RELATIONS AMONG SEVERAL SETS OF VARIATES

by

Jon R. Kettenring

University of North Carolina

Institute of Statistics Mimeo Series No. 538

This research was supported by the Air Force Office of Scientific Research Contract No. AF-AFOSR-760-65

Department of Statistics
University of North Carolina
Chapel Hill, N. C.

Summary

This paper commences with a discussion of the relations, particularly those represented by canonical correlations, between two sets of variates. The main purpose, however, is to investigate two extensions of canonical correlation theory to deal with several sets of variates. An old criterion of minimal generalized variance is redeveloped, and a new criterion based on sums of squared correlations is formulated. Solutions in both cases are obtained by an iterative procedure which can be readily carried out on a computer.

CHAPTER I

Relations Between Two Sets of Variates

Preliminary Remarks

The study of relations between two sets of random variables was begun by Hotelling in his brief 1935 paper [6], "The Most Predictable Criterion." His definitive study of the problem appeared the next year in a fifty seven page paper [7] which still stands as a key reference in multivariate literature.

Let $\underline{X}_1$ $(P_1 \times 1)$ and $\underline{X}_2$ $(P_2 \times 1)$, $P_1 \leq P_2$, represent the two sets of variates. At this point the only assumption about $\underline{X}$ is that

$$\text{var}\ (\underline{X}) = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{matrix} P_1 \\ P_2 \end{matrix}$$
$$\qquad\qquad\qquad\quad P_1 \quad\ P_2$$

exists and is positive definite. The relations between the two sets are found through analysis of Σ.

An illustration given by Hotelling [7] compared a set of mental test scores with a set of physical measurements on the same persons. "The questions then arise of determining the number and nature of the independent relations of mind and body . . . and of extracting from the multiplicity of correlations in the system suitable characterizations of these independent relations." Of particular interest might be a pair of linear combinations of scores and measurements yielding the maximum or minimum squared correlation.

The formal extraction procedure will be loosely referred to as a "canonical analysis." We start by finding two new variables, each with unit variance, formed from linear combinations of $\underline{X}_1$ and $\underline{X}_2$ yielding the largest possible correlation. The new variables are known as canonical variables and together they form the first pair of canonical variables. The associated correlation is called the first canonical correlation. When one of the initial sets consists of criterion variables and the other of predictor variables, then the canonical variable derived from the criterion set corresponds to Hotelling's most predictable criterion [6].

The analysis may be continued until P_1 pairs of canonical variables and P_1 canonical correlations have been identified. At each stage it is required that the variable derived from each set be uncorrelated with other canonical variables formed from that set, that the correlation between the derived variables be as large as possible, and finally that the new variables have unit variance. If $P_1 < P_2$, a formal completion of the problem may be made by finding an additional $(P_2 - P_1)$ canonical variables as linear combinations of the second set; besides having unit variance these variables must be uncorrelated amongst themselves as well as with the other canonical variables from the second set. These latter $(P_2 - P_1)$ variables are usually of little interest.

A consequence of the canonical analysis is that there is zero correlation between a canonical variable from the first set and a canonical variable from the second set whenever they do not constitute one of the canonical pairs. In addition some of the canonical correlations may be zero, the number being equal to the difference between P_1 and the rank of Σ_{12}.

Already it should be clear that the canonical correlations are invariant under internal nonsingular transformations of either set; this fact was pointed out by Hotelling [7] and used advantageously by Horst ([3], [4], and [5]), Steel [13], and others. Hence, without any loss of generality, one may replace Σ by

$$R = \begin{pmatrix} I & R_{12} \\ R_{21} & I \end{pmatrix} \begin{matrix} P_1 \\ P_2 \end{matrix}$$
$$\quad\; P_1 \quad P_2$$

where R is obtained from Σ by the transformation

$$R = D_T^{-1} \Sigma D_T'^{-1}$$

with

$$D_T = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \begin{matrix} P_1 \\ P_2 \end{matrix} ,$$
$$\quad\; P_1 \quad P_2$$

$$\Sigma_{11} = T_1 T_1' ,$$

and $\Sigma_{22} = T_2 T_2'$.

T_1 and T_2 may be taken to be lower triangular, a computationally convenient form, or symmetric, in which case it is customary to write

$$T_i = \Sigma_{ii}^{1/2}, \; i = 1,2.$$

A canonical analysis of two sets of variates with variance-covariance matrix R simply involves the selection of appropriate internal orthogonal transformations for each set. Thus if a variable is dropped from the first set, the induced canonical variables for the second set will be related to the original canonical variables for the second set by an orthogonal transformation.

The Basic Theorem of Canonical Analysis

Hotelling's derivation of the results needed to determine the canonical variables and correlations was based on calculus and Lagrange multipliers. It is possible, however, to give a rigorous derivation using only matrix theory. Lancaster [8] has done essentially this. Other writers have presented or discussed results which could be expanded into such a proof; see, for example, Vinograde ([15], pp. 160-161), Roy ([12], pp. 142, 153), and Horst ([3], pp. 131-132). The proof given here is similar to Lancaster's. It illustrates, more clearly than the calculus approach, the structural significance of the canonical correlations.

Theorem 1·1 Suppose
$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{matrix} P_1 \\ P_2 \end{matrix} > 0$$
$$\begin{matrix} P_1 & P_2 \end{matrix}$$

and $P_1 \leq P_2$. Then there exists a matrix

$$D_B = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix} \begin{matrix} P_1 \\ P_2 \end{matrix}$$
$$\begin{matrix} P_1 & P_2 \end{matrix}$$

and a unique matrix

$$\Phi = D_B \Sigma D_B'$$

such that

$$\Phi = \begin{pmatrix} I & \Phi_{12} \\ \Phi_{21} & I \end{pmatrix},$$

$$\Phi_{12} = \underset{P_1}{(D} \quad \underset{P_2-P_1}{0)} \; P_1 \, ,$$

$$D = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_{P_1}),\; 1 > \lambda_1 \geq \ldots \geq \lambda_{P_1} \geq 0,$$

and $\Phi_{21} = \Phi_{12}'$.

$\left((\lambda_1^2, \lambda_2^2, \ldots, \lambda_{P_1}^2)\right.$ must be the eigen values of $\left.\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.\right)$

Also $B_1 = B_1^*\Sigma_{11}^{-1/2}$,

where B_1^* is any orthogonal matrix such that

$$B_1^*\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}B_1^{*\prime} = D^2,$$

and $B_2 = B_2^*\Sigma_{22}^{-1/2}$,

where B_2^* is an orthogonal matrix satisfying

$$B_1^*\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2} = \Phi_{12}B_2^*.$$

This equation will determine the first j rows of B_2^*, j being the number of non zero eigen values λ; the remaining (P_2-j) rows are any orthogonal completion of B_2^*.

<u>proof</u>

Write $\Sigma_{11} = \Sigma_{11}^{1/2}\Sigma_{11}^{1/2}$ and $\Sigma_{22} = \Sigma_{22}^{1/2}\Sigma_{22}^{1/2}$.

Let $D_T = \begin{pmatrix} \Sigma_{11}^{1/2} & 0 \\ 0 & \Sigma_{22}^{1/2} \end{pmatrix}$ and

form $R = D_T^{-1}\Sigma D_T^{-1\prime}$

$$= \begin{pmatrix} I & R_{12} \\ R_{21} & I \end{pmatrix}$$

where $R_{12} = \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}$

and $R_{21} = R_{12}'$.

By lemma A2 (see Appendix), there exist orthogonal matrices B_1^* and B_2^* such that

$$B_1^* R_{12} B_2^{*\prime} = \Phi_{12}$$

where $\Phi_{12} = (D \quad 0)$

$$D = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_{P_1}),\ 1 > \lambda_1 \geq \ldots \geq \lambda_{P_1} \geq 0,$$

and $(\lambda_1^2, \lambda_2^2, \ldots, \lambda_{P_1}^2)$ are the eigen values of $R_{12}R_{21}$. Moreover, B_1^* may be any orthogonal matrix such that

$$B_1^* R_{12} R_{21} B_1^{*\prime} = D^2.$$

Note that the eigen values of $R_{12}R_{21}$ are the same as the eigen values of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Let $D_{B^*} = \begin{pmatrix} B_1^* & 0 \\ 0 & B_2^* \end{pmatrix}$

and $D_B = D_{B^*} D_T^{-1}$.

This definition of D_B yields the desired decomposition Φ.

Suppose now $D_{\tilde{B}}$ is any other matrix yielding a decomposition of Σ in the form Φ. Then

$$\Phi_{12} = \tilde{B}_1 \Sigma_{12} \tilde{B}_2'$$

$$= (\tilde{B}_1 \Sigma_{11}^{1/2}) \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} (\Sigma_{22}^{1/2} \tilde{B}_2')$$

$$\therefore \quad D^2 = \tilde{B}_1^* \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2} \tilde{B}_1^{*\prime}$$

where $\tilde{B}_1^* = \tilde{B}_1 \Sigma_{11}^{1/2}$ is orthogonal.

$\therefore$ $(\lambda_1^2, \lambda_2^2, \ldots, \lambda_{P_1}^2)$ must be the eigen values of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and Φ itself is unique.

The properties of B_2^* follow from the two equations

$$B_1^* R_{12} = \Phi_{12} B_2^*$$

$$B_2^* R_{21} R_{12} B_2^{*\prime} = \begin{pmatrix} D_2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Corollary 1·1·1 B_1 is any matrix such that

$$B_1 \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} B_1' = D^2$$

and $$B_1 \Sigma_{11} B_1' = I.$$

Alternatively, the rows of B_1 are any complete set of eigen vectors of $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$, such that the ith row of B_1 is associated with ith largest eigen value and $B_1 \Sigma_{11} B_1' = I$. The first j rows of B_2 are determined by the equation

$$B_1 \Sigma_{12} \Sigma_{22}^{-1} = \Phi_{12} B_2.$$

The last $(P_2 - j)$ rows of B_2 are chosen so that

$$B_2 \Sigma_{22} B_2' = I.$$

proof

Suppose $B_1 \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} B_1' = D^2$

and $$B_1 \Sigma_{11} B_1' = I.$$

Then $B_1 \Sigma_{11}^{1/2} (\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2}) \Sigma_{11}^{1/2} B_1' = D^2$

so that $B_1 \Sigma_{11}^{1/2}$ is orthogonal and may be taken as B_1^* in the theorem.

Suppose now $(\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} - \lambda I) B_1' = 0$

and $$B_1 \Sigma_{11} B_1' = I.$$

These hold if and only if

$$(\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2} - \lambda I)(B_1 \Sigma_{11}^{1/2})' = 0$$

and $$(B_1 \Sigma_{11}^{1/2})(B_1 \Sigma_{11}^{1/2})' = I$$

so that, again, $B_1\Sigma_{11}^{1/2}$ is orthogonal and may be taken as B_1^* in the theorem.

The last two statements of the corollary are immediate consequences of the theorem.

Corollary 1·1·2 Suppose $\tilde{B}_1^*$ and $\tilde{B}_2^*$ are any two orthogonal matrices producing a decomposition of the form

$$\tilde{B}_1^* R_{12} \tilde{B}_2^{*\prime} = \tilde{\Phi}_{12} = \underset{P_1 \quad P_2-P_1}{(\tilde{D} \quad 0)}\ P_1$$

where $\tilde{D} = ((d_{ij}))$ is not necessarily a diagonal matrix but with diagonal elements

$$d_{11} \geq d_{22} \geq \ldots \geq d_{P_1P_1} \geq 0 .$$

Then it is impossible for

$d_{ii} \geq \lambda_i$, $i = 1, 2, \ldots, j$, unless

$d_{ik} = 0$, $i = 1, 2, \ldots, j$, $k = 1, 2, \ldots, P_1 (k \neq i)$

in which case there is equality for $i = 1, 2, \ldots, j$. This is true for $j = 1, 2, \ldots, P_1$.

proof

$\tilde{B}_1^* R_{12} R_{21} \tilde{B}_1^{*\prime} = \tilde{D}\tilde{D}^{\prime} = F$ (say) where $F = ((f_{ij}))$.

$B_1^* R_{12} R_{21} B_1^{*\prime} = D^2 = \text{diag}\ (\lambda_1^2, \lambda_2^2, \ldots, \lambda_{P_1}^2)$.

$d_{ii}^2 \leq f_{ii}$ with equality if and only if $d_{ij} = 0$, $j = 1, 2, \ldots, P_1 (j \neq i)$.

$\lambda_1^2 \geq f_{11}$ by lemma A1; hence if $\lambda_1 = d_{11}$ then $d_{1j} = 0$ for $j = 2, 3, \ldots, P_1$, and the first rows of $\tilde{B}_1^*$ and B_1^* are in the same eigen space. Now $\lambda^2 \geq f_{22}$ by lemma A1; hence if $\lambda_2 = d_{22}$, then $d_{2j} = 0$ for

$j = 1, 2, \ldots, P_1$ ($j \neq 2$), and the second rows of $\tilde{B}_2^*$ and B_2^* are in the same eigen space. Continuing in this way, it follows that $d_{ii} \geq \lambda_i$ for $i = 1, 2, \ldots, j$; $j = 1, 2, \ldots, P_1$ if and only if $\tilde{D}$ is diagonal (in which case $d_{ii} = \lambda_i$, $i = 1, 2, \ldots, P_1$).

Theorem 1·1 together with corollary 1·1·1 provide the essential information for carrying out a canonical analysis. The corollary, in particular, ties the theorem to the procedures given in Anderson [1]. The second corollary is needed to show that the λ's are in fact the canonical correlations. Now it follows that there is zero correlation between canonical variables not belonging to the same canonical pair and that the number of non zero canonical correlations is the same as the rank of Σ_{12}.

There are two other useful consequences of theorem 1·1. First suppose $P_1 = P_2$. Take any P_1 linear combinations of variables from the first set and pair them with an equal number of linear combinations of variables from the second. Then if any two derived variables not in the same pair are uncorrelated, the correlations between the paired variables must, apart from sign, be canonical. Here the smallest canonical correlation is the minimal non negative correlation attainable between any pair. Second, when $P_1 < P_2$ there will always be $(P_2 - P_1)$ uncorrelated variables, linear combinations of the variates in the second set, which are uncorrelated with any variable derived as a linear combination of the first set.

It is sometimes said that "the ensemble of canonical variables accounts for all of the existing relations between the two sets of variables." In what sense is this so? Writing $R_{12} = ((r_{ij}))$, the

r_{ij} represent all that may be explained by analysis of Σ or R.

Let $\lambda^2 = \sum_{i=1}^{P_1} \lambda_i^2$.

$$\text{Then } \lambda^2 = \text{tr}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$= \text{tr}R_{12}R_{21}$$

$$= \sum_{i=1}^{P_1} \sum_{j=1}^{P_2} r_{ij}^2 .$$

By dint of the last equation, it is reasonable to call λ^2 a measure of the "total relationship" between the two sets. Hence the ith canonical pair isolates and accounts for a proportion λ_i^2/λ^2 of this total. And, together, the P_1 pairs account for all of the total.

The next corollary relates the eigen values of R to the canonical correlations. This result underlies certain generalizations to be discussed in chapter two.

<u>Corollary 1·1·3</u> The largest P_1 eigen values of R are $1 + \lambda_1$, $1 + \lambda_2, \ldots, 1 + \lambda_{P_1}$. A corresponding set of eigen vectors consists of $({}_1\underline{b}_i^{*\prime}, {}_2\underline{b}_i^{*\prime})$, $i = 1, 2, \ldots, P_1$, where ${}_j\underline{b}_i^{*\prime}$ is the ith row of B_j^*, $j = 1, 2$.

<u>proof</u>

$$\left[\begin{pmatrix} I & R_{12} \\ R_{21} & I \end{pmatrix} - \rho I\right]\begin{pmatrix} {}_1\underline{b}^* \\ {}_2\underline{b}^* \end{pmatrix} = \underline{0}$$

$$\Leftrightarrow \begin{bmatrix} (1-\rho)I & R_{12} \\ R_{21} & (1-\rho)\underline{I} \end{bmatrix}\begin{pmatrix} {}_1\underline{b}^* \\ {}_2\underline{b}^* \end{pmatrix} = \underline{0}$$

$$\Leftrightarrow \begin{cases} (1-\rho)\ {}_1\underline{b}^* + R_{12}\ {}_2\underline{b}^* = \underline{0} \\ R_{21}\ {}_1\underline{b}^* + (1-\rho)\ {}_2\underline{b}^* = \underline{0}\ . \end{cases}$$

Suppose $\rho \neq 1$; then

$${}_1\underline{b}^* = (\rho-1)^{-1} R_{12}\ {}_2\underline{b}^* \text{ and } {}_2\underline{b}^* = (\rho-1)^{-1} R_{21}\ {}_1\underline{b}^*\ .$$

Substituting

$$[R_{12}R_{21} - (1-\rho)^2 I]\ {}_1\underline{b}^* = \underline{0}$$

$$[R_{21}R_{12} - (1-\rho)^2 I]\ {}_2\underline{b}^* = \underline{0}$$

Hence $\rho = 1 \pm \lambda$ and, for each positive λ, $1 + \lambda$ is one of the P_1 largest eigen values of R and ${}_1\underline{b}^{*\prime}$ and ${}_2\underline{b}^{*\prime}$ may be taken as the corresponding rows of B_1^* and B_2^*. When $\lambda = 0$ is a root of multiplicity K, then $\rho = 1$ is a root of R of multiplicity at least 2K; thus the remaining of the P_1 largest eigen values of R (if any) are all unity. For each $\lambda = 0$, ${}_1\underline{b}^{*\prime}$ and ${}_2\underline{b}^{*\prime}$ may be taken as the corresponding rows of B_1^* and B_2^* since

$$R_{12}R_{21}\ {}_1\underline{b}^* = \underline{0}$$

$$R_{21}R_{12}\ {}_2\underline{b}^* = \underline{0}$$

<u>Some Special Results</u>

Let V(M) designate the vector space generated by the row vectors of the matrix M. If

$$B_1 = \begin{pmatrix} {}_1\underline{b}_1' \\ {}_1\underline{b}_2' \\ \vdots \\ {}_1\underline{b}_{P_1}' \end{pmatrix} \text{ and } B_2 = \begin{pmatrix} {}_2\underline{b}_1' \\ {}_2\underline{b}_2' \\ \vdots \\ {}_2\underline{b}_{P_2}' \end{pmatrix}$$

then the standard canonical analysis involves the selection of

${}_1\underline{b}_i' \varepsilon V(I_{P_1})$, $i = 1, 2, \ldots, P_1$, and ${}_2\underline{b}_j' \varepsilon V(I_{P_2})$, $j = 1, 2, \ldots, P_2$, subject to the conditions of theorem 1·1. The intent now is to show how to modify this procedure to allow for arbitrary homogeneous linear restrictions on the coefficient vectors.

Consider the (possibly) more restrictive setting where

$${}_1\underline{b}_i' \varepsilon V(M_1),\ M_1(q_1 \times P_1),\ q_1 \leq P_1,\ \text{rank } (M_1) = q_1$$

$${}_2\underline{b}_j' \varepsilon V(M_2),\ M_2(q_2 \times P_2),\ q_2 \leq P_2,\ \text{rank } (M_2) = q_2$$

with $q_1 \leq q_2$, dropping the assumption $P_1 \leq P_2$.

Theorem 1·2 Let Σ be as in theorem 1·1. Then there exist

$${}_1\underline{b}_i \varepsilon V(M_1)\quad i = 1, 2, \ldots, q_1$$

$${}_2\underline{b}_j \varepsilon V(M_2)\quad j = 1, 2, \ldots, q_2$$

such that

$$B_1 \Sigma_{11} B_1' = I_{q_1},\ B_2 \Sigma_{22} B_2' = I_{q_2},\ \text{and } B_1 \Sigma_{12} B_2' = \Phi_{12}$$

where

$$B_1 = \begin{pmatrix} {}_1\underline{b}_1' \\ {}_1\underline{b}_2' \\ \vdots \\ {}_1\underline{b}_{q_1}' \end{pmatrix},\quad B_2 = \begin{pmatrix} {}_2\underline{b}_1' \\ {}_2\underline{b}_2' \\ \vdots \\ {}_2\underline{b}_{q_2}' \end{pmatrix},$$

$$\Phi_{12} = \underset{\quad q_1 \qquad q_2 - q_1}{(D \qquad 0\ )\,q_1},$$

and $D = \text{diag }(\lambda_1, \lambda_2, \ldots, \lambda_{q_1})$, $1 > \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{q_1} \geq 0$.

$(\lambda_1^2, \lambda_2^2, \ldots, \lambda_{q_1}^2)$ are the eigen values of $(M_1 \Sigma_{11} M_1')^{-1} M_1 \Sigma_{12} M_2' (M_2 \Sigma_{22} M_2')^{-1} M_2 \Sigma_{21} M_1'$. If the rows of $\tilde{B}_1$ are any complete set of eigen vectors of this latter matrix, subject to $\tilde{B}_1 M_1 \Sigma_{11} M_1' \tilde{B}_1' = I$, then

B_1 may be taken as

$$B_1 = \tilde{B}_1 M_1.$$

Also $B_2 = \tilde{B}_2 M_2$ where the first j rows of $\tilde{B}_2$ are solutions of the equations

$$\tilde{B}_1 M_1 \Sigma_{12} M_2' (M_2 \Sigma_{22} M_2')^{-1} = \Phi_{12} \tilde{B}_2$$

(j being the number of non zero eigen values λ).

The last $(q_2 - j)$ rows of $\tilde{B}_2$ are chosen so that

$$\tilde{B}_2 M_2 \Sigma_{22} M_2' \tilde{B}_2' = I.$$

Suppose S and T are non singular transformations such that

$$S : V(M_1) \longrightarrow V(SM_1)$$

$$T : V(M_2) \longrightarrow V(TM_2);$$

then $S : \tilde{B}_1 \longrightarrow \tilde{B}_1 S^{-1}$, $B_1 \longrightarrow B_1$

$$T : \tilde{B}_2 \longrightarrow \tilde{B}_2 T^{-1}, \; B_2 \longrightarrow B_2$$

$$(S,T) : D \longrightarrow D.$$

<u>proof</u>

Form $(M_1 \Sigma_{11} M_1')^{-1} M_1 \Sigma_{12} M_2' (M_2 \Sigma_{22} M_2')^{-1} M_2 \Sigma_{21} M_1'$.

The eigen values of this matrix, to wit, $\lambda_1^2, \lambda_2^2, \ldots, \lambda_{q_1}^2$, are those obtained by applying theorem 1·1 to the positive definite matrix

$$\begin{pmatrix} M_1 \Sigma_{11} M_1' & M_1 \Sigma_{12} M_2' \\ M_2 \Sigma_{21} M_1' & M_2 \Sigma_{22} M_2' \end{pmatrix} = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \Sigma \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}'.$$

By the same theorem and corollary 1·1·1, there exist $\tilde{B}_1$ amd $\tilde{B}_2$ such that

$$\tilde{B}_1 M_1 \Sigma_{11} M_1' \tilde{B}_1' = I_{q_1} \; ,$$

$$\tilde{B}_2 M_2 \Sigma_{22} M_2' \tilde{B}_2' = I_{q_2} \; ,$$

$$\tilde{B}_1 M_1 \Sigma_{12} M_2' \tilde{B}_2' = \Phi_{12},$$

and $\tilde{B}_1 M_1 \Sigma_{12} M_2' (M_2 \Sigma_{22} M_2')^{-1} = \Phi_{12} \tilde{B}_2$.

Hence let $B_1 = \tilde{B}_1 M_1$ and $B_2 = \tilde{B}_2 M_2$.

For the transformation S,

$$[S'^{-1}(M_1\Sigma_{11}M_1')^{-1}M_1\Sigma_{12}M_2'(M_2\Sigma_{22}M_2')^{-1}M_2\Sigma_{21}M_1'S' - \lambda I](\tilde{B}_1 S^{-1})' = 0$$

$$\Longleftrightarrow [(M_1\Sigma_{11}M_1')^{-1}M_1\Sigma_{12}M_2'(M_2\Sigma_{22}M_2')^{-1}M_2\Sigma_{21}M_1' - \lambda I]\tilde{B}_1' = 0.$$

$\therefore S \longrightarrow \tilde{B}_1 \longrightarrow \tilde{B}_1 S^{-1}$. Since $V(M_1)$ and $V(SM_1)$ are isomorphic, $S{:}B_1 \longrightarrow B_1$.

The effect of the transformation T is seen similarly.

Now $(SM_1\Sigma_{11}M_1'S')^{-1}SM_1\Sigma_{12}M_2'T'(TM_2\Sigma_{22}M_2'T')^{-1}TM_2\Sigma_{21}M_1'S'$

$$= S'^{-1}(M_1\Sigma_{11}M_1')^{-1}M_1\Sigma_{12}M_2'(M_2\Sigma_{22}M_2')^{-1}M_2\Sigma_{21}M_1'S'$$

which has eigen values the same as

$(M_1\Sigma_{11}M_1')^{-1}M_1\Sigma_{12}M_2'(M_2\Sigma_{22}M_2')^{-1}M_2\Sigma_{21}M_1'$ so that $(S,T){:}D \rightarrow D$.

The assumption $\Sigma > 0$ is not necessary for the theorem to hold. It is sufficient that

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \Sigma \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}' > 0.$$

Hence, it is easy to see how one would complete a canonical analysis for singular sets of variables. Suppose rank $(\Sigma_{11}) = r_1 \leq P_1$ and rank $(\Sigma_{22}) = r_2 \leq P_2$. Then there exist matrices M_1 $(r_1 \times P_1)$ and $M_2 (r_2 \times P_2)$ of full rank such that $M_1\Sigma_{11}M_1' = I_{r_1}$ and $M_2\Sigma_{22}M_2' = I_{r_2}$. Now, working with the nonsingular sets,

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \Sigma \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}' > 0$$

from which point theorem 1·1 may be applied. An implication of theorem 1·2 is that the reduction is invariant under choice of M_1 and M_2.

The next theorem was motivated originally by problems to be discussed in chapter two, but is of interest here for the further insight it gives to the relations between two sets of variates. In essence, one set of variables will be held fixed, after which attention is directed to certain invariants of the system under internal nonsingular transformations of the other set.

Revising the original assumptions on $\underline{X}_2' = ({}_2X_1, {}_2X_2, \ldots, {}_2X_{P_2})$ slightly, var $({}_2X_i) = 1$, $i = 1, 2, \ldots, P_2$. Let A, B, and C be nonsingular matrices

$$A = \begin{pmatrix} \underline{a}_1' \\ \underline{a}_2' \\ \vdots \\ \underline{a}_{P_1}' \end{pmatrix}, \quad B = \begin{pmatrix} \underline{b}_1' \\ \underline{b}_2' \\ \vdots \\ \underline{b}_{P_1}' \end{pmatrix}, \text{ and } C = \begin{pmatrix} \underline{c}_1' \\ \underline{c}_2' \\ \vdots \\ \underline{c}_{P_1}' \end{pmatrix}$$

satisfying

$$A\Sigma_{11}A' = B\Sigma_{11}B' = C\Sigma_{11}C' = I. \tag{1·1}$$

Define new variables

$$\underline{U} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_{P_1} \end{pmatrix} = A\underline{X}_1,$$

$$\underline{V} = \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_{P_1} \end{pmatrix} = B\underline{X}_1,$$

$$\underline{W} = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_{P_1} \end{pmatrix} = C\underline{X}_1.$$

Let $\alpha_i = \sum_{j=1}^{P_2} \{\text{corr}(U_i,\ {}_2X_j)\}^2$,

$$\beta_i = \left\{\sum_{j=1}^{P_2} \text{corr}(V_i,\ \ X_j)\right\}^2,$$

and $\gamma_i = \left|\text{var}\begin{pmatrix} W_i \\ \underline{X}_2 \end{pmatrix}\right|$ for $i = 1, 2, \ldots, P_1$.

Let $\underline{1}$ be a vector of ones and $\Delta_i(\cdot)$ the ith largest eigen value of the matrix in the argument.

Now for $i = 1, 2, \ldots, P_1$ in turn, select $\underline{a}_i$, $\underline{b}_i$, and $\underline{c}_i$ to maximize α_i and β_i and to minimize γ_i satisfying at each step the restrictions in (1·1). The technique for doing this and the values attained are given in theorem 1·3. [Theorem 1·3(i) was discussed recently by Fortier [2] and earlier by Rao[10].]

<u>Theorem 1·3</u> (i) $\alpha_i = \Delta_i(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{21})$, $i = 1, 2, \ldots, P_1$;

(ii) $\beta_i = \begin{cases} \underline{1}'\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\underline{1}, & \text{for } i = 1, \\ 0, & i = 2, 3, \ldots, P_1; \end{cases}$

(iii) $\gamma_i = \left|\Sigma_{22}\right|\left\{1 - \Delta_i(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\right\}$, $i = 1, 2, \ldots, P_1$.

These values are attained for each i when $\underline{a}_i$, $\underline{b}_i$, and $\underline{c}_i$ are eigen vectors (subject to 1·1) corresponding to the ith largest eigen values of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{21}$, $\Sigma_{11}^{-1}\Sigma_{12}\underline{1}\,\underline{1}'\Sigma_{21}$, and $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ respectively.

proof

Making repeated reference to lemma A1,

$$\alpha_1 = \sup_{\underline{a}} \frac{\underline{a}'\Sigma_{12}\Sigma_{21}\underline{a}}{\underline{a}'\Sigma_{11}\underline{a}} = \Delta_1(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{21})$$

$$\underline{a}_1 = \Sigma_{11}^{-1/2}\underline{e}_1$$

$$\underline{o} = (\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{21}\Sigma_{11}^{-1/2} - \alpha_1 I)\underline{e}_1$$

$$= (\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{21} - \alpha_1 I)\Sigma_{11}^{-1/2}\underline{e}_1$$

$$= (\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{21} - \alpha_1 I)\underline{a}_1$$

$$\alpha_2 = \sup_{\substack{\underline{a} \\ \underline{a}_1'\Sigma_{11}\underline{a}=o}} \frac{\underline{a}'\Sigma_{12}\Sigma_{21}\underline{a}}{\underline{a}'\Sigma_{11}\underline{a}} = \sup_{\substack{\underline{a} \\ \underline{e}_1'\Sigma_{11}^{1/2}\underline{a}=o}} \frac{\underline{a}'\Sigma_{12}\Sigma_{21}\underline{a}}{\underline{a}'\Sigma_{11}\underline{a}} = \Delta_2(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{21})$$

$$\underline{a}_2 = \Sigma_{11}^{-1/2}\underline{e}_2$$

$$\underline{o} = (\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{21}\Sigma_{11}^{-1/2} - \alpha_2 I)\underline{e}_2$$

$$= (\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{21} - \alpha_2 I)\underline{a}_2$$

$$\vdots \qquad \vdots$$

$$\alpha_{p_1} = \sup_{\substack{\underline{a} \\ \underline{a}_j'\Sigma_{11}\underline{a}=o \\ j=1,2,\ldots,P_1-1}} \frac{\underline{a}'\Sigma_{12}\Sigma_{21}\underline{a}}{\underline{a}'\Sigma_{11}\underline{a}} = \sup_{\substack{\underline{a} \\ \underline{e}_j'\Sigma_{11}^{1/2}\underline{a}=o \\ j=1,2,\ldots,P_1-1}} \frac{\underline{a}'\Sigma_{12}\Sigma_{21}\underline{a}}{\underline{a}'\Sigma_{11}\underline{a}} = \Delta_{p_1}(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{21})$$

$$\underline{a}_{p_1} = \Sigma_{11}^{-1/2}\,\underline{e}_{p_1}$$

$$\underline{o} = (\Sigma_{11}^{-1/2}\,\Sigma_{12}\,\Sigma_{21}\,\Sigma_{11}^{-1/2} - \alpha_{p_1} I)\,\underline{e}_{p_1}$$

$$= (\Sigma_{11}^{-1}\,\Sigma_{12}\,\Sigma_{21} - \alpha_{p_1} I)\,\underline{a}_{p_1}\,.$$

The other results follow in precisely the same manner. Note that to obtain results for the β_i 's one needs to consider variation of

$$\frac{\underline{b}'\,\Sigma_{12}\underline{1}\,\underline{1}'\Sigma_{21}\underline{b}}{\underline{b}'\,\Sigma_{11}\,\underline{b}}\,;$$

for the γ_i's, one needs to consider variation of

$$\frac{\begin{vmatrix} \underline{c}'\Sigma_{11}\underline{c} & \underline{c}'\Sigma_{12} \\ \Sigma_{21}\underline{c} & \Sigma_{22} \end{vmatrix}}{\underline{c}'\Sigma_{11}\underline{c}} = |\Sigma_{22}|\frac{(\underline{c}'\Sigma_{11}\underline{c} - \underline{c}'\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\underline{c})}{\underline{c}'\Sigma_{11}\underline{c}}$$

$$= |\Sigma_{22}| - |\Sigma_{22}| \frac{\underline{c}'\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\underline{c}}{\underline{c}'\Sigma_{11}\underline{c}}$$

with respect to $\underline{c}$. For the latter problem, since the object is to seek the minimum value for γ_i, the lemma may be applied directly to

$$\frac{\underline{c}'\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\underline{c}}{\underline{c}'\Sigma_{11}\underline{c}} .$$

The γ_i's are particularly noteworthy because of their intimate relation to the canonical correlations. Moreover, the W's are a set of canonical variables for the first set.

CHAPTER II

Relations Among Several Sets of Variates

Preliminary Remarks

Hotelling [7], in addition to solving the two set problem, recognized the need to generalize his solution to deal with more than two sets of variates. Only a small amount of research has been done in this direction, however. A resume of this work will follow the introduction of necessary notation and assumptions.

Let $\underline{X}_1$ $(P_1 \times 1)$, $\underline{X}_2$ $(P_2 \times 1)$, ..., $\underline{X}_m$ $(P_m \times 1)$ represent the $m(>2)$ sets of random variables, arranged so that $P_1 \leq P_2 \leq \ldots \leq P_m$. $\underline{X}' = (\underline{X}_1', \underline{X}_2', \ldots, \underline{X}_m')$. Assume that

$$\text{var}(\underline{X}) = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2m} \\ \vdots & \vdots & & \vdots \\ \Sigma_{m1} & \Sigma_{m2} & \cdots & \Sigma_{mm} \end{pmatrix} \begin{matrix} P_1 \\ P_2 \\ \\ P_m \end{matrix}$$
$$\qquad\qquad\qquad P_1 \quad P_2 \qquad\quad P_m$$

is positive definite. Then there exist nonsingular matrices T_i such that $\Sigma_{ii} = T_i T_i'$, $i = 1, 2, \ldots, m$.

If $\underline{U}_i = T_i^{-1} \underline{X}_i$ and $\underline{U}' = (\underline{U}_1', \underline{U}_2', \ldots, \underline{U}_m')$, then

$$\mathrm{var}(\underline{U}) = R = \begin{pmatrix} I & R_{12} & \cdots & R_{1m} \\ R_{21} & I & \cdots & R_{2m} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ R_{m1} & R_{m2} & \cdots & I \end{pmatrix} .$$

The final variables of interest are $\underline{Z}_i = B_i^* \underline{U}_i$ where B_i^* is an orthogonal matrix and

$\underline{Z}_i' = ({}_iZ_1, {}_iZ_2, \ldots, {}_iZ_{P_i}), \; i = 1, 2, \ldots, m.$

$\underline{Z}' = (\underline{Z}_1', \underline{Z}_2', \ldots, \underline{Z}_m').$

$$\mathrm{var}(\underline{Z}) = \Phi = \begin{pmatrix} I & \Phi_{12} & \cdots & \Phi_{1m} \\ \Phi_{21} & I & \cdots & \Phi_{2m} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \Phi_{m1} & \Phi_{m2} & \cdots & I \end{pmatrix} .$$

The matrices $B_i^* = \begin{pmatrix} {}_i\underline{b}_1^{*\prime} \\ {}_i\underline{b}_2^{*\prime} \\ \vdots \\ {}_i\underline{b}_{P_i}^{*\prime} \end{pmatrix}$ and $B_i = \begin{pmatrix} {}_i\underline{b}_1' \\ {}_i\underline{b}_2' \\ \vdots \\ {}_i\underline{b}_{P_i}' \end{pmatrix}$

are related by $B_i^* = B_i T_i$

so that $\underline{Z}_i = B_i^* \underline{U}_i = B_i T_i T_i^{-1} \underline{X}_i = B_i \underline{X}_i .$

Introducing quasi-diagonal matrices

$D_T = \mathrm{diag}(T_1 \mid T \mid \ldots \mid T_m),$

$D_B = \mathrm{diag}(B_1 \mid B_2 \mid \ldots \mid B_m),$

and $D_{B^*} = \mathrm{diag}(B_1^* \mid B_2^* \mid \ldots \mid B_m^*),$

(2·1) we have $D_{B^*}D'_{B^*} = D'_{B^*}D_{B^*} = I$,

$$D_{B^*} = D_B D_T,$$

$$R = D_T^{-1} \Sigma D_T'^{-1},$$

and $$\Phi = D_B \Sigma D'_B = D_{B^*} R D'_{B^*}.$$

Two more quasi-diagonal matrices

$$D_{B(i)} = \text{diag}({}_1\underline{b}'_i \mid {}_2\underline{b}'_i \mid \ldots \mid {}_m\underline{b}'_i)$$

and $$D_{B^*(i)} = \text{diag}({}_1\underline{b}^{*\prime}_i \mid {}_2\underline{b}^{*\prime}_i \mid \ldots \mid {}_m\underline{b}^{*\prime}_i)$$

with $D_{B^*(i)} = D_{B(i)} D_T$

define $\Phi_{(i)} = D_{B(i)} \Sigma D'_{B(i)} = D_{B^*(i)} R D'_{B^*(i)}$

for $i = 1, 2, \ldots, P_1$. $\Phi_{(i)}$ is just the variance-covariance (or correlation) matrix of $({}_1Z_i, {}_2Z_i, \ldots, {}_mZ_i)$. It follows from (2·1)

(2·2) that $D_{B^*(i)}D'_{B^*(i)} = I$,

(2·3) and $D_{B^*(i)}D'_{B^*(j)} = 0$ for $i = 1, 2, \ldots, P_1$; $j = 1, 2, \ldots, i-1$.

From the point of view of the present research, a bonafide generalization of Hotelling's procedure must

(i) define a logical criterion for $\Phi_{(i)}$, $i = 1, 2, \ldots, P_1$;

(ii) satisfy (2·2) and (2·3);

(iii) reduce to Hotelling's solution when $m = 2$.

When all the P_i's are not equal, the remaining elements of D_{B^*} may be specified by an appropriate orthogonal completion process. But this point is of little importance and will not be considered further. (Even if all the P_i's are equal D_{B^*} need not be unique.) Here the term "canonical analysis" will refer to the formal extraction of relations among the sets as defined by the criterion.

Previous work on generalizations has been done by Vinograde (1950), Steel (1951), Horst (1961), and McKeon (1966). Not all of their proposals satisfy conditions (i) and (ii) of the last paragraph although all satisfy condition (iii).

In 1935 Wilks [16] developed the likelihood ratio test for independence of sets of variates. The parametric analogue of his test statistic is just $|R|$ or the product of the eigen values of R. Hotelling [7] observed that the eigen values of (R-I) (and hence of R) are invariant under internal nonsingular transformations of the $\underline{X}_i$'s.

McKeon [9] has defined the generalized canonical correlation as $(\gamma-1)/(m-1)$ where γ is the largest eigen value of R. Recall from corollary 1·1·3 that when $m = 2$ this is just the first canonical correlation. However when $m > 2$ there will usually not be an associated eigen vector whose components generate m standardized variates.

Nevertheless, if these m derived variates are subsequently standardized, then the resulting variance-covariance matrix, $\Phi_{(1)}$, will give the best least square or Euclidean norm approximation to a rank one matrix in the class of possible $\Phi_{(1)}$ matrices associated with R. This criterion was suggested by Horst ([4], [5]) and called by him the rank one approximation method. It extends logically to P_1 stages in accordance with (2·2) and (2·3).

Two similar techniques, the oblique maximum variance method and the orthogonal maximum variance method, have also been suggested by Horst ([4], [5]). The former is based upon the P_1 largest eigen values and associated vectors of R. The induced variates (P_1 per set) are subsequently standardized, but they usually will not be uncorrelated within sets. The latter technique is a revision of the former which incorporates (2·3) as a constraint.

Horst named his most recognized contribution the maximum correlation method ([2], [4], and [5]). A better name from our point of view would be the "sum of correlations" (SUMCOR) method. One starts by selecting ${}_1Z_1, {}_2Z_1, \ldots, {}_mZ_1$ (subject to(2·2)) which maximize the sum of the correlations $(\underline{1}'\Phi_{(1)}\underline{1} - m)$or, equivalently, minimize $\sum_{i=1}^{m}\sum_{\substack{j=1 \\ i\neq j}}^{m} \mathrm{var}({}_iZ_1 - {}_jZ_1)$. To continue, a $D_{B^*(j)}$ which maximizes $\underline{1}'\Phi_{(j)}\underline{1}$, subject to (2·2) and(2·3), is found for j = 2, 3, ..., P_1.

At the sth stage one must solve

$$(2\cdot 4)\quad (I - \sum_{i=1}^{s-1} D'_{B^*(i)}D_{B^*(i)})(R - I)(I - \sum_{i=1}^{s-1} D'_{B^*(i)}D_{B^*(i)})D'_{B^*(s)}\underline{1} = D'_{B^*(s)}\underline{\lambda}_s$$

simultaneously for $D_{B^*(s)}$ and $\underline{\lambda}_s$ such that $\underline{1}'\underline{\lambda}_s$ $(= \underline{1}'\Phi_{(s)}\underline{1} - m)$is a maximum. Horst suggests an intuitively appealing iterative procedure to do this; the properties of the procedure, however, are unknown.

Horst's other three methods, being essentially eigen value problems, admit straightforward solutions. For the rank one approximation method, the largest eigen value of $\Phi_{(s)}$ is, as Horst remarked, a measure of the fit to a rank one matrix. If all the off-diagonal elements of $\Phi_{(s)}$ are near unity, then the largest eigen value will be near m and $\Phi_{(s)}$ will, loosely speaking, be "nearly of rank one." This suggests that the rank one method may yield useful approximations for the more difficult SUMCOR problem.

Vinograde [1] proposed a quite different method. The crux of his idea is to select a B_i^* (i = 1, 2, ..., m) which diagonalizes $\sum_{j=1}^{m} R_{ij}R_{ji}$. Referring to theorem 1·1 this is easily seen to be a generalization of the two set canonical analysis. This approach is mainly of mathematical interest in that it provides a simple canonical form for partitioned positive definite matrices.

The earliest idea of statistical interest came from Steel [13], a student of Vinograde. He established a system of equations from which one, in theory, could find a $D_{B^*(1)}$ yielding a minimum value of $|\Phi_{(1)}|$. In other words, the resulting standardized variables ${}_1Z_1, {}_2Z_1, \ldots, {}_mZ_1$ possess minimum generalized variance (GENVAR).

Steel argued that, for any D_{B^*}, $|\Phi_{(1)}|$ would appear as a diagonal element of the mth compound matrix of Φ. Fixing attention on this element, derivative equations are found which, along with orthogonality restrictions, are $\sum_{i=1}^{m} P_i^2$ in number with as many unknowns, the elements of $B_1^*, B_2^*, \ldots, B_m^*$. These equations are non linear and difficult to solve. Some solution will minimize $|\Phi_{(1)}|$.

Finally, we supplement this list with a new technique, the sum of squares of correlations (SSQCOR) method. ${}_1Z_1, {}_2Z_1, \ldots, {}_mZ_1$ are generated by a $D_{B^*(1)}$ which maximizes $(\mathrm{tr}\Phi_{(1)}^2 - m)$ or $\sum_{\substack{i=1 \\ i \neq j}}^{m} \sum_{j=1}^{m} \{\mathrm{corr}({}_iZ_1, {}_jZ_1)\}^2$. For $j = 2, 3, \ldots, P_1$, the same criterion determines $D_{B^*(j)}$ subject to (2·2) and (2·3).

For statistical purposes, the SUMCOR, GENVAR and SSQCOR criteria seem to be the most relevant. The choice of criterion should, of course, be related to the problem at hand.

The present research is primarily concerned with the GENVAR and SSQCOR criteria. To see that they may be superior in certain situations to the SUMCOR criterion, consider the following example.

$$(2{\cdot}5) \qquad R = \left[\begin{array}{cc|cc|cc} 1.0 & 0.0 & 0.3 & 0.0 & 0.3 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.1 & 0.0 & 0.1 \\ \hline 0.3 & 0.0 & 1.0 & 0.0 & -0.3 & 0.0 \\ 0.0 & 0.1 & 0.0 & 1.0 & 0.0 & 0.1 \\ \hline 0.3 & 0.0 & -0.3 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 1.0 \end{array}\right]$$

$$(2.6) \quad \text{Then} \quad D_{B^*(1)} = \left[\begin{array}{cc|cc|cc} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ \hline 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ \hline 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \end{array}\right]$$

satisfies (2·4) with

$$\underline{1}'\Phi_{(1)}\underline{1} = 3.6$$

as does

$$(2{\cdot}7) \qquad D_{B^*(1)} = \left[\begin{array}{cc|cc|cc} 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ \hline 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ \hline 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{array}\right]$$

with $\underline{1}'\Phi_{(1)}\underline{1} = 3.6$. Evidently there is indifference between the two weighting systems (2·6) and (2·7) with respect to the SUMCOR criterion. Intuitively one may feel that the first system of weights provides a better summary of the relations among the three sets. Both the GENVAR and the SSQCOR criteria would corroborate this feeling.

Other situations may favor SUMCOR over GENVAR or SSQCOR. More often, in the selection of ${}_1Z_1$, ${}_2Z_1$, ..., ${}_mZ_1$, the three criteria produce similar results. This tendency has been observed in a number of real data studies.

The Generalized Variance Criterion

Steel's problem may be handled without recourse to compound matrices by repeated use of a simple determinantal expansion. The final equations are in a form amenable to canonical analysis.

Writing $[A]_{(ij)}$ for the matrix obtained from A by deleting the ith row and jth column, define

$$f_i = |\Phi_{(i)}|,$$

$$\underline{\lambda}_i' = ({}_1\lambda_i, {}_2\lambda_i, \ldots, {}_m\lambda_i),$$

$$\underline{\gamma}_{ik}' = ({}_1\gamma_{ik}, {}_2\gamma_{ik}, \ldots, {}_m\gamma_{ik}),$$

$${}_jM_i = [\Phi_{(i)}]_{(jj)},$$

$${}_jN_i = (\Sigma_{j1}\ {}_1\underline{b}_i | \cdots | \Sigma_{jj-1}\ {}_{j-1}\underline{b}_i | \Sigma_{jj+1}\ {}_{j+1}\underline{b}_i | \cdots | \Sigma_{jm}\ {}_m\underline{b}_i),$$

$${}_jN_i^* = (R_{j1}\ {}_1\underline{b}_i^* | \cdots | R_{jj-1}\ {}_{j-1}\underline{b}_i^* | R_{jj+1}\ {}_{j+1}\underline{b}_i^* | \cdots | R_{jm}\ {}_m\underline{b}_i^*),$$

$${}_jP_i = {}_jN_i\ {}_jM_i^{-1}\ {}_jN_i',$$

and $${}_jP_i^* = {}_jN_i^*\ {}_jM_i^{-1}\ {}_jN_i^{*\prime};\quad i = 1,2,\ldots,P_1;\ j = 1,2,\ldots m;\ k = 1,2,\ldots,i-1.$$

The λ's and γ's will be Lagrange multipliers corresponding to conditions (2·2) and (2·3) respectively.

f_i may now be expanded into the form

(2·8) $$f_i = |{}_jM_i|(1 - {}_j\underline{b}_i^{*\prime}\ {}_jP_i^*\ {}_j\underline{b}_i^*),\ j = 1, 2, \ldots, m.$$

The salient feature of (2·8) is that ${}_jM_i$ and ${}_jP_i^*$ do not involve ${}_j\underline{b}_i^*$.

At the first stage ($i = 1$), form the Lagrangian equation

$$g_1 = f_1 - \sum_{j=1}^{m}(1 - {}_j\underline{b}_1^{*\prime}\ {}_j\underline{b}_1^*)|{}_jM_1|\,{}_j\lambda_1.$$

Differentiating symbolically with respect to ${}_j\underline{b}_1^*$

$$\frac{\partial g_1}{\partial_j\underline{b}_1^*} = -2|{}_jM_1|\,{}_jP_1^*\ {}_j\underline{b}_1^* + 2|{}_jM_1|\,{}_j\lambda_1\ {}_j\underline{b}_1^*;$$

equating the derivative to zero yields

(2·9) $$\underline{0} = ({}_jP_1^* - {}_j\lambda_1 I)\,{}_j\underline{b}_1^*,$$

$$ {}_j\lambda_1 = {}_j\underline{b}_1^{*\prime}\,{}_jP_1^*\,{}_j\underline{b}_1^*,$$

and $f_1 = |{}_jM_1|(1 - {}_j\lambda_1)$ for j = 1, 2, ..., m. An optimal solution of these m "eigen value" equations involves the determination of $\underline{1}\,D_{B^*(1)}$ and $\underline{\lambda}_1$ such that (2·9) holds with f_1 as small as possible.

Moving to the second stage, the Lagrangian equation is

$$g_2 = f_2 - \sum_{j=1}^{m}(1 - {}_j\underline{b}_2^{*\prime}\,{}_j\underline{b}_2^*)\,|{}_jM_2|\,{}_j\lambda_2 + 2\sum_{j=1}^{m}{}_j\underline{b}_2^{*\prime}\,{}_j\underline{b}_1^*\,|{}_jM_2|\,{}_j\gamma_{21}.$$

$$\frac{\partial g_2}{\partial\,{}_j\underline{b}_2^*} = -2|{}_jM_2|\,{}_jP_2^*\,{}_j\underline{b}_2^* + 2|{}_jM_2|\,{}_j\lambda_2\,{}_j\underline{b}_2^* + 2|{}_jM_2|\,{}_j\underline{b}_1^*\,{}_j\gamma_{21}$$

which, equating to zero, becomes

(2·10) $$\underline{0} = {}_jP_2^*\,{}_j\underline{b}_2^* - {}_j\lambda_2\,{}_j\underline{b}_2^* - {}_j\underline{b}_1^*\,{}_j\gamma_{21}$$

with $${}_j\gamma_{21} = {}_j\underline{b}_1^{*\prime}\,{}_jP_2^*\,{}_j\underline{b}_2^*,$$

$${}_j\lambda_2 = {}_j\underline{b}_2^{*\prime}\,{}_jP_2^*\,{}_j\underline{b}_2^*,$$

and $$f_2 = |{}_jM_2|(1 - {}_j\lambda_2) \text{ for } j = 1, 2, \ldots, m.$$

Rewriting (2·10),

$$\underline{0} = [(I - {}_j\underline{b}_1^*\,{}_j\underline{b}_1^{*\prime})\,{}_jP_2^* - {}_j\lambda_2 I]\,{}_j\underline{b}_2^*$$

$$= [(I - {}_j\underline{b}_1^*\,{}_j\underline{b}_1^{*\prime})\,{}_jP_2^*(I - {}_j\underline{b}_1^*\,{}_j\underline{b}_1^{*\prime}) - {}_j\lambda_2 I]\,{}_j\underline{b}_2^*$$

for j = 1, 2, ..., m.

For higher stages s (up to P_1) the situation is completely analogous. Only the key equations are given:

$$g_s = f_s - \sum_{j=1}^{m}(1 - {}_j\underline{b}_s^{*\prime}\,{}_j\underline{b}_s^*)\,|{}_jM_s|\,{}_j\lambda_s + 2\sum_{i=1}^{s-1}\sum_{j=1}^{m}{}_j\underline{b}_s^{*\prime}\,{}_j\underline{b}_i^*\,|{}_jM_s|\,{}_j\gamma_{si};$$

$${}_j\gamma_{si} = {}_j\underline{b}_i^{*\prime}\,{}_jP_s^*\,{}_j\underline{b}_s^*\ ;$$

$$ {}_j\lambda_s = {}_j\underline{b}_s^{*\prime}\, {}_jP_s^*\, {}_j\underline{b}_s^*; $$

$$ f_s = |{}_jM_s|(1 - {}_j\lambda_s); $$

(2·11) $$ \underline{0} = [(I - \sum_{k=1}^{s-1} {}_j\underline{b}_k^*\, {}_j\underline{b}_k^{*\prime})\, {}_jP_s^* - {}_j\lambda_s I]\, {}_j\underline{b}_s^* $$

(2·12) $$ = [(I - \sum_{k=1}^{s-1} {}_j\underline{b}_k^*\, {}_j\underline{b}_k^{*\prime})\, {}_jP_s^*(I - \sum_{k=1}^{s-1} {}_j\underline{b}_k^*\, {}_j\underline{b}_k^{*\prime}) - {}_j\lambda_s I]\, {}_j\underline{b}_s^*; $$

$i = 1, 2, \ldots, s-1$; $j = 1, 2, \ldots, m$.

(2·11) and (2·12) may be reformulated in terms of the original variables; then, for $j = 1, 2, \ldots, m$,

(2·13) $$ \underline{0} = [(\Sigma_{jj}^{-1} - \sum_{k=1}^{s-1} {}_j\underline{b}_k\, {}_j\underline{b}_k^{\prime})\, {}_jP_s - {}_j\lambda_s I]\, {}_j\underline{b}_s $$

(2·14) $$ = [\Sigma_{jj}^{1/2}(\Sigma_{jj}^{-1} - \sum_{k=1}^{s-1} {}_j\underline{b}_k\, {}_j\underline{b}_k^{\prime})\, {}_jP_s(\Sigma_{jj}^{-1} - \sum_{k=1}^{s-1} {}_j\underline{b}_k\, {}_j\underline{b}_k^{\prime})\Sigma_{jj}^{1/2} - {}_j\lambda_s I]\, \Sigma_{jj}^{1/2}\, {}_j\underline{b}_s. $$

In practice one would solve either (2·11) or (2·13). An iterative procedure for solving (2·11) will be described. The procedure, with only minor modification accounting for the transformation of variables, is equally applicable to (2·13).

The following argument should provide some insight into the procedure. Suppose $\underline{1}^{\prime} D_{B^*(s)}$ and $\underline{\lambda}_s$ comprise an optimal solution of (2·11). Then, for any j, ${}_j\lambda_s$ and ${}_j\underline{b}_s^*$ are consistent with a canonical analysis of

$$ \begin{pmatrix} I & {}_jN_s^* \\ {}_jN_s^{*\prime} & {}_jM_s \end{pmatrix} $$

subject to (s-1) independent linear homogeneous constraints -- defined by (s-1) independent rows of $\sum_{k=1}^{s-1} {}_j\underline{b}_k^*\, {}_j\underline{b}_k^{*\prime}$ -- on the coefficient vectors of the first set (cf. theorem 1·2). This may be seen by rewriting (2·12) in the form

$$(2\cdot15)\qquad {}_j\lambda_s = \sup_{\substack{\underline{b}^{*\prime} \varepsilon V_{P_j+1-s} \\ \underline{b}^{*\prime}\underline{b}^* = 1}} \underline{b}^{*\prime}\,{}_jP^*_s\,\underline{b}^*$$

where $V_{P_j+1-s} = V(I - \sum_{k=1}^{s-1} {}_j\underline{b}^*_k\,{}_j\underline{b}^{*\prime}_k)$

and $\underline{b}^*$ is in the same eigen space as $(I - \sum_{k=1}^{s-1} {}_j\underline{b}^*_k\,{}_j\underline{b}^{*\prime}_k)\,{}_j\underline{b}^*_s$.

Also $(I - \sum_{k=1}^{s-1} {}_j\underline{b}^*_k\,{}_j\underline{b}^{*\prime}_k)\,{}_j\underline{b}^*_i = \underline{0}$ for $i = 1, 2, \ldots, s-1$;

and rank $(I - \sum_{k=1}^{s-1} {}_j\underline{b}^*_k\,{}_j\underline{b}^{*\prime}_k) = \mathrm{tr}(I - \sum_{k=1}^{s-1} {}_j\underline{b}^*_k\,{}_j\underline{b}^{*\prime}_k)$

$$= P_j - \mathrm{tr}(\sum_{k=1}^{s-1} {}_j\underline{b}^*_k\,{}_j\underline{b}^{*\prime}_k)$$

$$= P_j - \mathrm{tr}(\sum_{k=1}^{s-1} {}_j\underline{b}^{*\prime}_k\,{}_j\underline{b}^*_k)$$

$$= P_j + 1 - s$$

{the first step is a consequence of the idempotency of $(I - \sum_{k=1}^{s-1} {}_j\underline{b}^*_k\,{}_j\underline{b}^{*\prime}_k)$}. The restriction $\underline{b}^{*\prime} \varepsilon V_{P_j+1-s}$ is equivalent to (2·3). A final observation is that $0 \leq {}_j\lambda_s < 1$ from which one may show that $0 < f_s \leq 1$.

Suppose now that variables have been chosen for the first (s-1) stages and consider the following iterative procedure for generating an optimal sth stage solution:

(i) specify initial variables ${}_jZ^{(o)}_s$ by vectors ${}_j\underline{b}^{*(o)}_s$ which satisfy orthogonality conditions like (2·2) and (2·3) for the required ${}_j\underline{b}^*_s$ vectors;

(ii) for n = 1, 2, ... solve

$$[(I - \sum_{k=1}^{s-1} {}_j\underline{b}^*_k\,{}_j\underline{b}^{*\prime}_k)\,{}_jP^{*(n)}_s - {}_j\lambda^{(n)}_s I]\,{}_j\underline{b}^{*(n)}_s = \underline{0};\ j = 1,2,\ldots,m;$$

obtaining the largest eigen value ${}_j\lambda_s^{(n)}$ and an associated (normalized) eigen vector ${}_j\underline{b}_s^{*(n)}$.

${}_jP_s^{*(n)}$ and ${}_jM_s^{(n)}$ (used below) are computed as ${}_jP_s^*$ and ${}_jM_s$ using ${}_1\underline{b}_s^{*(n)}, \ldots, {}_{j-1}\underline{b}_s^{*(n)}, {}_{j+1}\underline{b}_s^{*(n-1)}, \ldots, {}_m\underline{b}_s^{*(n-1)}$. The generalized variance of ${}_1Z_s^{(n)}, \ldots, {}_jZ_s^{(n)}, {}_{j+1}Z_s^{(n-1)}, \ldots, {}_mZ_s^{(n-1)}$ is then

$${}_jf_s^{(n)} = |{}_jM_s^{(n)}|(1 - {}_j\lambda_s^{(n)}).$$

This procedure, as will be shown, must converge (monotonically in ${}_jf_s^{(n)}$) to a solution of (2·11). The solution will be optimal if the initial variables are appropriately chosen. A sufficient condition for the solution to be optimal is for the generalized variance of the ${}_jZ_s^{(o)}$ variables to be less than the generalized variance of any non-optimal ${}_jZ_s$ variables yielding a solution to (2·11).

The nature of the calculations is such that (cf. theorems 1·2 and 1·3(iii))

$${}_1f_s^{(1)} \geq {}_2f_s^{(1)} \geq \ldots \geq {}_mf_s^{(1)} \geq {}_1f_s^{(2)} \geq \ldots .$$

Thus since ${}_jf_s^{(n)}$ is positive

(2·17) $$\lim_{n\to\infty} {}_jf_s^{(n)} = \tilde{f}_s(\text{say}),\ j = 1, 2, \ldots, m.$$

Let $${}_j\hat{\lambda}_s^{(n)} = {}_j\underline{b}_s^{*(n-1)'}\ {}_jP_s^{*(n)}\ {}_j\underline{b}_s^{*(n-1)}$$

and $${}_of_s^{(n)} = {}_mf_s^{(n)} .$$

Then $$\lim_{n\to\infty}\left(\frac{{}_{j-1}f_s^{(n)}}{{}_jf_s^{(n)}}\right) = \lim_{n\to\infty}\left\{\frac{|{}_jM_s^{(n)}|(1 - {}_j\hat{\lambda}_s^{(n)})}{|{}_jM_s^{(n)}|(1 - {}_j\lambda_s^{(n)})}\right\}$$
$$= 1,$$

(2·18) and $\lim_{n\to\infty} ({}_j\lambda_s^{(n)} - {}_j\hat{\lambda}_s^{(n)}) = 0$ (since $0 < 1 - {}_j\lambda_s^{(n)} \leq 1$)

for j = 1, 2, ..., m. Hence, in the limit, ${}_j\hat{\lambda}_s^{(n)} = {}_j\lambda_s^{(n)}$ and ${}_j\underline{b}_s^{*(n-1)}$ and ${}_j\underline{b}_s^{*(n)}$ are in the same eigen space -- the space corresponding to the largest eigen value ${}_j\lambda_s^{(n)}$ of ${}_jP_s^{*(n)}$.

Suppose ${}_j\tilde{Z}_s$ generated by ${}_j\underline{\tilde{b}}_s^*$, j = 1, 2, ..., m, have generalized variance $\tilde{f}_s$. Use ${}_1\underline{\tilde{b}}^*_s$, ${}_2\underline{\tilde{b}}^*_s$, ..., ${}_m\underline{\tilde{b}}^*_s$ to form ${}_j\tilde{M}_s$, ${}_j\tilde{P}^*_s$, and $D_{\tilde{B}^*(s)}$.

Let $\quad {}_j\tilde{\lambda}_s = {}_j\underline{\tilde{b}}_s^{*\prime}\, {}_j\tilde{P}_s\, {}_j\underline{\tilde{b}}_s^*$;

then $\quad \tilde{f}_s = |{}_j\tilde{M}_s|(1 - {}_j\lambda_s)$, j = 1, 2, ..., m.

Furthermore $\underline{1}' D_{\tilde{B}^*(s)}$ and $\underline{\tilde{\lambda}}$ comprise a solution to the partial derivative equations (2·11) since $\tilde{f}_s$ cannot be decreased by changing any one of the ${}_j\underline{\tilde{b}}_s^*$.

In practice -- where ${}_j\lambda_s^{(n)}$ will always be of multiplicity one -- ${}_j\underline{b}_s^{*(n-1)}$ will equal ${}_j\underline{b}_s^{*(n)}$, apart from a scalar factor (-1), in the limit. With no loss in generality the ${}_j\underline{b}_s^{*(n)}$ may be adjusted in sign so that

(2·19) $$\lim_{n\to\infty} {}_j\underline{b}_s^{*(n)} = {}_j\underline{\tilde{b}}_s^* \text{ (to be used as above),}$$

(2·20) $$\lim_{n\to\infty} {}_jM_s^{(n)} = {}_j\tilde{M}_s,$$

(2·21) $$\lim_{n\to\infty} {}_j\lambda_s^{(n)} = {}_j\tilde{\lambda}_s,$$

(2·17) holds, and $\underline{1}' D_{\tilde{B}^*(s)}$ and $\underline{\tilde{\lambda}}$ form a solution of (2·11).

The Sum of Squares of Correlations Criterion

The SSQCOR method is developed in much the same way as the GENVAR method. The main changes are in the definition of the criterion, the interpretation of the Lagrange multipliers, and the composition of the matrices of the "eigen value" equations.

Redefining some notation from the last section,

$$f_i = \mathrm{tr}\Phi^2_{(i)} - m$$

$$= \sum_{\substack{j=1 \\ j\neq k}}^{m} \sum_{k=1}^{m} ({}_j\underline{b}^{*\prime}_i R_{jk}\, {}_k\underline{b}^*_i)^2,$$

$${}_jP_i = {}_jN_i\, {}_jN_i^{\prime},$$

and $${}_jP^*_i = {}_jN^*_i\, {}_jN^{*\prime}_i \text{ for } i = 1, 2, \ldots, P_1 \text{ and } j = 1, 2, \ldots, m.$$

At the sth stage ($1 \leq s \leq P_1$) the Lagrangian equation

is $$g_s = f_s + 2\sum_{j=1}^{m} (1 - {}_j\underline{b}^{*\prime}_s\, {}_j\underline{b}^*_s)\, {}_j\lambda_s - 4\sum_{i=1}^{s-1} \sum_{j=1}^{m} {}_j\underline{b}^{*\prime}_s\, {}_j\underline{b}^*_i\, {}_j\gamma_{si}.$$

$$\frac{\partial g_s}{\partial\, {}_j\underline{b}^*_s} = 4\, {}_jP^*_s\, {}_j\underline{b}^*_s - 4\, {}_j\lambda_s\, {}_j\underline{b}^*_s - 4\sum_{i=1}^{s-1} {}_j\gamma_{si}\, {}_j\underline{b}^*_i$$

which, equating to zero, gives

$$\underline{0} = {}_jP^*_s\, {}_j\underline{b}^*_s - {}_j\lambda_s\, {}_j\underline{b}^*_s - \sum_{i=1}^{s-1} {}_j\underline{b}^*_i\, {}_j\gamma_{si},$$

$${}_j\gamma_{si} = {}_j\underline{b}^{*\prime}_i\, {}_jP^*_s\, {}_j\underline{b}^*_s$$

$$= \sum_{\substack{k=1 \\ k\neq j}}^{m} ({}_j\underline{b}^{*\prime}_i R_{jk}\, {}_k\underline{b}^*_s)^2,$$

$${}_j\lambda_s = {}_j\underline{b}^{*\prime}_s\, {}_jP^*_s\, {}_j\underline{b}^*_s$$

$$= \sum_{\substack{k=1 \\ k\neq j}}^{m} ({}_j\underline{b}^{*\prime}_s R_{jk}\, {}_k\underline{b}^*_s)^2,$$

and $f_s = \underline{1}'\underline{\lambda}_s$; i = 1, 2, ..., s-1; j = 1, 2, ..., m.

The key equations (2·11) - (2·15) are also valid here with, of course, the new definitions of ${}_jP_s$ and ${}_jP^*_s$. To obtain an optimal solution one finds some $\underline{1}'D_{B^*(s)}$ (or $\underline{1}'D_{B(s)}$) and $\underline{\lambda}_s$ such that (2·11) [or (2·13)] holds with f_s as large as possible. The iterative procedure described in the last section may be used to produce these solutions.

In this context

$$\begin{aligned} {}_j\lambda_s^{(n)} &= {}_j\underline{b}_s^{*(n)\prime}\,{}_jP_s^{*(n)}\,{}_j\underline{b}_s^{*(n)} \\ &= \sum_{k=1}^{j-1}({}_j\underline{b}_s^{*(n)\prime}R_{jk}\,{}_k\underline{b}_s^{*(n)})^2 + \sum_{k=j+1}^{m}({}_j\underline{b}_s^{*(n)\prime}R_{jk}\,{}_k\underline{b}_s^{*(n-1)})^2 \end{aligned}$$

and $${}_jf_s^{(n)} = \mathrm{tr}({}_jM_s^{(n)})^2 - m + 1 + 2\,{}_j\lambda_s^{(n)},\; j = 1, 2, \ldots, m.$$

The calculations guarantee that (cf. theorem 1·3(i))

$${}_1f_s^{(1)} \le {}_2f_s^{(1)} \le \ldots \le {}_mf_s^{(1)} \le {}_1f_s^{(2)} \le \ldots .$$

Since $0 \le {}_jf_s^{(n)} < m(m-1)$, equation (2·17) is valid here.

Also ${}_jf_s^{(n)} - {}_{j-1}f_s^{(n)} = 2({}_j\lambda_s^{(n)} - {}_j\hat{\lambda}_s^{(n)})$ so that (2·18) holds too.

Now it is apparent that the SSQCOR iterative procedure has convergence properties like those mentioned for the GENVAR procedure. In particular (2·19), (2·20), and (2·21) hold, as well as (2·17); $\underline{1}'D_{\tilde{B}^*(s)}$ and $\underline{\tilde{\lambda}}$ make up a solution of (2·11). A sufficient condition for the solution to be optimal is for $\sum_{\substack{i=1\\i\neq j}}^{m}\sum_{j=1}^{m}\{\mathrm{corr}({}_iZ_s^{(o)}, {}_jZ_s^{(o)})\}^2$ to be greater than the corresponding sum for any sub-optimal ${}_jZ_s$ variables connected with a solution of (2·11).

Some Calculations and Comparisons

Horst ([3] and [4]; see also [5]) illustrated his methods using data from Thurstone and Thurstone [4]. The variables consisted of nine test items broken down into three sets of three items each. Each item measured one of three abilities, and each set contained one measure of each ability. Horst transformed the original Σ matrix into

R =

1.000	0.000	0.000	0.636	0.126	0.059	0.626	0.195	0.059
0.000	1.000	0.000	-0.021	0.633	0.049	0.035	0.459	0.129
0.000	0.000	1.000	0.016	0.157	0.521	0.048	0.238	0.426
0.636	-0.021	0.016	1.000	0.000	0.000	0.709	0.050	-0.002
0.126	0.633	0.157	0.000	1.000	0.000	0.039	0.532	0.190
0.059	0.049	0.521	0.000	0.000	1.000	0.067	0.258	0.299
0.626	0.035	0.048	0.709	0.039	0.067	1.000	0.000	0.000
0.195	0.459	0.238	0.050	0.532	0.258	0.000	1.000	0.000
0.059	0.129	0.426	-0.002	0.190	0.299	0.000	0.000	1.000

The iterative procedure for the SSQCOR criterion was applied to R using five different sets of initial vectors:

(i) equal weight on all variables

(2·22) $$ {}_j\underline{b}_1^{*(o)\prime} = (0.5735,\ 0.5735,\ 0.5735),\ j = 1, 2, 3; $$

(ii) all weight on first variables

(2·23) $$ {}_j\underline{b}_1^{*(o)\prime} = (1.0,\ 0.0,\ 0.0),\ j = 1, 2, 3; $$

(iii) all weight on last variables

(2·24) $$ {}_j\underline{b}_1^{*(o)\prime} = (0.0,\ 0.0,\ 1.0),\ j = 1, 2, 3; $$

(iv) Horst's second stage (s=2) solution to (2·4)

(2·25)
$$\begin{cases} {}_1\underline{b}_1^{*(o)\prime} = (-0.6806,\ 0.5743,\ 0.4550); \\ {}_2\underline{b}_1^{*(o)\prime} = (-0.7524,\ 0.5557,\ 0.3536); \\ {}_3\underline{b}_1^{*(o)\prime} = (-0.7324,\ 0.5477,\ 0.4045); \end{cases}$$

(v) Horst's third stage (s=3) solution to (2·4)

(2·26)
$$\begin{cases} {}_1\underline{b}_1^{*(o)\prime} = (0.0228,\ -0.6372,\ 0.7703); \\ {}_2\underline{b}_1^{*(o)\prime} = (-0.0122,\ -0.5485,\ 0.8361); \\ {}_3\underline{b}_1^{*(o)\prime} = (0.0604,\ -0.5395,\ 0.8398). \end{cases}$$

The iterative procedure for the GENVAR criterion was tested using (2·22) as a starting point. The iterative procedure was terminated in each case as soon as $\sum_{j=1}^{3} \left| {}_j\lambda_1^{(n)} - {}_j\lambda_1^{(n-1)} \right| < 0.0001$.

The values of ${}_3f_1^{(n)}$ are shown in Table 1. (The SSQCOR figures in this table are subject to error in the sixth decimal place.) Table 2 displays the final ${}_j\underline{b}_1^{*(n)}$ vectors including those obtained by Horst for the SUMCOR criterion. Table 3 gives the values of the three criteria for each of the solutions listed in Table 2.

It is noteworthy that the five different sets of initial vectors generated essentially the same solution to the SSQCOR problem. And apparently this solution is an optimal one. [To see that the procedure does not always work independent of the initial vectors, apply the SSQCOR procedure to (2·5) starting with (2·7).]

In fact the solutions are nearly the same for all of the criteria. This is not surprising in view of the high positive correlations on the diagonals of the R_{ij} blocks and the generally low absolute correlations elsewhere in the R_{ij} blocks.

Computations for this example were done on an IBM 360 Model 75 computer. The time required to compile the program and investigate three or four different starting points was about one minute.

Table 1

Values of ${}_3f_1^{(n)}$

Criterion:	SSQCOR	SSQCOR	SSQCOR	SSQCOR	SSQCOR	GENVAR
Starting Point:	(2·22)	(2·23)	(2·24)	(2·25)	(2·26)	(2·22)
n = 1	3.303480	2.799538	2.118418	2.046118	0.635250	0.165024
2	3.324824	3.051978	3.044878	2.047182	0.686164	0.162227
3	3.328192	3.209252	3.235992	2.049262	1.686334	0.161835
4	3.329200	3.281344	3.293624	2.054866	3.182884	0.161694
5	3.329578	3.311030	3.315908	2.069552	3.323596	0.161640
6	3.329720	3.322640	3.324514	2.107198	3.329160	0.161619
7	3.329778	3.327086	3.327802	2.197832	3.329620	0.161611
8	3.329796	3.328780	3.329044	2.385994	3.329738	0.161608
9	3.329804	3.329422	3.329522	2.677556	3.329786	0.161607
10	3.329810	3.329666	3.329700	2.972386	3.329798	0.161607
11	3.329810	3.329752	3.329770	3.167168	3.329802	0.161606
12		3.329792	3.329792	3.263168	3.329808	
13		3.329802	3.329804	3.303786	3.329810	
14		3.329808	3.329808	3.319844		
15		3.329808	3.329808	3.326020		
16		3.329806		3.328374		
17				3.329264		
18				3.329602		
19				3.329734		
20				3.329780		
21				3.329800		
22				3.329808		
23				3.329804		
24				3.329806		

Table 2

Final Iterated Vectors

Criterion:	SUMCOR	SSQCOR	SSQCOR	SSQCOR	SSQCOR	SSQCOR	GENVAR
Starting Point:	--	(2·22)	(2·23)	(2·24)	(2·25)	(2·26)	(2·22)
$_1\underline{b}_1^*$	0.7323	0.7338	0.7347	0.7338	0.7338	0.7348	0.7371
	0.5139	0.5123	0.5115	0.5124	0.5123	0.5115	0.5076
	0.4468	0.4462	0.4456	0.4462	0.4461	0.4455	0.4460
$_2\underline{b}_1^*$	0.6586	0.6603	0.6612	0.6603	0.6604	0.6613	0.6636
	0.6247	0.6233	0.6226	0.6233	0.6233	0.6226	0.6197
	0.4195	0.4189	0.4185	0.4190	0.4189	0.4185	0.4192
$_3\underline{b}_1^*$	0.6781	0.6795	0.6802	0.6795	0.6796	0.6803	0.6811
	0.6395	0.6383	0.6378	0.6383	0.6383	0.6378	0.6361
	0.3621	0.3616	0.3612	0.3617	0.3616	0.3612	0.3626

Table 3

Comparison of Criteria Values

Solution Criterion:	Value of SUMCOR	Value of SSQCOR	Value of GENVAR
SUMCOR	4.4693	3.32926	0.161680
SSQCOR (2·17)	4.4695	3.32981	0.161615
SSQCOR (2·18)	4.4695	3.32981	0.161613
SSQCOR (2·19)	4.4695	3.32981	0.161615
SSQCOR (2·20)	4.4695	3.32981	0.161615
SSQCOR (2·21)	4.4695	3.32981	0.161613
GENVAR	4.4694	3.32976	0.161606

Appendix

Lemma A1 Let A and B be (p x p) matrices with A symmetric and B positive definite. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ be the eigen values of $B^{-1/2}AB^{-1/2}$ and $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_p$ a corresponding set of orthonormal eigen vectors. Then

$$\sup_{\underline{y} \neq \underline{o}} \frac{\underline{y}'A\underline{y}}{\underline{y}'B\underline{y}} = \lambda_1 \quad \text{and} \quad \sup_{\substack{\underline{y} \neq \underline{o} \\ \underline{e}_i'B^{1/2}\underline{y}=o \\ i=1,2,\ldots,k}} \frac{\underline{y}'A\underline{y}}{\underline{y}'B\underline{y}} = \lambda_{k+1} \text{ for } k = 1,2,\ldots,p-1.$$

proof

$$\sup_{\underline{y} \neq \underline{o}} \frac{\underline{y}'A\underline{y}}{\underline{y}'B\underline{y}} = \sup_{\underline{x} \neq \underline{o}} \frac{\underline{x}'B^{-1/2}AB^{-1/2}\underline{x}}{\underline{x}'\underline{x}} .$$

$$\sup_{\substack{\underline{y} \neq \underline{o} \\ \underline{e}_i'B^{1/2}\underline{y}=o \\ i=1,2,\ldots,k}} \frac{\underline{y}'A\underline{y}}{\underline{y}'B\underline{y}} = \sup_{\substack{\underline{x} \neq \underline{o} \\ \underline{e}_i'\underline{x}=o \\ i=1,2,\ldots,k}} \frac{\underline{x}'B^{-1/2}AB^{-1/2}\underline{x}}{\underline{x}'\underline{x}} .$$

The lemma now follows from [11], 1f.2.1 and 1f.2.3, page 50.

Lemma A2 Let R be any (p x q) matrix with $p \leq q$. Then there exist orthogonal matrices Γ and Δ such that

$$\Gamma R \Delta' = (D \quad 0)$$

where $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$

is of dimension (p x p) and $\lambda_1^2, \lambda_2^2, \ldots, \lambda_p^2$ are the eigen values of RR'.

proof

See Vinograde [15].

References

[1] Anderson, T. W. (1958), Introduction to Multivariate Statistical Analysis. Wiley, New York.

[2] Fortier, J. J. (1966), "Simultaneous linear prediction," Psychometrika, 31, 369-381.

[3] Horst, Paul (1961), "Relations among m sets of measures," Psychometrika, 26, 129-149.

[4] ______ (1961), "Generalized canonical correlations and their applications to experimental data," J. Clinical Psych. (monograph supplement), 14, 331-347.

[5] ______ (1965), Factor Analysis of Data Matrices. Holt, Rinehart, and Winston, New York.

[6] Hotelling, Harold (1935), "The most predictable criterion," J. Educ. Psych., 26, 139-142.

[7] ______ (1936), "Relations between two sets of variates," Biometrika, 28, 321-377.

[8] Lancaster, H. O. (1966), "Kolmogorov's remark on the Hotelling canonical correlations," Biometrika, 53, 585-588.

[9] McKeon, J. J. (1966), "Canonical analysis: some relations between canonical correlation, factor analysis, discriminant function analysis, and scaling theory," Psychometric Monographs, 13.

[10] Rao, C. Radhakrishna (1964), "Use and interpretation of principle components," Sankhyā, 26, 329-358.

[11] ______ (1965), Linear Statistical Inference and its Applications. Wiley, New York.

[12] Roy, S. N. (1957), Some Aspects of Multivariate Analysis. Wiley, New York.

[13] Steel, Robert G. D. (1951), "Minimum generalized variance for a set of linear functions," Am. Math. Statist., 22, 456-460.

[14] Thurstone, L. L. and Thurstone, Thelma Gwinn (1941), "Factorial studies of intelligence," Psychometric Monographs, 2.

[15] Vinograde, Bernard (1950), "Canonical positive definite matrices under internal linear transformations," Proc. Amer. Math. Soc., 1, 159-161.

[16] Wilks, S. S. (1935), "On the independence of k sets of normally distributed statistical variables," Econometrica, 3, 309-326.