QÜESTIIÓ, Vol 19, 1, 2, 3 pp. 23-63, 1995

A REVIEW OF CANONICAL COORDINATES AND AN ALTERNATIVE TO CORRESPONDENCE ANALYSIS USING HELLINGER DISTANCE

C. RADHAKRISHNA RAO*
Pennsylvania State University

In this paper, a general theory of canonical coordinates is developed for reduction of dimensionality in multivariate data, assessing the loss of information and plotting higher dimensional data in two or three dimensions for visual displays. The theory is applied to data in two way tables with variables in one category and samples (individual or populations) in the other. Two types of data are considered, one with continuous measurements on the variables and another with frequencies of attributes. An alternative to the usual correspondence analysis of contingency tables based on Hellinger rather than the chisquare distance is suggested. The new method has some attractive features and does not suffer from some inherent drawbacks resulting from the use of the chi-square distance and variable sample sizes for the populations in the correspondence analysis. The technique of biplots where the populations and the variables are represented on the same chart is discussed.

Key words: Canonical coordinates, Chisquare distance, Contingency tables, Correspondence analysis, Hellinger distance, Matrix approximations, Power divergence tests, Principal component analysis.

AMS classification index: 62H30, 62H17.

*C. Radhakrishna Rao. Department of Statistics. Pennsylvania State University. University Park, PA 16802.

–Article rebut el gener de 1995.

–Acceptat el maig de 1995.

1. INTRODUCTION

The concept of canonical variates (coordinates) was introduced in an early paper by the author (Rao (1948)) for graphical representation of taxonomical units characterized by multiple measurements. This was, perhaps, the first attempt to reduce high dimensional data to two or three dimensions using an objective criterion for purposes of graphical displays. Since then, graphical representation of multivariate data for visual examination of clusters, outliers and other structures in the data has been an active field of research. Some of the developments are biplots (Gabriel (1971), Gifi (1990), Gower (1993), Greenacre (1993)) multidimensional scaling (Kruskal and Wish (1978)), correspondence analysis (Benzécri (1992), Greenacre (1984)), Chernoff's faces (Chernoff (1993)) and parallel coordinates (Mahalanobis, Mazumdar and Rao (1949), Wegman (1990)). Cavalli-Sforza (1991) uses canonical coordinates (variables) in interpreting the evolution of human populations.

The object of the present paper is to briefly review the concept of canonical coordinates as originally introduced in 1948 and later elaborated in Rao (1964, 1979, 1980, 1985) in the light of modern developments and present an alternative to the current practice of correspondence analysis, which seems to have some attractive properties.

In Section 2 we consider the general problem of transforming the points of a p-dimensional vector space endowed with a specified inner product to a lower dimensional Euclidean space with the usual definition of inner product and distance. The solution to the problem is considered in a more general set up than what is possible through the use of Eckart and Young (1936) theorem. In Section 3, some measures are introduced to assess the loss of information in reduction of dimensionality. The role of biplots and their interpretation are also discussed. An example with continuous measurements is considered in Section 4. An alternative to correspondence analysis applied to contingency tables based on Hellinger rather than the chisquare distance is also given in Section 4. Some results on optimization problems and chi-square goodness-of-fit tests are discussed in the Appendix.

2. REDUCTION OF DIMENSIONALITY

The problem we consider may be stated as follows. Let $X = (X_1 : \cdots : X_m)$ be a $p \times m$ data matrix, with the i-th column vector X_i representing measurements of p variables made on the i-th population (individual or unit). The column vector X_i will be referred to as the i-th population profile (PP). The PP's can be represented as m points in a p-dimensional vector space R^p with a specified inner product and the associated norm

$$(2.1) \qquad (x,y) = x'My, \quad x,y \in R^p$$

$$(2.2) \qquad \|x\| = (x,x)^{1/2}, \quad x \in R^p$$

where M is a positive definite matrix. We may call this the Mahalanobis or M-space. In practical situations, it may be necessary to attach a weight $w_i \geq 0$ to the i-th PP, the exact use of which will be detailed in the following discussion. We represent the vector $(w_1,\ldots,w_m)'$ by w and the diagonal matrix with w_i as the i-th diagonal element by W. The M-space with weight as an additional dimension will be referred to as WM-space. [In our treatment we consider W as a general positive definite matrix to cover more general applications].

The problem is to find a $k \times m$ matrix

$$(2.3) \qquad Y = (Y_1 : \cdots : Y_m)$$

with $k < p$ for representing the PP's in a k-dimensional Euclidean space (E^k) with the usual inner product, $x'y$ for $x,y \in E^k$, and the k-vector Y_i as the profile of the i-th population, in such a way that the relative positions of the PP's in the M-space (in terms of distances between profiles) are preserved to the extent possible in E^k. For this purpose, we need to have a criterion for measuring the loss of information in reducing the dimension of the profile space, by minimizing which we obtain an optimum solution for (2.3).

The relative positions of the PP's in the M-space can be described by what may called a *configuration matrix*

$$(2.4) \qquad C = (X - \xi 1')'M(X - \xi 1') = ((X_i - \xi)'M(X_j - \xi)) = (c_{ij})$$

where ξ is some chosen reference (profile) vector, 1 is the column vector of unities and the c_{ij}'s represent the distances and angles between profiles as explained in the Figure 1.

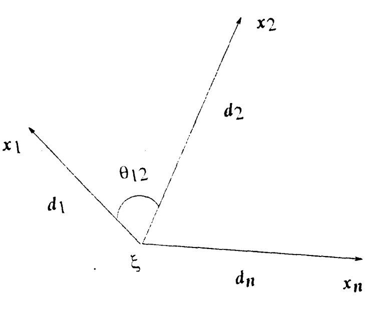

Figure 1.
Configuration of the profiles in M-space ($d_i = \sqrt{c_{ii}}, \quad d_i d_j \cos\theta_{ij} = c_{ij}$)

The corresponding configuration about the origin in the reduced space E^k is $Y'Y$. The problem then reduces to minimizing

$$\|C - Y'Y\| \tag{2.5}$$

with respect to Y, a $k \times m$ matrix as defined in (2.3), for a suitably chosen matrix norm. The following theorem provides the solution.

Theorem 1

Consider the s.v.d. (singular value decomposition)

$$M^{1/2}(X - \xi 1')W^{1/2} = \lambda_1 U_1 V_1' + \ldots + \lambda_p U_p V_p' \tag{2.6}$$

with singular values $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$, where $M^{1/2}$ and $W^{1/2}$ are symmetric square roots of M and W. Then the choice

$$Y = Y_{(k)} = \begin{pmatrix} \lambda_1 V_1' W^{-1/2} \\ \vdots \\ \lambda_k V_k' W^{-1/2} \end{pmatrix} \tag{2.7a}$$

or conventionally written in the transposed form

$$\lambda_1 W^{-1/2} V_1, \lambda_2 W^{-1/2} V_2, \ldots, \lambda_k W^{-1/2} V_k \tag{2.7b}$$

where the components of the i-th m-vector are the i-th canonical coordinates (i.e., the coordinates in the i-th dimension of the reduced space) for the different populations, minimizes (2.5) for any (W, W)-invariant norm as defined in Note 2.1. We call these coordinates the *canonical coordinates* for populations or profiles (CCP).

The result (2.7a) follows from Theorems A.1 and A.2 given in the Appendix.

Note 2.1. (A, B)-invariant norm of an $m \times n$ matrix is the usual norm (satisfying the postulates of a norm) with the additional property

$$\|C \cdot D\| = \| \cdot \| \quad \text{for any} \quad C, D \text{ such that} \qquad C'AC = A, D'BD = B \tag{2.8}$$

where C is an $m \times m$ matrix, D is an $n \times n$ matrix, and A and B are positive definite matrices of orders m and n respectively. This is a generalization given in Rao (1980) of a unitarily invariant norm defined by von Neumann (1937) with A and B as unit matrices.

Note 2.2. In our applications, we indicate some choices of the reference vector ξ. However, we note that a further minimization of (2.5) with respect to ξ leads to the choice

$$\hat{\xi} = (1'W1)^{-1}XW1 \tag{2.9}$$

where 1 is the column vector of unities.

Note 2.3. Using the notation

$$\begin{aligned} \Lambda_{(i)} &= \text{Diag}(\lambda_1, \ldots, \lambda_i) \\ U_{(i)} &= (U_1 : \ldots : U_i), V_{(i)} = (V_1 : \ldots : V_i) \end{aligned}$$

we may write the solution Y given in (2.7a) in the concise form

$$Y_{(k)} = \Lambda_{(k)} V'_{(k)} W^{-1/2}. \tag{2.10}$$

Note 2.4. In the expression (2.6), a symmetric square root of a positive definite matrix is used. It can be computed in a simple way as follows. If A is a positive definite matrix of order p with the spectral decomposition

$$A = \Sigma\lambda_i^2 Q_i Q_i' = Q\Lambda^2 Q'$$

where $Q = (Q_1 : \ldots : Q_p)$, then

$$\begin{aligned} A^{1/2} &= \Sigma\lambda_i Q_i Q_i' = Q\Lambda Q' \\ A^{-1/2} &= \Sigma(\lambda_i)^{-1} Q_i Q_i' = Q\Lambda^{-1} Q'. \end{aligned} \tag{2.11}$$

We may look at the problem in a slightly different way by defining what is called the dispersion matrix between profiles

$$B = (X - \xi 1')W(X - \xi 1')' = (b_{ij}) \tag{2.12}$$

where b_{ii} is the weighted variance of the i-th variable and b_{ij} is the weighted covariance between the i-th and j-th variables across the profiles. Consider an approximation, $Z_i \in R^p$ to $(X_i - \xi)$, with the restriction that $Z_1, \ldots, Z_m$ lie in a k dimensional subspace of R^p, in which case we have the representation

$$Z = (Z_1 : \ldots : Z_m) = AC \tag{2.13}$$

where A is a $p \times k$ matrix whose columns span the subspace and C is a $k \times m$ matrix. Without loss of generality we may choose A to satisfy the condition $A'MA = I$ (i.e., the columns of A are orthonormal in the M-space). The dispersion matrix between profiles in the reduced space is $ACWC'A'$, and we choose A and C such that

$$||B - ACWC'A'|| \tag{2.14}$$

is a minimum for an appropriate norm of the matrix. The solution is given in Theorem 2.

Theorem 2

Consider the same s.v.d. as in Theorem 1

$$M^{1/2}(X - \xi 1')W^{1/2} = \lambda_1 U_1 V_1' + \ldots + \lambda_p U_p V_p'.$$

Then the optimum choice of AC which minimizes (2.14) for any (M,M)-invariant norm is

$$AC_{(k)} = M^{-1/2}(\lambda_1 U_1 V_1' + \ldots + \lambda_k U_k V_k')W^{-1/2} \tag{2.15}$$

where the suffix (k) is introduced to indicate the dimension of the reduced space. We may choose

$$\begin{aligned} A &= M^{-1/2}(U_1 : \ldots : U_k) = M^{-1/2}U_{(k)} \\ C_{(k)} &= \begin{pmatrix} \lambda_1 V_1' W^{-1/2} \\ \vdots \\ \lambda_k V_k' W^{-1/2} \end{pmatrix} = \Lambda_{(k)} V_{(k)}' W^{-1/2}. \end{aligned} \tag{2.16}$$

The results follow by applying Theorems A.1 and A.2 given in the Appendix.

Note 2.5. We may represent the profiles by plotting the columns of C in *a* k-dimensional Euclidean space, which is the same solution as that obtained in Theorem 1.

A geometric approach to the problem of reduction of dimensionality is to fit a k-dimensional plane to the data. A set of m points on a k-plane can be written as

$$\xi 1' + AC \tag{2.17}$$

where A is a $p \times k$ matrix and C is a $k \times m$ matrix. We determine A, C, ξ such that

$$||X - \xi 1' - AC|| \tag{2.18}$$

is a minimum for a suitably chosen norm. The solution is given in Theorem 3.

Theorem 3

Consider the same s.v.d. as in Theorem 1. Then the choices of A and C as in Theorem 2 and $\xi = (1'W1)^{-1}XW1$ as in (2.9) minimize any (M,W)-invariant norm of (2.18).

The result follows by a direct application of Theorem A.1 of the Appendix.

Note 2.6. We may also look at the problem in some other ways. Let T be a $k \times p$ matrix providing a transformation of the column vectors of X to $Y = TX$ in a k-dimensional space with the induced inner product matrix $TM^{-1}T'$. The squared distance between the i-th and j-th profiles is

$$D_{ij}^2 = (X_i - X_j)'M(X_i - X_j) \tag{2.19}$$

in the full space, and

$$D_{ij(k)}^2 = (X_i - X_j)'T'(TM^{-1}T')^{-1}T(X_i - X_j) \tag{2.20}$$

in the reduced space. By definition $D_{ij(k)}^2 \leq D_{ij}^2$. We may then choose T by minimizing some function of the differences or ratios of D_{ij}^2 and $D_{ij(k)}^2$.

One of the functions suggested by Rao (1948) was the difference in the weighted sum of all possible differences

$$\Sigma\Sigma w_i w_j (D_{ij}^2 - D_{ij(k)}^2) \tag{2.21}$$

which leads to the same solution for $Y = TX$ as in Theorems 1, 2 and 3.

Another method is to choose T by maximizing the minimum of $D_{ij(k)}^2$ over all i and j as suggested by Eslava-Gomez and Marriott (1993), or by maximizing the minimum of the ratios $D_{ij(k)}^2/D_{ij}^2$. Both these methods are computationally very complex, but can be managed when p is small.

Note 2.7. The choices of M and W as inputs in the analysis for canonical coordinates need some discussion. The choice of M is related to the distance measure between profiles appropriate to a given investigation. In taxonomical classification, M is generally chosen as the inverse of the variance-covariance (dispersion) matrix of the measurements on units within taxa leading to Mahalanobis distance (see Rao (1945, 1947)). The matrix W is taken to be diagonal with the i-th diagonal element w_i proportional to the number of individuals sampled from the i-th taxa to estimate its profile. For a chosen M, the configuration of the profiles in the reduced space will depend on W, but is likely to be robust provided the w_i's are not widely different. In the study reported in Rao (1948), all the w_i's were chosen as equal although the sample sizes for different populations were different. However, the choice of w_i's as proportional to sample sizes enables us to test hypotheses on goodness of fit of lower dimensional planes to the observed profiles. For details, the reader is referred to Rao (1973, pp. 556–560, 1985).

If we desire that the configuration of a subset of profiles to be better preserved in the reduced space than the others, then we have to give bigger weights to those profiles.

Note 2.8. In many situations we have a data matrix X giving the measurements of p variables made on m individuals without any further information to guide us in the choices of the M and W matrices. In such cases, the usual choices of M and W are the unit matrices and the resulting canonical coordinate analysis is the Principal Component Analysis (PCA) introduced by Hotelling. Some characterizations of the principal components and their applications are given in papers by Rao (1958, 1964, 1987). It is also the practice to apply PCA on CX, i.e., after a suitable scaling of the measurements. One choice of C is a diagonal matrix with the i-th diagonal element $c_i = s_{ii}^{-1/2}$, where s_{ii} is the i-th diagonal element of the matrix

$$(X - \bar{X}1')(X - \bar{X}1')'.$$

This procedure is equivalent to using the canonical coordinate analysis choosing $M = C$ and $W = I$. Another possibility which has not been considered before is the choice, $c_i = 1/m_i$ where m_i is a measure of location such as the mean or median of the measurements on the i-th variable.

Note 2.9. A more general problem not considered in this paper is as follows. The basic space is somewhat general with a specified nonnegative proximity index between any two points. Given a set of points with the matrix of proximity indices between points, the problem is to transform the points to a low dimensional Euclidean space such that the inequality relationships between proximity indices are maintained to the extent possible in the corresponding Euclidean distances. Such a transformation is achieved through the algorithm for multidimensional scaling as developed by Kruskal and Wish (1978).

3. LOSS OF INFORMATION

The representation of the PP's in a lower dimensional space will entail some loss of information depending on the object of statistical analysis. However, we provide some general criteria for assessing the amount of distortion in the configuration of the profiles due to reduction of dimensionality.

In Theorems 1 and 2 of Section 2, it is shown that the best approximation to X in the reduced space is

$$\hat{X} = \xi 1' + M^{-1/2} U_{(k)} \Lambda_{(k)} V'_{(k)} W^{-1/2} \tag{3.1}$$

so that the matrix

$$D_1 = X - \hat{X} = M^{-1/2}(\lambda_{k+1} U_{k+1} V'_{k+1} + \ldots + \lambda_p U_p V'_p) W^{-1/2}. \tag{3.2}$$

gives a complete account of the errors in individual profiles due to reduction.

The configuration of the profiles in the reduced space is

$$C_{(k)} = W^{-1/2} V_{(k)} \Lambda^2_{(k)} V'_{(k)} W^{-1/2} \tag{3.3}$$

so that the matrix

$$D_2 = C_{(p)} - C_{(k)} = W^{-1/2}(\lambda^2_{k+1} V_{k+1} V'_{k+1} + \ldots + \lambda^2_p V_p V'_p) W^{-1/2}. \tag{3.4}$$

measures the distortion in the configuration, where $C_{(p)} = C$ as defined in (2.4). An overall (weighted) measure of loss of information is the ratio of

$$\text{trace} \quad W^{1/2} D_2 W^{1/2} = \lambda^2_{k+1} + \ldots + \lambda^2_p, \tag{3.5}$$

to the total variation $(\lambda^2_1 + \ldots + \lambda^2_p)$, which can be written as

$$1 - (\sum_1^k \lambda^2_i)/(\sum_1^p \lambda^2_i). \tag{3.6}$$

It is more important to assess the distortions in the inter profile squared distances. The matrix of these squared distances denoted by S can be computed from the configuration matrix C using the formula (A 1.16) given in the Appendix as

$$S = c1' + 1c' - 2C \tag{3.7}$$

where c is the vector of the diagonal elements of C. The corresponding matrix in the reduced space is

$$S_{(k)} = c_{(k)} 1' + 1 c'_{(k)} - 2C_{(k)} \tag{3.8}$$

so that the matrix

$$D_3 = S - S_{(k)} = (d^*_{ij}) \tag{3.9}$$

measures the deficiencies in the distances due to reduction of dimensionality. An over all measure of deficiency is

$$\Sigma\Sigma w_i w_j d_{ij}^2 = \lambda_{k+1}^2 + \ldots + \lambda_p^2 \tag{3.10}$$

which is the same as in (3.5).

The dispersion matrix between profiles in the whole space, as introduced in (2.12) is

$$B = (X - \xi 1')W(X - \xi 1')' = (b_{ij}) \tag{3.11}$$

while the corresponding matrix in the reduced k-dimensional space is

$$B_{(k)} = M^{-1/2}U_{(k)}\Lambda_{(k)}^2 U_{(k)}' M^{-1/2} = (b_{ij(k)}). \tag{3.12}$$

The proportion of the between profile variance in the i-th variable explained by the first k canonical variates (coordinates) is

$$b_{ii(k)}/b_{ii}, \; i = 1, \ldots, p. \tag{3.13}$$

For an interpretation of the canonical coordinates in different dimensions it would be useful to compute the proportion of variance in each variable explained by each of the canonical variates, i.e., to obtain a decomposition of (3.13) in terms of canonical variates. For this purpose, we introduce the matrices

$$E_1 = M^{-1/2}(\lambda_1 U_1 : \ldots : \lambda_p U_p) = (e_{ij}) \tag{3.14}$$

$$E_2 = (e_{ij}/\sqrt{b_{ii}}) = (f_{ij}) \tag{3.15}$$

where b_{ii} is as defined in (3.11). Let $E_{i(k)}$ be the matrix obtained by retaining only the first k columns in E_i for i=1,2. Then it is seen that

$$E_1 E_1' = B, \; E_{1(k)} E_{1(k)}' = B_{(k)}. \tag{3.16}$$

Let us consider the matrix $E_{1(k)}$ and define what may be called canonical coordinates for variables (CCV) in k dimensions as follows.

Table 3.1.

Canonical coordinates for variables

Variable	dim 1	dim 2	$\cdots$	dim k
1	e_{11}	e_{12}	$\cdots$	e_{1k}
2	e_{21}	e_{22}	$\cdots$	e_{2k}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
p	e_{p1}	e_{p2}	$\cdots$	e_{pk}

If we plot the variables as points in E^k using the row coordinates in different dimensions, then the scalar products of the vectors representing the variables are the elements of $B_{(k)}$, the best k-dimensional approximation to B.

There is some advantage in plotting the variables using the standardized coordinates (f_{ij}) defined in (3.15) as shown in Table 3.2.

Table 3.2.

Standardized CCV's and the variance explained by each canonical variate

Variable	Standardized coordinates dim 1 . . . dim k			Proportion of variance explained dim 1 . . . dim k			total
1	f_{11}	$\cdots$	f_{1k}	f_{11}^2	$\cdots$	f_{1k}^2	$\sum f_{1i}^2$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
p	f_{p1}	$\cdots$	f_{pk}	f_{p1}^2	$\cdots$	f_{pk}^2	$\sum f_{pi}^2$

The magnitudes in the right hand block of Table 3.2 indicate the influence of different variables in each dimension (canonical variate) in the reduced space. This may enable us to associate each dimension with certain variables.

We may plot the variables using the standardized CCV's in the same chart as the canonical coordinates for the profiles. It is seen that all variable points lie inside the unit sphere in E^k, and the variables close to the surface of the sphere have greater influence on the canonical variates.

It may also be mentioned that it is the usual practice in a biplot to represent the i-th variable as a directed line using the direction cosines proportional to the i-th row elements in the matrix

$$E_{1(k)} = M^{-1/2}(U_1 : \ldots : U_k) \tag{3.17}$$

in which case the projections of a profile point in these directions are proportional to the approximate coordinates of the profile in the original space (see Gabriel (1971) and Greenacre (1993)).

Note 3.1. We may consider the k columns in Table 3.1 of the CCV's as k points in the p-dimensional variable space. These points were termed as *typical profiles* in Rao (1964), in the sense that the variance-covariance matrix of the variables computed from them provides the best approximation to that computed from all the original profiles.

Table 4.1.1.

Mean value by groups and characters

Caste groupe	Symbol	Sample size	Stature st	Sitting height sh	Nasal depth nd	Nasal height nl	Head length hl	Frontal breadth fb	Bizygomatic breadth bb	Head breadth hb	Nasal breadth nb
1. Brahmin	(B_1)	85	164.51	86.43	25.49	51.24	191.92	104.74	133.86	139.88	36.55
2. Brahmin	(B_2)	92	165.07	86.25	24.74	50.40	191.35	104.46	132.68	139.50	36.13
3. Chatri	(C)	139	163.33	82.25	24.73	52.72	192.58	103.98	131.70	131.72	35.64
4. Muslim	(M)	168	162.45	81.83	24.49	51.38	190.78	103.28	131.52	137.40	36.36
5. Bhatru	(Bh)	149	163.38	84.49	25.09	52.06	186.10	99.34	133.55	138.58	35.65
6. Habru	(H)	124	164.91	85.53	24.19	50.30	186.94	100.18	131.16	137.40	35.82
7. Dom	(D)	113	166.53	84.19	25.33	50.34	186.40	104.16	132.64	137.52	38.11
8. Ahir	(A_1)	67	161.37	84.35	24.29	48.98	187.45	102.76	131.70	138.12	35.60
9. Kurmi	(A_2)	94	161.35	83.41	24.03	49.22	188.86	102.62	131.82	137.86	36.21
10. Artesan	(A_3)	172	161.34	83.09	23.73	48.72	187.69	102.44	131.30	136.84	36.27
11. Kahar	(A_4)	57	160.53	81.47	23.84	48.62	188.83	101.68	130.70	136.28	36.61
12. Tharu	(Th)	191	163.33	83.57	21.72	49.94	187.78	100.00	131.68	135.90	37.90
13. Chamar	(Ch)	158	161.88	80.39	23.27	48.50	186.85	102.62	130.28	136.40	36.35
14. Chero	(T_1)	100	162.03	79.67	22.64	47.88	187.09	101.40	130.00	135.48	38.88
15. Majhi	(T_2)	153	162.96	79.89	22.37	46.70	187.12	102.56	131.34	135.92	39.24
16. Panika	(T_3)	156	159.76	78.71	22.40	47.22	186.34	102.14	130.44	135.10	38.05
17. Kharwar	(T_4)	197	160.79	80.31	22.19	48.34	185.92	102.34	131.22	136.74	38.47

Note 3.2. The standardized CCV's are not the coordinates for row profiles. They are used for interpreting the CC's of column profiles. If a representation of row profiles is needed, we consider the matrix X' with appropriate choice of the M and W matrices (which may be different from those used for column profiles) and repeat the analysis indicated in (2.6) — (2.7b).

4. APPLICATIONS

4.1. Continuous measurements

In practical applications, we have at least two types of data matrices, one where the profiles are vectors of measurements on continuous variables and another where the profiles are vectors of relative frequencies of attributes. We discuss these two cases with some examples.

Table 4.1.1 gives the sample sizes and mean values of 9 anthropometric measurements made on samples of individuals from 17 populations. This is our data matrix X' of order $m \times p$ with m=17 and p=9. The mean values in each row of Table 4.1.1 represent a population profile which can be plotted as a point in a p-dimensional vector space. (Note that in the Table 4.1.1, rows represent populations and columns the variables). We would like to represent the populations in a lower (2 or 3) dimensional space for a visual examination to find clusters and outliers and other structures in the data. Besides the data, we need some other inputs to apply the methods of data reduction developed in the previous sections.

First is the choice of a metric or an inner product in the vector space of profiles, in terms of which the configuration of a set of profiles can be described. The choice naturally depends on the object of investigation. In the context of anthropometric measurements a natural choice is the Mahalanobis distance (see Mahalanobis (1936) and Rao (1948)). If X_i' and X_j' are the row vectors (mean values) representing the profiles of the i-th and j-th populations respectively, then the square of the Mahalanobis distance between populations i and j is

$$(X_i - X_j)'\Sigma^{-1}(X_i - X_j)$$

where Σ is the common (or the average) variance-covariance (dispersion) matrix of the variables within a population. We may choose $M = \Sigma^{-1}$. In the above example we have samples from each population with a total number of 2215 individuals. We use MANOVA (or Analysis of dispersion, Rao (1973, p. 570)) to obtain the decomposition of T the total sum of squares and products matrix of order 9x9 as between B and within W. Dividing W by the degrees of freedom (2215-17=2198), we have an estimate of Σ as shown in Table 4.1.2.

Table 4.1.2.

The estimated variance-covariance matrix $\hat{\Sigma}$ within populations

	st	*sh*	*nd*	*nl*	*hl*	*fb*	*bb*	*hb*	*nb*
st	32.9476	10.7434	1.7820	3.9658	10.2211	4.8894	7.6002	3.6472	3.1023
sh		10.2400	1.1726	2.4304	5.5989	3.7782	4.3895	2.9794	0.9721
nd			3.0625	1.7824	1.7752	0.8527	1.2624	1.0301	0.5123
nh				12.2500	4.0610	1.5627	2.9687	2.7326	0.3940
hl					43.5600	5.8729	8.4397	5.8865	3.2737
fb						15.3664	8.8511	7.8692	1.8446
bb							20.9764	11.1438	3.2122
hb								20.2500	1.6341
nb									6.6049

Second is the choice of the weights to be attached to the profiles. A natural choice of the weight to be attached to the i-th population is its sample size n_i expressed as a proportion of the total sample size n ($w_i = n_i/n$). Such a choice may not always be the best, especially when the sample sizes are widely different, although there is an advantage when tests of significance are also considered. (In an earlier analysis, Rao (1948) used equal weights although the sample sizes were different. Usually the reduced configuration is robust to different choices of weights).

Choosing $M^{-1} = \hat{\Sigma}$ as in Table 4.1.2. and $w_i = n_i/n$, i.e.,

$$2215W = \text{diag}(85, 92, 139, 168, 149, 124, 113, 67, 94, 172, 57, 191, 158, 100, 153, 156, 197)$$

where the numbers are the sample sizes, we compute the s.v.d. of

$$\hat{\Sigma}^{-\frac{1}{2}}(X - \xi 1')W^{1/2} = \lambda_1 U_1 V_1' + \ldots + \lambda_p U_p V_p'$$

with $\xi = XW1/1'W1$. The first three canonical coordinates as defined in (2.7) are as shown in Table 4.1.3. The three columns in the table are the vectors $\lambda_1 W^{-1/2} V_1$, $\lambda_2 W^{-1/2} V_2$ and $\lambda_3 W^{-1/2} V_3$.

The squares of the singular values are 2165.6, 1012.2, 478.1, 265.5, 207.6, 93.9, 64.5, 38.6 and 17.0 with a total of 4343.1. The first three canonical coordinates explain about 84% of variation between populations. Figure 2 gives a plot of the first two canonical coordinates

Table 4.1.3.

Canonical coordinates in three dimensions

Group	dim 1	dim 2	dim 3
B_1	1.505584255	0.151968922	0.49580195
B_2	1.268677020	−0.008997802	0.43108765
C	0.376888549	1.864144347	−1.05099680
M	0.271688813	0.739815783	0.04093115
Bh	1.497884705	−0.638438490	−0.23425360
H	1.188939615	−0.741844231	−0.08910818
D	0.458655511	0.268433534	0.63562071
A_1	0.962270474	−0.186013502	0.41809923
A_2	0.570135707	−0.064868650	0.24783429
A_3	0.352468454	−0.066568893	0.18233164
A_4	−0.006945018	0.191180399	0.17306344
Th	−0.301523213	−1.119103947	−0.96252679
Ch	−0.503912465	0.368649078	0.29020494
T_1	−1.306812703	−0.088500525	0.11166126
T_2	−1.525798049	−0.125746569	0.20196936
T_3	−1.412895439	0.197121678	0.11349265
T_4	−1.125431499	−0.338872721	0.01751354
% Variance explained	49.86	23.30	11.00

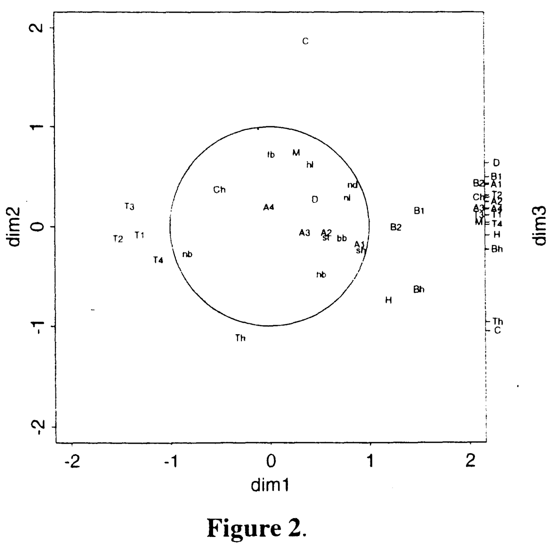

Figure 2.
Configuration of canonical coordinates for caste groups and variables. The coordinates in the third dimension for the caste groups are represented on a separate line.

The standardized first three canonical coordinates for the variables as defined in (3.15) and the between group variance in each variable explained are given in Table 4.1.4. They are the column vectors $\lambda_1\Delta^{-1}\Sigma^{1/2}U_1, \lambda_2\Delta^{-1}\Sigma^{1/2}U_2$ and $\lambda_3\Delta^{-1}\Sigma^{1/2}U_3$ where Δ is the diagonal matrix formed by the square roots of the diagonal elements of the matrix $(X-\xi 1')W(X-\xi 1')'$.

Table 4.1.4.

Standardized CC's for variables and between group variance explained

Variable	dim 1	dim 2	dim 3	% variance
st	0.56447484	−0.04964491	−0.02136700	32.15530
sh	0.91478661	−0.24173510	−0.03048027	89.61995
nd	0.83822736	0.41848655	0.25011364	94.03129
nl	0.77824697	0.28621522	−0.43125314	87.35668
hl	0.40934124	0.62262950	−0.22874285	60.75510
fb	0.02029308	0.72195951	0.43410258	71.00824
bb	0.72821282	−0.12153872	−0.02456127	54.56688
hb	0.51704122	−0.48947063	0.60153126	86.87530
nb	−0.82527584	−0.27455517	0.11745159	77.02556

The first two CC's for the variables are plotted in the same Figure 2. It is seen that they all lie within the unit circle. The nasal measurements (nl, nd and hb) and sitting height (sh) are close to the circumference of the circle indicating that they are well represented in the first two CC's for the populations.

Note 4.1. A two dimensional representation may not show the exact relative positions of the populations in the original space but the distortion may be large in particular cases. In general it will be wise to take into account the differences in the third as well as in the higher dimensions while interpreting the distances in the two dimensional representation. An innovation in Figure 2 is the representation of the coordinates in the third dimension on a separate line, which shows the additional differences among the groups. [Note that the square of the distance between any two groups in the three dimensional representation is the square of the distance in the two dimensional representation plus the square of the difference in the coordinates of the third dimension.]

A general recommendation is to consider a two dimensional plot supplemented by representation of the points in other dimensions on separate lines as parallel coordinates (see Wegman (1990) for details).

4.2. Two way contingency tables

We consider dichotomous categorical data with s rows and m columns and n_{ij} observations in the (i, j)-th cell. Define

$$N = (n_{ij}), n_{i.} = \sum_{j=1}^{m} n_{ij},\ n_{.j} = \sum_{i=1}^{s} n_{ij},\ n = \sum_{1}^{s}\sum_{1}^{m} n_{ij}$$

$$R = \text{Diag}\,(n_{1.}/n, \ldots, n_{s.}/n),\ C = \text{Diag}\,(n_{.1}/n, \ldots, n_{.m}/n)$$

(4.2.1) $$P = n^{-1}NC^{-1} = \begin{pmatrix} p_{1|1} & \cdots & p_{1|m} \\ \vdots & \ddots & \vdots \\ p_{s|1} & \cdots & p_{s|m} \end{pmatrix}, \quad \text{column profiles,}$$

(4.2.2) $$Q = n^{-1}R^{-1}N = \begin{pmatrix} q_{1|1} & \cdots & q_{m|1} \\ \vdots & \ddots & \vdots \\ q_{1|s} & \cdots & q_{m|s} \end{pmatrix}, \quad \text{row profiles}$$

$$p = R1,\ q = C1.$$

The problem is to represent the column (row) profiles as points in $E^k, k < s$, such that the Euclidean distances between points reflect specified affinities between the corresponding column (row) profiles.

The technique developed for this purpose by Benzécri (1992) is known as correspondence analysis (CA) which can be identified as canonical coordinate analysis. For instance, for representing the column profiles by this method, one chooses

(4.2.3) $$X = P,\ M = R^{-1},\ W = C$$

and applies the analysis described in Theorem 1 (equation (2.6)). Thus one finds the s.v.d. of

(4.2.4a) $$R^{-1/2}(P - p1')C^{1/2} = \lambda_1 U_1 V_1' + \cdots + \lambda_s U_s V_s'$$

giving the coordinates for the column profiles in E^k

(4.2.4b) $$\lambda_1 C^{-1/2}V_1, \lambda_2 C^{-1/2}V_2, \ldots, \lambda_k C^{-1/2}V_k$$

where the components of i-th vector are the coordinates of the profiles in the i-th dimension. The standardized canonical coordinates in E^k for the rows, as described in (3.15), obtained from the same s.v.d. as in (4.2.4a) are

(4.2.4c) $$\lambda_1 \Delta^{-1}R^{1/2}U_1, \lambda_2 \Delta^{-1}R^{1/2}U_2, \ldots, \lambda_k \Delta^{-1}R^{1/2}U_k$$

where the components of the i-th vector are the coordinates of the rows in the i-th dimension and Δ is a diagonal matrix with the i-th diagonal element as the square root of the i-th diagonal element of the matrix

(4.2.4d)

$$R^{1/2}(\lambda_1^2 U_1 U_1' + \cdots + \lambda_s^2 U_s U_s')R^{1/2} = (P - p1')C(P - p1')'.$$

The coordinates (4.2.4c) do not represent the row profiles but are useful in interpreting the different dimensions of the column profiles. The coordinates for representing the row profiles in correspondence analysis are given in (4.2.6c).

Implicit in this analysis is the choice of measure of affinity between the i-th and j-th profiles as the squared distance

$$d_{ij}^2 = \frac{(p_{1|i} - p_{1|j})^2}{p_1} + \cdots + \frac{(p_{s|i} - p_{s|j})^2}{p_s}. \tag{4.2.5}$$

which is the chisquare distance. The squared Euclidean distance in E^k, the reduced space, between the points representing the i-th and j-th profiles is an approximation to (4.2.5). Thus the clusters we see in the Euclidean representation is based on the affinities measured by the chisquare distance (4.2.5).

Why should one choose the chisquare distance to measure the affinities between profiles? Some of the advantages mentioned by Benzécri and Greenacre are as follows.

1. Note that the expression in (4.2.4a)

$$R^{-1/2}(P - p1')C^{1/2} = R^{1/2}(Q - 1q')C^{-1/2} = T \tag{4.2.6a}$$

so that if we need a representation of the row (as population) profiles in E^k, we use the same s.v.d. as in (4.2.4a)

$$R^{1/2}(Q - 1q')C^{-1/2} = \lambda_1 U_1 V_1' + \cdots + \lambda_s U_s V_s' \tag{4.2.6b}$$

leading to the row (population) coordinates

$$(\lambda_1 R^{-1/2} U_1 : \ldots : \lambda_k R^{-1/2} U_k) \tag{4.2.6c}$$

so that no extra computations are needed if we want a representation of the row profiles also. In correspondence analysis it is customary to plot the points (4.2.4b) and (4.2.6c) in the same chart. Then the standardized coordinates for the columns (as variables) are

$$\lambda_1 \Delta_1^{-1} C^{1/2} V_1, \ldots \lambda_k \Delta_1^{-1} C^{1/2} V_k \tag{4.2.6d}$$

where $\triangle_1$ is a diagonal matrix with the i-th diagonal element as the square root of the i-th diagonal element of $(Q - 1q')'R(Q - 1q')$.

2. It is easy to see that

$$\begin{aligned} n(\lambda_1^2 + \cdots + \lambda_k^2) &= n \text{ trace } TT', \quad \text{with } T \text{ as in (4.2.6a)} \\ &= \sum\sum \frac{(n_{ij} - np_i q_j)^2}{np_i q_j} \end{aligned}$$

which is the Pearson chisquare statistic for testing independence between the attributes in a contingency table. Thus the computations involved in CA automatically allow us to test for independence, and also test for the dimensionality of the space of profiles using statistics of the type

$$n(\lambda_i^2 + \cdots + \lambda_s^2), i = 1, 2, \ldots \tag{4.2.7}$$

as discussed in Rao (1973, pp. 556-560).

3. CA is only an exploratory data analysis to examine the configuration of row and column profiles in a general way, so that a particular convenient choice of the distance measure can serve the purpose.

On the other hand, there seem to be some drawbacks in using the chisquare distance.

1. The chisquare distance (4.2.5) is not a function of the i-th and j-th column profiles only. It involves the marginal profile which is a weighted average of the individual column profiles. The weights depend on the observed numbers of individuals in the column categories. These numbers may not have any relevance to the problem under study, especially when the columns represent different populations from each of which some individuals are chosen and classified according to row categories. In such a case the marginal profile depends on the actual sample sizes chosen or realized for different populations. The examples discussed in the sequel show that the derived configurations in the reduced space may be sensitive to the sample numbers.

2. The marginal profile depends on the set of populations included in CA. The CA's based on a given set of populations (S_1) and an extended set of populations (S_1, S_2) may provide different configurations to the subset S_1 as shown in Example 4.2.1.

3. There is no particular advantage in plotting the row and column profiles in the same chart. Indeed one could use different distance measures for column and row profiles and study configurations of the column and row profiles separately.
4. Since the chisquare distance uses the marginal proportions in the denominator, undue emphasis is given to the categories with low frequencies in measuring affinities between profiles.

An alternative to the chisquare distance which has some advantages is the Hellinger Distance (HD) between the i-th and j-th column profiles defined by

$$d_{ij}^2 = (\sqrt{p_{1|i}} - \sqrt{p_{1|j}})^2 + \cdots + (\sqrt{p_{s|i}} - \sqrt{p_{s|j}})^2 \tag{4.2.8}$$

which depends only on the i-th and j-th column profiles. In such a case, the Euclidean distance in the reduced space between the i-th and j-th column profiles is an approximation to (4.2.8). For the derivation of canonical coordinates of the column profiles (considered as population) we choose

$$X = \begin{pmatrix} \sqrt{p_{1|1}} & \cdots & \sqrt{p_{1|m}} \\ \vdots & \ddots & \vdots \\ \sqrt{p_{s|1}} & \cdots & \sqrt{p_{s|m}} \end{pmatrix}$$

$$M = I, \quad W = C = \text{Diag}\,(n_{.1}/n, \ldots, n_{.m}/n) \tag{4.2.9}$$

and consider the s.v.d.

$$(X - \xi 1')C^{1/2} = \lambda_1 U_1 V_1' + \cdots + \lambda_s U_s V_s'. \tag{4.2.10}$$

We may choose $\xi' = (\xi_1, \ldots, \xi_s)$ as

$$\xi_i = \sqrt{p_i} = \sqrt{n_{i.}/n}, \quad \text{or} \tag{4.2.11}$$

$$= n^{-1}(n_{.1}\sqrt{p_{i|1}} + \cdots + n_{.m}\sqrt{p_{i|m}}). \tag{4.2.12}$$

The canonical coordinates in E^k for the column profiles choosing ξ as in (4.2.11) or (4.2.12) are

$$\lambda_1 C^{-1/2} V_1, \lambda_2 C^{-1/2} V_2, \ldots, \lambda_k C^{-1/2} V_k \tag{4.2.13a}$$

where the components of the i-th vector are the coordinates of the m column (population) profiles in the i-th dimension. The standardized coordinates in E^k for the variables, i.e., the row categories, obtained as described in (3.15) from the same s.v.d. as in (4.2.10) are

$$\lambda_1 \Delta^{-1} U_1, \lambda_2 \Delta^{-1} U_2, \ldots, \lambda_k \Delta^{-1} U_k \tag{4.2.13b}$$

where Δ is a diagonal matrix with the i-th diagonal element as the square root of the i-th diagonal element of

(4.2.13c) $$\lambda_1^2 U_1 U_1' + \cdots + \lambda_s^2 U_s U_s' = (X - \xi 1')C(X - \xi 1')'.$$

The s components of $\lambda_i \Delta^{-1} U_i$ in (4.2.13b) are the coordinates of the s variables in the i-th dimension.

It can be shown that the statistic

(4.2.14) $$4n(\lambda_1^2 + \cdots + \lambda_s^2)$$

is distributed asymptotically as chisquare on $(s-1)(m-1)$ degrees of freedom to test independence in the two way contingency table. Further, hypotheses specifying the dimensions of the subspace in which the profiles can be represented can also be tested in the same way as in (4.2.7) using the residual singular values.

The advantages in using HD between profiles are the following.

1. The measure depends only on the profiles of the concerned pair. It is not altered when an extended set of profiles is considered.
2. The measure does not depend on the sample sizes on which the profiles are estimated.
3. If a representation of the row profiles is also needed we take $X = sqrt(Q')$, i.e., the elements of X are the square roots of the elements of Q' where Q is the matrix defined in (4.2.2) and compute the s.v.d.

(4.2.15a) $$(X - \eta 1')R^{1/2} = \mu_1 A_1 B_1' + \cdots + \mu_s A_s B_s'$$

leading to the canonical coordinates for row profiles (considered as populations)

(4.2.15b) $$\mu_1 R^{-1/2} B_1, \mu_2 R^{-1/2} B_2, \ldots, \mu_k R^{-1/2} B_k.$$

The corresponding standardized coordinates for the columns considered as variables are

(4.2.15c) $$\mu_1 \Delta_c^{-1} A_1, \mu_2 \Delta_c^{-1} A_2, \ldots$$

where Δ_c is a diagonal matrix with the i-th diagonal element as the square root of the i-th diagonal element of

(4.2.15d) $$\mu_1^2 A_1 A_1' + \cdots + \mu_s^2 A_s A_s'.$$

4. If we choose ξ as in (4.2.11), then the matrix in (4.2.10) is

$$(X - \xi 1')C^{1/2} = \left(\sqrt{\frac{n_{ij}}{n}} - \sqrt{\frac{n_{i.}}{n}\frac{n_{.j}}{n}} \right)$$

which is symmetric in i and j. Then, the same s.v.d. as in (4.2.10) could be used for computing the canonical coordinates

$$\lambda_1 R^{-1/2} U_1, \lambda_2 R^{-1/2} U_2, \ldots, \lambda_k R^{-1/2} U_k$$

for the row profiles, as in the case of CA.

Example 4.2.1

Consider for example the data on smoking habits (s_1, s_2, s_3) of different categories of employees (A, B, C, D, E) given in Tables 4.2.1 and 4.2.2, one of which has data on two additional categories of employees.

Table 4.2.1.

	Frequencies				Profiles			
	A	B	C	Total	A	B	C	Average
s_1	5	0	0	5	1	0	0	5/11
s_2	0	5	0	5	0	1	0	5/11
s_3	0	0	1	1	0	0	1	1/11
Total	5	5	1	11	1	1	1	1

Table 4.2.2.

	Frequencies						Profiles					
	A	B	C	D	E	Total	A	B	C	D	E	Average
s_1	5	0	0	4	1	10	1	0	0	.8	.2	10/21
s_2	0	5	0	1	4	10	0	1	0	.2	.8	10/21
s_3	0	0	1	0	0	1	0	0	1	0	0	1/21
Total	5	5	1	5	5	21	1	1	1	1	1	1

The two dimensional representations of the employee categories obtained through correspondence analysis using the formula (4.2.4b) for the data in the above tables are given in Figure 3, where the lower case letters are used for the data of Table 4.2.2. We make the following observations.

1. The relative positions of A, B and C change when additional employee categories are introduced, although their individual profiles remain the same.
2. The profiles for A, B, C in Table 4.2.1 suggest that they are equally distant between each other, and the correspondence analysis would have revealed this if the sample size for category C had been the same as that for A and B.

Thus, with chisquare distances, there is instability of the configuration of the populations in the reduced space with changes in the sample sizes and the addition or deletion of populations.

The two dimensional representation of the employee categories based on Hellinger distance using the formula (4.2.13a) is given in Figures 3 and 4. It is seen that the graph based on Table 4.2.1 shows that A, B and C are equidistant from each other and the graph based on Table 4.2.2 shows that the positions of A, B, C represented by lower case letters are not changed when the additional categories d and e are introduced. Thus, the use of Hellinger distance seems to provide a more satisfactory solution.

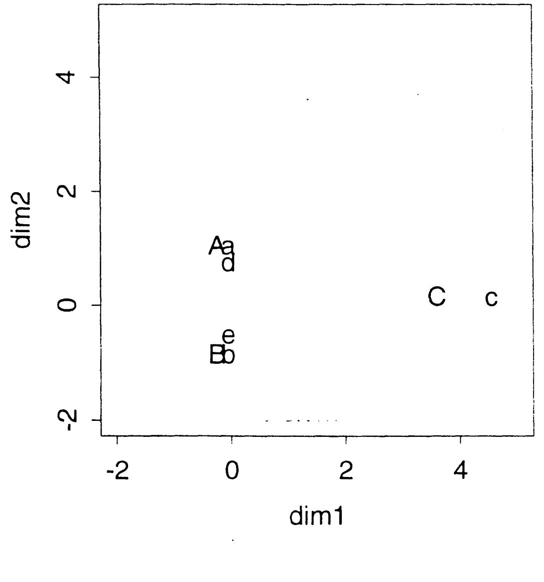

Figure 3.
Configuration of canonical coordinates based on chisquare distance (Lower case letters are used when the analysis is based on all the categories).

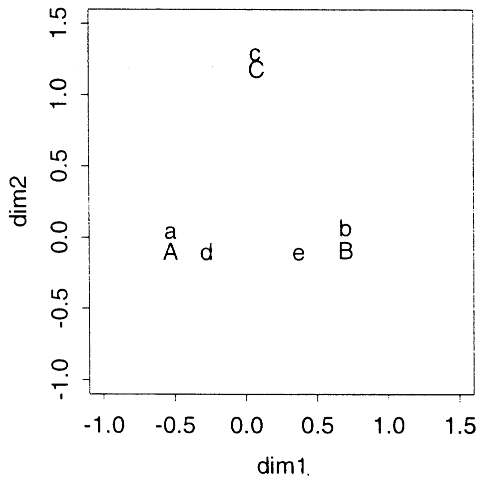

Figure 4.

Configuration of canonical coordinates based on Hellinger distance (Lower case letters are used when the analysis is based on all the categories).

Example 4.2.2

We consider the data (from Greenacre (1993)) on 796 scientific researchers classified according to their scientific discipline (as populations) and funding category (as variables) as shown in Table 4.2.3.

Table 4.2.3.

Scientific disciplines by research funding categories

		Funding category					
Scientific discipline		a	b	c	d	e	Total
Geology	G	3	19	39	14	10	85
Biochemistry	B_1	1	2	13	1	12	29
Chemistry	C	6	25	49	21	29	130
Zoology	Z	3	15	41	35	26	120
Physics	P	10	22	47	9	26	114
Engineering	E	3	11	25	15	34	88
Microbiology	M_1	1	6	14	5	11	37
Botany	B_2	0	12	34	17	23	86
Statistics	S	2	5	11	4	7	29
Mathematics	M_2	2	11	37	8	20	78
Total		31	128	310	129	198	796

The canonical coordinates for the scientific disciplines (considered as populations) in the first three dimensions and percentage of variance explained by each are given in Table 4.2.4 for the analyses based on the chisquare distance (correspondence analysis) and the Hellinger distance (alternative). The formula (4.2.6b) is used for the analysis based on chisquare and the formula (4.2.15a) for that based on Hellinger distance. For Hellinger distance analysis, the central point is chosen according to the formula (4.2.12).

Table 4.2.4.

Canonical coordinates for the scientific disciplines in the first three dimensions

Subjects	Chisquare Distance			Hellinger distance		
	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3
G	.076401	.302569	−.087749	−.031140	.167408	−.048245
B_1	.179892	−.454996	−.151716	−.129374	−.242174	−.077614
C	.037644	.073353	.042371	−.021144	.040433	.028254
Z	−.327365	.102283	.064515	.138850	.045255	.056894
P	.315552	.026997	.108688	−.165340	.010679	.023844
E	−.117495	−.291712	.107330	.049451	−.129906	.082901
M_1	.012766	−.109656	−.041435	−.004913	−.052588	−.008439
B_2	−.178695	−.038501	−.129055	.151404	−.036559	−.108025
S	.124638	.014162	.107190	−.066639	.011763	.052571
M_2	.106751	−.061316	−.175688	−.050307	−.037572	−.078006
% var.	47.20	36.66	13.11	45.87	34.10	16.57

The plots of the scientific disciplines (subjects) using the canonical coordinates based on the chisquare and Hellinger distances are given in Figures 5 and 6 respectively. The coordinates in the third dimension are plotted on a line on the right hand side of the two dimensional plot. This will be of help in visualizing the plot in three dimensions and in interpreting the distances in the two dimensional plot. Thus, although B_2 and E appear to be close to each other in the two dimensional chart, they are clearly separated in the third dimension. No additional distances in the third dimension are involved in the case of P, C, S, Z and E.

It is of interest to note in this example that the configuration of the scientific disciplines in three dimensions obtained by both the methods are very similar. The percentage of variance explained in each dimension is nearly the same for both the methods.

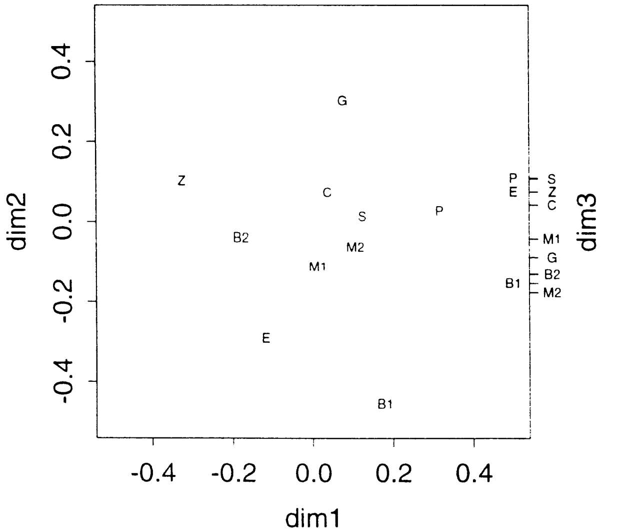

Figure 5.
Configuration of scientific disciplines using Chisquare distance (Correspondence Analysis).

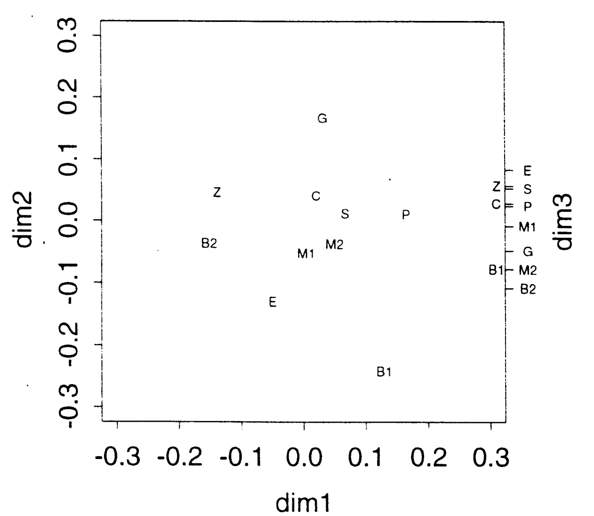

Figure 6.
Configuration of scientific disciplines using Hellinger distance (Alternative to Correspondence Analysis).

The standardized canonical coordinates for the funding categories (considered as variables) are computed using the formula (4.2.6d) for the chisquare analysis and the formula (4.2.15b) for the Hellinger distance analysis. These are obtained from the *same s.v.d.* used to compute the canonical coordinates for the scientific disciplines. Table 4.2.5 gives the standardized canonical coordinates for the funding categories, a, b, c, d, e, using the two methods.

Table 4.2.5.

Standardized canonical coordinates for funding categories (variables) in the first three dimensions

Funding category	Chisquare Distance				Hellinger Distance			
	dim1	dim2	dim3	%var	dim1	dim2	dim3	%var
a	.758	.114	−.619	97.1	.796	.164	−.573	98.9
b	.535	.728	−.137	83.5	.438	.766	−.008	77.9
c	.583	.352	.694	94.6	.501	.327	.759	93.4
d	−.426	.331	−.172	99.8	−.888	.358	−.285	99.7
e	−.108	−.909	−.081	99.6	−.088	−.978	−.159	98.9

The standardized canonical coordinates for the funding categories are plotted in Figure 7 (for chisquare distance) and in Figure 8 (for Hellinger distance). It may be noted that all the points lie within the unit circle. It is customary to represent the canonical coordinates for the subjects and variables in one chart. We are using separate charts in order to explain the salient features of the configuration of the variables. The following interpretations emerge from the study of Table 4.2.5 and Figures 7 and 8.

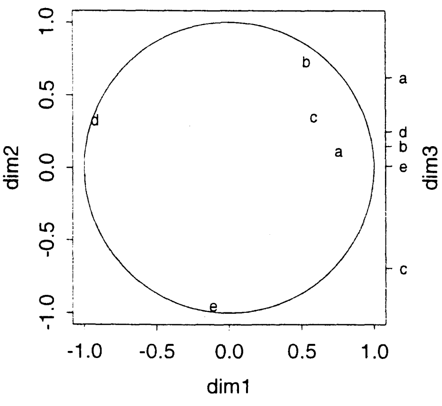

Figure 7.
Configuration of funding categories using standardized canonical coordinates based on Chisquare distance.

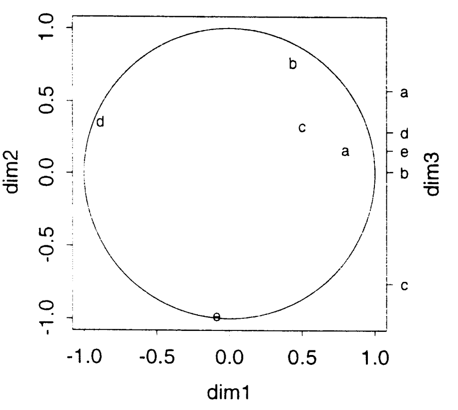

Figure 8.
Configuration of funding categories using standardized canonical coordinates based on Hellinger distance.

1. The configurations of the funding categories as exhibited by Figures 7 and 8 obtained by using chisquare and Hellinger distances are very similar.

2. All most all the variation in the funding categories a, d and e is captured in the first three canonical coordinates of the scientific disciplines. A large percentage of variation in b and c is explained by the first three coordinates.

3. The first dimension is strongly influenced by a, d the second dimension by b, e and the third dimension by a, c.

Thus the use of standardized coordinates for variables enables us to interpret the different dimensions in terms of observed variables. There are other ways of plotting the coordinates of the variables as mentioned in the paragraphs below Table 3.2. Such biplots having a different interpretation are discussed in Gabriel (1971), Gifi (1990), Gower (1993) and Greenacre (1993).

Note 4.2. In computing the canonical coordinates based on Hellinger distance (HD) using the formula (4.2.10), we chose the relative sample sizes as the weights to be attached to the populations. We could have chosen an alternative set of weights if we wanted distances between a specified subset of populations to be better preserved in the reduced space than the others. In particular, we could have chosen uniform weights for all populations. In fact such an option could be exercised if the sample sizes of different populations were widely different. Unfortunately no such options are available in correspondence analysis.

Example 4.2.3

In the example 4.2.2, there was a perfect match between the plots based on CA and HD. This probably demonstrates that the method of derivation of canonical coordinates is somewhat robust to the choice of the distance measure as well as to the weights. However the choice of HD provides as insurance against possible distortion due to variations in sample sizes for the populations as the following example shows.

Table 4.2.6, reproduced from Gifi (1981), gives the distributions of the pages devoted to different topics denoted by A, B, C, D, E, F and G in 20 books on Multivariate analysis designated as $a, b, \ldots, t$. Gifi (1981) did correspondence analysis on the data and drew some conclusions based on the first three canonical coordinates which explain a high percentage of variation.

Table 4.2.6.

Number of pages by topics

Books	*A*	*B*	*C*	*D*	*E*	*F*	*G*
a	31	0	0	0	0	164	11
b	0	16	54	18	27	13	14
c	0	40	32	10	42	60	0
d	19	0	35	19	28	163	52
e	14	7	35	22	17	0	56
f	20	69	72	33	55	0	32
g	74	0	86	14	0	84	48
h	78	0	80	5	17	105	60
i	74	19	33	12	26	0	0
j	80	68	67	15	29	0	0
k	108	48	4	10	46	108	0
l	109	13	5	17	39	32	46
m	16	35	69	24	0	26	41
n	26	86	60	6	48	48	28
o	290	10	6	0	8	0	2
p	184	48	82	42	134	0	0
q	29	0	0	0	41	211	32
r	0	19	56	0	39	75	0
s	0	22	45	42	60	230	59
t	30	128	90	28	48	0	0

The first three canonical coordinates for the profiles of the books based on CA and HD approaches are given in Table 4.2.7.

Table 4.2.7.

	Chisquare Distance			Hellinger Distance		
	dim1	dim2	dim3	dim1	dim2	dim3
a	−1.10857	−0.61445	−0.33902	0.64632	0.36299	0.12879
b	0.07397	0.70254	0.25265	−0.01661	−0.48923	−0.12388
c	−0.21153	0.46054	−0.49228	0.10998	−0.42185	0.32822
d	−0.77795	−0.11074	0.15556	0.46658	−0.01597	−0.10284
e	0.02781	0.40651	1.06135	−0.19193	−0.15180	−0.45570
f	0.35780	0.69602	0.09284	−0.37016	−0.29359	−0.14451
g	−0.16412	−0.15719	0.46353	0.23979	0.16911	−0.35829
h	−0.25023	−0.19626	0.39002	0.26103	0.14730	−0.23804
i	0.72788	−0.19452	−0.04749	−0.50899	0.14292	0.02936
j	0.68403	0.24337	−0.17956	−0.53320	−0.01242	0.04724
k	0.02729	−0.36648	−0.44297	0.03996	0.21098	0.36189
l	0.26802	−0.44749	0.28287	−0.00524	0.27070	−0.06481
m	0.02188	0.50893	0.51719	0.01506	−0.19266	−0.34080
n	0.12052	0.48459	−0.19476	−0.04555	−0.20966	0.04945
o	1.08308	−1.32602	0.03206	−0.39476	0.66357	−0.00090
p	0.64959	−0.07081	−0.13268	−0.49299	0.09097	0.08510
q	−0.98347	−0.39273	−0.25019	0.58910	0.21379	0.19442
r	−0.40006	0.32919	−0.33826	0.21605	−0.35929	0.28764
s	−0.74726	0.08101	−0.00508	0.43349	−0.30134	0.03139
t	0.56547	0.81454	−0.35256	−0.51167	−0.27874	0.10162

It may be noted that the total number of pages of a book depends on the font size of the print, while its profile in terms of proportions of pages used on different topics remain the same for all sizes. Table 4.2.8 gives the data on books having the same profiles as in Table 4.2.6 with the total number of pages altered for the books d, f, g, h, j and n.

Table 4.2.8.

Number of pages by topics

Books	*A*	*B*	*C*	*D*	*E*	*F*	*G*
a	31	0	0	0	0	164	11
b	0	16	54	18	27	13	14
c	0	40	32	10	42	60	0
d	190	0	350	190	280	1630	520
e	14	7	35	22	17	0	56
f	10	34	36	17	28	0	16
g	740	0	860	140	0	840	480
h	780	0	800	50	170	1050	600
i	74	19	33	12	26	0	0
j	40	34	33	8	15	0	0
k	108	48	4	10	46	108	0
l	109	13	5	17	39	32	46
m	16	35	69	24	0	26	41
n	13	43	30	3	24	24	14
o	290	10	6	0	8	0	2
p	184	48	82	42	134	0	0
q	29	0	0	0	41	211	32
r	0	19	56	0	39	75	0
s	0	22	45	42	60	230	59
t	30	128	90	28	48	0	0

The three dimensional canonical coordinates based on CA and HD approaches are given in Table 4.2.9. Using the coordinates one can obtain the mutual distances between the books in the three dimensional reduced Euclidean space. Figure 9 gives a plot comparing the squared distances between books based on CA using the data of Tables 4.2.6 and 4.2.8. Figure 10 gives the corresponding plot for the squared distances based on the HD approach. It is seen that the three dimensional representation of the data of Tables 4.2.6 and 4.2.8 are more similar under HD analysis than that under CA. The relative positions of the books are influenced by the font size in printing when CA is used, although the profiles of the books are not altered. There appears to be greater stability with the HD analysis which provides insurance against different choice of sample sizes. Further, one can exercise the option of using a common weight for all the books in the HD analysis when the differences in book sizes are large.

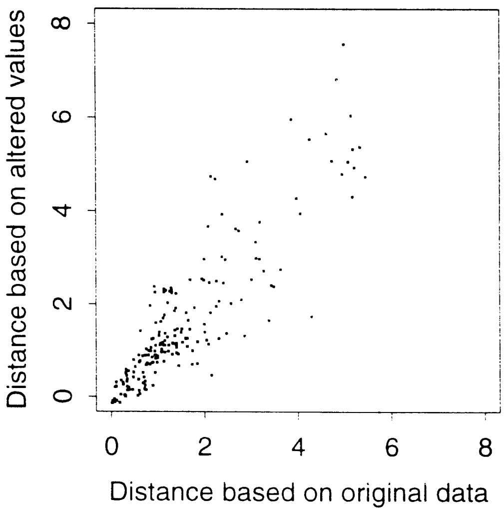

Figure 9.
A comparative plot of squared distances between all pairs of books in the reduced spaces based on correspondence analysis.

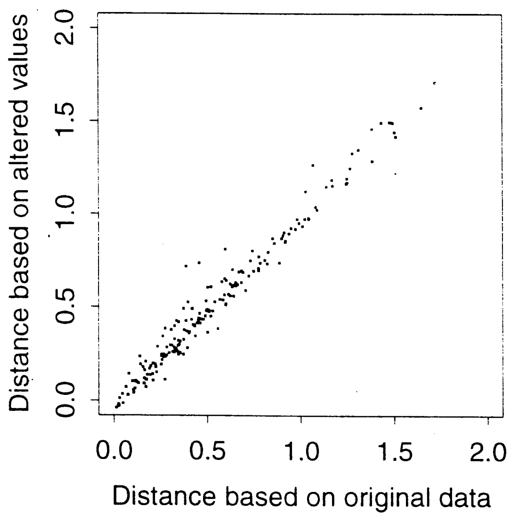

Figure 10.
A comparative plot of squared distances between all pairs of books in the reduced spaces based on Hellinger distance analysis.

Table 4.2.9.

	Chisquare Distance			Hellinger Distance		
	dim1	dim2	dim3	dim1	dim2	dim3
a	−0.62310	0.30413	−0.53463	−0.35082	−0.04632	0.42925
b	0.63345	0.41500	0.44316	0.25096	−0.37625	−0.40565
c	0.90486	0.78379	−0.17802	0.22985	−0.60374	−0.08540
d	−0.36427	0.36470	−0.12611	−0.23454	−0.16742	0.01739
e	0.20621	0.05647	0.54414	0.34853	0.01585	−0.32928
f	1.23299	0.45214	0.27035	0.59591	−0.20402	−0.28683
g	−0.16974	−0.27626	0.25537	−0.08445	0.24917	−0.09573
h	−0.18352	−0.18729	0.09783	−0.07908	0.11818	0.00148
i	0.85607	−0.49586	−0.21672	0.73058	0.07206	0.07026
j	1.35943	−0.06808	0.07459	0.77422	−0.01788	−0.04783
k	0.61122	0.01365	−0.60327	0.28646	−0.20830	0.40764
l	0.32350	−0.41447	−0.39537	0.23929	0.02213	0.24724
m	0.58680	0.20860	0.65792	0.17707	0.01371	−0.36016
n	1.20448	0.51816	0.07444	0.32243	−0.27238	−0.09222
o	0.47616	−1.60929	−0.73075	0.61206	0.41339	0.46017
p	0.87199	−0.26044	−0.37985	0.72806	−0.03491	0.09187
q	−0.44497	0.44631	−0.57047	−0.27936	−0.25639	0.41829
r	0.34209	0.56913	−0.04981	0.10278	−0.49844	−0.07991
s	−0.10036	0.60181	−0.15749	−0.14779	−0.45659	−0.10069
t	1.87687	0.57376	0.28038	0.78353	−0.24748	−0.19823

5. CONCLUDING REMARKS

A general theory is developed for plotting high dimensional "population by variable" data, i.e. measurements made on a set of characteristics of given populations, in a low dimensional Euclidean space. A first step in such a problem is the specification of the basic metric space in which the populations can be represented as points using the entire data, and a characterization of the configuration of the points in terms of distances between points. The second is the development of methodology for transforming the points from the basic space to a low dimensional Euclidean space with the usual definition of distance preserving the configuration of points to the extent possible. The choices of the basic space and the distance function between points have to be made on practical considerations depending on the problem under investigation. A closed form solution is obtained when the basic space is a vector space endowed with an inner product and the associated norm. Some examples are given involving measurements on continuous and discrete variables.

When we have data in the form of frequencies of individuals of a population under different categories of an attribute, a well known method for dimensionality reduction for representing, say the populations, is correspondence analysis. The basic space in this case is a vector space where each population is represented by the vector of relative frequencies of the different categories of an attribute and distance between vectors is defined by a chisquare type formula. Such a distance function is not an intrinsic measure of difference between two populations as it depends not only on the differences between their relative frequencies, but also on the average relative frequencies computed from the set of populations under study. Thus the configuration of any subset of populations depends on what other populations are included in the analysis, and also on the relative numbers of individuals observed from each population. An alternative approach of representing a population by the vector of the square roots of relative frequencies and defining distance between two populations by the Hellinger formula does not have the drawbacks associated with the chisquare type formula. In addition, the new analysis has the same advantage of providing tests of significance for homogeneity of the populations as in correspondence analysis based on the chisquare formula.

It may be contended that CA is meant to be used for the analysis of contingency tables with dichotomized data using two attributes like hair color and eye color (as originally demonstrated by R.A. Fisher), and not for the analysis of population by variable data where anomalies of the type described in the paper may occur. However, one finds in published literature more examples of the latter type of data analyzed through CA. Further, even with attribute data, if the configurations of the column (or row) profiles for two different populations (with possibly different marginal distributions) are to be compared, HD analysis is more appropriate than the CA. It is the author's opinion that the choice of a distance measure between populations (row or column profiles) must depend on the nature of the data and the purpose of analysis. Prescription to use a particular distance as in the CA may be misleading. Distance measures other than the chisquare and Hellinger types may be more appropriate in some situations. For a purely exploratory data analysis, it is possible that a wide variety of distance measures reveal similar configurations of the populations in terms of clustering and inter cluster relationships.

Between the choices of chisquare and Hellinger distances, the latter seems to offer some advantages, as the latter has similar theoretical properties as the former and in addition it is defined as an intrinsic function of two population profiles independent of what other populations are included in a study.

A recent technical report by Rios, Villarroya and Oller (1994) discusses the same problem as in the present paper, viz., simultaneous representation of populations and random variables, under the assumption of an underlying parametric model.

The method, referred there as *Intrinsic Data Analysis*, is based on the Riemannian structure given by the Fisher information metric and its corresponding distance, the *Rao distance*. The statistical populations are viewed as points on a Riemannian manifold and the random variables with finite expectation, as vector fields, namely, the gradient of the random variable mean value, or, by integration, a bundle of curves on the manifold.

Then, assuming certain additional regularity conditions, a reference point on the manifold is selected as the statistical populations Riemannian center of mass, and the points representing the populations and the curves representing the variables are mapped, through the inverse of the Riemannian exponential map, into the tangent space at the center of mass, which has a Euclidean vector space structure. Then, classical dimension reduction techniques such as principal component analysis can be used to obtain a low dimensional Euclidean space which allows an optimal population representation. Finally, the curves in the tangent space are projected into the low dimensional space obtained.

This method is applied to multivariate normal and multinomial distributions. In the multinomial case, the Rao distance, ρ, between two populations $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$, is proportional to the Bhattacharyya distance

$$\rho = 2 arccos \sum_{j=1}^{n} \sqrt{p_j q_j}$$

which is a monotone transformation of the Hellinger distance, and thus this method will share some properties with the latter.

APPENDIX

A1. Theorems on optimization.

Let A be an $r_1 \times r_2$ matrix, X_1 be an $r_1 \times s_1$ matrix of rank s_1, X_2 be an $r_2 \times s_2$ matrix of rank s_2 and Σ_1 and Σ_2 be positive definite matrices of orders r_1 and r_2

respectively. Further let

(A 1.1) $$P_1 = X_1(X_1'\Sigma_1 X_1)^{-1}X_1'\Sigma_1$$

(A 1.2) $$P_2 = X_2(X_2'\Sigma_2 X_2)^{-1}X_2'\Sigma_2$$

be orthogonal projection matrices and

$$F = \Sigma_1^{1/2}(I - P_1)A(I - P_2')\Sigma_2^{1/2}$$

with the s.v.d. (singular value decomposition)

(A 1.3) $$F = \lambda_1 U_1 V_1' + \ldots + \lambda_m U_m V_m', \quad \lambda_i > 0, i = 1, \ldots, m$$

Theorem A.1

Let A, X_1, X_2, Σ_1 and Σ_2 be as defined above. Then, for any (Σ_1, Σ_2)-invariant norm $||\cdot||$, as defined in (2.8),

(A 1.4) $$\min_{\Psi_1, \Psi_2, \Gamma} ||A - X_1\Psi_1 - \Psi_2 X_2' - \Gamma||$$

with rank $\Gamma = r \leq \min(r_1 - s_1, r_2 - s_2)$ is attained at

(A 1.5) $$X_1\Psi_1 = P_1 A, \; \Psi X_2' = (I - P_1)AP_2'$$

(A 1.6) $$\Gamma = \Sigma_1^{-1/2}(\lambda_1 U_1 V_1' + \ldots + \lambda_r U_r V_r')\Sigma_2^{-1/2}$$

where λ_i, U_i and V_i are as defined in (A 1.3).

For details of the proof, the reader to referred to Rao (1980, 1985, p.175), where a number of applications of the above theorem are given.

Theorem A.2

Let B an $r_1 \times r_2$ matrix. Then

(A 1.7) $$\min_C ||B'\Sigma_1 B - C'C||$$

with rank $C = s \leq$ rank $B'\Sigma_1 B$ is attained for any (Σ_2, Σ_2)-invariant norm at

(A 1.8) $$C' = \Sigma_2^{-1/2}(\lambda_1 V_1 : \ldots : \lambda_s V_s)$$

where λ_i and V_i are the singular values and vectors from the s.v.d.

(A 1.9) $$\Sigma_1^{1/2} B \Sigma_2^{1/2} = \lambda_1 U_1 V_1' + \ldots + \lambda_m U_m V_m'$$

Proof

A direct application of Theorem A.1 gives the optimum choice of $C'C$ as

(A 1.10) $$C'C = \Sigma_2^{-1/2}(\lambda_1^2 V_1 V_1' + \ldots + \lambda_s^2 V_s V_s')\Sigma_2^{-1/2}$$

where λ_i^2 and V_i are from the spectral decomposition

(A 1.11) $$\Sigma_2^{1/2} B' \Sigma_1 B \Sigma_2^{1/2} = \lambda_1^2 V_1 V_1' + \ldots + \lambda_m^2 V_m V_m'$$

which can be computed from the s.v.d. (A 1.9). From (A 1.10), it is seen that one choice of C' is as given in (A 1.8).

■

Theorem A.3

Let $X = (X_1 : \ldots : X_m)$ be a $p \times m$ matrix with the $m \times m$ configuration matrix

(A 1.12) $$F = (X - \xi 1')' M (X - \xi 1') = (f_{ij})$$

and the $m \times m$ inter square distance matrix

(A 1.13) $$D = (d_{ij}),\ d_{ij} = (X_i - X_j)' M (X_i - X_j)$$

where M is a positive definite matrix and

(A 1.14) $$\xi = \Sigma w_i X_i,\ w_i \geq 0 \quad \text{and} \quad \Sigma w_i = 1$$

Then

(i)

$$-2F = (d_{ij} - d_{.j} - d_{i.} + d_{..})$$

(A 1.15) where $$d_{i.} = \Sigma w_j d_{ij},\ d_{.j} = \Sigma w_i d_{ji} \quad \text{and} \quad d_{..} = \Sigma\Sigma w_i w_j d_{ij}.$$

(ii)

(A 1.16) $$D = f1' + 1f' - 2F$$

where f is the vector of diagonal elements of F.

(iii)

$$\text{Trace } W^{1/2} F W^{1/2} = \Sigma w_i (X_i - \xi)' M (X_i - \xi)$$

(A 1.17) $$= \Sigma\Sigma w_i w_j d_{ij} = d_{..}$$

where

$$W = \text{Diag}(w_1, \ldots, w_m)$$

The results (A 1.15), (A 1.16) and (A 1.17) are easy to establish.

A2. Homogeneity tests in a contingency table

Let n_{ij} be the number of individuals belonging to category A_i of an attribute out of $n_{.j}$ individuals sampled from population j for $i = 1, \ldots, p$ and $j = 1, \ldots, m$. Then the data can be exhibited as a contingency table (population by attribute) with fixed column totals.

Population by attribute frequencies

Attribute	Populations				Total
	1	2	$\cdots$	m	
A_1	n_{11}	n_{12}	$\cdots$	n_{1m}	$n_{.1}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_p	n_{p1}	n_{p2}	$\cdots$	n_{pm}	$n_{.m}$
Total	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.m}$	$n_{..}$

Let $\pi_{i|j}$ be the true proportion of individuals with attribute A_i in population j. We would like to test the homogeneity hypothesis

(A 2.1) $$\pi_{i|1} = \pi_{i|2} = \ldots = \pi_{i|m}, \; i = 1, \ldots, p$$

using the statistic

(A 2.2) $$H = 4 \sum_i \sum_j n_{.j} (\sqrt{p_{i|j}} - \sqrt{\hat{\pi}_i})^2$$

where $p_{i|j} = n_{ij}/n_{.j}$ and $\hat{\pi}_i$ is an estimate of the common value π_i of (A 2.1), $i = 1, \ldots, p$. The statistic (A 2.2) based on Hellinger distance (see Freeman and Tukey (1950), and Matusita (1965)) falls in the category of the power divergence tests discussed in great detail by Read and Cressie (1988). The estimates of $\hat{\pi}_i$ generally recommended are

(i) the maximum likelihood (ML) estimate

(A 2.3) $$\hat{\pi}_i = \sum n_{.j} p_{i|j} / n_{..} = n_{i.}/n_{..}$$

and

(ii) the minimum distance (MD) estimate

$$\text{(A 2.4)} \qquad \hat{\pi}_i = \frac{(\sum_j n_{.j}\sqrt{p_{i|j}})^2}{\sum_i(\sum_j n_{.j}\sqrt{p_{i|j}})^2}$$

which ensure asymptotic χ^2 distribution on $(p-1)(m-1)$ degrees of freedom for H. We suggest another estimate

(iii) the unrestricted MD estimate

$$\text{(A 2.5)} \qquad \hat{\pi}_i = (\sum n_{.j}\sqrt{p_{i.j}}/n_{..})^2$$

which although does not satisfy the natural restriction $\sum\hat{\pi}_i = 1$, ensures the same asymptotic distribution for H as in (i) and (ii) above. The only condition needed is that $n_{.j} \to \infty$ for each j. When $\hat{\pi}_i$ as in (A 2.5) is used in H of (A 2.2), H ceases to be in the class of power divergence tests.

All the stated results follow from the general theorems proved in Rao (1973, pp. 382-388) and Read and Cressie (1988), or simply by noting that

$$\text{(A 2.6)} \qquad (\sqrt{p}-\sqrt{\hat{\pi}})^2 = \frac{(p-\hat{\pi})^2}{\hat{\pi}}\left(1+\frac{\sqrt{p}}{\sqrt{\hat{\pi}}}\right)^{-2}$$

which has the Pearson χ^2 component as the dominant factor with the other factor tending to 1/4 as $p/\hat{\pi} \to 1$. We need only use the expression (A 2.6) for each term in (A 2.2) to show the equivalence of H with Pearson's χ^2.

Note 1. With the choice of the ML estimate (A 2.3) for $\hat{\pi}$, the statistic H can be written as

$$\text{(A 2.7)} \qquad H = \sum_i\sum_j(\sqrt{n_{ij}} - \sqrt{n_{i.}n_{.j}/n_{..}})$$

which is symmetric with respect to interchange of rows and columns. The H based on the estimates (A 2.4) or (A 2.5) does not have this property.

Note 2. The choice of $\hat{\pi}_i$ as in (A 2.5), which is used in the analysis of Section 4.2 seems to be appropriate even in tests of significance. For instance, a natural test for homogeneity when $m = 2$ is

$$\text{(A 2.8)} \qquad 4\frac{n_{.1}n_{.2}}{n_{.1}+n_{.2}}\sum_{i=1}^{p}(\sqrt{p_{i|1}}-\sqrt{p_{i|2}})^2$$

which is equal to

(A 2.9) $$H = 4[n_{.1} \sum_{i=1}^{p} (\sqrt{p_{i|1}} - \hat{\pi}_i)^2 + n_{.2} \sum_{i=1}^{p} (\sqrt{p_{i|2}} - \sqrt{\hat{\pi}_i})^2]$$

only when $\hat{\pi}_i$ is as chosen in (A 2.5).

The research work of this paper is sponsored by the Army Research Office under Grant DAAH04-93-G-0030.

REFERENCES

[1] **Benzécri, J.P.** (1992). *Correspondence Analysis Handbook.* Marcel Dekker, Inc., New York.

[2] **Cavallis-Sforza, L.L.** (1991). "Genes, peoples and languages". *Scientific American*, **265**, 104–110.

[3] **Chernoff, H.** (1973). "The use of faces to represent points in k-dimensional space graphically". *J. Amer. Statist. Assoc.*, **68**, 361–368.

[4] **Cressie, N.A.C.** and **Read, T.R.C.** (1984). "Multinomial goodness-of-fit tests". *J. Roy. Statist. Soc.*, **46**, 440–464.

[5] **Eckart, C.** and **Young, G.** (1936). "The approximation of one matrix by another of lower rank". *Psychometrika*, **1**, 211–308.

[6] **Eslava-Gomez, G.** and **Marriott, F.H.C.** (1993). "Criteria to represent groups in the plane when the grouping is unknown". *Biometrics*, **49**, 1088–1098.

[7] **Freeman, M.F.** and **Tukey, J.W.** (1950). "Transformations related to the angular and the square root". *Ann. Math. Stat.*, **21**, 607–611.

[8] **Gabriel, K.R.** (1971). "The biplot graphical display of matrices with applications to principal component analysis". *Biometrika*, **58**, 453–467.

[9] **Gifi, A.** (1981, 1990). *Nonlinear Multivariate Analysis.* New York: John Wiley.

[10] **Gower, J.C.** (1993). "Recent advances in biplot methodology". In *Multivariate Analysis: Future Directions 2* (Eds. C.M. Cuadras and C.R. Rao), North Holland, 295–325.

[11] **Greenacre, M.J.** (1984). *Theory and Applications of Correspondence Analysis.* London: Academic Press.

[12] **Greenacre, M.J.** (1993). "Biplots in correspondence analysis". *J. Applied Statistics*, **20**, 251–269.

[13] **Kruskal, J.B.** and **Wish, M.** (1978). *Multidimensional Scaling.* Sage Publications.

[14] **Mahalanobis, P.C.** (1936). "On the generalized distance in statistics". *Proc. Nat. Inst. Sci. India*, **12**, 49–55.

[15] **Mahalanobis, P.C., Mazumdar, D.N.** and **Rao, C.R.** (1949). "Anthropometric survey of United Provinces, 1941. A statistical study". *Sankhya*, **9**, 90–324.

[16] **Matusita, Kameo** (1955). "Decision rules, based on the distance, for problems of fit, two samples, and estimation". *Ann. Math. Stat.*, **26**, 631–640.

[17] **Nishisato, S.** (1980). *Analysis of Categorical Data: Dual Scaling and its Applications*. University of Toronto Press, Toronto, Canada.

[18] **Rao, C.R.** (1945). "Information and accuracy attainable in the estimation of statistical parameters". *Bull. Cal. Math. Soc.*, **37**, 81–91.

[19] **Rao, C.R.** (1947). "The problem of classification and distance between two populations". *Nature*, **159**, 30.

[20] **Rao, C.R.** (1948). "The utilization of multiple measurements in problems of biological classification (with discussion)". *J. Roy. Statist. Soc.*, **Series B10**, 159–193.

[21] **Rao, C.R.** (1958). "Some statistical methods for comparison of growth curves". *Biometrics*, **14**, 1–17.

[22] **Rao, C.R.** (1964). "The use and interpretation of principal component analysis in applied research". *Sankhya*, **26**, 329–357.

[23] **Rao, C.R.** (1973). *Linear Statistical Inference and its Applications*, 2nd Edition, New York : Wiley.

[24] **Rao, C.R.** (1979). "Separation theorems for singular values of matrices and their applications in multivariate analysis". *J. Multivariate Analysis*, **9**, 362–377.

[25] **Rao, C.R.** (1980). "Matrix approximations and reduction of dimensionality in multivariate statistical analysis". In *Multivariate Analysis V* (Ed. P.R. Krishnaiah), Amsterdam: North Holland, 3–22.

[26] **Rao, C.R.** (1985). "Tests for dimensionality and interaction of mean vectors under general and reducible covariance structures". *J. Multivariate Analysis*, **16**, 173–184.

[27] **Rao, C.R.** (1987). "Prediction of future observations in growth curve type models". *J. Statistical Science*, **2**, 434–471.

[28] **Read, T.R.C.** and **Cressie, N.A.C.** (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. New York: Springer-Verlag.

[29] **Rios, M., Villarroya, A.** and **Oller, J.M.** (1994). "Intrinsic Data Analysis: A method for the simultaneous representation of populations and variables". *Mathematics preprint series*, **160**. Universitat de Barcelona.

[30] **Von Neumann, J.** (1937). "Some matrix inequalities and metrization of metric spaces". *Tomsk. Univ. Rev.*, **1**, 286–299.

[31] **Wegman, E.J.** (1990). "Hyperdimensional data analysis using parallel coordinates". *J. Amer. Statist. Assoc.*, **85**, 664–675.