The Geometry of Canonical Variate Analysis
Author(s): N. A. Campbell and William R. Atchley
Source: *Systematic Zoology*, Vol. 30, No. 3 (Sep., 1981), pp. 268-280
Published by: Taylor & Francis, Ltd. for the Society of Systematic Biologists
Stable URL: https://www.jstor.org/stable/2413249
Accessed: 23-09-2024 10:00 UTC

Syst. Zool., 30(3), 1981, pp. 268–280

THE GEOMETRY OF CANONICAL VARIATE ANALYSIS

N. A. CAMPBELL AND WILLIAM R. ATCHLEY

Abstract

Campbell, N. A. (Division of Mathematics and Statistics, CSIRO, Wembley 6014, Western Australia) and W. R. Atchley (Department of Entomology, University of Wisconsin, Madison, Wisconsin 53706) 1981. The geometry of canonical variate analysis. Syst. Zool., 30:268–280.— The geometry of canonical variate analysis is described as a two-stage orthogonal rotation. The first stage involves a principal component analysis of the original variables. The second stage involves a principal component analysis of the group means for the orthonormal variables from the first-stage eigenanalysis. The geometry of principal component analysis is also outlined. Algebraic aspects of canonical variate analysis are discussed and these are related to the geometrical description. Some practical implications of the geometrical approach for stability of the canonical vectors and variable selection are presented. [Multivariate analysis; canonical variate analysis; discriminant analysis; principal component analysis.]

Canonical variate analysis is one of the most important and widely used multivariate statistical techniques in biological research. The procedure was developed by R. A. Fisher (1936) and further expanded by M. S. Bartlett, P. C. Mahalanobis, and C. R. Rao to examine several significant problems relevant to systematic biology. These include separation of groups of morphologically similar organisms; ascertainment of patterns of character covariation, such as size and shape patterns, between groups; assessment of intergroup affinities; and the allocation of individuals to pre-existing groups.

Canonical variate analysis is discussed widely in modern textbooks on multivariate analysis (e.g., Kshirsagar, 1972: Ch. 9). However, most treatments stress algebraic, computational and inferential aspects, rather than geometrical understanding (see also Dempster, 1969).

In this paper, we describe the geometry of canonical variate analysis, Mahalanobis D^2, and principal component analysis. The algebra underlying this geometrical discussion is provided. Some practical implications of the geometrical approach are presented.

EIGENANALYSIS AND PRINCIPAL COMPONENT ANALYSIS

Canonical variate analysis can be considered as a two-stage rotation. The first stage involves a principal component analysis or eigenanalysis of the original variables. The second stage involves an eigenanalysis of the variation between the group means for the variables from the first-stage principal component analysis.

The eigenanalysis of a symmetric matrix is a fundamental notion in multivariate analysis. It forms the basis of the calculations for a principal component analysis. The ideas and concepts in principal component analysis are important for both the geometric and algebraic presentations of canonical variate analysis given later.

A principal component analysis can be considered as a rotation of the axes of the original variable coordinate system to new orthogonal axes, called principal axes, such that the new axes coincide with directions of maximum variation of the original observations. Consider the line or axis passing through the ends of the elliptical cluster of points in Figure 1. Project the original data points onto this axis. The point y_{1m} is the projection of the point (x_{1m}, x_{2m}) onto the axis defined by the direction Y_1. This axis has the property that the variance of the projected points y_{1m}, $m = 1, \ldots, n$, is greater than the variance of the points when projected onto any other line or axis passing through $(\bar{x}_1, \bar{x}_2)$. Any line parallel to Y_1 also has the property of maximum variance of the projected points. It is however con-

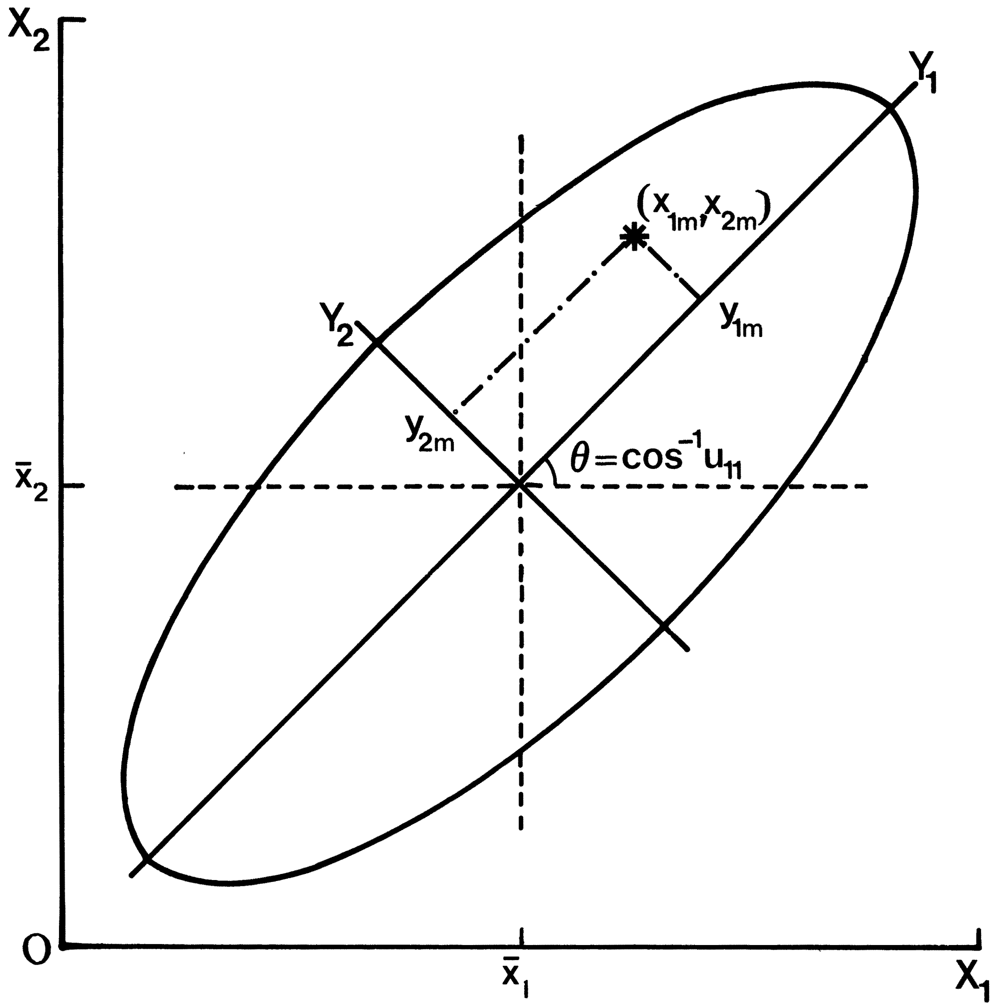


FIG. 1.—Idealized representation of scatter diagram for two variables, showing the mean for each variable ($\bar{x}_1$ and $\bar{x}_2$), 95% concentration ellipse, and principal axes Y_1 and Y_2. The points y_{1m} and y_{2m} give the principal component scores for the observation $\mathbf{x}_1 = (x_{1m}, x_{2m})^T$. The cosine of the angle θ between Y_1 and X_1 gives the first component u_{11} of the eigenvector corresponding to Y_1.

venient geometrically to use the first representation.

The property of maximum variation of the projected points defines the first principal axis; it is the line or direction with maximum variation of the projected values of the original data points. The projected values corresponding to this direction of maximum variation are the *principal component scores*. The first principal axis is often called the line of best fit since the sum of squares (SSQ) of the perpendicular deviations of the original data points from the line is a minimum. Successive principal axes are determined with the property that they are orthogonal to the previous principal axes and that they maximize the variation of the projected points subject to these constraints. For two variables, only one more axis or direction can be determined; this second axis is represented by Y_2 in Figure 1.

In practice, a principal component analysis consists initially of finding the eigenvalues e_i and eigenvectors $\mathbf{u}_i$ of the sample covariance or correlation matrix. The *eigenvalue* is simply the usual sample variance of the projected data points. The components of the *eigenvector* are the cosines of the angles between the original variable axes and the corresponding principal axis. These cosines are often referred to as direction cosines. In Figure 1, the cosine of the angle between the original variable axis X_1 and the first principal axis Y_1 gives the first component u_{11} of the first eigenvector $\mathbf{u}_1$, while the cosine of the angle between the ordinate variable X_2 and Y_1 gives u_{12}. Similarly, the cosines of the angles between the second principal axis Y_2 and the original coordinate axes give the components u_{21} and u_{22} of $\mathbf{u}_2$.

An essential notion in multivariate analysis is that of a *linear combination* of variables; it is fundamental to both canonical variate analysis and principal component analysis. Consider v variables $x_1, \ldots, x_v$, written as the vector $\mathbf{x} = (x_1, \ldots, x_v)^T$, and the coefficients $c_1, \ldots, c_v$, written as the vector $\mathbf{c}$. Then a linear combination is defined by

$$\begin{aligned} y = c_1x_1 + \ldots + c_vx_v &= \sum_{i=1}^{v} c_ix_i \\ &= \mathbf{c}^T\mathbf{x}, \end{aligned}$$

where y is the new variable defined by the linear combination of the original variables. For example, if the coefficients are all unity ($c_i = 1$ for all i), then $\mathbf{c}^T\mathbf{x} = \sum_{i=1}^{v} x_i$, which is just the sum of the variables. This can be written in matrix notation as $\mathbf{1}^T\mathbf{x}$, where $\mathbf{1}$ denotes a vector of 1's.

A principal component analysis seeks a linear combination of the original variables such that the usual sample variance of the resulting values is a maximum. The components of the eigenvectors $\mathbf{u}_1$ (Fig. 1) provide the coefficients which define the linear combination, while the resulting values or scores are the projected points y_{im}. That is, $y_{1m} = u_{11}x_{1m} + u_{12}x_{2m}$, and $y_{2m} = u_{21}x_{1m} + u_{22}x_{2m}$. In matrix notation, $y_{im} = \mathbf{u}_i^T\mathbf{x}_m$, where $\mathbf{x}_m = (x_{1m}, \ldots, x_{vm})^T$ denotes the mth observation vector. The sample variance of the projected points y_{1m} gives the first eigenvalue e_1. Some constraint on the components of $\mathbf{u}_1$ is necessary, otherwise the variance can be made arbitrarily large. The usual one to adopt is that $\sum_{i=1}^{v} u_{1i}^2 = 1$ or that $\mathbf{u}_1^T\mathbf{u}_1 = 1$. Maximization of the variance of the y_{1m} subject to the given constraint leads to the eigenequation

$$(\mathbf{V} - e\mathbf{I})\,\mathbf{u} = \mathbf{o} \qquad (1)$$

or

$$\mathbf{V}\,\mathbf{u} = \mathbf{u}e$$

where $\mathbf{V}$ denotes the within-group covariance matrix. Let

$$\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_v)$$

denote the matrix of eigenvectors, and let the diagonal matrix

$$\mathbf{E} = \text{diag}(e_1, \ldots, e_v)$$

denote the matrix of eigenvalues. Then the eigenequation becomes

$$\begin{aligned} \mathbf{V} &= \mathbf{U}\mathbf{E}\mathbf{U}^T \qquad (2) \\ &= \sum_{i=1}^{v} e_i\mathbf{u}_i\mathbf{u}_i^T. \end{aligned}$$

The eigenvectors satisfy $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{U}\mathbf{U}^T = \mathbf{I}$.

An important result, which follows by taking the trace of both sides of (2), is that the sum of the variances of the original variables is equal to the sum of the eigenvalues. Since each successive principal component accounts for a maximum amount of the variation, subject to being uncorrelated with the previous components, $e_1 > e_2 > \ldots > e_v$.

Principal component analysis is considered to be a useful tool when the first few principal components explain much of the variation, so that a few bivariate scatter plots of the scores summarize the multivariate data. For morphometric data, it is often found that the elements of the first eigenvector are all positive; an increase in each variable results in a general increase in the value of the principal component score. For this reason, the first component is often termed a size component (e.g., Jolicoeur and Mosimann, 1960).

CANONICAL VARIATE ANALYSIS—GENERAL IDEAS

In a canonical variate analysis, linear combinations of the original variables are determined in such a way that the differences between a number of reference groups are maximized relative to the variation within groups. It is hoped that the group configuration can be adequately represented in a two or three dimensional subspace defined by the first two or three canonical vectors. The first *canonical vector* is given by the coefficients of the linear combination which maximizes the ratio of the between- to within-groups SSQ's for the resulting *canonical variate.* The corresponding ratio is referred to as the *canonical root.* Successive linear combinations of the original variables are chosen to be uncorrelated both within and between groups. Pythagorean distance is then appropriate for interpreting a scatter plot of the group means, with the important canonical variates as the coordinates.

Figure 2 depicts a typical situation for two variables. The concentration ellipses reflect the clustering of the observations in the main body of the data. The points $\mathbf{x}_{1m} = (x_{11m}, x_{21m})^T$ and $\mathbf{x}_{2m}$ represent typical observations. The vector $\mathbf{c}$ represents the direction of the calculated canonical vector.

The point representing the observation y_{1m} gives the projection of the observation $\mathbf{x}_{1m}$ onto the canonical vector. For convenience, $\mathbf{x}_{km}$ and y_{km} will be used to denote both the observation and the point representing the observation. The observation y_{1m} is given by the linear combination $c_1x_{11m} + \ldots + c_vx_{1vm} = \mathbf{c}^T\mathbf{x}_{1m}$. The observation y_{km} is the canonical variate score for the mth observation for the kth group. Hence the point y_{2m} represents the projection of the observation $\mathbf{x}_{2m}$ onto the canonical vector. Similarly the points $\bar{y}_k$ represent the projections of the group means onto the canonical vector.

When all observations $\mathbf{x}_{km}$ are projected onto the canonical vector, a distribution of scores for each group will result. If the underlying distribution of the vectors of observations is multivariate Gaussian, then the histograms of canonical variate scores will follow the familiar bell-shaped appearance of a univariate Gaussian density. It is important to realize that the actual canonical variate scores do not follow a univariate Gaussian distribution, since the components of the vector of coefficients $\mathbf{c}$ are themselves realizations of random variables (e.g., Kshirsagar, 1972:197).

The orientation or direction of the canonical vector $\mathbf{c}$ is such that the ratio of the between- to within-groups SSQ from the one-way analysis of variance of the projected points y_{1m}, $m = 1, \ldots, n_1$; y_{2m}, $m = 1, \ldots, n_2$; $\ldots$; y_{gm}, $m = 1, \ldots, n_g$, is greater than that for any other orientation of the canonical vector.

The ratio of the between-groups to the within-groups SSQ gives the canonical root. The cosines of the angles between the canonical vector and the original coordinate axes give the components of the canonical vector. The projected points or observations are the canonical variate scores.

The property of maximum ratio of between- to within-groups variation defines the first canonical vector. This first axis is again a line of best fit, though the fit is now to the group means, and the shape of the concentration ellipsoids must be taken into consideration. A geometrical explanation is given in the next section.

In canonical variate analysis, the degree of correlation between, and the variances of, the original variables determine the degree and direction of maximum between- to within-group variation. Variables with high positive within-groups correlation, and negative between-groups correlation, provide maximum discrimination (e.g., Lubischew, 1962:fig. 1(a)); the reverse is also true. A very slight shift in the ratio of the two variables will provide almost complete discrimination. The lower the absolute value of the within-groups correlation, the poorer is the discrimination (Lubischew, 1962:fig. 1(a)).

The within-group variation is taken as the appropriate measure against which to judge between-group variation. The distance between the groups, or between individuals, is judged relative to the variances and correlations between the variables.

Phillips, Campbell, and Wilson (1973: figs. 5, 6) show group centroids and concentration ellipses for three groups and two variables with (a) the same variances but differing degrees of correlation within groups; and (b) differing variances but the same degree of correlation between the two variables. The degree of overlap on the first canonical variate increases as the within-groups correlation decreases, so that relative between-groups dispersion is less marked. The orientation of the canonical vector also changes. As the within-groups correlation tends to zero, the first canonical vector becomes more closely oriented with the abscissa.

As the within-groups variances change in figure 6 of Phillips, Campbell and Wilson (1973), the orientation of the canonical vectors changes to maintain maximum relative between-groups variation. While the changes in orientation of the first canonical vector are relatively small when compared with the changes due to different correlation, the effect on the degree of separation of two of the groups is marked. For example, with within-groups variances of 1.0 and 3.0, there is effectively complete separation of all three groups, or marked separation of group I and considerable overlap of groups II and III, depending on the ratio of the variances.

The canonical variates provide a simplified description of the group configurations. A related statistic, Mahalanobis D^2, provides a measure of the distance between the groups in the total variable space. Traditionally, the squared distance between any two groups in Figure 2 would be measured by their Euclidean or Pythagorean distance, i.e. by taking the difference between the group means for each coordinate, squaring the difference and summing. However, such a measure fails to take account of the correlations between the variables. Mahalanobis D^2 incorporates the effect of variable correlations.

GEOMETRY OF CANONICAL VARIATE ANALYSIS AND MAHALANOBIS D^2

Canonical variate analysis can be considered as a two-stage rotation procedure. The first stage involves description of the variation within groups, by orthogonal rotation of the original variables to new uncorrelated variables. One of the most common ways to achieve the first-stage rotation is from a principal component analysis. The new uncorrelated principal component variables are then scaled by the square roots of the corresponding eigenvalues to have unit variance within groups, so that the resulting variables are orthonormal. The rotation and scaling has the effect of transforming the within-groups concentration ellipsoid to a sphere.

Figure 3(a) shows the group means $\bar{\mathbf{x}}_k$ and associated concentration ellipses for two variables. Figure 3(b) shows the same configuration of means, with the individual concentration ellipses replaced by the concentration ellipse corresponding to the pooled within-groups SSQPR matrix. The first-stage principal component analysis corresponds to finding the principal axes of the pooled within-groups concentration ellipse. The eigenanalysis of the within-groups SSQPR matrix gives the principal component scores p_{ikm}. Figure 3(c) shows the initial config-

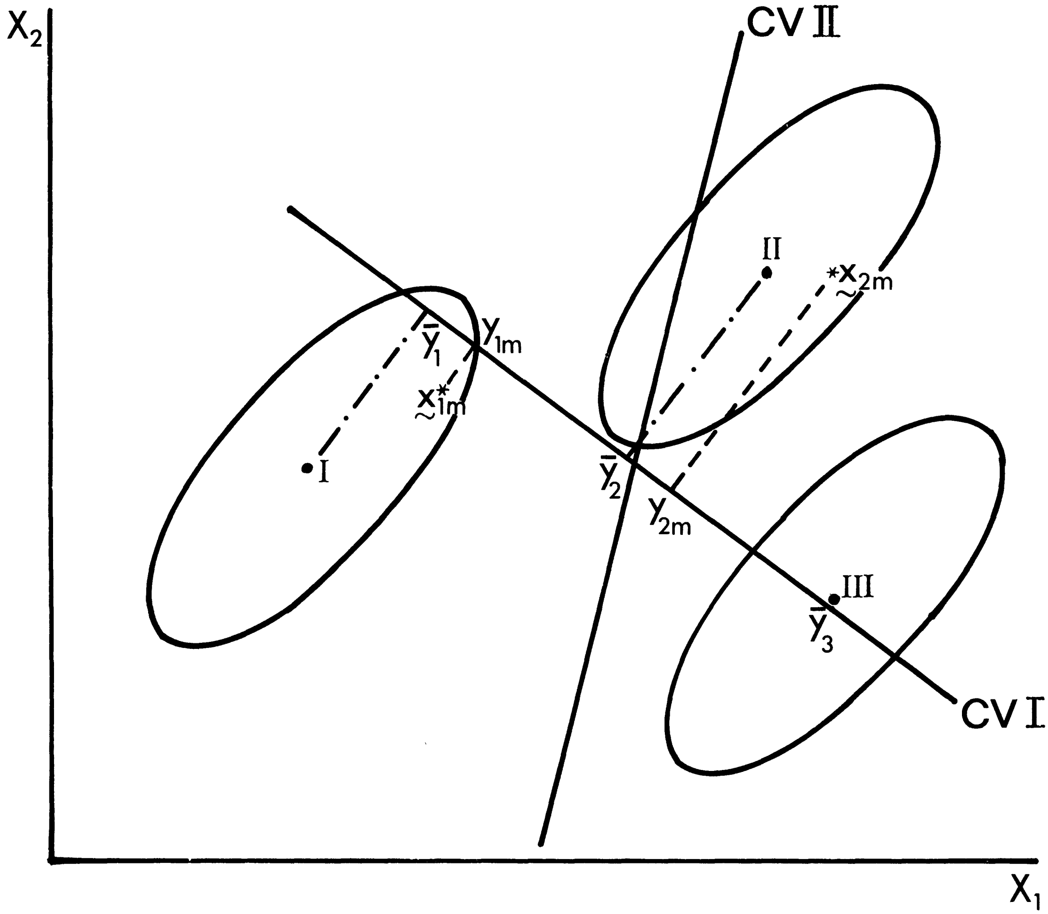


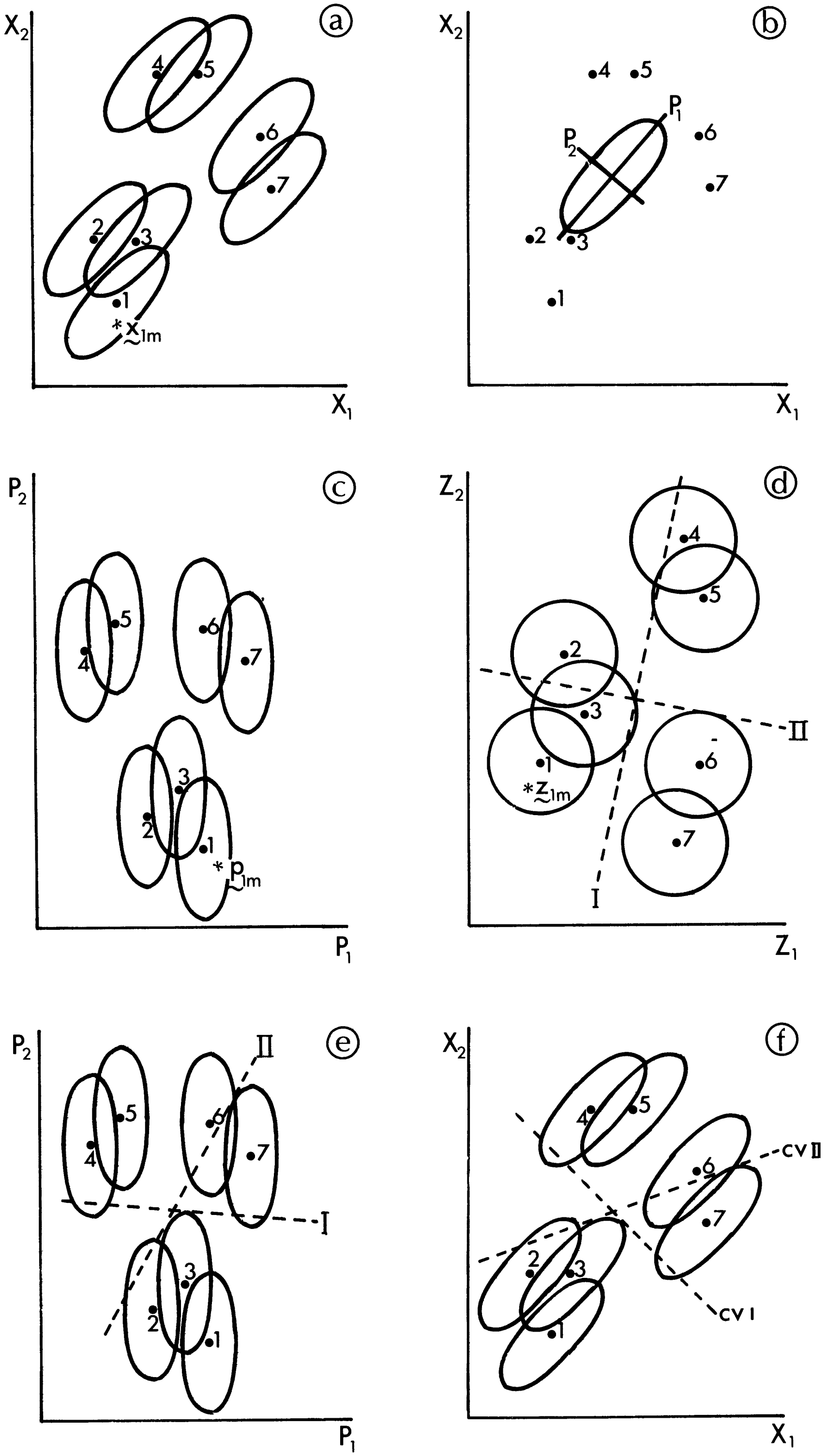
FIG. 2.—Representation of the canonical vectors for three groups and two variables. The group means (I, II and III) and 95% concentration ellipses are shown. The vectors CVI and CVII are the two canonical vectors. In the text, CVI = **c**. The points y_{1m} and y_{2m} represent the canonical variate scores corresponding to the first canonical vector for the observations $\mathbf{x}_{1m}$ and $\mathbf{x}_{2m}$.

uration with the principal components P_1 and P_2 as the coordinate axes.

The first-stage analysis involves rotation and scaling, from concentration ellipsoids to concentration spheres. Since the sample variance of the variable p_{ikm} is the eigenvalue, dividing the p_{ikm} by the square root of the eigenvalue will give a new variable z_{ikm} having unit variance. Figure 3(d) shows the effect of scaling each orthogonal principal component to produce orthonormal variables. The scaling transforms the within-groups concentration ellipse to a concentration circle.

The relative positions of the group means are now changed. In Figures 3(a) to 3(c), the means are associated with elliptical concentration contours, and so Mahalanobis D^2 is the appropriate distance between any pair of groups. In Figure 3(d), the concentration contours are now circular, indicating that the new variables are uncorrelated, with unit variance. The usual Euclidean or Pythagorean distance can now be used to determine distances. In particular the squared Mahalanobis distance between any pair of groups is simply the square of the distance between the group means in the rotated and scaled space depicted in Figure 3(d).

It can be shown that the rotation and



a
X_2
X_1
4
5
6
7
2
3
1
$*x_{1m}$
b
X_2
X_1
4
5
P_1
P_2
6
7
2
3
1
c
P_2
P_1
5
4
6
7
3
2
1
$*p_{1m}$
d
Z_2
Z_1
4
5
2
3
II
1
$*z_{1m}$
6
7
I
e
P_2
P_1
II
5
4
6
7
I
3
2
1
f
X_2
X_1
4
5
6
cv II
7
2
3
1
cv I

scaling is equivalent to expressing the axes of the original rectangular coordinate system as oblique axes. The cosine of the angle between any two of the oblique axes is equivalent to the partial correlation coefficient between the variables, and each variable is expressed on a scale on which one unit is equal to one standard deviation.

The rotated and scaled axes, which reflect patterns of within-group variation, now become the reference coordinate axes for the second stage of the analysis. The original group means are considered relative to these axes, as in Figure 3(d).

The second-stage rotation is again accomplished by a principal component analysis, this time of the group means $\bar{\mathbf{z}}_k$ for the new orthonormal variables. This provides an examination of the between-groups variation, relative to the patterns of within-group variation defined by the first-stage principal components. The eigenvalues give the usual sample canonical roots f while the eigenvectors give the canonical vectors $\mathbf{a}_i$ for the orthonormal variables.

Note that the second-stage principal component analysis is carried out with the group means for the orthonormal variables weighted by the corresponding numbers in each group. This use of a weighted between-groups SSQPR matrix gives the maximum likelihood solution. An alternative is to calculate an unweighted between-groups SSQPR matrix, in which the sample sizes are ignored.

The canonical vectors $\mathbf{c}_i$ for the original variables are found by reversing the scaling and rotation of the first-stage analysis, as shown in Figures 3(e) to 3(f). While the canonical vectors $\mathbf{a}_i$ for the orthonormal variables are orthogonal, the canonical vectors $\mathbf{c}_i$ for the original variables will not, in general, be orthogonal, as shown in Figure 3(f). However, the canonical variate scores $\mathbf{c}_i^T\mathbf{x}$ are uncorrelated with the scores $\mathbf{c}_j^T\mathbf{x}$ within each group, since by the nature of the rotation, the canonical vectors are orthogonal with respect to the within-groups covariance matrix $\mathbf{V}$.

The data represented in Figure 3(a) can first be scaled by the pooled within-groups standard deviations to unit standard deviation along each coordinate axis. The first-stage principal component analysis is then based on the correlation matrix derived from the pooled within-groups SSQPR matrix; this correlation matrix will be referred to subsequently as the pooled (within-groups) correlation matrix. The geometry of canonical variate analysis then follows as above, though the resulting canonical vectors $\mathbf{c}_s$ are those for standardized variables, and will be referred to as the standardized canonical vectors. The components of the vector $\mathbf{c}_s$ are given by multiplying the components of the vector $\mathbf{c}$ by the corresponding pooled standard deviations.

Consider again the orthonormal variable space, in which concentration ellipsoids are transformed to concentration spheres. The first canonical variate for the orthonormal variables is the line of closest fit to the group means in this space. The second canonical variate is

←

FIG. 3.—Illustration of the rotation and scaling implicit in the calculation of the canonical vectors. 3(a)—group means and associated 95% concentration ellipses for two variables and seven groups. Idealized observation $\mathbf{x}_{1m}$ is indicated; 3(b)—group means, with concentration ellipses centred at overall mean. Principal axes P_1 and P_2 are indicated; 3(c)—rotation to principal axes P_1 and P_2 of the common covariance matrix. The point $\mathbf{p}_{1m}$ gives the principal component scores for the observation $\mathbf{x}_{1m}$; 3(d)—scaling from orthogonal variables to orthonormal variables, so that concentration ellipses become concentration circles. The point $\mathbf{z}_{1m}$ represents the observation $\mathbf{x}_{1m}$ in these new coordinates. The axes I and II are the principal axes for the group means; 3(e)—the scaling from orthogonal to orthonormal variables is reversed. The coordinates P_1 and P_2 are as in 3(c); 3(f)—the rotation from the original variables to the orthogonal variables is reversed. CVI and CVII represent the canonical vectors.

orthogonal to the first in this space. Since the orthonormal variables reflect patterns of within-groups variation, the orthogonality in the original variable space is with respect to the corresponding within-groups covariance matrix.

AN ALGEBRAIC APPROACH

Canonical variate analysis seeks a linear combination $y_{km} = \mathbf{c}^T\mathbf{x}_{km}$ of the original observations $\mathbf{x}_{km}$ such that the ratio of the between-groups to the within-groups SSQ for a one-way analysis of variance of the y_{km} is a maximum.

A one-way analysis of variance of the univariate canonical variate scores y_{km} involves the usual within-groups SSQ

$$\sum_{k=1}^{g}\sum_{m=1}^{n_k}(y_{km} - \bar{y}_k)^2$$

and the between-groups SSQ

$$\sum_{k=1}^{g} n_k(\bar{y}_k - \bar{y}_T)^2,$$

with $\bar{y}_k = n^{-1}\sum_{m=1}^{n_k} y_{km}$, $n_T = \sum_{k=1}^{g} n_k$, and $\bar{y}_T = n_T^{-1}\sum_{k=1}^{g} n_k\bar{y}_k$.

Since the canonical variate score is given algebraically by $y_{km} = \mathbf{c}^T\mathbf{x}_{km}$, the within-groups SSQ can be rewritten as

$$\sum_{k=1}^{g}\sum_{m=1}^{n_k}(y_{km} - \bar{y}_k)^2 = \sum_{k=1}^{g}\sum_{m=1}^{n_k}(\mathbf{c}^T\mathbf{x}_{km} - \mathbf{c}^T\bar{\mathbf{x}}_k)^2$$

and this is the same as

$$\sum_{k=1}^{g}\sum_{m=1}^{n_k}\{\mathbf{c}^T(\mathbf{x}_{km} - \bar{\mathbf{x}}_k)\}^2.$$

The term $\mathbf{c}^T(\mathbf{x}_{km} - \bar{\mathbf{x}}_k)$ inside the $\{\ldots\}$ is a scalar quantity, and can also be written as $(\mathbf{x}_{km} - \bar{\mathbf{x}}_k)^T\mathbf{c}$, so that the within-groups SSQ becomes

$$\sum_{k=1}^{g}\sum_{m=1}^{n_k}\mathbf{c}^T(\mathbf{x}_{km} - \bar{\mathbf{x}}_k)(\mathbf{x}_{km} - \bar{\mathbf{x}}_k)^T\mathbf{c}.$$

Since the canonical vector $\mathbf{c}$ is the same for all observations for all groups, the within-groups SSQ may also be written as

$$\mathbf{c}^T\left\{\sum_{k=1}^{g}\sum_{m=1}^{n_k}(\mathbf{x}_{km} - \bar{\mathbf{x}}_k)(\mathbf{x}_{km} - \bar{\mathbf{x}}_k)^T\right\}\mathbf{c}.$$

But the term in $\{\ldots\}$ is the familiar form of the pooled within-groups SSQPR matrix, $\mathbf{W}$; it reflects the squared deviations and cross deviations of each observation from the mean of its corresponding group. To see this, note that the entry for the ith variable is

$$\sum_{k=1}^{g}\sum_{m=1}^{n_k}(x_{kim} - \bar{x}_{ki})^2$$

while that for the ith and jth variables is

$$\sum_{k=1}^{g}\sum_{m=1}^{n_k}(x_{kim} - \bar{x}_{ki})(x_{kjm} - \bar{x}_{kj});$$

these are the within-groups or error terms in analysis of covariance.

Hence the within-groups SSQ can be written as $\mathbf{c}^T\mathbf{W}\mathbf{c}$.

The between-groups SSQ, $\sum_{k=1}^{g} n_k(\bar{y}_k - \bar{y}_T)^2$, for the canonical variate scores can be written in a similar way to the within-groups SSQ. Condensing the steps gives

$$\begin{aligned}\sum_{k=1}^{g} n_k(\bar{y}_k - \bar{y}_T)^2 &= \sum_{k=1}^{g} n_k(\mathbf{c}^T\bar{\mathbf{x}}_k - \mathbf{c}^T\bar{\mathbf{x}}_T)^2\\ &= \sum_{k=1}^{g} n_k\{\mathbf{c}^T(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_T)\}^2\\ &= \sum_{k=1}^{g} n_k\mathbf{c}^T(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_T)(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_T)^T\mathbf{c}\\ &= \mathbf{c}^T\left\{\sum_{k=1}^{g} n_k(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_T)(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_T)^T\right\}\mathbf{c}.\end{aligned}$$

The term in $\{\ldots\}$ is the between-groups SSQPR matrix; it reflects the squared deviations and cross deviations of each mean from the mean of the means.

Hence the between-groups SSQ can be written as $\mathbf{c}^T\mathbf{B}\mathbf{c}$.

The canonical vector $\mathbf{c}$ is chosen to maximize the ratio of the between- to within-groups SSQ of the resulting linear combination, i.e. to maximize the ratio $f = \mathbf{c}^T\mathbf{B}\mathbf{c}/\mathbf{c}^T\mathbf{W}\mathbf{c}$. The vector $\mathbf{c}$ is usually scaled so that the average within-groups variance of the canonical variate scores is unity. With $n_w = \sum_{k=1}^{g} (n_k - 1)$, this requirement becomes

$$n_w^{-1} \sum_{k=1}^{g} \sum_{m=1}^{n_k} (y_{km} - \bar{y}_k)^2 = 1.$$

In matrix notation, this is equivalent to specifying that

$$n_W^{-1}\mathbf{c}^T\mathbf{W}\mathbf{c} = 1, \quad \text{or,}$$

$$\text{with} \quad \mathbf{V} = n_w^{-1}\mathbf{W}, \quad \text{that} \quad \mathbf{c}^T\mathbf{V}\mathbf{c} = 1.$$

Choosing $\mathbf{c}$ to maximize the ratio f leads to the fundamental canonical variate eigenequation

$$(\mathbf{B} - f\mathbf{W})\mathbf{c} = \mathbf{o} \qquad (3)$$

or

$$\mathbf{B}\mathbf{c} = \mathbf{W}\mathbf{c}f.$$

For g groups and v variables, there are $h = \min(v, g - 1)$ canonical vectors with associated non-zero canonical roots. When $g - 1 < v$, the sample group means lie in a $h = g - 1$ dimensional subspace. The canonical vectors provide an alternative description of the h-dimensional space.

Write

$$\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_h)$$

and

$$\mathbf{F} = \mathrm{diag}(f_1, \ldots, f_h).$$

Then the eigenanalysis in (3) becomes

$$\mathbf{BC} = \mathbf{WCF} \qquad (4)$$

with the scaling

$$\mathbf{C}^T\mathbf{WC} = n_w\mathbf{I}$$

and

$$\mathbf{C}^T\mathbf{BC} = n_w\mathbf{F}\ .$$

COMPUTATIONAL ASPECTS

The geometrical approach given above may be expressed algebraically as follows. The first-stage principal component analysis corresponds to finding the principal axes of the pooled within-groups concentration ellipsoid. This is achieved algebraically by an eigenanalysis of the within-groups SSQPR matrix $\mathbf{W}$. Write $\mathbf{W}$ in terms of its eigenvectors $\mathbf{U}$ and eigenvalues $\mathbf{E}$, viz.

$$\mathbf{W} = \mathbf{UEU}^T,$$

with $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_v)$ and $\mathbf{E} = \mathrm{diag}(e_1, \ldots, e_v)$.

The principal component scores in Figure 3(c) are then given by $p_{ikm} = \mathbf{u}_i^T\mathbf{x}_{km}$, or $\mathbf{p}_{km} = \mathbf{U}^T\mathbf{x}_{km}$. The pooled within-groups variance $(p_{ikm} - \bar{p}_{ik})^2$ of the scores p_{ikm} for the ith principal component is simply the corresponding eigenvalue e_i. To see this, follow the same steps as for the derivation of the within-groups SSQ in the previous, viz.

$$\begin{aligned}
&\sum_{k=1}^{g} \sum_{m=1}^{n_k} (p_{ikm} - p_{ik})^2 \\
&= \sum_{k=1}^{g} \sum_{m=1}^{n_k} \{\mathbf{u}_i^T(\mathbf{x}_{km} - \bar{\mathbf{x}}_k)\}^2 \\
&= \mathbf{u}_i^T\left\{ \sum_{k=1}^{g} \sum_{m=1}^{n_k} (\mathbf{x}_{km} - \bar{\mathbf{x}}_k)(\mathbf{x}_{km} - \bar{\mathbf{x}}_k)^T \right\}\mathbf{u}_i \\
&= \mathbf{u}_i^T\mathbf{W}\mathbf{u}_i = e_i.
\end{aligned}$$

The transformation from concentration ellipsoids to concentration spheres in Figure 3(d) is given by $z_{ikm} = e_i^{-1/2}p_{ikm} = e_i^{-1/2}\mathbf{u}_i^T\mathbf{x}_{km}$ or $\mathbf{z}_{km} = \mathbf{E}^{-1/2}\mathbf{U}^T\mathbf{x}_{km}$. For the second-stage analysis, the group means for the original variables are expressed in terms of these new orthonormal variables. The rotated and scaled vector of means for the kth group is $\bar{\mathbf{z}}_k = \mathbf{E}^{-1/2}\mathbf{U}^T\bar{\mathbf{x}}_k$.

Let $\bar{\mathbf{X}}$ denote the gxv matrix of group means, centered so that the mean of the means is zero, with each vector of means weighted by the corresponding sample size, viz.

$$\bar{\mathbf{X}}^T = \{n_1^{1/2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_T), \ldots, n_g^{1/2}(\bar{\mathbf{x}}_g - \bar{\mathbf{x}}_T)\}.$$

The between-groups SSQPR matrix is then $\bar{\mathbf{X}}^T\bar{\mathbf{X}}$. The matrix of group means for the orthonormal variables is $\bar{\mathbf{Z}}^T = \mathbf{E}^{-1/2}\mathbf{U}^T\bar{\mathbf{X}}^T$. The between-groups matrix for the orthonormal variables is then given by

$$\begin{aligned}
\mathbf{Z}^T\mathbf{Z} &= \mathbf{E}^{-1/2}\mathbf{U}^T\bar{\mathbf{X}}^T\bar{\mathbf{X}}\mathbf{U}\mathbf{E}^{-1/2} \\
&= \mathbf{E}^{-1/2}\mathbf{U}^T\mathbf{B}\mathbf{U}\mathbf{E}^{-1/2}. \qquad (5)
\end{aligned}$$

The second-stage rotation results from an eigenanalysis of this between-groups

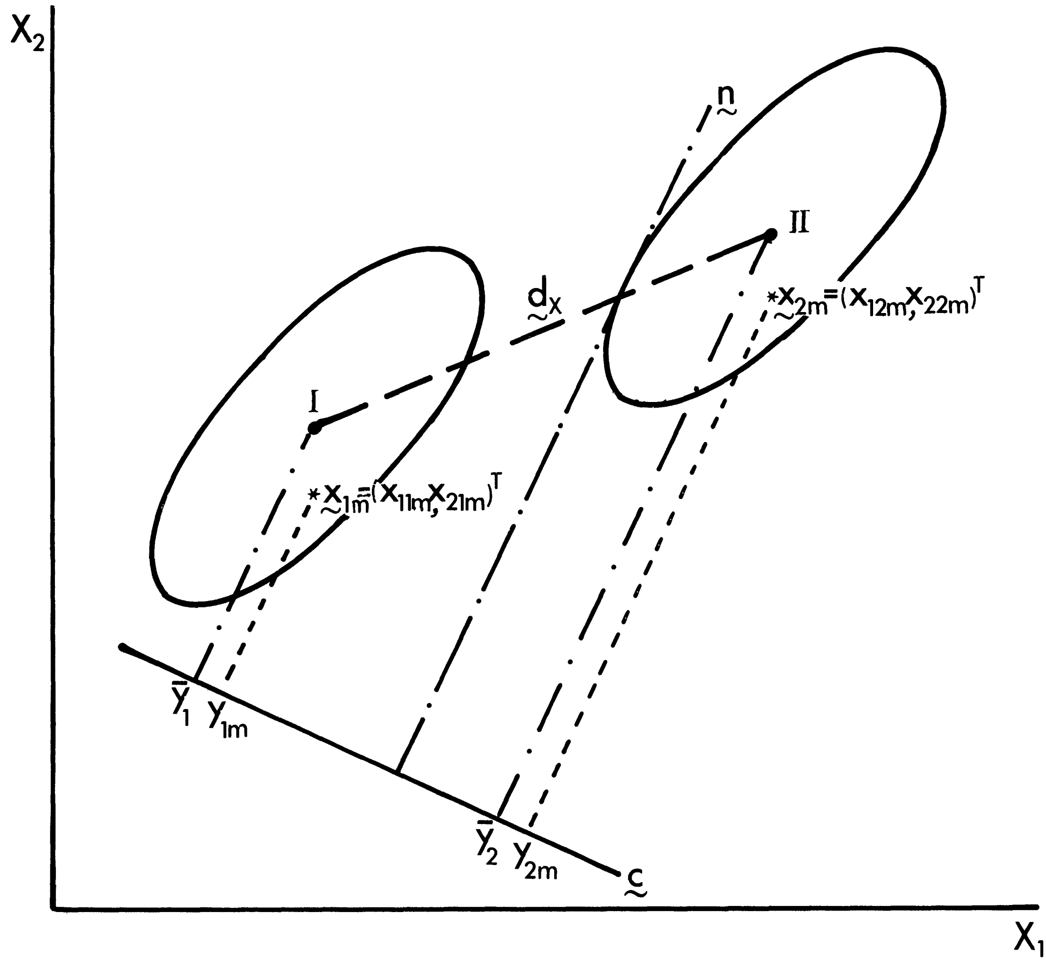


FIG. 4.—Representation of the discriminant function for two groups and two variables, showing the group means and associated 95% concentration ellipses. The vector **c** is the discriminant vector. The points $\bar{y}_1$ and $\bar{y}_2$ represent the discriminant means for the two groups.

The discriminant vector can be constructed by drawing the tangent **n** to the concentration ellipse at the point of intersection with the line **d** joining the group means; the discriminant vector is orthogonal to the tangent **n**.

matrix. The second-stage principal component analysis is

$$(\mathbf{E}^{-1/2}\,\mathbf{U}^T\mathbf{B}\mathbf{U}\mathbf{E}^{-1/2} - \mathrm{f}\,\mathbf{I})\,\mathbf{a} = \mathbf{o}, \qquad (6)$$

giving the canonical roots f_i and canonical vectors $\mathbf{a}_i$ for the orthonormal variables. Premultiplication by $\mathbf{UE}^{-1/2}$ shows that the canonical vectors $\mathbf{c}_i$ for the original variables $\mathbf{x}$ are found from the $\mathbf{a}_i$ by

$$\mathbf{c}_i = \mathbf{UE}^{-1/2}\,\mathbf{a}_i\ .$$

The computational aspects described above are those followed in many computer programs. The advantages of a first-stage principal component rotation in morphometric studies are illustrated in the last section.

THE TWO-GROUP DISCRIMINANT FUNCTION

When there are only two groups, a canonical variate analysis simplifies to the linear discriminant of Fisher (1936). The two-group case is both conceptually and computationally simpler than the multiple-group canonical variate analysis.

Figure 4 depicts a typical situation for two groups and two variables. The basic approach follows that outlined in the third section. The discriminant vector $\mathbf{c}$ defines that direction which gives maximum between- to within-groups variation of the discriminant scores y_{1m}, $m = 1, \ldots, n_1$; y_{2m}, $m = 1, \ldots, n_2$.

Given the group means and associated concentration ellipses, there is a simple geometrical construction for the discriminant vector: (i) join the group means to give the vector $\mathbf{d}_x$; (ii) construct the tangent vector $\mathbf{n}$ at the point of intersection with the concentration ellipse; and (iii) construct the discriminant vector $\mathbf{c}$, orthogonal to the vector $\mathbf{n}$. This procedure can be simplified further, by determining concentration ellipses with increased probability levels. The vector joining the points of intersection of the overlapping ellipses is again the normal vector $\mathbf{n}$. The position of this latter vector is such that it passes through the mean of the means.

For two groups, the squared distance between the canonical variate or discriminant means $\bar{y}_1$ and $\bar{y}_2$ is the squared Mahalanobis distance. This is defined as

$$D^2 = \mathbf{d}_x^T \mathbf{V}^{-1} \mathbf{d}_x,$$

where $\mathbf{d}_x = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$ is the vector of differences between the group means.

The discriminant vector $\mathbf{c}$ is then given by

$$\mathbf{c} = D^{-1} \mathbf{V}^{-1} \mathbf{d}_x.$$

The component D^{-1} does not usually enter the definition of the discriminant vector. With the definition $\mathbf{c}_u = \mathbf{V}^{-1}\mathbf{d}_x$, the within-groups variance is then equal to D^2, while the squared difference between the means for the unscaled discriminant scores is D^4. The ratio of squared difference between the means to the within-groups variance is D^2.

The canonical or discriminant root is given by

$$f = n_w^{-1} n_1 n_2 n_T^{-1} D^2$$

and the between-groups SSQPR matrix is

$$\mathbf{B} = n_1 n_2 n_T^{-1} \mathbf{d}_x \mathbf{d}_x^T .$$

DETERMINING THE IMPORTANT VARIABLES

Various approaches have been proposed to determine the variables which contribute most to the group separation. Probably the most widely used approach is that based on the relative magnitudes of the canonical variate coefficients for the variables standardized to unit standard deviation within groups. The standardized coefficients are given by multiplying the original coefficients by the pooled within-groups standard deviations. Variables with the larger absolute values of the standardized coefficients are often taken to be the more important ones.

Variables with small standardized coefficients can nearly always be eliminated. However when some of the variables are highly correlated within groups, those variables with the larger absolute coefficients are not necessarily the more important ones. With the presence of highly correlated variables, it is important to examine the stability of the coefficients. When there is little variation between the group means for the orthonormal variables along a particular within-groups direction, and the corresponding within-groups eigenvalue is also small, marked instability can be expected in some of the coefficients defining the canonical variates. To be more specific, those variables with large loadings for the corresponding within-groups eigenvector may have unstable coefficients for the canonical variates. The degree of instability will depend on the contribution of the corresponding orthonormal variable to the discrimination and on the magnitude of the within-groups eigenvalue. As a practical guideline, when the between-groups SSQ for a particular orthonormal variable is small (say, less than 5–10% of the total between-groups variation), and the corresponding eigenvalue is also small (say, less than 1–2% of the sum of the eigenvalues), then some instability can be expected.

One approach to the potential problem of unstable coefficients is to introduce shrunken estimators (Campbell and Reyment, 1978). In practice, this involves adding shrinkage constants to some or all of the within-groups eigenvalues. This modification is done before these eigenvalues are used to scale the uncorrelated first-stage principal component variables to produce the first-stage orthonormal variables.

It is often observed that while some of the coefficients of the canonical variates of interest change in magnitude and often in sign when these shrunken estimator procedures are introduced, shrinking the contribution of a within-groups eigenvector/value combination has little effect on the corresponding canonical roots. This indicates that little or no discriminatory information has been lost. When this occurs, the obvious conclusion is that one or some of the variables contributing most to the orthonormal variable whose effect has been shrunk have little influence on the discrimination. The variables involved are those that make the greatest contribution to the corresponding eigenvector. In general, one or some of these variables can then be eliminated.

A further advantage of this type of procedure is that the computational routine involved can be used to assess the contribution of each of the first-stage principal components to the discrimination. This is useful in morphometric studies, since the various eigenvectors can often be associated with patterns of growth.

ACKNOWLEDGMENTS

Thanks are due to Richard Litchfield for drafting the figures. Atchley was supported by the College of Agriculture and Life Sciences, University of Wisconsin, Madison and by NSF, DEB 7923012 and DEB 7906058.

REFERENCES

CAMPBELL, N. A., AND REYMENT, R. A. 1978. Discriminant analysis of a Cretaceous foraminifer using shrunken estimators. Math. Geol., 10:347–359.

DEMPSTER, A. P. 1969. Elements of continuous multivatiate analysis. Addison-Wesley, Reading, Mass. 338 pp.

FISHER, R. A. 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen., 7:179–188.

JOLICOEUR, P., AND MOSIMANN, J. E. 1960. Size and shape variation in the painted turtle, a principal component analysis. Growth, 24:339–354.

KSHIRSAGAR, A. M. 1972. Multivariate analysis. Marcell Dekker, New York, 534 pp.

LUBISCHEW, A. A. 1962. On the use of discriminant functions in taxonomy. Biometrics, 18:455–477.

PHILLIPS, B. F., CAMPBELL, N. A., AND WILSON, B. R. 1973. A multivariate study of geographic variation in the whelk *Dicathais*. J. Exp. Mar. Biol. Ecol., 11:29–63.

Manuscript received May 1980
Revised February 1981