Growth-free canonical variates and generalized inverses

J. C. GOWER

Gower, J. C. 1976 12 31: Growth-free canonical variates and generalized inverses. *Bulletin of the Geological Institutions of the University of Uppsala*, N. S., Vol. 7, pp. 1—10. Uppsala. ISSN 0302-2749.

Often interesting aspects of the differences between two populations are obscured by growth or other age effects. If there are k such effects all represented as direction cosines given as the columns of the $v \times k$ matrix $\mathbf{K}$, where v is the number of variates then, as is well known, $\mathbf{Q} = \mathbf{I} - \mathbf{K}(\mathbf{K'K})^{-1}\mathbf{K'}$ projects every sample value onto a space orthogonal to $\mathbf{K}$ and these projected values are free from the age effects. Hence, when there are p populations all with the same $\mathbf{K}$, canonical variates $\mathbf{l}$ can be found by solving

$$\mathbf{Q}(\mathbf{B} - \lambda\mathbf{W})\mathbf{Q}\mathbf{l} = \mathbf{0} \qquad \text{(I)}$$

where $\mathbf{B}$ and $\mathbf{W}$ are the between population and pooled within population dispersion matrices. It is shown that the solutions to (I) are consistent with a somewhat different formulation of the problem put forward by Burnaby (1966). This requires the solution of

$$(\mathbf{CB} - \lambda\mathbf{I})\mathbf{l} = \mathbf{0} \qquad \text{(II)}$$

where $\mathbf{C} = \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{K}(\mathbf{K'W}^{-1}\mathbf{K})^{-1}\mathbf{K'W}^{-1}$.

The approach given here allows the theory to be unified with the classical theory (i.e. when $\mathbf{Q} = \mathbf{I}$) but leads to some computational problems because $\mathbf{QWQ}$ is only of rank $v - k$ and has no ordinary inverse. Equation (I) may be solved by using any generalized inverse (in Rao's sense) of $\mathbf{QWQ}$ (of which $\mathbf{C}$ is a special case).

In many applications, it is more appropriate to replace $\mathbf{B}$ by $\mathbf{G'G}$, the *unweighted* between population dispersion matrix, where $\mathbf{G}$ is the $p \times v$ matrix of the population means. In this case, the canonical variates and their mean values can be computed by first finding the roots and vectors of the symmetric matrix $\mathbf{GCG'}$. This has certain computational and statistical advantages over other methods.

The results are illustrated by a simple example.

For these results to be of practical value, some means of estimating $\mathbf{K}$ is required. Two situations can be recognised; (a) when concomitant variables, such as age itself or variables whose values are known to be associated with age, are available; (b) when estimation must be based only on internal evidence. Two possible estimation processes for both situations are discussed.

John C. Gower, Rothamsted Experimental Station, Harpenden, Herts., United Kingdom, 1st March, 1976.

Prefatory note

The following paper is a slightly revised version of a draft written in 1967 when I was visiting the Bell Telephone Laboratories, Murray Hill, N.J. This was shortly after the use of generalised inverses, which are now commonplace in statistical work, had become widely accepted. The first part of the paper shows how Burnaby's (1966) results can be recast in terms of generalised inverses. His matrix $\mathbf{C}$ (see equation 2) is just one form of the generalised inverse of the usual within-groups dispersion matrix, corrected for linear growth effects given as the columns of a matrix $\mathbf{K}$. This use of generalised inverses unifies the theory with that of classical canonical variate analysis but is otherwise of little value, unless the elements of $\mathbf{K}$ can be estimated. The second part of the paper discusses several approaches to this estimation problem.

Rao (1966) published a theoretical paper, also to some extent inspired by Burnaby's work, but which concentrates on discrimination aspects and hence on two populations. He shows that Burnaby's form can be derived in three different ways using (a) a sufficient statistic, (b) an ancillary statistic and (c) the method of maximum likelihood ratio. These results all conform with that given by maximising the ratio of two quadratic forms, the method also used by Burnaby. Rao does not discuss estimation problems.

I have not attempted to publish this work earlier because of reservations concerning the estimation methods proposed. At the very least, I felt that the methods should be tried out on real data but none has hitherto been accessible to me. In the companion paper, Reyment & Banfield (1976) now show how the methods behave with palaeontological data. This is an ideal field of study because samples from fossil populations are likely to contain individuals at various stages of growth. They are also likely to contain individuals from the different sexes whose differences might be eliminated by associating dummy "growth" variates with the sexes.

In the years between the original preparation of this paper and the present, I know of no serious attempts to deal with these problems. Now that suitable data have become available, it seems desirable to put forward this material in a more generally available form so that readers may judge for themselves whether these methods have yet reached a stage where they are useful, and perhaps be stimulated to offer improvements.

1. Introduction

Burnaby (1966) considered the problem of finding canonical variates when the desired comparisons amongst populations are partially confounded with growth effects, or similar components, representable as gradients. With k components and v variates, the effects which are to be eliminated may be represented by a $v \times k$ matrix $\mathbf{K}$ whose r^{th} column consists of elements proportionate to the direction cosines of the r^{th} component. If there are p populations with common dispersion matrix $\mathbf{W}$ and between population matrix $\mathbf{B}$ then, using Lagrange multipliers, Burnaby found the linear combination of the original variates, whose coefficients $\mathbf{l}$ maximize the ratio of between and within sums of squares $\mathbf{l'Bl/l'Wl}$, subject to the restrictions $\mathbf{l'Wl} =$ constant and $\mathbf{K'l} = \mathbf{0}$. His solution satisfies

$$(\mathbf{CB} - \lambda\mathbf{I})\mathbf{l} = \mathbf{0} \quad (1)$$

where

$$\mathbf{C} = \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{K}(\mathbf{K'W}^{-1}\mathbf{K})^{-1}\mathbf{K'W}^{-1} \quad (2)$$

and the ratio $\mathbf{l'Bl/l'Wl} = \lambda$.

Thus λ and $\mathbf{l}$ are a latent root and vector pair of $\mathbf{CB}$, where the vectors $\mathbf{l}$ have been constrained to be orthogonal to the space spanned by the directions $\mathbf{K}$ representing growth factors.

The idempotent symmetric matrix $\mathbf{M} = \mathbf{K}(\mathbf{K'K})^{-1}\mathbf{K'}$ projects any $v \times l$ vector $\mathbf{y}$ onto the space spanned by the vector $\mathbf{K}$, and $\mathbf{Q} = \mathbf{I} - \mathbf{M}$ projects onto a space orthogonal to $\mathbf{K}$. Thus $\mathbf{y}$ can be resolved into components $\mathbf{My}$ confounded with growth factors and components $\mathbf{Qy}$ free from the effects of growth factors. To find canonical variates free from the growth effects, the intuitive statistical approach is to work in terms of the unconfounded components $\mathbf{Qy}$. This is effectively what Delany & Healy (1964) did, except that they used a regression technique to eliminate growth effects and estimated the regression coefficients from the data (see section 7.3, below). Working with the unconfounded components, the between and within population dispersion matrices are $\mathbf{QBQ}$ and $\mathbf{QWQ}$ so the canonical variate coefficients are the vectors $\mathbf{l}$ satisfying

$$\begin{aligned} (\mathbf{QBQ} - \lambda\mathbf{QWQ})\mathbf{l} &= \mathbf{0} \\ \text{i.e. } \mathbf{Q}(\mathbf{B} - \lambda\mathbf{W})\mathbf{Ql} &= \mathbf{0}. \end{aligned} \quad (3)$$

Below, a simple relation between the solutions of (1) and (3) is demonstrated which allows the special theory for the elimination of growth effects to be absorbed in the usual theory, provided generalized inverses of matrices are freely used. Then some results I obtained previously (Gower 1966 a, b) are extended to fit into this more general theory and used as a basis for the calculation of D^2 and canonical variate values. Finally, four methods are discussed that might be used for estimating $\mathbf{K}$ from data.

2. Equivalence of the two equations

We require the following algebraic identities

$$\mathbf{QK} = \mathbf{K'Q} = \mathbf{0} \quad (4)$$

$$\mathbf{QC} = \mathbf{CQ} = \mathbf{C} \quad (5)$$

$$\mathbf{CWQ} = \mathbf{QWC} = \mathbf{Q}. \quad (6)$$

(4) follows directly from the definition of $\mathbf{Q}$ and (6) follows from (4) and (2); (5) follows from the multiplication of $\mathbf{Q}$ and $\mathbf{C}$ and some algebraic manipulation.

Pre-multiplying (3) by $\mathbf{C}$ and using (5) and (6) gives:

$$(\mathbf{CB} - \lambda\mathbf{I})\mathbf{Ql} = \mathbf{0}. \quad (7)$$

This is of the same form as (1) but with $\mathbf{Ql}$ in place of $\mathbf{l}$ showing that if λ and $\mathbf{l}$ are a root and vector pair of (3), then λ, $\mathbf{Ql}$ are a corresponding pair of (1). Using (4), we have $\mathbf{K'}(\mathbf{Ql}) = \mathbf{0}$, so whatever the vectors of (3), $\mathbf{l}$ must be orthogonal to $\mathbf{K}$. This implies that the vectors of (3) may have components in the space spanned by $\mathbf{K}$ but these components can be removed by pre-multiplying by $\mathbf{Q}$. In fact, because $\mathbf{QM}=\mathbf{0}$, it is easily seen that the vectors of (3) may have arbitrary components, for if $\mathbf{l}$ is a solution so is $\mathbf{l}+\mathbf{Mm}$, where $\mathbf{m}$ is an arbitrary $v \times l$ vector.

The above remarks are very closely related to problems arising in the solution of sets of linear equations $\mathbf{Al}=\mathbf{0}$ when $\mathbf{A}$ is not of full rank. Rao (1962) discusses these problems and defines a generalized inverse (or g-inverse) of an $m \times n$ matrix $\mathbf{A}$ as an $n \times m$ matrix $\mathbf{A}^-$, such that for any $\mathbf{y}$ for which $\mathbf{Ax}=\mathbf{y}$ is consistent, $\mathbf{x}=\mathbf{A}^-\mathbf{y}$ is a solution. Rao establishes many properties of g-inverses amongst which he shows that if $\mathbf{A}^-$ is a g-inverse then $\mathbf{AA}^-\mathbf{A}=\mathbf{A}$, and conversely. We note that $\mathbf{C}$ is a g-inverse of $\mathbf{QWQ}$ and vice versa, for using (5) and (6) repeatedly gives:

$$(\mathbf{QWQ})\mathbf{C}(\mathbf{QWQ} = \mathbf{QW}[\mathbf{QCQ}]\mathbf{WQ} = \mathbf{QW}[\mathbf{CWQ}] = \mathbf{QWQ}$$

and

$$\mathbf{C}(\mathbf{QWQ})\mathbf{C} = [\mathbf{CQ}]\mathbf{WQC} = [\mathbf{CWQ}]\mathbf{C} = \mathbf{QC} = \mathbf{C}.$$

3. Interpretation in terms of D^2

In this section, we require the most general form of g-inverse of a matrix $\mathbf{A}$, given a particular g-inverse $\mathbf{A}^-$. Rao (1962) shows that if $\mathbf{ax}=\mathbf{y}$ admits a solution $\mathbf{x}=\mathbf{A}^-\mathbf{y}$ for all $\mathbf{y}$ consistent with the equations, then all solutions can be expressed in the form $\mathbf{x}=\mathbf{A}^-\mathbf{y}+(\mathbf{I}-\mathbf{H})\mathbf{z}$ where $\mathbf{H}=\mathbf{A}^-\mathbf{A}$ and $\mathbf{z}$ is arbitrary. If $\mathbf{R}$ is a g-inverse of $\mathbf{A}$, then $\mathbf{ARA}=\mathbf{A}$, showing that $\mathbf{x}=\mathbf{RA}$, $\mathbf{y}=\mathbf{A}$ are consistent solutions of $\mathbf{Ax}=\mathbf{y}$, and hence:

$\mathbf{RA} = \mathbf{A}^-\mathbf{A}+(\mathbf{I}-\mathbf{H})\mathbf{Z}_1$ where $\mathbf{Z}_1$ is an arbitrary matrix.

Similarly $\mathbf{AR}=\mathbf{AA}^-+\mathbf{Z}_2(\mathbf{I}-\mathbf{J})$ where $\mathbf{J}=\mathbf{AA}^-$ and $\mathbf{Z}_2$ is arbitrary.
This last equation may be written

$$\mathbf{AR}=\mathbf{I}+(\mathbf{Z}_2-\mathbf{I})(\mathbf{I}-\mathbf{J}).$$

Repeating the argument with the columns of $\mathbf{R}$ as successive variables gives:

$$\mathbf{R}=\mathbf{A}^-+\mathbf{A}^-(\mathbf{Z}_2-\mathbf{I})(\mathbf{I}-\mathbf{J})+(\mathbf{I}-\mathbf{H})\mathbf{Z}_3.$$

Thus

$$\mathbf{R}=\mathbf{A}^-+\mathbf{U}(\mathbf{I}-\mathbf{J})+(\mathbf{I}-\mathbf{H})\mathbf{V} \qquad (8)$$

where $\mathbf{U}$ and $\mathbf{V}$ are arbitrary is a general expression containing all g-inverses of $\mathbf{A}$[1]. It can be verified directly that $\mathbf{ARA}=\mathbf{A}$ for all $\mathbf{U}$ and $\mathbf{V}$.

We have shown above the $\mathbf{C}$ is a g-inverse of $\mathbf{QWQ}$, and conversely. In this case, $\mathbf{H}=\mathbf{CQWQ}=\mathbf{Q}$, and $\mathbf{J}=\mathbf{QWQC}=\mathbf{Q}$, by equations (5) and (6). Thus, the general g-inverse is of the form

$$(\mathbf{QWQ})^- = \mathbf{C}+\mathbf{UM}+\mathbf{MV}, \qquad (9)$$

and Burnaby's special case is obtained by setting $\mathbf{U}$ and $\mathbf{V}$ to zero. Simple algebraic manipulation establishes that:

$$\text{Trace}\,(\mathbf{C}+\mathbf{UM}+\mathbf{MV})'(\mathbf{C}+\mathbf{UM}+\mathbf{MV}) = \text{Trace}\,(\mathbf{C}'\mathbf{C})+\text{Trace}\,(\mathbf{UM}+\mathbf{MV})'(\mathbf{UM}+\mathbf{MV})$$

showing that Burnaby's inverse has the smallest L_2-norm amongst all generalized inverses of $\mathbf{QWQ}$; this suggests that $\mathbf{C}$ has good properties for computation.

Suppose the mean values for p populations are given in a $p \times v$ matrix $\mathbf{G}$ whose i^{th} row gives the mean $\mathbf{g}_i$ of the i^{th} population, then $\mathbf{GQ}$ is the projection of the means onto the $\mathbf{Q}$-space. Following the appendix to Gower (1966 a), the Mahalanobis D^2 distance between the projections of the i^{th} and j^{th} populations onto the $\mathbf{Q}$-space is:

$$D_Q^2 = (\mathbf{g}_i-\mathbf{g}_j)\mathbf{Q}(\mathbf{QWQ})^-\mathbf{Q}(\mathbf{g}_i-\mathbf{g}_j)' = (\mathbf{g}_i-\mathbf{g}_j)\mathbf{C}(\mathbf{g}_i-\mathbf{g}_j)'. \qquad (10)$$

This is identical to the form proposed by Burnaby (1966) but we have shown here that the value of D_Q^2 does not depend on the particular choice of the g-inverse of $\mathbf{QWQ}$.

As might be expected, this result can be put into more general terms, for if $\mathbf{W}_G^- = \mathbf{W}^- + (\mathbf{I}-\mathbf{H})\mathbf{L}+\mathbf{M}(\mathbf{I}-\mathbf{J})$ is a general g-inverse of $\mathbf{W}$ then for $\boldsymbol{\delta}'\mathbf{W}_G^-\boldsymbol{\delta}$ (where $\boldsymbol{\delta}'=\mathbf{g}_i-\mathbf{g}_j$) to be invariant we must have:

$$\boldsymbol{\delta}'(\mathbf{I}-\mathbf{H}) = (\mathbf{I}-\mathbf{J})\boldsymbol{\delta}=\mathbf{0}.$$

If the symmetric inverse $\mathbf{C}$ is used, the second condition is merely the transpose of the first, so necessary and sufficient conditions for the uniqueness of D^2 are that

$$\boldsymbol{\delta}=\mathbf{WC}\ . \qquad (11)$$

When $\mathbf{W}$ is not of full rank, there will be

[1] Rao (1967) gives another form of general g-inverse:

$$\mathbf{R}_r = \mathbf{A}^- + \mathbf{W}-\mathbf{HWJ}$$

which uses only one arbitrary matrix $\mathbf{W}$, and may therefore seem to conflict with (8). However setting $\mathbf{W}=\mathbf{U}(\mathbf{I}-\mathbf{J})+(\mathbf{I}-\mathbf{H})\mathbf{V}$ shows that $\mathbf{R}_r$ contains the solution of form $\mathbf{R}$. Also setting $\mathbf{U}=\mathbf{W}$, $\mathbf{V}=\mathbf{WJ}$ shows that $\mathbf{R}$ contains the solutions of form $\mathbf{R}_r$. Therefore, both forms are equivalent.

linear constraints $\mathbf{WK}=\mathbf{0}$ on its rows and therefore $\mathbf{k}=(\mathbf{I}-\mathbf{H})\mathbf{z}$ ($\mathbf{z}$ arbitrary). Hence, $\mathbf{k}'\boldsymbol{\delta}=\mathbf{z}'(\boldsymbol{\delta}-\mathbf{H}'\boldsymbol{\delta})$ and, therefore, if $\boldsymbol{\delta}$ is to be chosen so that $\boldsymbol{\delta}'\mathbf{W}^{-}\boldsymbol{\delta}$ is unique, we must have by (11), $\mathbf{k}'\boldsymbol{\delta}=\mathbf{0}$. This condition is satisfied when, as is usual, the mean values are the sums of variate values which are constrained in the same way as the rows of $\mathbf{W}$. The columns of $\mathbf{K}$ specify a set of linear constraints on corrected variate values $\mathbf{XQ}$, which are of the required form because $\mathbf{K}'\mathbf{Q}=\mathbf{0}$.

Burnaby defines $D_M^2=D^2-D_Q^2$. A causal reading of his paper may suggest that D^2 has been resolved into orthogonal components D_Q^2 and D_M^2 but this is not so; the additivity merely reflects the definition of D_M^2 which is interpreted as the amount of distance lost through working only in the $\mathbf{Q}$-space.

If $\mathbf{R}$ is any g-inverse of $\mathbf{QWQ}$, and $\mathbf{l}$ is any vector satisfying $(\mathbf{RQBQ}-\lambda\mathbf{I})\mathbf{l}=\mathbf{0}$, then it is easy to see (by substituting (9) for $\mathbf{R}$ and then pre-multiplying by $\mathbf{Q}$) that $\mathbf{Ql}$ also satisfies (7). Thus if computer programs are available for computing g-inverses, it is not necessary to use Burnaby's particular g-inverse $\mathbf{C}$, although this is a convenient analytical form. In fact, g-inverses can be conveniently computed by using most of the standard matrix inversion algorithms, modified so that any prospective division by zero is ignored. Programs modified in this way will still provide the regular inverse of matrices of full rank. If $\mathbf{C}$ is required, it can always b computed from $\mathbf{R}$ by evaluating $\mathbf{QRQ}$.

4. Reference of means to canonical axes

Gower (1966 b) pointed out that when canonical variates are used for descriptive purposes, the equation $(\mathbf{G}'\mathbf{G}-\lambda\mathbf{W})\mathbf{l}=\mathbf{0}$ should be solved rather than $(\mathbf{B}-\lambda\mathbf{W})\mathbf{l}=\mathbf{0}$. The columns of $\mathbf{G}$ are as sumed to be measured from an origin representing the unweighted overall mean of the population means (i.e., the column sums of $\mathbf{G}$ are zero and its rank is $\mathrm{Min}(p-1,v)$). Thus, $\mathbf{G}'\mathbf{G}$ is the matrix of unweighted sums of squares and products between populations, whereas $\mathbf{B}$ is the corresponding weighted sums of squares and products matrix. The reason for preferring the unweighted matrix will become clear below.

Replacing $\mathbf{B}$ by $\mathbf{G}'\mathbf{G}$ in (3) gives

$$\mathbf{Q}(\mathbf{G}'\mathbf{G}-\lambda\mathbf{W})\mathbf{Ql}=\mathbf{0}. \qquad (12)$$

which has solution vectors $\mathbf{X}$ and roots $\mathbf{L}$ (say) so that

$$\mathbf{QG}'\mathbf{GQX}=\mathbf{QWQXL} \qquad (13)$$

and the means $\mathbf{P}$ of the canonical variates are

$$\mathbf{GQX}=\mathbf{P}. \qquad (14)$$

The rows of $\mathbf{P}$ may be regarded as the coordinates of the means referred to canonical variate axes. These coordinates may be found by solving $(\mathbf{CG}'\mathbf{G}-\lambda\mathbf{I})\mathbf{l}=\mathbf{0}$ to give $\mathbf{L}$ and $\mathbf{X}$ and then substituting in (14). If the vectors are scaled such that

$$\mathbf{XQQ}'\mathbf{Q}=\mathbf{C}, \qquad (15)$$

then the squared distance between a pair of means is the corresponding value of D_Q^2. Gower (1966 a) showed that the canonical axes have the property that the vectors corresponding to the first r roots are the r principal components of a set of p points, representing the population means, whose $(1/2)p(p-1)$ Euclidean distances are the values of D^2 but that this is not true if $\mathbf{B}$ is used in place of $\mathbf{G}'\mathbf{G}$.

A shorter more direct proof of this property is given below, adapted to the situation where growth effects are to be eliminated.

Pre-multipliyng (14) by $\mathbf{GCG}'$ gives

$$\begin{aligned}\mathbf{GCG}'\mathbf{P} &= \mathbf{GCG}'\mathbf{GQX} \\ &= \mathbf{GC}(\mathbf{QG}'\mathbf{GQX}) && [\text{by } (5)] \\ &= \mathbf{GC}(\mathbf{QWQXL}) && [\text{by } (13)] \\ &= \mathbf{GQXL} && [\text{by } (5) \text{ and } (6)] \\ &= \mathbf{PL} && [\text{by } (14)] \qquad (16)\end{aligned}$$

Thus $\mathbf{P}$ are the latent vectors of $\mathbf{GCG}'=\mathbf{T}$ (say) and $\mathbf{L}$ are the corresponding latent roots. $\mathbf{T}$ is a $p\times p$ matrix, $\mathbf{P}$ is $p\times v$ and $\mathbf{L}$ is a diagonal $v\times v$ matrix. The ranks of $\mathbf{G}$ and $\mathbf{C}$ are $\mathrm{Min}(p-1,v)$ and $(v-k)$, respectively, (unless there are so few sample values that the ranks are further reduced). Thus there are $\mathrm{Min}(p-1,v-k)$ non-zero latent roots appearing in the diagonal elements of $\mathbf{L}$. Because the column sums of $\mathbf{G}$ are zero, the row (and column) sums of $\mathbf{T}$ are all zero and therefore $\mathbf{T}$ has a least one zero root, with corresponding unit latent vector. The remaining columns of $\mathbf{P}$ being orthogonal to this unit vector must sum to zero. Scaling the vectors so that

$$\mathbf{P}'\mathbf{P}=\mathbf{L} \qquad (17)$$

therefore ensures (a) that the rows of $\mathbf{P}$ are the coordinates of points referred to principal axes and (b) that

$$\mathbf{I}=\mathbf{PP}'. \qquad (18)$$

The squared distances between the i^{th} and j^{th} points are therefore

$$t_{ii}+t_{jj}-2t_{ij}. \tag{19}$$

Substituting $\mathbf{g}_i\mathbf{C}\mathbf{g}_j'$ for t_{ij} in (19) shows the squared distances are D_Q^2 as required. The scaling (17) ensures that the coordinates on an axis corresponding to a zero root are all zero so they need never be computed.

It remains to show that the two sets of scaling given by (15) and (17) are consistent. Pre- and post-multiplying (15) by $\mathbf{G}$ and $\mathbf{G}'$ and using (14) gives

$$\mathbf{PP}'=\mathbf{GCG}'=\mathbf{T}$$

as in (18). Thus the two sets of scaling are equivalent.

5. An alternative method of calculating the canonical variates

The above results are not only of academic interest for they provide an alternative method of calculation to that given by Burnaby (1966). Most methods of solving (1) aim at providing a symmetric matrix for the latent root and vector process. This has the computational advantage that a symmetric and positive semi-definite matrix, as in the present and most other cases in multivariate statistics, has non-negative roots and real vectors. This property saves special programming to cope with arithmetic operations on complex numbers and therefore allows more efficient algorithms to be used.

$\mathbf{T}$ is already in the desirable symmetric form and can be readily computed by the standard processes of matrix inversion, multiplication and subtraction. The vectors of $\mathbf{T}$, scaled as in (17), immediately give the values of the means referred to canonical axes and we now show that vectors $\mathbf{QX}$ satisfying (12) are also easy to compute from a knowledge of the roots and vectors of $\mathbf{T}$. Multiplying (14) by $\mathbf{QG}'$ and using (13) gives

$$\mathbf{QG'P}=\mathbf{QWQXL}. \tag{20}$$

Pre-multiplying (20) by $\mathbf{C}$ and using (5) and (6) gives

$$\mathbf{CG'P}=\mathbf{QXL}.$$

Hence

$$\mathbf{QX}=\mathbf{CG'PL}^{-} \tag{21}$$

where $\mathbf{L}^{-}$ is the g-inverse of $\mathbf{L}$ obtained by inverting all non-zero (diagonal) elements of $\mathbf{L}$. Thus (21) enables the required projections of $\mathbf{X}$ onto the $\mathbf{Q}$-space to be calculated.

Other advantages of using $\mathbf{T}$ are (i) that by using (19) it gives a simple direct way of evaluating D_Q^2 which requires hardly any special computation and (ii) that often $\mathbf{T}$ is a smaller matrix than $\mathbf{C}$ and $\mathbf{B}$.

When $\mathbf{T}$ is used, the estimated values of D^2 may be corrected for bias. Gower (1966 b) outlines how this can be done when $\mathbf{C}=\mathbf{W}^{-1}$. In the present case, the formulae for bias given by Rao (1952, p. 364) have to be slightly modified. When all variates are observed in each population, the bias in D_{ij}^2 is approximately $(1/n_i - 1/n_j)$ Trace ($\mathbf{CQWQ}$), where n_i is the sample size of population i. We have

$$\text{Trace}\,(\mathbf{CQWQ})=\text{Trace}\,(\mathbf{CWQ})=\text{Trace}\,(\mathbf{Q}) \\ =v-k.$$

The first two steps above follow from (5) and (6) and the last because $\mathbf{Q}$ is idempotent with rank $(v-k)$. Thus, to correct for bias, subtract $(v-k)(1/n_i+1/n_j)$ from D_{ij}^2. The simplest way of effecting this correction is to subtract $(v-k)/n_i$ from the i^{th} diagonal element of $\mathbf{T}$ and then evaluate a new matrix $\mathbf{T}^*$ with elements $t_{ij}-t_{i.}-t_{.j}+t_{..}$ $\mathbf{T}^*$ will have rank $(p-1)$, even though $\mathbf{T}$ itself may have lower rank, and it may not be positive semi-definite.

6. Example

The calculations are illustrated using Burnaby's (1966) dummy example in which there are three variates $(v=3)$, two populations $(p=2)$, and two sets of constraints $(k=2)$. We are given

$$\mathbf{W}=\begin{pmatrix}1&0&2\\0&4&2\\2&2&6\end{pmatrix}$$

$$\mathbf{G}=\begin{pmatrix}1\tfrac{1}{2}&2&1\\-1\tfrac{1}{2}&-2&-1\end{pmatrix}$$

$$\mathbf{K}'=\begin{pmatrix}1&4&3\\1&2&2\end{pmatrix}.$$

In the above, the origin of the general mean has been shifted so that the column sums of $\mathbf{G}$ are zero.

We can compute

$$\mathbf{C}=1/8\begin{pmatrix}4&2&-4\\2&1&-2\\-4&-2&4\end{pmatrix}$$

which has rank $(v-k)=1$.

Hence

$$\mathbf{T} = \mathbf{GCG}' = 9/8 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

The latent roots of $\mathbf{T}$ are 9/4 and 0 and the vectors scaled so that their sums of squares are equal to these roots are the columns of

$$\mathbf{P} = \begin{pmatrix} 3/2\sqrt{2} & 0 \\ -3/2\sqrt{2} & 0 \end{pmatrix}.$$

The rows of $\mathbf{P}$ are the coordinates of the means referred to canonical variate axes (in this case there is only one non-null axis). The distance between the two points represented by $\mathbf{P}$ is $D_0^2 = [3/2\sqrt{2}) - (-3/\sqrt{2}))]^2 = 4\frac{1}{2}$, as obtained by Burnaby; the same value can be obtained from $\mathbf{T}$ because $D_0^2 = t_{11} + t_{22} - 2t_{12} = 4\frac{1}{2}$. The canonical variate loading coefficients are obtained by calculating $\mathbf{QX} = \mathbf{CG'PL^-}$. We have shown that $\mathbf{L} = \text{diag}\ (9/4,0)$ and therefore $\mathbf{L^-} = \text{diag}\ (4/9,0)$. In fact, only the first column of $\mathbf{L^-}$ is relevant because zero roots correspond to dimensions in which there is no variation and may therefore be ignored. The computed value of $(\mathbf{CG'PL})'$ is $(2,1,-2)/(2\sqrt{2})$. This value is proportional to that obtained by Burnaby which is sufficient for direction cosine purposes, but the present value gives the correct coordinate values for D^2 purposes at can be verified by calculating $\mathbf{GXQ}$. This agreement is rather fortuitous as Burnaby has implicitly defined the between population sums of squares and products matrix $\mathbf{B}$ by $\mathbf{G'G}$; if he had introduced unequal population sizes and used the usual weighted sums of squares and products, the two results would have differed. The extent of this difference would depend on the disparity of the two sample sizes.

A final check on the calculations can be obtained by verifying that the scaling agrees with eqpation (15), i.e., $(\mathbf{QX})(\mathbf{QX})' = \mathbf{C}$.

7. Estimation of K

So far we have assumed that the same values $\mathbf{K}$ apply to all populations. In practice this is not likely to be a reasonable assumption and it would usually be wise to examine each population separately before estimating $\mathbf{K}$ from the pooled populations. Significance tests require to be developed to help judge whether two matrices $\mathbf{K}_i$ and $\mathbf{K}_j$ extracted by the methods given below from populations i and j may be regarded as spanning the same space; note that the individual elements of $\mathbf{K}_i$ and $\mathbf{K}_j$ need not agree.

If separate correction matrices are applicable to each population, but the corrected dispersion matrices are homogeneous, we should evaluate $\sum_{i=1}^{p} \mathbf{Q}_i\mathbf{W}_i\mathbf{Q}_i$ as the pooled within-population dispersion matrix after correcting for growth effects. The population means $\mathbf{G}_i$ of the i^{th} population would be similarly converted to $\mathbf{G}_i\mathbf{Q}_i$ and used to evaluate a corrected between-population dispersion matrix, using a weighted or unweighted version, as was felt appropriate. With this approach, it is unlikely that $\sum_{i=1}^{p} \mathbf{Q}_i\mathbf{W}_i\mathbf{Q}_i$ would be singular, so no special problems of the kind discussed above would occur.

In the remainder of this section, the estimation of $\mathbf{K}$ is discussed. The procedures may be regarded as appropriate to a single population or, when matrices are pooled over populations, as appropriate to the combined populations.

Two distinct situations occur. Either the matrix $\mathbf{K}$ has to be estimated from the data on the original set of v variates (*internal estimation*) or, rather in the manner of covariance analysis, data on concomitant variables are available (*external estimation*). For example, the age of each sample or the value of a variate highly correlated with age might be recorded or in botanical problems, pH and moisture in the soil around each plant and distance from shelter might be recorded as concomitant variables for external estimation. The variates and concomitant variables might need transformation to approximate the linear relationships assumed here.

7.1. *Internal estimation by principal components*

The results of Jolicoeur (1963) suggest that $\mathbf{K}$ can be estimated as the first k latent vectors of $\mathbf{W}$. Associated with this method is the difficulty that the principal components depend on the scales of measurement of the different variates. This problem disappears when all scales are the same or if normalized variates are used. Jolicoeur (1963) points out that in biological problems, the logarithms of the variates are usually related linearly, and that taking logarithms is one way in which the principal components can be made scale-free. Under these circumstances, the proposed method of estimation seems intuitively reasonable, provided that the growth effects are the major source of variation within each population. In this case, the elliptical cloud of points representing within-population samples would be spread out around an elongated major axis (the direction of the principal component).

If this method of estimation is used, the matrix $\mathbf{C}_p$ (corresponding to $\mathbf{C}$) can be simplified. Suppose the columns of $\mathbf{K}$ are the first k latent vectors of $\mathbf{W}$ then

$$\mathbf{WK} = \mathbf{KL}_k$$

where $\mathbf{L}_k$ is the diagonal matrix containing the first k latent roots of $\mathbf{W}$. We can assume the vectors are scaled so that $\mathbf{K'K} = \mathbf{I}_k$ hence

$$\mathbf{K'W}^{-1}\mathbf{K} = \mathbf{L}_k^{-1}$$

and therefore

$$\begin{aligned}\mathbf{C}_p &= \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{KL}_k\mathbf{K'W}^{-1}\\ &= \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{WKK'W}^{-1}\\ &= (\mathbf{I} - \mathbf{KK'})\mathbf{W}^{-1}\\ &= \mathbf{QW}^{-1}.\end{aligned}$$

To emphasise the symmetric nature of $\mathbf{C}_p$, notice that

$$\mathbf{C}_p = \mathbf{QW}^{-1} = \mathbf{W}^{-1}\mathbf{Q}$$

and therefore

$$\mathbf{QW}^{-1}\mathbf{Q} = \mathbf{Q}(\mathbf{QW}^{-1}) = \mathbf{QW}^{-1} = \mathbf{C}_p.$$

Thus we may write

$$\mathbf{C}_p = \mathbf{QW}^{-1} = \mathbf{W}^{-1}\mathbf{Q} = \mathbf{QW}^{-1}\mathbf{Q}. \quad (22)$$

This is a particularly simple form of g-inverse. Several methods of computing its value suggest themselves:

(i) Compute $\mathbf{QW}^{-1}$.

(ii) $\mathbf{QW}^{-1}\mathbf{Q} = \mathbf{W}^{-1} - \mathbf{KL}_n^{-1}\mathbf{K'}$

(iii) If all the vectors of $\mathbf{W}$ have been computed and those remaining after $\mathbf{K}$ have been removed are $\mathbf{K}_{v-k}$, with associated roots $\mathbf{L}_{v-k}$ use $\mathbf{QW}^{-1}\mathbf{Q} = \mathbf{K}_{v-k}\mathbf{L}_{v-k}^{-1}\mathbf{K}'_{v-k}$

(iv) Compute $\mathbf{R}$, any g-inverse of $\mathbf{QWQ}$, then $\mathbf{QW}^{-1}\mathbf{Q} = \mathbf{QRQ}$.

Method (iv) is as suggested before, and has the advantage that it is independent of the form of $\mathbf{Q}$. One of the other methods may be useful in investigations when only the principal component estimation of $\mathbf{K}$ is of interest. We also have from (22)

$$\mathbf{QW}^{-1}\mathbf{M} = \mathbf{MW}^{-1}\mathbf{Q} = \mathbf{0}. \quad (23)$$

The results (22) and (23) can be used to show that in this instance, D^2 may be resolved into components representing growth effects and components orthogonal to growth. We consider

$$\begin{aligned}\mathbf{GW}^{-1}\mathbf{G'} &= \mathbf{G}(\mathbf{Q}+\mathbf{M})\mathbf{W}^{-1}(\mathbf{Q}+\mathbf{M})\mathbf{G'}\\ &= \mathbf{G}(\mathbf{QW}^{-1}\mathbf{Q})\mathbf{G'} + \mathbf{G}(\mathbf{MW}^{-1}\mathbf{M})\mathbf{G'}.\end{aligned} \quad (24)$$

Using equation (19) shows that

$$D^2 = D_Q^2 + D_M^2.$$

7.2. Internal estimation by factor analysis

An alternative method of estimation suggested by Tessier (1955), and recently investigated further by Hopkins (1966), is to take $\mathbf{K}$ as the factor loadings associated with the factor analysis of $\mathbf{W}$ with k factors. The rationale behind this method is that if growth is assumed linear for an individual, then the value x_i of the i^{th} variate can be written

$$x_i - \bar{x}_i = \lambda l_i. \quad (25)$$

Here l_i is the cosine of the angle between the direction of growth and the i^{th} variate, and $\bar{x}_i$ is some arbitrary reference point, here taken as the mean value of the i^{th} variate in the whole population; λ represents the distance along the line of growth from the reference point. Each member of the population will have its own reference point and the variation of these reference points with respect to an origin can be represented by an additional term in (25) to give

$$x_i - \bar{x}_i = \lambda l_i + e_i.$$

When there are k age effects, this equation becomes

$$x_i - \bar{x}_i = \sum_{j=1}^{k} l_{ij}\lambda_j + e_i. \quad (26)$$

This has the general form of the fundamental equation of factor analysis and may be written in vector notation as

$$\mathbf{x} = \mathbf{L}\boldsymbol{\lambda} + \mathbf{e}$$

where $\mathbf{L}$ is the $v \times k$ matrix of factor loadings.

The values of λ_j will vary in the population (representing the age structure) and correspond to factors in the factor-analytical model. In allometric problems, it is the values of l_i (the factor loadings) which require estimation, but in the case under discussion we only require to eliminate these effects from the e_i (the specific factors). In factor analysis, it is usual to assume that the factors λ_j

and the e_i are independent and this seems a reasonable assumption here.

However, in factor analysis, the different e_i are assumed independently distributed with diagonal dispersion matrix $\mathbf{V}$. This assumption is unjustified here as it amounts to assuming that after eliminating age effects, all variates are independent of each other. With $\mathbf{V}$ not diagonal, the maximum liklihood estimates of $\mathbf{L}$ and $\mathbf{V}$ (on the assumption that the x_i are distributed in multivariate normal form) become degenerate, admitting any $v \times k$ matrix $\mathbf{L}$ of the rank k and $\mathbf{V}=\mathbf{W}-\mathbf{LL}'$ as solutions. This solution includes the principal component case where the elements of $\mathbf{L}$ are taken as the first k latent vectors of $\mathbf{W}$.

The above suggests that even in allometry it may not be sufficient to estimate $\mathbf{L}$, assuming the specific factors are independent and that $k=1$, as this may involve making the unrealistic assumption that all variates are independently distributed after eliminating the allometric effect. It might be better to fit as many factors as seem to be consistent with the data, presumably being guided by appropriate significance tests (see, for example, Lawley & Maxwell (1971 p. 34). The first factor so found can be taken to estimate allometric effects and the others describe the correlational structure after eliminating growth effects; the variances amongst the specific factors will be additional. Thus, as in the principal component method, the maximum variation in the population is attributed to the allometric effects.

The maximum likelihood estimate of the dispersion matrix under these circumstances is

$$\hat{\mathbf{W}}=\mathbf{KK}'+\mathbf{LL}'+\mathbf{V}. \qquad (28)$$

and not the sample dispersion matrix $\mathbf{W}$. Here $\mathbf{K}$ gives the loadings for the first k factors which are to be associated with growth effects and $\mathbf{L}$ gives the loadings for the remaining factors. We have

$$\mathbf{Q}\hat{\mathbf{W}}\mathbf{Q}=\mathbf{LL}'+\mathbf{QVQ}. \qquad (29)$$

The matrix $\mathbf{C}_f$ corresponding to $\mathbf{C}$ can again be somewhat simplified. If the factor loadings satisfy the maximum likelihood equations then (see, for example, Lawley & Maxwell (1971))

$$(\mathbf{W}-\mathbf{V})\mathbf{V}^{-1}(\mathbf{K,L})=(\mathbf{K,L})\mathbf{J} \qquad (30)$$

where

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

To obtain a unique solution, $\mathbf{J}$ is restricted to being diagonal (i.e., $(\mathbf{K,L})$ are latent vectors of (30) and so the diagonal elements of $\mathbf{J}$ are the latent roots). We require to evaluate $\mathbf{C}_f$, and first evaluate $(\mathbf{K}'\hat{\mathbf{W}}^{-1}\mathbf{K})$ which is the same as $\mathbf{K}'\mathbf{W}^{-1}\mathbf{K}$ (see, Lawley & Maxwell (1971 p. 27)). Writing $\mathbf{J}_k=\mathbf{K}'\mathbf{V}^{-1}\mathbf{K}$ and $\mathbf{J}_l=\mathbf{L}'\mathbf{V}^{-}\ \mathbf{L}$, (30) gives:

$$\mathbf{WV}^{-1}\mathbf{K}=\mathbf{K}(\mathbf{I}+\mathbf{J}_k)$$

and hence

$$\begin{aligned}\mathbf{K}'\mathbf{W}^{-1}\mathbf{K}&=\mathbf{K}'\mathbf{V}^{-1}\mathbf{K}(\mathbf{I}+\mathbf{J}_k)^{-1}\\&=(\mathbf{I}+\mathbf{J}_k^{-1})^{-1}.\end{aligned} \qquad (31)$$

Therefore

$$\begin{aligned}\mathbf{C}_f&=\hat{\mathbf{W}}^{-1}-\mathbf{W}^{-1}\mathbf{K}(\mathbf{I}+\mathbf{J}_k^{-1})\mathbf{K}'\mathbf{W}^{-1}\\&=\hat{\mathbf{W}}^{-1}-\mathbf{V}^{-1}\mathbf{K}(\mathbf{I}+\mathbf{J}_k)^{-1}(\mathbf{I}+\mathbf{J}_k^{-1})(\mathbf{I}+\mathbf{J}_k)^{-1}\mathbf{K}'\mathbf{V}^{-1}\\&=\hat{\mathbf{W}}^{-1}-\mathbf{V}^{-1}\mathbf{K}\mathbf{J}_k^{-1}(\mathbf{I}+\mathbf{J}_k)^{-1}\mathbf{K}'\mathbf{V}^{-1}.\end{aligned} \qquad (32)$$

Now it can be verified from direct multiplication by (28) that

$$\begin{aligned}\hat{\mathbf{W}}^{-1}=\mathbf{V}^{-1}&-\mathbf{V}^{-1}\mathbf{K}(\mathbf{I}+\mathbf{J}_k)^{-1}\mathbf{K}'\mathbf{V}^{-1}\\&-\mathbf{V}^{-1}\mathbf{L}(\mathbf{I}+\mathbf{J}_l)^{-1}\mathbf{L}'\mathbf{V}^{-1}.\end{aligned}$$

Substituting into (32) gives

$$\mathbf{C}_f=\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{K}\mathbf{J}_k^{-1}\mathbf{K}'\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{L}(\mathbf{I}+\mathbf{J}_l)^{-1}\mathbf{L}'\mathbf{V}^{-1} \qquad (33)$$

which in somewhat simpler to compute than the form (32) which involves $\hat{\mathbf{W}}$, because the only matrices to be inverted are diagonal.

7.3. *External estimation by regression*

Suppose k concomitant variables have been observed and are represented by the $n \times k$ matrix $\mathbf{X}$ (n is the sample size). The corresponding data on the other variates are represented by the $n \times v$ matrix $\mathbf{Y}$. The observed variance and covariance matrices between the two sets of variates may be written in partitioned form as

$$\begin{array}{cc} & \begin{array}{cc} k & v \end{array} \\ \begin{array}{c} k \\ v \end{array} & \begin{pmatrix} \mathbf{V} & \mathbf{U} \\ \mathbf{U}' & \mathbf{W} \end{pmatrix}. \end{array} \qquad (34)$$

Ordinary multiple regression techniques estimate the residuals after eliminating the regression of $\mathbf{Y}$ on the concomitant variables as

$$(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y} \qquad (35)$$

which may be written as $\mathbf{Q}_1\mathbf{Y}$.

Although $\mathbf{Q}_1$ is idempotent and of similar form to the matrix $\mathbf{Q}$, previously discussed, the situation differs because the dispersion matrix amongst the residuals is

$$\mathbf{Y'Q_1Y} \tag{36}$$

which is, in general, nonsingular; its rank is Min $(v,n-k-1)$. Thus problems arising from the singularity of **QWQ** do not occur. This was the method used by Delany & Healy (1964), in their study of the Long Tailed Field Mouse, with one concomitant variable, a measure of tooth-wear, thought to be a good indicator of age.

A point of computational interest is that (36) may be written $(\mathbf{W}-\mathbf{U'V}^{-1}\mathbf{U})$ whose inverse occurs as the lower right-hand corner matrix in the inversion of (34). This inverse matrix is required when computing D^2 values or when using the method given in Section 5.

7.4. *External estimation by canonical correlation*

Suppose there is one concomitant variable x and two y-variates (y_1 and y_2). If it were noticed that $x=y_1+y_2$ for all samples it would be normal to work in the direction orthogonal to y_1+y_2. Thus, we are lead to ask which linear combination of the y's best predicts x. When there are several concomitant variables, we look for the set of linear combinations of the y's which best predicts the space spanned by the x's or, what is the same thing, predicts a set of linear combinations of the x's which span the x-space. The theory of canonical correlations was designed to answer questions of this sort. Suppose that $k \leqq v$, as is likely to be the case, and that the coefficient of the appropriate linear combinations of the x's and y's are given in matrices $\mathbf{L}(k \times k)$ and $\mathbf{M}(v \times v)$, respectively. Using the notation of (34), the equations to be solved are then

$$\begin{aligned} \mathbf{UM} &= \mathbf{VLR'} \\ \mathbf{U'L} &= \mathbf{WMR} \end{aligned} \tag{38}$$

where $\mathbf{R}$ is a $v \times k$ matrix whose first k diagonal elements are known as the canonical correlations and all others elements are zero; **L** and **M** satisfy:

$$\begin{aligned} \mathbf{U'V}^{-1}\mathbf{UM} &= \mathbf{WM}(\mathbf{RR'}) \\ \mathbf{UW}^{-1}\mathbf{U'L} &= \mathbf{VL}(\mathbf{R'R}) \end{aligned} \tag{39}$$

Only k canonical correlations will be non-zero and our estimate $\mathbf{K}_R$ of the matrix **K** is the first k columns of the $v \times v$ matrix **M**.

The matrix $\mathbf{C}_R$ corresponding to **C** does not seem to have a simple form although **MM′** may be substituted for $\mathbf{W}^{-1}$ in equation (2); the equivalence of these two forms can be derived from equation (39).

7.5. *Comparison of the two methods of external estimation*

When a unit canonical correlation exists, the corresponding canonical loadings m derived from equation (38) satisfy

$$(\mathbf{U'V}^{-1}\mathbf{U}-\mathbf{W})\mathbf{m} = \mathbf{0}. \tag{40}$$

Thus, in this case, the dispersion matrix obtained from the residuals after correcting for the regression of the y's on the x's by the first external estimation method, will not be of full rank and g-inverses have to be used. Further, if $k=1$ in (34), then **V** and **L** are scalars and $\mathbf{UW}^{-1}\mathbf{U'}=\mathbf{V}$ so that the first row of the matrix (34) is equal to the linear combination of the v remaining rows, obtained by pre-multiplying by $\mathbf{UW}^{-1}$ and this leads us to suspect that $\mathbf{W}^{-1}$ is then a g-inverse of $\mathbf{W}-\mathbf{U'V}^{-1}\mathbf{U}$. This is easily verified for

$$\begin{aligned} &(\mathbf{W}-\mathbf{U'V}^{-1}\mathbf{U})\mathbf{W}^{-1}(\mathbf{W}-\mathbf{U'V}^{-1}\mathbf{U}) \\ &= \mathbf{W}-2\mathbf{U'V}^{-1}\mathbf{U}+\mathbf{U'V}^{-1}(\mathbf{UW}^{-1}\mathbf{U'})\mathbf{V}^{-1}\mathbf{U} \\ &= \mathbf{W}-\mathbf{U'V}^{-1}\mathbf{U} \end{aligned}$$

because

$$\mathbf{UW}^{-1}\mathbf{U'} = \mathbf{V}.$$

We also note that the corrected values of the matrix **G** of population means satisfy $\mathbf{Gm}=\mathbf{0}$, because if $\mathbf{G}=\bar{\mathbf{Y}}-\bar{\mathbf{X}}\mathbf{V}^{-1}\mathbf{U}$, the regression correction, and $\bar{\mathbf{X}}=\bar{\mathbf{Y}}\mathbf{W}^{-1}\mathbf{U'}$, implied by the unit correlation, then $\mathbf{G}=\bar{\mathbf{Y}}(\mathbf{I}-\mathbf{W}^{-1}\mathbf{U'V}^{-1}\mathbf{U})$. Now from (37), $\mathbf{m}=\mathbf{W}^{-1}\mathbf{U'L}$, so that

$$\begin{aligned} \mathbf{Gm} &= \bar{\mathbf{Y}}(\mathbf{m}-\mathbf{W}^{-1}\mathbf{U'V}^{-1}(\mathbf{UW}^{-1}\mathbf{U'})\mathbf{L}) \\ &= \bar{\mathbf{Y}}(\mathbf{m}-\mathbf{m}) = \mathbf{0}. \end{aligned}$$

Thus since the same linear restrictions apply to the rows of **G** as to the rows of $(\mathbf{W}-\mathbf{Y'V}^{-1}\mathbf{U})$, then the values of D^2 are, by Section 3, obtained uniquely from $\mathbf{GW}^{-1}\mathbf{G'}$, because $\mathbf{W}^{-1}$ is a g-inverse of $\mathbf{W}-\mathbf{U'V}^{-1}\mathbf{U}$. This shows that the age effects need to be eliminated from **G** but not from $\mathbf{W}^{-1}$. The corresponding result when canonical correlation estimation is used, is given by $\mathbf{GC}_R\mathbf{G'}$ where $\mathbf{C}_R$ is equation (2) with $\mathbf{K}=\mathbf{m}$, but in this case the two methods must be equivalent.

When $k=1$, and the first canonical correlation is not unity, we still have that **m** is proportionate to $\mathbf{W}^{-1}\mathbf{U'}$. These are the estimates of the regression coefficients obtained when the single x-variate is expressed as a multiple linear regression on the y-variates. In this sense, the linear combination amongst the y's is the best linear predictor of the x-variate. When there are several x-variates, the canonical variates which are linear combinations of the y's will still be the best linear predictors of the space spanned by the x's. Therefore, the canonical correlation method is equivalent to the regression method if there is just one concomitant variate. With two or more concomitant variates, the canonical correlation method is preferable to adjusting each y-variate by multiple regression because this ignores their intercorrelations.

REFERENCES

Burnaby, T. P. 1966: Growth-invariant discriminant functions and generalized distances. *Biometrics 22*, 96—110.

Delany, M. J. & Healy, M. J. R. 1964: Variation in the long-tailed field mouse (*Apodemus sylvaticus* (L.)) in north-west Scotland. II. Simultaneous examination of all characters. *Proc. R. Soc. B, 161*, 200—207.

Gower, J. C. 1966 a: Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika 53*, 325—338.

Gower, J. C. 1966 b: A Q-technique for the calculation of canonical variates. *Biometrika 53*, 588—589.

Hopkins, J. W. 1966: Some considerations in multivariate allometry. *Biometrics 22*, 747—760.

Jolicoeur, P. 1963: The multivariate generalization of the allometry equation. *Biometrics 19*, 497—499.

Lawley, D. N. & Maxwell, A. E. 1971: Factor analysis as a statistical method (2nd edition). London, Butterworths.

Rao, C. R. 1952: *Advanced statistical methods in biometric research*. 390 pp. John Wiley & Sons, Inc., New York.

Rao, C. R. 1962: A note on a generalized inverse of a matrix with applications to problems in mathematical statistics, *J. R. Stat. Soc. B, 24*, 152—158.

Rao, C. R. 1966: Discriminant function between composite hypotheses and related problems. *Biometrika 53*, 339—345.

Rao, C. R. 1967: Calculus of generalized inverses of matrices: Part 1 — General theory. *Sankhya*, *Series A*, *29*, 317—342.

Tessier, G. 1955: Sur la determination de l'axe d'un image rectiligne de points. *Biometrics 11*, 344—357.

Printed by Ekenäs Tryckeri Ab, Ekenäs, Finland December, 1976.