1+n_2-2)\ \underset{\sim}{S} + \frac{n_1 n_2}{n_1+n_2}\ \underset{\sim}{d}\underset{\sim}{d}']\ \underset{\sim}{b} = \frac{n_1 n_2}{n_1+n_2}\ \underset{\sim}{d}, \tag{3.2}$$

where $\underset{\sim}{S}$ is the pooled covariance matrix of both samples, and $\underset{\sim}{d} = \bar{\underset{\sim}{x}} - \bar{\underset{\sim}{y}}$ is the sample mean difference. The key idea is, of course, to write the vector of regression coefficients as a function of $\underset{\sim}{S}$ and $\underset{\sim}{d}$. Now we can apply a formula for inversion of sums of matrices (K 18, MKB 459, R 33) to the matrix multiplying $\underset{\sim}{b}$ in (3.2) and get, after some simplification, the least squares solution

$$\underset{\sim}{b^*} = \frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-2) + n_1 n_2\ \underset{\sim}{d}'\underset{\sim}{S}^{-1}\underset{\sim}{d}}\ \underset{\sim}{S}^{-1}\underset{\sim}{d}. \tag{3.3}$$

With the notation of section 2, we have $\underset{\sim}{a} = \underset{\sim}{S}^{-1}\underset{\sim}{d}$ and $D_p^2 = \underset{\sim}{d}'\underset{\sim}{S}^{-1}\underset{\sim}{d}$, and the constant of proportionality between $\underset{\sim}{b^*}$ and $\underset{\sim}{a}$ is established as

$$k = \frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-2) + n_1 n_2\ D_p^2}\ . \tag{3.4}$$

Let R_p^2 denote the coefficient of determination of the regression (where the index p indicates again that we are using p variables). Then it is easy to

show that

$$R_p^2 = \underset{\sim}{d}'\underset{\sim}{b}^* \tag{3.5}$$

holds. Multiplication of (3.3) from the left by $\underset{\sim}{d}'$ gives therefore the important relations

$$R_p^2 = \frac{n_1 n_2 \, D_p^2}{(n_1+n_2)(n_1+n_2-2) + n_1 n_2 \, D_p^2} \tag{3.6}$$

and

$$D_p^2 = \frac{(n_1+n_2)(n_1+n_2-2)}{n_1 n_2} \cdot \frac{R_p^2}{1-R_p^2} \, . \tag{3.7}$$

Using (3.7), the proportionality factor k can be expressed as a function of n_1, n_2 and R_p^2, and all statistics used in the two sample problem (section 2.3) can therefore be written as functions of the regression quantities.

We can now switch back from the centered variables as used in (3.1) to the usual form of the linear regression model, which includes an additional coefficient b_0 (intercept). The coefficients $\underset{\sim}{b}^*$ remain unaffected by this change. Moreover, any choice of the code values c_1 and c_2 ($c_1 \neq c_2$) can be viewed as a nonsingular linear transformation of Fisher's code, leaving $\underset{\sim}{b}^*$ unaffected except for multiplication by a proportionality constant. Most important of all, the coefficient of determination R_p^2 remains invariant under all these transformations. Therefore (3.7) and $\underset{\sim}{a} = k^* \underset{\sim}{b}^*$ always hold, where k^* is a scalar constant, and $\underset{\sim}{b}^*$ is the vector of regression coefficients, ignoring the intercept.

Though Fisher's code has the advantage of simplifying the proof, we can as well choose the values 0 and 1 in practical applications. For higher numerical accuracy in the regression coefficients it is often better to choose a larger difference between c_1 and c_2.

As a marginal note, let us state that the regression vector $\underset{\sim}{b}^*$ and the discriminant function vector $\underset{\sim}{a}$ are identical (that is, the proportionality factor is ± 1) precisely if $|c_2-c_1| = D_p^2/R_p^2$, but these two quantities are of course not known prior to the numerical analysis.

3.2. The Case when only the Mean is Available in One Group

It can easily be checked that the derivations of section 3.1 hold also if $n_2 = 1$, that is, the second sample consists of only one observation. By (3.3), the regression vector $\underset{\sim}{b}^*$ is proportional to the vector of discriminant function coefficients:

$$\underset{\sim}{b}^* = k^* \underset{\sim}{S}^{-1}(\underset{\sim}{\bar{x}}-\underset{\sim}{y}) \tag{3.8}$$

for some factor k^*, where $\underset{\sim}{\bar{x}}$ is the mean vector of the first sample, $\underset{\sim}{y}$ is the single observation in the second sample, and $\underset{\sim}{S}$ is based on the n_1 observations of the first sample only. From (3.7) we have therefore

$$D_p^2 = (\underset{\sim}{\bar{x}}-\underset{\sim}{y})'\underset{\sim}{S}^{-1}(\underset{\sim}{\bar{x}}-\underset{\sim}{y}) = \frac{n_1^2-1}{n_1}\,\frac{R_p^2}{1-R_p^2}\,. \tag{3.9}$$

Now note that (3.9) does not depend on the fact that $\underset{\sim}{y}$ is the "mean vector" of only one observation - in fact, if we replace $\underset{\sim}{y}$ in the data matrix by any vector $\underset{\sim}{v}$, we will get

$$b^* = k^* S^{-1}(\bar{x}-v) \tag{3.10}$$

for some factor k*, and

$$\frac{n_1^2-1}{n_1} \cdot \frac{R_p^2}{1-R_p^2} = (\bar{x}-v)'S^{-1}(\bar{x}-v). \tag{3.11}$$

This suggests the following procedure for the general case of "only the mean available in one sample". Add the mean vector $\bar{y}$ from the second sample as the (n_1+1)-st observation to the n_1 observations from the first sample. Code the observations by

$$w_j = \begin{cases} c_1 & \text{if } 1 \le j \le n_1 \\ c_2 & \text{if } j = n_1 + 1 \end{cases} \tag{3.12}$$

Perform a LR of w on the measured variables, using all n_1+1 observations. Denote the vector of regression coefficients by b^* (ignoring the intercept), then

$$b^* = k^* S^{-1}(\bar{x}-\bar{y}) \tag{3.13}$$

for some factor k*, and

$$D_p^2 = (\bar{x}-\bar{y})'S^{-1}(\bar{x}-\bar{y}) = \frac{n_1^2-1}{n_1} \cdot \frac{R_p^2}{1-R_p^2} \cdot \tag{3.14}$$

Since S is based only on the first sample, (3.13) and (3.14) can be used to compute discriminant function coefficients and Mahalanobis distances in all

cases of section 2.5, including the one sample case! The test statistics listed in section 2 can therefore all be computed using the regression approach.

The practical application is very simple, since only one row and one column must be added to the data matrix. The completed data matrix takes the form

$$
\begin{array}{cccc}
x_{11} & x_{21} & \cdots & x_{p1} \quad c_1 \\
x_{12} & x_{22} & \cdots & x_{p2} \quad c_1 \\
x_{13} & x_{23} & \cdots & x_{p3} \quad c_1 \\
\vdots & \vdots & & \vdots \quad \vdots \\
x_{1n_1} & x_{2n_1} & \cdots & x_{pn_1} \quad c_1 \\
\bar{y}_1 & \bar{y}_2 & \cdots & \bar{y}_p \quad c_2
\end{array}
$$

The regression equation based on this matrix will in general include an intercept which is not zero, but which can be ignored.

In our example, we used the code $c_1 = 0$ (for the $n_1 = 44$ individuals of the first sample) and $c_2 = 100$ (for the mean vector of the second sample). We computed an LR of this code variable on (a) all five variables, and (b) variables ST, WZA and MFH. The results are summarized in table 2, where all columns except those labeled as "coefficients" should be ignored for the moment. The proportionality of the regression solution to the discriminant function calculated in section 5 and the correctness of formula (3.14) for computing D_p^2 can easily be checked.

4. Tests of hypotheses about the linear discriminant function using the regression technique

4.1. Differences between LR and DA

When encouraging the use of LR to perform DA and T^2-tests, it must be stressed that there are important differences between the two models. In LR it is assumed that

$$\underset{\sim}{w} = \underset{\sim\sim}{X}\beta + \underset{\sim}{\varepsilon}, \tag{4.1}$$

where $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0},\sigma^2\underset{\sim}{I}_n)$, and $\underset{\sim}{X}$ is regarded as fixed. Then

$$\underset{\sim}{\hat{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{w} \sim N_{p+1}(\underset{\sim}{\beta},\sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}). \tag{4.2}$$

In the regression case, $\underset{\sim}{w}$ is random and measured on an interval scale, and all the standard errors and F-(or t-) tests presented for $\underset{\sim}{\beta}$ in the LR output depend on these assumptions. In using the LR procedure for DA, $\underset{\sim}{w}$ is a fixed dichotomy, and $\underset{\sim}{X}$ is distributed as the mixture of two multivariate normal distributions. It is therefore rather surprising that even the F-tests given in LR are valid for the DA-case, as will be shown in the following paragraphs.

4.2. The two sample case with full information in both groups

Let us denote by $F_{reg}(H_q)$ the usual F-statistic used in LR to test whether p-q coefficients of the linear regression function are zero (JW 306, M 108, see also any text on linear regression). If R_p^2 and R_q^2 denote the coefficients of determination from the regressions with p and q variables respectively, then $F_{reg}(H_q)$ can be written as

$$F_{reg}(H_q) = \frac{n_1+n_2-p-1}{p-q} \cdot \frac{R_p^2-R_q^2}{1-R_p^2}$$

$$= \frac{n_1+n_2-p-1}{p-q} \cdot \frac{D_p^2-D_q^2}{\frac{(n_1+n_2-2)(n_1+n_2)}{n_1 n_2} + D_q^2}, \qquad (4.3)$$

where the second equality follows from (3.6). This is the same as formula (2.18) with $m = n_1+n_2-2$ and $r = (n_1+n_2)/n_1n_2$. Thus, in this case, all F_{reg}-statistics (and particularly the F- or t-statistics) are correct also in the DA-situation. Since even the degrees of freedom associated with (2.18) and (4.3) are identical, tests can be performed directly without correcting the regression output.

4.3. The case when only the mean is available in one group

Using the regression approach as described in section 3.2 leads to the following F_{reg}-statistic for testing redundancy of p-q variables:

$$F_{reg}(H_q) = \frac{n_1 p}{p-q} \cdot \frac{R_p^2-R_q^2}{1-R_p^2}$$

$$= \frac{n_1-p}{p-q} \cdot \frac{D_p^2-D_q^2}{\frac{n_1^2-1}{n_1} + D_q^2} . \qquad (4.4)$$

The degrees of freedom computed by the LR approach are p-q and n_1-p, which is correct also for the T^2-situation. However, putting $r = (n_1+n_2)/n_1n_2$ and $m = n_1-1$ into (2.18) shows that the correct F-statistic for the discriminant

function hypothesis would be

$$F(H_q) = \frac{n_1-p}{p-q} \quad \frac{D_p^2-D_q^2}{\frac{(n_1+n_2)(n_1-1)}{n_1n_2} + D_q^2} \quad . \tag{4.5}$$

This can be computed from the corresponding LR-statistic by

$$F(H_q) = \frac{(n_1^2-1)n_2 + n_1n_2D_q^2}{(n_1-1)(n_1+n_2)+n_1n_2D_q^2} \quad F_{reg}(H_q) \tag{4.6}$$

or, using (3.7):

$$F(H_q) = \frac{(n_1+1)n_2}{n_1+n_2+n_1(n_2-1)R_q^2} \quad F_{reg}(H_q). \tag{4.7}$$

This shows that for given R_q^2, F is an increasing linear function of F_{reg}. Moreover, $F(H_q)$ can be computed from R_p^2 and R_q^2 according to

$$F(H_q) = \frac{(n_1+1)n_2(n_1-p)(R_p^2-R_q^2)}{[n_1+n_2+n_1(n_2-1)R_q^2]\ (p-q)(1-R_p^2)}. \tag{4.8}$$

If a standard LR program is used, relation (4.7) is most useful. The F_{reg}-statistics computed by the regression program are simply multiplied by the factor

$$c_q = \frac{(n_1+1)n_2}{n_1+n_2+n_1(n_2-1)R_q^2} \quad , \tag{4.9}$$

which can easily be done using a pocket calculator. If R_q^2 is not available from the regression output, it can first be computed from $F_{reg}(H_q)$ and R_p^2 by

$$R_q^2 = R_p^2 - \frac{p-q}{n_1-p} (1-R_p^2) F_{reg}(H_q). \qquad (4.10)$$

In addition, putting (4.10) into (4.7) shows that for given R_p^2, $F(H_q)$ is a monotonically increasing function of $F_{reg}(H_q)$. Therefore the forward selection and backward elimination procedures based on partial F_{reg}-statistics, as used in many LR-programs, yield a correct order of selection, when applied to this situation. The stopping criterion, however, must in general be modified.

In practical applications, we recommend augmenting the list of partial F_{reg}-statistics by the factors c_{p-1} (formula 4.9) and the corrected partial F-statistics (formula 4.7). If the LR program gives rather partial t- than F-statistics, these should be multiplied with the square root of c_{p-1}.

In our example, the partial F_{reg} statistics given by the standard LR program were as displayed in the corresponding columns of table 2. Three additional columns show the correction factors, the corrected partial F-statistics, and the degrees of freedom. From the corrected partial F's it becomes clear that some of the five variables are redundant for discrimination. The comparison of models (a) and (b) yields $F_{reg}(H_3) = .105$ with 2 and 39 degrees of freedom. Using the correction factor (4.9) $c_3 = 11.69$, we get again the correct statistic $F(H_3) = 1.2$.

In some cases the correction factor c_q turns out to be especially simple, as shown in table 3. A remarkably memorizable correction factor is given in the case $n_2 = \infty$ and $q = 0$ (Hotelling's T^2 in the one sample case). Clearly, no correction is needed if $n_2 = 1$, since this can also be considered as a special case of the usual two group situation.

5. Remarks and Conclusions

5.1. More about the relations between LR and DA

The relations between LR and DA as described in the previous sections seem in a way like "lucky algebraic coincidences" - there is no obvious intuitive reason why they should hold. The use of LR to perform DA and T^2-tests has therefore the fame of a trick. However, some deeper reasons for the relations between the two models can be found if we relate them both to a multivariate regression model (A 178, JW 318, K 97, M 170, MKB 157, R 543, SC 139). To explain this, let us change the notation used so far a little, and denote by $\begin{pmatrix} \underset{\sim}{X} \\ \underset{\sim}{Y} \end{pmatrix}$ the p-dimensional vector of random variables to be used for discrimination, where $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ have q and p-q components, respectively. (Here we do not distinguish notationally between the variables of the first and those of the second group). Suppose furthermore that a binary variable w is coded such as to indicate the group membership. Then we can study the multivariate regression of $\underset{\sim}{Y}$ on the set of q + 1 variables $\{\underset{\sim}{X},w\}$. Since $\underset{\sim}{Y}$ consists of p-q variables, this model can be interpreted as p-q simultaneous multiple linear regressions on q+1 variables. Alternatively, since w is binary, we can look at it as a multivariate regression model for two groups, thus representing 2(p-q) multiple regressions on q variables, the two regression hyperplanes associated with each Y-variable being parallel.

In this multivariate linear model we have (allowing for intercepts) a total of (p-q)(q+2) regression parameters. Among these, p-q are associated with the code-variable w, one for each of the Y-variables. The maximum likelihood estimates of these p-q parameters turn out to be identical with the (sample) conditional mean difference of $\underset{\sim}{Y}$, given $\underset{\sim}{X}$, if Fisher's code is used. However,

in (2.11) of this paper we have already seen that discriminant function coefficients are strongly related to conditional mean differences. On the other hand, the (univariate) linear regression of w on {X,Y} is also strongly related to the above multivariate model (which is not very surprising).

Let us summarize the most important facts as follows. We are dealing with three models:

(1) the linear discriminant analysis of two groups which are defined by a binary variable w, using the set of variables $\{\underset{\sim}{X},\underset{\sim}{Y}\}$.

(2) the linear "pseudo"-regression of w on $\{\underset{\sim}{X},\underset{\sim}{Y}\}$

(3) the multivariate linear regression of $\underset{\sim}{Y}$ on $\{\underset{\sim}{X},w\}$.

Although the relations between models (1) and (2) are fairly easy to establish - see sections 3 and 4 -, the deeper reasons for their existence can be seen in the mutual relationship of (1) and (2) to (3). More specifically:

- The coefficients of the linear discriminant function (model 1) are strongly related to the coefficients associated with w in model 3, due to the fact that the latter ones are actually conditional mean differences. These coefficients are in turn closely related to the coefficients of the "pseudo"-regression model 2.

- Testing for redundancy of $\underset{\sim}{Y}$ in the DA-model (1) is, by (2.11), the same as testing for redundancy of w in model 3. This is, in turn, again the same as testing for redundancy of $\underset{\sim}{Y}$ in model 2.

For proofs and more details about the relationship between the three models see Flury (1983).

In order to help the reader who is not familiar with the multivariate linear model, we are now going to illustrate the above statements for p = 2 and q = 1,

in which case model 3 reduces to two parallel regression lines. Suppose the random variables X and Y have the joint bivariate normal distribution $N_2(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$ in population 1 and $N_2(\underset{\sim}{\nu},\underset{\sim}{\Sigma})$ in population 2, with $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\nu = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ and $\underset{\sim}{\Sigma} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. These two normal distributions are visualized in figure 1 by two ellipses of equal density. Also shown are the two (population) regression lines of Y on X, which are parallel because the covariance matrices are identical. The vector of discriminant function coefficients is $\underset{\sim}{\alpha} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$ by (2.1), and the conditional mean difference of Y, given X, is $\delta_{y.x} = 1$ by (2.7). The condition for redundancy of Y in the DA-model is the same as the condition that the two parallel regression lines coincide, and testing for redundancy of Y in the DA-model is the same as testing for redundancy of w in the regression of Y on {X,w}. This is, in turn, the same as testing for redundancy of Y in the regression of w on {X,Y}.

In the situation of figure 1, if the mean vector of the second population were $\underset{\sim}{\nu} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ instead of $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$, then the two regression lines would coincide and discrimination would be based on X only, as can easily be checked.

In practical applications, model 3 is less convenient than models 1 and 2, since it is algebraically more complicated and involves in general more parameters than just those we are interested in for the purpose of discrimination. However, it has considerable theoretical advantages.

- It shows that the relations between DA and LR (models 1 and 2) are more than just a lucky algebraic coincidence.

- It has a straightforward generalization to the k-sample case by using k - 1 indicator variables.

- It leads to an estimate for standard errors of discriminant function coefficients.

- Compared to the classical DA-model, it does not require p-dimensional normality. The assumptions in the multivariate regression model are rather (a) the mean of $\underset{\sim}{Y}$ is a linear function of $\underset{\sim}{X}$ and w; and (b) the residuals have a joint (p-q)-dimensional normal distribution. This is less restrictive than p-dimensional normality, and shows that it is not necessarily wrong to include non-normal or even discrete and binary variables into a linear discriminant function. This fact is often ignored when DA is being attacked (Breiman et al 1984, p. 16; Rubin 1984). Testing for redundancy of $\underset{\sim}{Y}$ in the DA-model does not require p-dimensional normality of $\begin{pmatrix} \underset{\sim}{X} \\ \underset{\sim}{Y} \end{pmatrix}$, but rather (p-q)-dimensional normality of the conditional distribution of $\underset{\sim}{Y}$, given $\underset{\sim}{X}$. (For an extensive discussion of discriminant models involving both continuous and discrete variables, see the papers by Krzanowski (1975, 1977, 1980) and their references.)

5.2. Standard errors of discriminant function coefficients

The problem of standard errors of discriminant function coefficients is a controversial topic. Many authors do not mention it at all (A, JW, K, L, M, MKB, SC). Others (R 569; Rao 1970, p. 587) take the point of view that estimates of standard errors for discriminant function coefficients are not meaningful, since every multiple of the linear discriminant function discriminates the groups as well.

On the other hand, if we define the vector of sample discriminant function coefficients by $\underset{\sim}{a} = \underset{\sim}{S}^{-1}\underset{\sim}{d}$, then the coefficients are uniquely defined. This is the point of view taken by Kendall and Stuart (1966, p. 331), who give a large sample estimate for the variance of discriminant function coefficients.

The regression approach gives also an answer to this problem. Let us denote by b_j the j-th regression coefficient, and by $F(\alpha_j=0)$ the partial F-value for the j-th variable. Then the standard error of b_j can be estimated by

$$s(b_j) = \frac{|b_j|}{\sqrt{F(\alpha_j=0)}} . \qquad (5.1)$$

This estimate can be derived from standard errors of conditional mean differences - that is, using again model 3 of section 5.1! It has recently been investigated by Haggstrom (1983) and by Flury and Riedwyl (1983b). The latter paper gives also a comparison between (5.1) and the estimate of Kendall and Stuart.

Of course it must be borne in mind that these estimates, as well as the coefficients b_j, are not absolute quantities. They are valid only in the metric given by the regression approach, that is, for a given constant of proportionality k (formula 3.4). This means in particular that standard errors obtained from different analyses cannot be compared directly, but within the same discriminant function such a comparison is correct.

It only remains to note that (5.1) is exactly the standard error given by the LR-approach in the usual two sample case. In this form it has already been used, although without sufficient theoretical motivation, in many practical applications. In the case when only the mean is available in one sample, formula (5.1) can be used to obtain the correct standard errors after having computed $F(\alpha_j=0)$ by the formulas of section 4.3.

In our example, the corrected standard errors have been computed and are given in the corresponding columns of table 2.

5.3. Advantages of the regression approach

Regression programs are probably the most used statistical software. They are nowadays even available for home- and pocket computers. Using the LR-approach makes it therefore easy to calculate T^2-statistics and discriminant functions when no special software for these techniques is available.

Besides this practical aspect, the regression approach has in our opinion some distinct didactic advantages:

- First, many scientists in different fields have a good knowledge of LR, but their mathematical background would not be strong enough to understand the theory of T^2-tests. With a basic understanding of the concept of discriminant function, they can be instructed to transfer their knowledge about testing hypotheses in LR to the DA-situation.

- Second, using the regression approach leads easily to the question of redundancy of variables, which is usually not treated in connection with T^2-tests. For practitioners, however, this problem is often as important as an overall test of significance. In courses with practitioners we found it particularly useful to stress the fundamental relations (3.7) and (3.14) between R^2 and D^2, after having introduced the notion of Mahalanobis-distance for the two-dimensional case (JW 19, Flury and Riedwyl 1983a, p. 100). We believe that the regression approach, together with an introduction to DA from the point of view of classification, is sufficient to evoke a correct understanding of the methods discussed in this paper.

- Third, knowing the relations between LR and DA, and particularly the mutual relationship of both methods to the multivariate linear model of section 5.1, is of some theoretical interest. Furthermore, this multivariate model shows that the linear discriminant function, thanks to its relation to conditional mean differences, is not necessarily "wrong" for non-normal data.

References

ANDERSON, T. W. (1958): An Introduction to Multivariate Statistical Analysis, New York: John Wiley.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984): Classification and Regression Trees, Belmont (CA): Wadsworth International Group.

CRAMER, E. M. (1967): "Equivalence of Two Methods of Computing Discriminant Function Coefficients," Biometrics, 23, 153.

FISHER, R. A. (1936): "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, VII, pt. 2, 179-188.

FLURY, B. (1983): "A Note on Multivariate Parallel Regression," Technical Report #83-42, Dept. of Statistics, Purdue University.

FLURY, B. and RIEDWYL, H. (1983a): Angewandte multivariate Statistik, Stuttgart & New York: Gustav Fischer.

_____ (1983b): "The Use of Multiple Regression to Perform Two Group Discriminant Analysis and T^2-Tests," Technical Report #13, Dept. of Statistics, University of Berne.

HAGGSTROM, G. W. (1983): "Logistic Regression and Discriminant Analysis by Ordinary Least Squares," Journal of Business and Economic Statistics, 1, 229-238.

HEALY, M. J. R. (1965): "Computing a Discriminant Function from Within-Sample Dispersion," Biometrics, 21, 1011-1012.

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

KARSON, M. J. (1982): Multivariate Statistical Methods, Ames (IA): The Iowa State University Press.

KENDALL, M. G. (1957): A Course in Multivariate Analysis, New York: Hafner Publishing Co.

KENDALL, M. G. and STUART, A. (1966): The Advanced Theory of Statistics, Vol. 3, New York: Hafner Press.

KRZANOWSKI, W. J. (1975): "Discrimination and Classification Using Both Binary and Continuous Variables," Journal of the American Statistical Association, 70, 782-790.

_____ (1977): "The Performance of Fisher's Linear Discriminant Function under Non-Optimal Conditions," Technometrics, 19, 191-200.

_____ (1980): "Mixtures of Continuous and Categorical Variables in Discriminant Analysis," Biometrics, 36, 493-499.

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

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

MORRISON, D. F. (1976): Multivariate Statistical Methods, New York: McGraw-Hill.

MULLIS, M. L. (1982): Personal Communication.

H. R. H. Prince PETER of Greece and Denmark et al (1966): "Anthropological Researches from the 3rd Danish Expedition to Central Asia," Hist. Filos. Skr. Dan. Vid. Selek., 4, no. 4.

RAO, C. R. (1970): "Inference on Discriminant Function Coefficients," in Essays in Probability and Statistics, ed. R. C. Bose et al, Chapel Hill: The University of North Carolina Press.

_____ (1973): Linear Statistical Inference and its Applications, 2nd ed., New York: John Wiley.

RIEDWYL, H. and KREUTER, U. (1976): "Identification," in Contributions to Applied Statistics, ed. J. Ziegler, Basel: Birkhäuser.

RUBIN, D. B. (1984): Comment on "Graphical Methods for Assessing Logistic Regression Models" by J. M. Landwehr, D. Pregibon and A. C. Shoemaker, Journal of the American Statistical Association, 79.

Table Titles

Table 1: Summary statistics for anthropometric example

Table 2: Computation of discriminant function and T^2-tests in the anthropometric example by linear regression

Table 3: Correction factor c_q when only the mean is available in one group. F_{reg}-statistics (computed by the regression approach) for testing sufficiency of q out of p variables should be multiplied by this factor. n_2 is the size of the sample in which only the mean is available. R_q^2 is the coefficient of determination obtained with q variables.

Figure Title

Figure 1: Parallel regression lines of Y on X in two normal populations with identical covariance matrices. The population means are $\underset{\sim}{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\underset{\sim}{\nu} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$, the common covariance matrix is $\underset{\sim}{\Sigma} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. The linear discriminant function is proportional to X-Y, that is, it is constant on straight lines parallel to the 45°-line. If the second population is shifted to $\underset{\sim}{\nu} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, the two regression lines coincide, and the discriminant function coefficient of Y is zero.

Table 1

	Data from Mullis (1982) (n_1=44)							Data from Prince Peter (1966)(n_2=51)
variable	mean	std. dev.	correlation matrix					mean
ST	164.70	6.106	1.00	.60	.13	.21	.36	166.68
LH	19.22	.493		1.00	-.01	.30	.44	19.08
WH	15.40	.564			1.00	.75	.09	15.07
WZA	14.63	.493				1.00	.07	14.07
MFH	12.51	.708					1.00	12.11

Table 2

(a) Solution based on p = 5 variables

Variable	Coefficient	Partial F_{reg}	Correction factor	partial F, corrected	degrees of freedom	std. error, corrected
(intercept)	46.22					
ST	.29	.3400	12.04	4.10	1,39	.14
LH	.81	.0123	11.03	.14	1,39	2.16
WH	3.15	.1944	11.57	2.25	1,39	2.10
WZA	-8.37	1.0081	14.83	14.95	1,39	2.16
MFH	-2.65	.4830	12.55	6.06	1,39	1.08

R_5^2 = .05169 $F_{reg}(H_0)$ = .4241 c_0 = 24.16 degrees of freedom = (5,39)

(b) Solution based on q = 3 variables

(intercept)	62.37					
ST	.31	.5692	13.62	7.75	1,41	.11
WZA	-5.54	1.3947	18.17	25.34	1,41	1.10
MFH	-2.40	.4806	13.26	6.37	1,41	.95

R_3^2 = .04666 $F_{reg}(H_0)$ = .6689 c_0 = 24.16 degrees of freedom = (3,41)

Table 3

	$n_2 = 1$	$1 \leq n_2 < \infty$	$n_2 = \infty$
$q = 0$	1	$\frac{(n_1+1)n_2}{n_1+n_2}$	n_1+1
$1 \leq q < p$	1	$\frac{(n_1+1)n_2}{n_1+n_2+n_1(n_2-1)R_q^2}$	$\frac{n_1+1}{n_1R_q^2+1}$

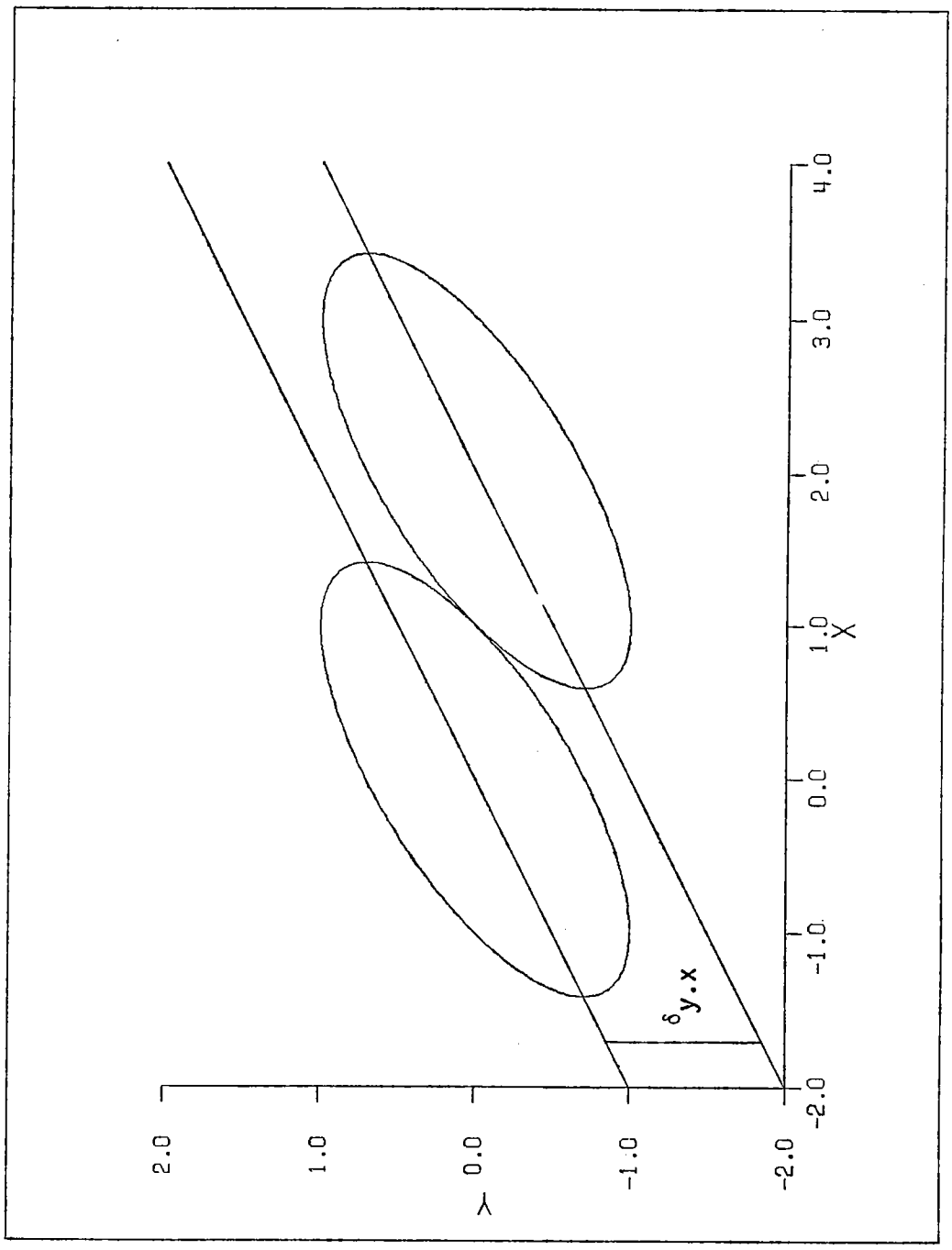

Figure 1