Biometrika (1966), **53**, 3 *and* 4, *p.* 325
Printed in Great Britain

Some distance properties of latent root and vector methods used in multivariate analysis

BY J. C. GOWER
Rothamsted Experimental Station

SUMMARY

This paper is concerned with the representation of a multivariate sample of size n as points $P_1, P_2, \ldots, P_n$ in a Euclidean space. The interpretation of the distance $\Delta(P_i, P_j)$ between the ith and jth members of the sample is discussed for some commonly used types of analysis, including both Q and R techniques. When all the distances between n points are known a method is derived which finds their co-ordinates referred to principal axes. A set of necessary and sufficient conditions for a solution to exist in real Euclidean space is found. Q and R techniques are defined as being dual to one another when they both lead to a set of n points with the same inter-point distances. Pairs of dual techniques are derived. In factor analysis the distances between points whose co-ordinates are the estimated factor scores can be interpreted as D^2 with a singular dispersion matrix.

1. INTRODUCTION

If we have information on v variates for each of n individuals, this can be set out in a two-way table with x_{ij} as the value of the jth variate for the ith individual. This table is the starting-point for most multivariate statistical methods such as discriminant analysis, factor analysis and principal components analysis.

Different methods require different assumptions about the structure of the sample and about the hypothetical multivariate probability distribution from which the sample is drawn. Thus, in discriminant analysis we must be able to assign each individual to one of a set of predetermined groups each with a known distributional form, and in factor analysis the dispersion matrix must have a particular structure.

Recently there has been widespread interest in the possibility of using numerical methods as an aid to classification. Sokal & Sneath (1963, p. 178) give references to investigations using a variety of different numerical methods; these methods all begin with a multivariate sample, but so far as is known each individual may come from a different biological population or all individuals may come from the same population. An additional complication is that many or all of the variates may be qualitative, so that product moment correlations between variates may be inappropriate. The techniques used are based on a $n \times n$ symmetric matrix $\mathbf{A}$ whose (i,j)th element, a_{ij}, is a coefficient of association between the ith and jth individuals. Sokal & Sneath (1963) give a list of commonly used coefficients, many of which are familiar to statisticians.

In one type of analysis the matrix $\mathbf{A}$ is first calculated and then a method of forming groups of individuals known as a cluster analysis follows. Individuals are assigned to the same group when their coefficients in $\mathbf{A}$ obey certain criteria, which depend on the method of cluster analysis being used. These criteria always have a simple geometric interpretation when a_{ij}, or some function of a_{ij}, is regarded as the distance between the ith and jth individuals. Thus an arbitrary distance δ may be chosen and a criterion defined by saying that all members of a group must be within distance δ of all other members, or that it must be possible to find a chain linking all members of the group such that the length of no link is greater than δ, or that the average distance between all members of the group is less than δ, etc. The usual technique is to let δ vary from a maximum to a minimum (sorting upwards) or from a minimum to a maximum (sorting downwards), and to examine how these groups split or combine at the different levels of δ. Not only do different criteria give different groups, but some criteria form different groups for the same value of δ when sorting upwards from when sorting downwards.

We are not primarily concerned with cluster analysis here but only wish to emphasize that the concept of distance between individuals is fundamental to these techniques.

Recognizing the metric nature of the matrix **A**, several workers have endeavoured to construct models in two or three dimensions which reflect the inter-individual distances; see, for example, Lysenko & Sneath (1959) and Bidwell & Hole (1964). One of the by-products of most standard multivariate statistical techniques is a representation of the multivariate sample in a small number of dimensions. There is a growing tendency to use these techniques formally on association matrices to derive multi-dimensional representations of the sample, although the standard underlying assumptions are not even approximately satisfied. In particular the association matrix is a $n \times n$ matrix formed from the comparison of all pairs of individuals, i.e. a so-called Q matrix, whilst standard techniques postulate a $v \times v$ dispersion or correlation matrix, i.e. a so-called R matrix, formed from comparisons between the variates. Despite this the techniques have been used successfully, in the sense that the expected relative magnitudes of inter-individual distances have been recovered. In this paper we shall investigate the extent to which this use is valid. To do this we shall need the method of principal components analysis, which is therefore discussed in §2.

2. Principal component analysis

When all the variates are quantitative, the method of principal components can be used to construct multi-dimensional models of the type just discussed. An interpretation for the special case of qualitative (0, 1) data is given in §4·1. Unlike other forms of multivariate analysis no assumption need be made about the distribution of the variates in the hypothetical population, except of course when significance tests are of interest. The multivariate sample is regarded as defining a set of n points P_i $(i = 1, 2, \ldots, n)$ in v space, where P_i has co-ordinates $(x_{i1}, x_{i2}, \ldots, x_{iv})$ referred to rectangular axes. Thus the implied distance d_{ij} between P_i and P_j is given by

$$d_{ij}^2 = \sum_{r=1}^{v} (x_{ir} - x_{jr})^2, \tag{1}$$

and the spatial configuration of the sample is of interest only if d_{ij} satisfactorily measures the similarity between the ith and jth individuals.

As a measure of similarity d_{ij} has the obvious defect that it depends in a complex manner on the scales of measurement of the different variates. When different variates are measured in different scales, d_{ij} has nonsensical physical dimensions. To evade this difficulty it is common practice to normalize variates by dividing each by its sample standard error, but other normalizers could be used, for example, the variate mean (when zero is not arbitrarily located), or the range or even the cube root of the sample third moment.

Formula (1) makes no attempt to allow for correlations. In this respect d_{ij} has similar properties to measures of distance based on various similarity coefficients currently in favour in classification work and contrasts with distances used in discriminant analysis, e.g. D^2.

Principal components are computed by evaluating the latent roots and vectors of the sums of squares and products matrix formed from the normalized variates. The vector corresponding to the largest root gives the direction cosines of a line through the sample mean, G, such that the sum of squares of the perpendicular distances from the P_i onto this line is a minimum. If the foot of the perpendicular from P_i to the line is H_i then since $GP_i^2 = GH_i^2 + H_iP_i^2$ we must have maximized the sum of squares ΣGH_i^2. In fact ΣGH_i^2 is equal to the largest latent root; principal components analysis is therefore often regarded as a means of finding the direction cosines of a line such that the sum of squares of the projections GH_i onto this line is a maximum. Having found one line with the above properties, we look for a second line, at right angles to the first and corresponding to the second largest latent root and vector, which minimizes the sum of squares of the perpendiculars onto the plane defined by the two lines. The sum of squares in this new direction is equal to the second largest latent root. Similar properties hold for further roots and vectors. Clearly if the sample fits exactly into $k(< v)$ dimensions the last $v-k$ latent roots must be zero. Any dimension which has a small latent root will contribute little to the original distances between the normalized co-ordinates and may be ignored. When nearly all the variation is in two or three dimensions the co-ordinates of the feet of the perpendiculars Π_i from P_i onto the reduced dimensions may be used to plot a graph or construct a model preserving as nearly as possible, in the defined sense, the inter-individual distances d_{ij}. The points found will have the property that the square of the distance from Π_i to Π_j summed over all pairs of individuals will be the maximum possible for the chosen number of dimensions. The possibility of associating definite properties, e.g. factors, with the reduced number of dimensions will not be discussed here.

3. Q-TECHNIQUES AND THE TREATMENT OF QUALITATIVE VARIATES

The method of principal components is often used, and misused, by statisticians. When unordered qualitative variates occur it is not applicable, except possibly for the special case of (0, 1) data; see § 4·1. A different axis can be assigned to each level of every qualitative

Table 1. *The latent roots and vectors of the symmetric matrix* $\mathbf{A}$

($\bar{c}_r$ is the mean value of the elements of the rth vector.)

		Root			
		λ_1	λ_2	$\cdots$	λ_n
Point	Q_1	c_{11}	c_{12}	$\cdots$	c_{1n}
	Q_2	c_{21}	c_{22}	$\cdots$	c_{2n}
	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	Q_n	c_{n1}	c_{n2}	$\cdots$	c_{nn}
Centroid	$\bar{Q}$	$\bar{c}_1$	$\bar{c}_2$	$\cdots$	$\bar{c}_n$

variate and a score of 1 given if this level occurs and 0 if it does not occur; the problem then reduces to that of (0, 1) data discussed in § 4·1. An alternative treatment frequently used for qualitative data, and sometimes for quantitative data, is to calculate a $n \times n$ Q matrix of coefficients of association between individuals. The co-ordinates of the point Q_i corresponding to the ith individual are now defined by the ith component of each of the n vectors of the Q matrix. It is hoped to get a low-dimensional representation of the sample by using co-ordinate axes corresponding to only a few large latent roots. Such methods have been found to lead to reasonable results, but the underlying theory is not so well understood as is that of principal components analysis. In particular two questions arise: what is the correct scaling for each latent vector and what is the distance $\Delta(Q_i, Q_j)$ when this scaling is used? In the remainder of this section we first answer these questions, then show how the method can be improved and give a few ancillary results.

Suppose $\mathbf{A}$ is a symmetric matrix of order n with latent roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ and associated column vectors $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$. We may write the vectors down in a square array as in Table 1 where the rth column is the vector $\mathbf{c}_r$. The technique outlined above is to take the elements of the ith row as the co-ordinates of a point Q_i in n space and the distance d_{ij} between Q_i and Q_j is then given by

$$d_{ij}^2 = \sum_{r=1}^{n} c_{ir}^2 + \sum_{r=1}^{n} c_{jr}^2 - 2\sum_{r=1}^{n} c_{ir}c_{jr}. \tag{2}$$

Now it is a well-known property of latent vectors that if they are normalized so that the sums of squares of their elements are equal to their corresponding latent roots then

$$\mathbf{A} = \mathbf{c}_1\mathbf{c}_1' + \mathbf{c}_2\mathbf{c}_2' + \ldots + \mathbf{c}_n\mathbf{c}_n'. \tag{3}$$

Thus $$a_{ii} = \sum_{r=1}^{n} c_{ir}^2 \quad \text{and} \quad a_{ij} = \sum_{r=1}^{n} c_{ir}c_{jr},$$

so that (2) becomes $d_{ij}^2 = a_{ii} + a_{jj} - 2a_{ij}$. Thus when the vectors of Table 1 have been normalized so that $\Sigma c_{ir}^2 = \lambda_r$,

$$\{\Delta(Q_i, Q_j)\}^2 = a_{ii} + a_{jj} - 2a_{ij}. \tag{4}$$

The points Q_i often have the right sort of metric properties for representing the interrelationships between the individuals when $\mathbf{A}$ is an association matrix. For if $\mathbf{A}$ is a similarity matrix or a formal product-moment correlation matrix between the individuals, $\Delta(Q_i, Q_j)$ will be zero for complete identification and will attain its maximum value for complete opposites. In both these cases the diagonal elements of $\mathbf{A}$ are unity and (4) becomes

$$\{\Delta(Q_i, Q_j)\}^2 = 2(1 - a_{ij}).$$

It also follows from (4) that if we put $a_{ij} = -\frac{1}{2}d_{ij}^2$ and $a_{ii} = 0$ then $\Delta(Q_i, Q_j) = d_{ij}$ and this gives a direct method of finding the co-ordinates of a set of points given their interdistances d_{ij}.

It is easy to see why a good representation can be obtained in a reduced number of dimensions when some of the λ_i are small. If λ_r is small then the contribution $(c_{ir} - c_{jr})^2$ to the distance between Q_i and Q_j will also be small; in fact the sum of squares of the co-ordinates along the λ_r axis is λ_r by definition. However, if λ_r is large but the c_{ir} corresponding to it are not very different then $(c_{ir} - c_{jr})^2$ will also be small. Thus, the only co-ordinates which contribute much to the distances are those for large λ_r which have wide variation in the elements of their vectors. In many applications it is found that the distances can be adequately expressed in terms of two or three such vectors. Very often the vector corresponding to the largest root has more or less constant elements which by (3) will contribute constant amounts to all the a_{ij} and may be regarded as allowing for the mean value of all the elements of **A**. Clearly such a mean value is unimportant, for if we add any constant to all the elements of **A** and go through the same numerical procedure we will leave the distances $\Delta(Q_i, Q_j)$ as given by (4) invariant. We will of course get different co-ordinate values, Q_i', in Table 1 and different roots but the distances will not have altered. These new points must be an orthogonal transformation of the first set after a change of origin and possibly a mirror image transformation, depending on the signs arbitrarily given to the vector elements; we can ask which of these transformations will give us the best fit with a reduced number of co-ordinates. This is precisely the situation for the use of a principal components analysis. If the elements of **A** are dimensionless, as is usual for association type matrices, then we have a set of co-ordinates given by Table 1, and principal components analysis will give the direction cosines of lines of best fit through $\bar{Q}$ in as many dimensions as required. These direction cosines are the latent vectors of a corrected sums of squares and products matrix **B** where the columns of Table 1 are now regarded as variates. We have

$$\sum_{i=1}^{n} c_{ij}^2 = \lambda_j, \quad \sum_{i=1}^{n} c_{ij}c_{ik} = 0,$$

so that

$$b_{ii} = \lambda_i - n\bar{c}_i^2, \quad b_{ij} = -n\bar{c}_i\bar{c}_j. \tag{5}$$

The characteristic equation for the latent roots μ_i of **B** is

$$f(\mu) \equiv \bar{c}_1^2 L_1 + \bar{c}_2^2 L_2 + \ldots + \bar{c}_n^2 L_n - L_0/n = 0, \tag{6}$$

where

$$L_0 = (\lambda_1 - \mu)(\lambda_2 - \mu)\ldots(\lambda_n - \mu)$$

and

$$L_i = (\lambda_1 - \mu)\ldots(\lambda_{i-1} - \mu)(\lambda_{i+1} - \mu)\ldots(\lambda_n - \mu) \quad \text{if} \quad i \neq 0.$$

Using equation (6) it is possible to relate the roots μ to the roots λ. First (6) has a zero root, because it is always possible to fit n points into $n-1$ dimensions, and secondly the remaining roots of (6) are separated by the roots λ. Thus a knowledge of the λ_i will tell us quite a lot about the best fit to be obtained from a principal components analysis of the Q_i.

If two, or more, of the λ_i are negative then we must have one, or more, negative values of μ, indicating that points cannot be found in real space such that $d_{ij}^2 = a_{ii} + a_{jj} - 2a_{ij}$. When the negative μ have small modulus they will have little effect on the values of d_{ij} as calculated from the real co-ordinates and will give no trouble. If, however, a large negative μ occurs it may still be possible to find a low-dimensional model to give the right order of magnitude of d_{ij} but this model will involve one or more purely imaginary sets of co-ordinate axes which will upset intuitive ideas of distance. Thus the points $Q_1(1, i)$ and $Q_2(-1, -i)$ will in fact be zero distance apart but in any real model will appear distant. Thus it becomes important to examine the conditions under which the matrix **A** will give rise to a real configuration.

Obviously it is sufficient for **A** to be positive semi-definite (p.s.d.) for then all λ and hence all μ are non-negative. On the other hand if there is more than one negative λ a real configuration is impossible because this implies at least one negative μ. If there is just one negative root λ_n then a real configuration is possible only when $\mu_{n-1} \geqslant 0$. If **A** has a zero root then a necessary and sufficient condition for a real configuration is that **A** is p.s.d., because if **A** has a negative root so will **B**. It is always possible to adjust **A** so that it has a zero root without altering the distances $\Delta(Q_i, Q_j)$ for if $\bar{a}_i$ is the mean value of the ith row (or column) of **A** and $\bar{a}$ is the overall mean we can define a matrix $\boldsymbol{\alpha}$ with elements given by

$$\alpha_{ij} = a_{ij} - \bar{a}_i - \bar{a}_j + \bar{a}. \tag{7}$$

It is easy to see that

$$\alpha_{ii} + \alpha_{jj} - 2\alpha_{ij} = a_{ii} + a_{jj} - 2a_{ij} \tag{8}$$

preserving the distance property. The sum of every row and column of $\boldsymbol{\alpha}$ is zero and consequently $\boldsymbol{\alpha}$ has a zero root.

A set of necessary and sufficient conditions for a real configuration of points Q_i to exist such that the distances d_{ij} are given by $d_{ij}^2 = a_{ii} + a_{jj} - 2a_{ij}$ is that $\boldsymbol{\alpha}$ be p.s.d. For many frequently used coefficients of association it can be shown that **A** has this property. An account of this work will be given elsewhere.

We have now established that if we are given an association matrix **A** we can find a set of points Q_i $(i = 1, \ldots, n)$ in n space such that the distance d_{ij} between Q_i and Q_j is given by $d_{ij}^2 = a_{ii} + a_{jj} - 2a_{ij}$ and that for many coefficients of association this is a desirable property. Furthermore, it is legitimate to use the method of principal components on the co-ordinates of the Q_i to find the best fit in fewer dimensions. This method requires a two-stage computational procedure and at each stage the latent roots and vectors of an $n \times n$ matrix are required. It would be advantageous if these two stages could be collapsed into one. From (5) we see that **B** will become the diagonal matrix diag $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ if $\bar{c}_i = 0$ for all i and under these conditions the points Q_i are themselves the principal component projections because the direction cosines defined by the vectors of any diagonal matrix are parallel to the original axes.

Because the rows of $\boldsymbol{\alpha}$ all sum to zero the vector $\mathbf{1} = (1, 1, \ldots, 1)'$ is the latent vector of $\boldsymbol{\alpha}$ corresponding to the zero root, so that if $\mathbf{v} = (v_1, v_2, \ldots, v_n)'$ is any other latent vector, then $\mathbf{1}'\mathbf{v} = 0$ and thus $\Sigma v_i = 0$.

If now Table 1 has arisen from $\boldsymbol{\alpha}$ rather than from **A** we have $\bar{c}_i = 0$, either because $\lambda_i = 0$ or by the above result. It follows from the previous paragraph that the points Q_i derived from $\boldsymbol{\alpha}$ are the principal component fit to the co-ordinates that would be arrived at after the two-stage calculation starting from **A**.

The computational procedure suggested is:

(i) Form the association matrix **A**.

(ii) Transform this to $\boldsymbol{\alpha}$ by equation (7).

(iii) Construct Table 1 by finding the latent roots and vectors of $\boldsymbol{\alpha}$ scaling each vector so that its sum of squares is equal to its corresponding latent root.

(iv) The ith row of Table 1 may now be regarded as the co-ordinates of a set of points Q_i whose distances apart are given by the best approximations to $(a_{ii} + a_{jj} - 2a_{ij})^{\frac{1}{2}}$ in the chosen number of dimensions.

(v) As in all principal components analyses, the sum of squares of the residuals (i.e. perpendiculars on to the reduced k-dimensional representation) will be the difference between the trace of $\boldsymbol{\alpha}$ and the sum of the k largest roots of $\boldsymbol{\alpha}$.

Results similar to those given above have previously been reported by Torgerson (1958, p. 123) but this author does not appear to realize the close connexion with principal components analysis and suggests that $\boldsymbol{\alpha}$ can be analysed by any factor analytical method; such an approach would only obscure interpretation. Rao (1964), in a paper surveying the different uses of principal components, has quoted Torgerson's results but overlooks the fact that the co-ordinates obtained from $\boldsymbol{\alpha}$ have their centroid as origin, for he advocates shifting the origin to the centroid by replacing $\boldsymbol{\alpha}$ by $(\mathbf{I}-\mathbf{1}\mathbf{1}'/n)\,\boldsymbol{\alpha}$. Because the column means of $\boldsymbol{\alpha}$ are zero, $\mathbf{1}'\boldsymbol{\alpha}$ is a null row vector and the formula leaves $\boldsymbol{\alpha}$ unchanged.

4. Duality of Q and R techniques

Q techniques are those which operate on a $n \times n$ matrix whose elements are measures of association between the individuals. R techniques operate on a $v \times v$ matrix defining relationships amongst the variates. Thus § 3 is concerned with a Q technique whilst principal components and most other statistical multivariate methods are R techniques. Both Q and R techniques give the co-ordinates of a set of n points in multidimensional space. If, given a multivariate sample, a Q and an R technique both give rise to the same set of points, in the sense that the distances between all pairs of points are duplicated, then we shall say that the Q and R techniques are dual to one another. In a few important cases we can find pairs of dual techniques.

4·1. *The dual of principal components analysis*

The Q-matrix defined by $q_{ij} = -\frac{1}{2}d_{ij}^2$, where d_{ij} is given by (1), must, using the method of § 3, give rise to the principal components fit to a set of points whose distances apart are $(q_{ii}+q_{jj}-2q_{ij})^{\frac{1}{2}} = d_{ij}$. These are the original distances and it follows that the method of § 3 operating on $-\frac{1}{2}d_{ij}^2$ is dual to principal components analysis of the sum of squares and products matrix between the variates. We now give an algebraic proof of this; we shall need some of the results later.

Let $\mathbf{X}$ be the $n \times v$ data matrix composed entirely of variates with zero sample means, then the sums of squares and products matrix required for a principal component analysis is $\mathbf{X}'\mathbf{X}$. Suppose this matrix has a latent root λ and corresponding vector $\mathbf{u}$. Thus

$$\mathbf{X}'\mathbf{X}\mathbf{u} = \lambda\mathbf{u}. \tag{9}$$

Consider the Q matrix $\mathbf{X}\mathbf{X}'$ which has a root μ and vector $\mathbf{v}$ then

$$\mathbf{X}\mathbf{X}'\mathbf{v} = \mu\mathbf{v}. \tag{10}$$

Pre-multiplying (9) by $\mathbf{X}$ we have

$$\mathbf{X}\mathbf{X}'(\mathbf{X}\mathbf{u}) = \lambda(\mathbf{X}\mathbf{u}), \tag{11}$$

so that $\lambda = \mu$ and $\mathbf{X}\mathbf{u} = k\mathbf{v}$, where k is a constant relating the scaling of the two sets of vectors. The direction cosines for principal components are normalized such that $\mathbf{u}'\mathbf{u} = 1$. From (11), $k^2\mathbf{v}'\mathbf{v} = \mathbf{u}'\mathbf{X}'\mathbf{X}\mathbf{u} = \lambda\mathbf{u}'\mathbf{u} = \lambda$. If we normalize the v vectors so that $\mathbf{v}'\mathbf{v} = \lambda$ as is required in § 3 then $k = 1$ and (11) becomes $\mathbf{v} = \mathbf{X}\mathbf{u}$.

The elements a_{ij} of the Q matrix $\mathbf{X}\mathbf{X}'$ are given by

$$a_{ii} = \sum_{r=1}^{v} x_{ir}^2, \quad a_{ij} = \sum_{r=1}^{v} x_{ir}x_{jr}. \tag{12}$$

The method of § 3 will give rise to a set of co-ordinates Q_i whose inter-co-ordinate distances are given by

$$d_{ij}^2 = a_{ii}+a_{jj}-2a_{ij} = \sum_{r=1}^{v}(x_{ir}-x_{jr})^2. \tag{13}$$

Thus a dual of principal components analysis is the method of § 3 operating on the Q matrix given by (12). Another Q matrix giving an identical solution is given by

$$q_{ij} = -\tfrac{1}{2}d_{ij}^2 = -\tfrac{1}{2}\sum_{r=1}^{v}(x_{ir}-x_{jr})^2$$

of which (12) is the $\boldsymbol{\alpha}$ matrix of § 3. The co-ordinates given by the first t vectors of (12) are the same as those obtained from the projections on to the first t direction cosine vectors obtained from the principal component analysis. This follows from the fact that the technique of § 3 gives the principal component fit to a set of points whose distances apart are given by (13) which are distances between the co-ordinate points used for the principal components analysis. There are two important special cases of this duality of principal components. These are dealt with in the following two subsections.

Sokal's measure of taxonomic distance. Suppose each x_{ij} is standardized by the variate standard error s_j, then principal components are derived from the correlation matrix. The distances are given by

$$l_{ij}^2 = \sum_{r=1}^{v} \left(\frac{x_{ir}-x_{jr}}{s_r}\right)^2,$$

which is the measure of taxonomic distance proposed by Sokal and others (Sokal & Sneath, 1963, p. 147). The technique of § 3 used on the matrix

$$d_{ij} = \sum_{r=1}^{v} x_{ir}\, x_{jr}/s_r^2$$

will lead to a reduced dimensional configuration identical to that obtained from a principal components analysis of the correlation matrix.

When $n \leqslant v$ the Q technique is clearly more convenient computationally because it requires the roots and vectors of a smaller matrix. When $n > v$ the Q technique may still be better because it leads directly to the co-ordinates of Q_i whilst the R technique needs a set of orthogonal transformations $\mathbf{u}'\mathbf{x}$ to derive these co-ordinates.

Matching coefficients. Suppose now that $\mathbf{X}$ is composed entirely of (0, 1) data. When comparing two individuals we have, in the usual notation, a (1, 1), b (0, 1), c (1, 0) and d (0, 0) pairs where $a+b+c+d = v$, the number of variates. Sokal & Sneath (1963, p. 125) give a list of different matching coefficients which have been used from time to time. We are here only concerned with $S_{ij} = (a+d)/v$ and note that

$$\sum_{r=1}^{v} (x_{ir}-x_{jr})^2 = b+c = v(1-S_{ij})$$

and that (12) becomes

$$\left.\begin{aligned} a_{ii} &= \text{number of characters present for } i\text{th individual,} \\ a_{ij} &= \text{number of characters common to } i\text{th and } j\text{th individuals.} \end{aligned}\right\} \qquad (14)$$

We shall ignore v in the following as it only has a constant multiplicative effect throughout and does not materially affect our results; in practice v may be important if there are many missing values. The technique of § 3 used on the matrix defined by (14) will give points distance $\{2(1-S_{ij})\}^{\frac{1}{2}}$ apart which are also the same as the distances given by $a_{ij} = S_{ij}$. The same points would be obtained by a principal components analysis on the matrix of corrected sums of squares and products between the variates. Although a conventional principal components analysis of (0, 1) data may seem of dubious validity, the above shows that it is exactly equivalent to assuming that the individuals are represented by points whose distances apart are proportional to $(1-S_{ij})^{\frac{1}{2}}$.

4·2. *Mahalanobis's D^2 statistic and discriminant analysis*

If we have k different multivariate normal populations with a common dispersion matrix $\mathbf{W}$, then the Mahalanobis measure of distance between the means $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}_j$ of the ith and jth populations is given by

$$D^2_{ij} = (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)\,\mathbf{W}^{-1}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)'.$$

Assuming, without loss of generality, that $\bar{x}_{1r} + \ldots + \bar{x}_{kr} = 0$ for $r = 1, 2, \ldots, v$, the $\boldsymbol{\alpha}$ matrix corresponding to D^2 is given by

$$\alpha_{ii} = \bar{\mathbf{x}}_i \mathbf{W}^{-1}\bar{\mathbf{x}}_i', \quad \alpha_{ij} = \bar{\mathbf{x}}_i \mathbf{W}^{-1}\bar{\mathbf{x}}_j'. \tag{15}$$

The method of § 3 will give a set of k points referred to rectangular principal axes with inter-distances D_{ij}. We shall show that the configuration using the space defined by some only of the co-ordinate axes is the same as that provided by the normal linear discriminant functions, or canonical variates, for the same dimensions. A closely related result is given by Rao (1952) where he showed that the first t canonical variates maximize the total D^2 in t dimensions.

The matrix $\boldsymbol{\alpha}$ of (15) may be written $\bar{\mathbf{X}}\mathbf{W}^{-1}\bar{\mathbf{X}}'$, where $\bar{\mathbf{X}}$ is the $k \times v$ matrix of the population means whose ith row gives the means for the ith population.

We put $\mathbf{W}^{-1} = \mathbf{U}\mathbf{U}'$. This can always be done and in fact $\mathbf{U}$ may be upper triangular. Thus $\boldsymbol{\alpha} = (\bar{\mathbf{X}}\mathbf{U})(\bar{\mathbf{X}}\mathbf{U})'$ and this is of the same form as (10) and consequently the analysis of § 3 gives the same results as a principal components analysis of $(\bar{\mathbf{X}}\mathbf{U})'(\bar{\mathbf{X}}\mathbf{U})$, that is $\mathbf{U}'\bar{\mathbf{X}}'\bar{\mathbf{X}}\mathbf{U}$. We require the latent vectors $\mathbf{l}$ of $\mathbf{U}'\bar{\mathbf{X}}'\bar{\mathbf{X}}\mathbf{U}$; that is

$$(\mathbf{U}'\bar{\mathbf{X}}'\bar{\mathbf{X}}\mathbf{U} - \lambda\mathbf{I})\mathbf{l} = 0, \quad \text{or} \quad \mathbf{U}'(\bar{\mathbf{X}}'\bar{\mathbf{X}} - \lambda\mathbf{U}'^{-1}\mathbf{U}^{-1})\,\mathbf{U}\mathbf{l} = 0.$$

Now $\mathbf{U}'^{-1}\mathbf{U}^{-1} = (\mathbf{U}\mathbf{U}')^{-1} = \mathbf{W}$ and therefore $\mathbf{U}'(\bar{\mathbf{X}}'\bar{\mathbf{X}} - \lambda\mathbf{W})\,\mathbf{U}\mathbf{l} = 0$.

This is the equation to be solved for finding the discriminant functions, whose coefficients are therefore given by $\mathbf{U}\mathbf{l}$.

The co-ordinates of the k means given by the discriminant coefficients are $\bar{\mathbf{X}}\mathbf{U}\mathbf{l}$ whilst those given by the principal components analysis are given by (11) and are also $\bar{\mathbf{X}}\mathbf{U}\mathbf{l}$. This proves the result.

It is instructive to see what happens if, given a single population, we define the distance d_{ij} between individuals by $d^2_{ij} = (\mathbf{x}_i - \mathbf{x}_j)\,\mathbf{W}^{-1}(\mathbf{x}_i - \mathbf{x}_j)'$, where $\mathbf{W}$ is the dispersion matrix. We may do this with D^2 in mind, hoping to allow for correlations between the variates in our measure of distance. The previous algebraic treatment follows except that $\bar{\mathbf{X}}'\bar{\mathbf{X}}$ is now $\mathbf{W}$. Thus we have to solve $(\mathbf{W} - \lambda\mathbf{W})\,\mathbf{U}\mathbf{l} = 0$, an equation which only has a non-null solution when $\lambda = 1$ in which case any vector $\mathbf{l}$ will suffice. This implies that variation in all directions is homogeneous, or, what amounts to the same thing, the points distance d_{ij} apart lie at the vertices of a regular simplex in $n-1$ dimensions projected onto a space of v dimensions. This result is satisfactory because no preference is given to any one individual out of the whole set which define the population and reflects the attitude that the group of individuals is homogeneous.

5. Factor analysis

In the last paragraph we showed that a reasonable measure of distance between sample numbers of a single population does not lead to a reduction in the number of dimensions of the sample from the smaller of $n-1$ and v. Factor analysis achieves this by assuming that each variate can be expressed as a linear compound of k ($< v$) hypothetical variates, the common factors, plus an additional term depending only on the particular variate and known as the specific factor. Algebraically the variates are related to each other by

$$x_i = \sum_{j=1}^{k} l_{ij} f_j + e_i \quad (i = 1, 2, \ldots, v).$$

The specifics e_i are supposed to be independent of each other and of the f_j. There is no loss of generality in assuming that the f_j are uncorrelated and with unit variance. The covariance matrix $\mathbf{C}$ of the x's is $\mathbf{C} = \mathbf{LL}' + \mathbf{V}$, where $\mathbf{V}$ is the diagonal matrix of the variances of the e_i and $\mathbf{L}$ is the $v \times k$ matrix of the coefficients l_{ij}, the factor loadings. The problem is to estimate $\mathbf{L}$ and $\mathbf{V}$ and also to estimate the factor scores f_j for each individual. The estimation procedure depends on what is assumed about the distribution of the x's.

5·1. *Distances between individuals*

It is usual in factor analysis to concentrate interest on the factor loadings but we are more concerned here with the distances between the individuals, regarding their factor scores as co-ordinates. We first note that the fundamental equation of factor analysis may be written

$$\mathbf{x} - \mathbf{e} = \mathbf{Lf}. \tag{16}$$

Thus for any estimates of $\mathbf{f}$ for a set of individuals the corrected scores $\mathbf{x} - \mathbf{e}$ lie in k-space as do the factor scores themselves. The factors $\mathbf{f}$ are uncorrelated with unit variance, though this is not true of the estimated factor scores using the methods discussed in the next section. Absence of correlation is retained for any orthogonal rotation of axes in factor space. Throughout this section we assume the theoretical unit dispersion matrix for the factor scores so that distances between the individuals with factor scores as co-ordinates can be calculated directly using D^2 with a unit-dispersion matrix. The dispersion matrix for the corrected scores is $\mathbf{C} - \mathbf{V}$ or $\mathbf{LL}'$ which has rank k and therefore no inverse in the ordinary sense. A slight extension to the definition will allow us to define D^2 for singular dispersion matrices; details are given in the Appendix.

Equation (16) relates the v variates $\mathbf{x} - \mathbf{e}$ to the k variates $\mathbf{f}$ as described in the Appendix and therefore, using the extended definition of D^2, the distances between the estimated factor scores are equal to the distances between the observed variates corrected for the estimated specifics. This result is independent of the particular method used for estimating the factor scores.

An alternative way of looking at the factor analysis model is to regard the v-space of the x-variates as embedded in a $(v + k)$-space whose axes are orthogonal, v of these representing the specifics and k the common factors. Correcting the x's for the specifics ensures that a k-space is attained.

Writing $\delta\mathbf{y}$ for $(\mathbf{x}_i - \mathbf{e}_i) - (\mathbf{x}_j - \mathbf{e}_j)$ and $\delta\mathbf{f}$ for $\mathbf{f}_i - \mathbf{f}_j$, the differences between the corrected scores and the estimated factor scores respectively of two individuals i and j, the equality between the two distances gives

$$\delta\mathbf{f}'\,\delta\mathbf{f} = \delta\mathbf{y}'(\mathbf{LL}')^{-}\,\delta\mathbf{y}, \tag{17}$$

where $(\mathbf{LL}')^{-}$ is a suitably chosen generalized inverse of $\mathbf{LL}'$ depending on the method used for estimating factor scores; see below.

5·2. *Distances using particular estimates of factor scores*

There are two closely related methods for estimating factor scores, one due to Bartlett (1938) and the other to Thomson (1951); see also Lawley & Maxwell (1963, p. 88). For each of these methods we can find an algebraic form for $\delta\mathbf{f}'\,\delta\mathbf{f}$ in terms of the original data and the matrices $\mathbf{C}$, $\mathbf{V}$, $\mathbf{L}$.

Bartlett's method estimates the factor scores $\mathbf{f}$ by

$$\tilde{\mathbf{f}} = \mathbf{J}^{-1}\mathbf{L}'\mathbf{V}^{-1}\mathbf{x}, \tag{18}$$

where

$$\mathbf{J} = \mathbf{L}'\mathbf{V}^{-1}\mathbf{L}. \tag{19}$$

Thus on noting that $\delta\mathbf{y} = \mathbf{L}\,\delta\mathbf{f}$ and writing $\delta\mathbf{x}$ for $\mathbf{x}_i - \mathbf{x}_j$, equation (17) becomes

$$\delta\mathbf{x}'\mathbf{V}^{-1}\mathbf{L}\mathbf{J}^{-1}\mathbf{J}^{-1}\mathbf{L}'\mathbf{V}^{-1}\delta\mathbf{x} = \delta\mathbf{x}'\mathbf{V}^{-1}\mathbf{L}\mathbf{J}^{-1}\mathbf{L}'(\mathbf{L}\mathbf{L}')^{-}\mathbf{L}\mathbf{J}^{-1}\mathbf{L}'\mathbf{V}^{-1}\delta\mathbf{x}. \tag{20}$$

If we set

$$(\mathbf{L}\mathbf{L}')^{-} = \mathbf{V}^{-1}\mathbf{L}\mathbf{J}^{-1}\mathbf{J}^{-1}\mathbf{L}'\mathbf{V}^{-1}, \tag{21}$$

we see that (20) is satisfied and it is easy to verify that this is a generalized inverse of $\mathbf{L}\mathbf{L}'$; for

$$(\mathbf{L}\mathbf{L}')\mathbf{V}^{-1}\mathbf{L}\mathbf{J}^{-2}\mathbf{L}'\mathbf{V}^{-1}(\mathbf{L}\mathbf{L}') = \mathbf{L}\mathbf{L}'$$

follows immediately from (19).

Thus for Bartlett's method of estimation the distance is given by

$$\delta\mathbf{x}'(\mathbf{L}\mathbf{L}')^{-}\delta\mathbf{x} \quad \text{which may be written} \quad \delta\mathbf{x}'(\mathbf{C}-\mathbf{V})^{-}\delta\mathbf{x}, \tag{22}$$

with the particular generalized inverse given by (21). Equation (22) is Mahalanobis's D^2 with $\mathbf{C}-\mathbf{V}$ in the role of the within population dispersion matrix.

Thomson's method of estimating factor scores gives

$$\tilde{\mathbf{f}} = \mathbf{L}'\mathbf{C}^{-1}\mathbf{x}. \tag{23}$$

Substituting into (17), we have that

$$\delta\mathbf{x}'\mathbf{C}^{-1}\mathbf{L}\mathbf{L}'\mathbf{C}^{-1}\delta\mathbf{x} = \delta\mathbf{x}'\mathbf{C}^{-1}\mathbf{L}\mathbf{L}'(\mathbf{L}\mathbf{L}')^{-}\mathbf{L}\mathbf{L}'\mathbf{C}^{-1}\delta\mathbf{x}, \tag{24}$$

a result which is true by definition of a generalized inverse.

Thus in Thomson's case the distance between the factor scores is given by

$$\delta\mathbf{x}'\mathbf{C}^{-1}(\mathbf{C}-\mathbf{V})\,\mathbf{C}^{-1}\delta\mathbf{x}, \tag{25}$$

which may be written $\delta\mathbf{x}'[\mathbf{C}(\mathbf{C}-\mathbf{V})^{-}\mathbf{C}]^{-}\delta\mathbf{x}$ so that $\mathbf{C}^{-1}(\mathbf{C}-\mathbf{V})\,\mathbf{C}^{-1}$ is a generalized inverse of $\mathbf{C}\mathbf{V}^{-1}\mathbf{L}\mathbf{J}^{-1}\mathbf{J}^{-1}\mathbf{L}'\mathbf{V}^{-1}\mathbf{C}$ which may be shown to be equivalent to $\mathbf{L}(\mathbf{I}+\mathbf{J}^{-1})^2\mathbf{L}'$. Thus (25) may be written

$$\delta\mathbf{x}'[\mathbf{L}(\mathbf{I}+\mathbf{J}^{-1})^2\mathbf{L}']^{-}\delta\mathbf{x}, \tag{26}$$

which is a more complex D^2 form than (22).

5·3. *Relation of factor analysis to principal components*

This investigation of distance in factor analysis should not be taken to imply that the writer recommends the use of the method in the type of situation considered earlier where the multivariate sample cannot be considered as homogeneous. The method can be legitimately used only in those situations where its underlying assumptions are at least approximately fulfilled; in many published applications this is not so.

Cattell (1965), a leading exponent of factor analysis, discussing the heterogeneous population problem recommends that a cluster analysis should precede factor analysis in an attempt to separate out the different populations or 'species'. We can then separately factor analyse the members of each species to find factors that define individuals within the species, and do an inter-species factor analysis on the species means to find the factors that differentiate species. This is contrary to common practice which uses factor analysis on the complete heterogeneous sample as an alternative to cluster analysis in the hope that the first factor will differentiate species; but as Cattell points out, 'the factors obtained will neither be clear species differentiators nor optimum individual differentiators'. In the model-making situations preliminary to classification, discussed earlier, there usually seems to be little justification for using factor analysis and the simpler computational methods given above in §3 are just as useful. It seems that the reason for obtaining meaningful results when using factor analysis in these situations is that the results obtained are often very close to the results obtained by a principal components analysis.

Some insight into the relationship between the two methods can be obtained by noting that the maximum likelihood estimates of factor loadings satisfy the equations

$$\mathbf{J}\mathbf{L}'\mathbf{V}^{-\frac{1}{2}} = \mathbf{L}'\mathbf{V}^{-\frac{1}{2}}(\mathbf{V}^{-\frac{1}{2}}\mathbf{A}\mathbf{V}^{-\frac{1}{2}}-\mathbf{I}). \tag{27}$$

This is equation (2·11) of Lawley & Maxwell (1963) in a rearranged form. Thus $\mathbf{L}'\mathbf{V}^{-\frac{1}{2}}$ are the first k latent vectors of $\mathbf{V}^{-\frac{1}{2}}\mathbf{A}\mathbf{V}^{-\frac{1}{2}}-\mathbf{I}$ and $\mathbf{J}$ is the diagonal matrix of the latent roots. By setting $\mathbf{J} = \mathbf{L}'\mathbf{V}^{-1}\mathbf{L}$ the vectors are scaled so that the sum of squares of their elements equal the roots. Alternatively $\mathbf{L}'\mathbf{V}^{-\frac{1}{2}}$ are the latent vectors of $\mathbf{V}^{-\frac{1}{2}}\mathbf{A}\mathbf{V}^{-\frac{1}{2}}$ with roots given by $\mathbf{I}+\mathbf{J}$. In either case $\mathbf{L}'\mathbf{V}^{-\frac{1}{2}}$ are the vectors of a matrix which is put into non-dimensional form by scaling by the diagonal matrix $\mathbf{V}^{-\frac{1}{2}}$. The situation is very close to the principal components analysis of the correlation matrix where $\mathbf{V}^{-\frac{1}{2}}$ is replaced by $\mathbf{S}^{-\frac{1}{2}}$, the matrix of standard errors given by $\mathbf{S} = \operatorname{diag}(\mathbf{A})$. Regarding $\mathbf{J}^{-\frac{1}{2}}\mathbf{L}'\mathbf{V}^{-\frac{1}{2}}$ as direction cosines derived from a principal components analysis of $\mathbf{V}^{-\frac{1}{2}}\mathbf{A}\mathbf{V}^{-\frac{1}{2}}$ we see that the projected values of scaled co-ordinates $\mathbf{V}^{-\frac{1}{2}}\mathbf{x}$ onto k dimensions are given by

$$\mathbf{v} = \mathbf{J}^{-\frac{1}{2}}\mathbf{L}'\mathbf{V}^{-1}\mathbf{x}. \tag{28}$$

This result differs from Bartlett's and Thomson's estimates of factor scores only in a scale factor along each of the k axes so that, except in exceptional circumstances, the same sort of spatial configuration can be expected from all the methods. Thus when $\mathbf{V}$ is proportional to $\mathbf{S}$ a principal components analysis can be expected to give similar results to a factor analysis.

6. Proximity analysis

An interesting technique has been described by Shepard (1962*a*, *b*) and further developed by Kruskal (1964*a*, *b*) who name their method Proximity Analysis. The method assumes the matrix $\mathbf{A}$ to be available and endeavours to find a set of points Q_i in a reduced number of dimensions k such that the distances $\Delta(Q_i, Q_j)$ are monotonically related to the a_{ij}. Rather empirical methods are given for deciding on the optimum value of k required to preserve the monotonic relationship without using too many dimensions. The relationship of the a_{ij} to the $\Delta(Q_i, Q_j)$ gives the monotonic transformation of a_{ij} required to reduce dimensionality. For example, the monotonic transformation from a_{ij} to $-\log a_{ij}$ may be such that points Q_i can be found in, say, two dimensions so that the ranking of the distances $\Delta(Q_i, Q_j)$ is very nearly that of the $-\log a_{ij}$ and if this is so the graph of a_{ij} against $\Delta(Q_i, Q_j)$ should recover this transformation. The transformation, although monotonic, need not have any simple functional form. Shepard has produced examples to show that the method does indeed recover a known transformation when it is used on dummy data in two or three dimensions. The main interest of Shepard's work is that the numerical values of the a_{ij} are not used; only the rank order of their sizes is important. Shepard argues that when all the $\frac{1}{2}n(n-1)$ distances are known, a knowledge of their rank order, rather than the actual distances, entails very little loss of information.

In § 3 and when all the diagonal elements are originally 1, before calculating the $\boldsymbol{\alpha}$ matrix, we have shown that the distances are given by $\{2(1-a_{ij})\}^{\frac{1}{2}}$ and this is a particular monotonic transformation of the a_{ij}. Consequently if we get a good fit in, say, two dimensions using this distance function, we would expect high correlation with Shepard's solution. However, Shepard may be able to find a good fit in a low number of dimensions where the method of § 3 cannot. We are in effect using a particular distance function, though there is nothing to stop us transforming the values of a_{ij} ourselves if we have good reason to think that such a transformation will better reflect the inter-relationships between the individuals. As all similarities are positive numbers not greater than one, the points given by § 3 must lie within an $n-1$ dimensional regular simplex. Thus the points for any three similarities s_{12}, s_{23}, s_{31} lie inside an equilateral triangle of side $\sqrt{2}$. The logarithmic transformation will convert this restricted space to one where a pair of completely dissimilar individuals will become points an infinite distance apart, but it may introduce difficulties should no solution exist in real space.

The relationships between Shepard's solution and the principal components solution needs investigation but as Shepard's computations are much more complex than those required to find latent roots and vectors of a symmetric matrix it is suggested that the method of § 3 be tried first and if this does not lead to a solution of sufficiently low dimensionality then the co-ordinates found in $n-1$ dimensions can be used as a starting-point for Shepard's iterative method.

7. Conclusion

The work reported in this paper grew from a dissatisfaction with the many reported applications of factor analysis and principal components analysis of Q matrices found in classification work, particularly in the biological literature. The interpretation of such methods can be better understood by examining the distances, suitably defined, between the individuals and we have given a method for finding co-ordinates for each individual referred to principal axes which preserve these distances. To distinguish the method from classical principal components analysis it might be useful to refer to a *principal co-ordinates* analysis.

Perhaps the most important aspect of this type of analysis is the suitable choice of a distance function. When identification is required D^2 has certain optimal properties for normal populations whilst for non-normal populations, functions similar to D^2 could be devised which would immediately provide a canonical analysis and might lead to a practical solution of the discriminant problem. Rao (1948) has suggested a general approach to the problem of distance which might be useful here. Unfortunately it is not always possible to recognize the different populations in the original sample of individuals and these must first be found by using an analysis based on similarities or distances which do not allow for within population correlation. Further examination of the properties of such distances is needed.

I thank Mr M. J. R. Healy for many helpful discussions and also for the neat proof that $\Sigma v_i = 0$ given at the end of §3. Much of this work was stimulated by Dr J. H. Rayner's concern with soil classification.

Appendix. Generalized distance when the dispersion matrix is singular

To define D^2 in the more general case we first note that if $\mathbf{M}$ is a non-singular $k \times k$ matrix and two sets of variates are related by $\mathbf{y} = \mathbf{Mx}$ then D^2 for the x's is the same as D^2 for the y's, because

$$\begin{aligned}(\mathbf{y}_i-\mathbf{y}_j)'\{E(\mathbf{yy}')\}^{-1}(\mathbf{y}_i-\mathbf{y}_j) &= (\mathbf{x}_i-\mathbf{x}_j)'\mathbf{M}'\{\mathbf{M}E(\mathbf{xx}')\mathbf{M}'\}^{-1}\mathbf{M}(\mathbf{x}_i-\mathbf{x}_j)\\ &= (\mathbf{x}_i-\mathbf{x}_j)'\{E(\mathbf{xx}')\}^{-1}(\mathbf{x}_i-\mathbf{x}_j).\end{aligned}$$

Thus a non-singular transformation from one set of k variates to any other set of k variates leaves D^2 unchanged. Now suppose $\mathbf{M}$ is a $v \times k$ matrix with $v > k$, then we derive v variates $\mathbf{y}$ from the k variates $\mathbf{x}$ and $E(\mathbf{yy}') = \mathbf{M}E(\mathbf{xx}')M'$ only has rank k. We may choose a subset of k variates out of the v y variates in many ways, and at least one such subset $\mathbf{y}_1$ will have a dispersion matrix of rank k. Suppose in general that $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_r$ are all different subsets of size k with dispersion matrices of rank k: then each subset is a non-singular transformation of the x variates and must give rise to the same distance D^2 as before. It is natural to choose this value of D^2 as the distance. The value of D^2 may be computed from a generalized inverse of the dispersion matrix (Rao, 1962).

References

Bartlett, M. S. (1938). Methods of estimating mental factors. *Nature, Lond.* **141**, 609–10.

Bidwell, O. W. & Hole, F. D. (1964). An experiment in the numerical classification of Kansas soils. *Soil sci. Soc. Amer. Proc.* **28**, 263–8.

Cattell, R. B. (1965). Factor analysis: An introduction to essentials. II. The role of factor analysis in research. *Biometrics* **21**, 405–35.

Kruskal, J. B. (1964*a*). Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27.

Kruskal, J. B. (1964*b*). Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29**, 115–29.

Lawley, D. N. & Maxwell, A. E. (1963). *Factor Analysis as a Statistical Method.* London: Butterworths.

Lysenko, O. & Sneath, P. H. (1959). The use of models in bacterial classification. *J. Gen. Microbiol.* **20**, 284–90.

Rao, C. R. (1948). On the distance between two populations. *Sankhyā* **9**, 246.

Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research.* New York: John Wiley.

Rao, C. R. (1962). A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *J. R. Statist. Soc.* B, **24**, 152–8.

Rao, C. R. (1964). The use and interpretation of principal components analysis in applied research. *Sankhyā* A, **26**, 329–58.

Shepard, R. N. (1962*a*) The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* **27**, 125–39.

Shepard, R. N. (1962*b*). The analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika* **27**, 219–46.

Sokal, R. R. & Sneath, P. H. (1963). *Principles of Numerical Taxonomy.* San Francisco and London: W. H. Freeman.

Thomson, G. H. (1951). *The Factorial Analysis of Human Ability.* 5th ed. London University Press.

Torgerson, W. S. (1958). *Theory and Methods of Scaling.* New York: John Wiley.