CANONICAL ANALYSIS OF SEVERAL SETS OF VARIABLES

by

JON ROBERTS KETTENRING
Department of Statistics

University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 615

APRIL 1969

The research was partially supported by the National Institutes of Health, Institute of General Medical Science, Grant No. GM-12865-05.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

It is a special pleasure to thank my adviser, Professor N. L. Johnson, for his guidance during the course of this research. Through numerous discussions, he has been a vital source of inspiration and encouragement.

Much of the work on Chapters II and III was accomplished at Bell Telephone Laboratories. While there, I received numerous helpful comments and suggestions from Dr. J. D. Carroll and Dr. R. Gnanadesikan.

I also would like to thank the other members of my doctoral committee, Professor I. M. Chakravarti, Professor R. L. Davis, Professor J. E. Grizzle, and Professor P. K. Sen, for their interest in my work.

In addition, I am grateful to Mr. James O. Kitchen for providing extensive advice and assistance on the computer programs which were required.

INTRODUCTION AND SUMMARY

The main purpose of this study is to investigate ways of extending the theory of canonical correlation or canonical analysis to deal with more than two sets of random variables.

The first chapter contains some new approaches to the classical concepts and theory for two nonsingular sets of variables. In addition, a number of results which are required for the proper treatment of more than two sets are also presented. The results in Section 1.5, on the minimal conditions for canonical analysis, and Section 1.6, on principal component models for pairs of canonical variables, are especially important in this connection.

The two set theory is further developed in the second chapter to cover situations where one or both of the sets may be singular. It is shown, in particular, that the analysis can be carried out without eliminating the singularities. The key theorem is given in Section 2.3, and three applications of it follow in the ensuing sections.

Five different techniques for the canonical analysis of several sets of variables are considered in Chapters III and IV. Two of these are due to Horst ([8][1], [9], and [10]), a third is due to Steel [21], and the remaining two are new.

---

1 The numbers in square brackets refer to the bibliography.

One important feature of these methods is that they all reduce to the classical procedure when the number of sets is only two. A second important common feature is that each method calls for the selection of (canonical) variables, one from each set, according to a criterion of optimizing some function of their correlation matrix.

It appears that the first attempt to extend the theory of canonical analysis to three or more sets was made by Vinograde [23]. His method, however, does not possess the above mentioned second feature and hence does not fall within the framework of this study.

In Chapter III, models of the general principal component type, extensions of those in Section 1.6, are constructed for each of the five methods. The models are useful in that they help to clarify the types of effects which the methods can detect.

The actual procedures for determining the canonical variables according to the five techniques are developed in Chapter IV. Three of the procedures turn out to be iterative in nature. A favorable property of these procedures is that each must converge. Furthermore, if the starting points are appropriately chosen, the derived variables will necessarily be the desired canonical variables. (Suggestions for starting points are made in Section 5.3.)

Horst's "maximum correlation" method (see [8] or [10]) is one of those which requires an iterative procedure. He, too, proposed an iterative procedure for finding the canonical variables, but his differs from the one recommended here (see Section 4.2). A strong drawback of Horst's procedure is that its convergence properties are not known.

Steel [21], using compound matrices, developed a complicated system of equations which one could solve, at least in theory, to obtain the canonical variables for his method. The equations in fact are difficult to handle; and, consequently, a new approach has been evolved (cf. Section 4.4). The resulting equations can be solved by an iterative procedure which is quite like the one proposed for the maximum correlation technique.

Three examples are given in the last chapter to illustrate the various methods. Some specific practical suggestions are also included.

CHAPTER I

RELATIONS BETWEEN TWO NONSINGULAR SETS OF VARIABLES

1.1 Preliminary Remarks.

The classical problem of identifying and measuring relations between two sets of variables was first tackled by Hotelling in a brief 1935 paper [11]. His definitive study of the problem appeared the next year in [12] and still stands as a key reference in multivariate statistical literature.

An example taken from [12] should help to pin down the problem which is being considered. In this example, a set of mental test variables are compared with a set of physical measurement variables. "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."

Hotelling's method is based on an analysis of the covariance matrix of the two sets of variates; and, consequently, the relations which can be found using it are necessarily of a linear type.

This chapter contains a rather thorough investigation of his method as it would be applied to an arbitrary positive definite covariance matrix. The notation needed to proceed further is set down in the next section. The focus in the third section shifts back to a description of the analysis. The next three sections contain technical results associated with Hotelling's technique. The last two sections include additional results which are relevant to some special types of relations between the sets. Much of this chapter has been motivated by and is directed towards problems to be taken up in Chapters III and IV.

1.2 Notation.

Let the two sets of random variables be

$$\underline{X}_1' = ({}_1X_1, {}_1X_2, \ldots, {}_1X_{p_1}) \quad \text{and} \quad \underline{X}_2' = ({}_2X_1, {}_2X_2, \ldots, {}_2X_{p_2}) \quad \text{with} \quad p_1 \leq p_2 .$$

Adjoining these together produces a $(p \times 1)$ vector

$$\underline{X}' = (\underline{X}_1', \underline{X}_2') \quad \text{with} \quad p = p_1 + p_2 .$$

The covariance matrix of $\underline{X}$, assumed positive definite throughout this chapter, is

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{matrix} p_1 \\ p_2 \end{matrix} = ((\sigma_{ij})) .$$
$$\qquad\quad p_1 \quad p_2$$

The rank of Σ_{12} is designated by r.

It is often convenient to make "preliminary" nonsingular linear transformations of the following type within each set:

$$\underline{Y}_1 = T_1^{-1}\underline{X}_1 \quad \text{and} \quad \underline{Y}_2 = T_2^{-1}\underline{X}_2$$

where T_1 and T_2 are matrices such that

$$\Sigma_{11} = T_1T_1' \quad \text{and} \quad \Sigma_{22} = T_2T_2'.$$

Let

$$\underline{Y}' = (\underline{Y}_1', \underline{Y}_2');$$

then the transformations can be expressed more concisely as

$$\underline{Y} = D_T^{-1}\underline{X}$$

where

$$D_T = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix}.$$

(T_1 and T_2 may be taken to be lower triangular, a computationally expedient form, or symmetric, in which case it is convenient to write $\Sigma_{ii}^{\frac{1}{2}}$ for T_i.) The covariance matrix of $\underline{Y}$ is

$$R = \begin{pmatrix} I & R_{12} \\ R_{21} & I \end{pmatrix} = D_T^{-1}\Sigma D_T^{-1\prime}.$$

Much of what is to follow will concern

$$\underline{Z}_i = B_i^* \underline{Y}_i = B_i \underline{X}_i, \qquad i = 1, 2,$$

where

$$\underline{Z}_i = \begin{pmatrix} {}_iZ_1 \\ {}_iZ_2 \\ \vdots \\ {}_iZ_{p_i} \end{pmatrix}, \quad B_i^* = \begin{pmatrix} {}_i\underline{b}_1^{*\prime} \\ {}_i\underline{b}_2^{*\prime} \\ \vdots \\ {}_i\underline{b}_{p_i}^{*\prime} \end{pmatrix}, \quad \text{and} \quad B_i = \begin{pmatrix} {}_i\underline{b}_1' \\ {}_i\underline{b}_2' \\ \vdots \\ {}_i\underline{b}_{p_i}' \end{pmatrix},$$

with B_i^* an orthogonal matrix. Writing

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

it follows that

$$\underline{Z} = D_{B*}\underline{Y} = D_B\underline{X}$$

where

$$D_{B*} = \begin{pmatrix} B_1^* & 0 \\ 0 & B_2^* \end{pmatrix} \quad \text{and} \quad D_B = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}.$$

The relation between D_{B*} and D_B is

$$D_{B*} = D_B D_T .$$

The covariance matrix of $\underline{Z}$ is

$$(1.2.1) \qquad \Phi = \begin{pmatrix} I & \Phi_{12} \\ \Phi_{21} & I \end{pmatrix} = D_{B*} R D_{B*}' = D_B \Sigma D_B'.$$

Relations between the two sets are most commonly studied through the analysis of an appropriately selected $\underline{Z}$. Special notation will not be introduced to distinguish $\underline{Z}$, or any of the quantities depending on $\underline{Z}$, when $\underline{Z}$ is the chosen one. Where a distinction is necessary, it should be clear from the context.

Most attention will be given to the pairs of variables,

$$\underline{Z}_{(i)}' = ({}_1Z_i, {}_2Z_i), \qquad i = 1, 2, \ldots, p_1.$$

The covariance matrix of $\underline{Z}_{(i)}$ is

$$\Phi_{(i)} = \begin{pmatrix} 1 & \phi_i \\ \phi_i & 1 \end{pmatrix}.$$

Denote the ordered eigenvalues of $\Phi_{(i)}$ by

$$ {}_1\lambda_i \geq {}_2\lambda_i > 0, \quad {}_1\lambda_i + {}_2\lambda_i = 2, $$

and the corresponding orthonormal eigenvectors by

$$ {}_1\underline{e}_i \text{ and } {}_2\underline{e}_i . $$

When discussing the variables $\underline{Z}_{(i)}$, there are implicitly in the background some $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(i-1)}$ such that the covariance structure of the i pairs taken together is compatible with (1.2.1).

Use will also be made of variables

$$ \underline{\tilde{Z}}'_{(i)} = ({}_1\tilde{Z}_i, {}_2\tilde{Z}_i), \quad i = 1, 2, \ldots, p_1 $$

which are general unit variance linear compounds of $\underline{X}_1$ and $\underline{X}_2$ respectively.

Additional notation will be introduced as needed.

Some special symbols to be used throughout are listed below:

$\|\|A\|\|$	for the Euclidean norm of a rectangular matrix A;
$\|A\|$	for the determinant of a square matrix A;
A^-	for a generalized inverse of a rectangular matrix A;
$tr(A)$	for the trace of a square matrix A;
$c_j(A)$	for the j-th largest eigenvalue of a matrix A with real eigenvalues;
$V(A)$	for the vector space generated by the columns of the matrix A (i.e., the range space of A);
$N(A)$	for the null space of the matrix A;
$\dim[V]$	for the dimension of the vector space V;
$E(\underline{X})$	for the expectation of $\underline{X}$;

var($\underline{X}$)	for the covariance matrix of $\underline{X}$;
corr($\underline{X}$, $\underline{Y}$)	for the correlation matrix of $\underline{X}$ and $\underline{Y}$;
cov($\underline{X}$, $\underline{Y}$)	for the covariance matrix of $\underline{X}$ and $\underline{Y}$;
$\underline{1}$	for a vector of ones;
δ_{ij}	for the Kronecker delta.

1.3 Introduction to Canonical Analysis

The formal extraction procedure referred to in Section 1.1 will be loosely called a "canonical analysis." It is convenient to describe the procedure in a sequential fashion. At the first stage, one looks for a pair $({}_1Z_1, {}_2Z_1)$ with the largest possible correlation. The derived variables are known as canonical variables and together they constitute the first pair of canonical variables. The associated correlation, denoted by ρ_1, 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 variate representing the criterion set corresponds to Hotelling's most predictable criterion [11].

The analysis may be continued until p_1 pairs of canonical variates and p_1 canonical correlations have been found. At the i-th stage $(i = 2, 3, \ldots, p_1)$, the i-th pair of canonical variates[1] $\underline{Z}_{(i)}$ is chosen so that the corresponding i-th canonical correlation, designated ρ_i, is the maximum correlation obtainable which is consistent with

(1.3.1) $\quad \text{corr}({}_1Z_j, {}_1Z_i) = 0, \quad j = 1, 2, \ldots, i-1;$

(1.3.2) $\quad \text{corr}({}_2Z_j, {}_2Z_i) = 0, \quad j = 1, 2, \ldots, i-1;$

(1.3.3) $\quad \text{corr}({}_1Z_j, {}_2Z_i) = 0, \quad j = 1, 2, \ldots, i-1;$

[1] The frequently appearing statement that $\underline{Z}_{(i)}$ is the i-th pair of canonical variates is an imprecise way of saying that $\underline{Z}_{(i)}$ is a particular one of the admissible i-th pairs of canonical variates.

(1.3.4) $\quad \mathrm{corr}({}_1Z_i, {}_2Z_j) = 0, \quad j = 1, 2, \ldots, i-1.$

It is not necessary to impose all of these constraints in order to arrive at the i-th pair of canonical variates. The derivations which follow the sequential formation of canonical variates, as outlined above, generally impose (1.3.1) and (1.3.2) in the process of determining the higher order pairs (see, for example, Anderson [1] or Lancaster [15]) and then proceed to show that (1.3.3) and (1.3.4) are also satisfied. In fact, even these two commonly used constraints ((1.3.1) and (1.3.2)) are more than is usually required. This point will be dealt with further in Section 1.5.

If $p_1 < p_2$, a formal completion of the analysis may be accomplished by finding an additional $(p_2 - p_1)$ canonical variables ${}_2Z_j$, $j = p_1 + 1, p_1 + 2, \ldots, p_2$, associated with the second set.[2] The only conditions imposed on these variables are that they be uncorrelated with each of the $2p_1$ other canonical variables and among themselves. Although it is the canonical pairs $\underline{Z}_{(i)}$ which receive most of the attention, the other $(p_2 - p_1)$ canonical variables can also be enlightening. In particular, the two sets $({}_1X_1, {}_1X_2, \ldots, {}_1X_{p_1})$ and $({}_2Z_{p_1+1}, {}_2Z_{p_1+2}, \ldots, {}_2Z_{p_2})$ are always uncorrelated.

The results of Section 1.4 establish that

(1.3.5) $$\sum_{i=1}^{p_1} \rho_i^2 = ||R_{12}||^2.$$

The elements of R_{12} are all that can be explained by the canonical analysis. (1.3.5) indicates in what sense the statement, "the ensemble of canonical variables accounts for all existing relations between the two sets," is valid.

[2] Any $\underline{Z}$ containing the p_1 canonical pairs plus these (p_2-p_1) additional canonical variables is referred to as a canonical $\underline{Z}$.

The number of non-zero canonical correlations, which is the same as r, the rank of Σ_{12}, R_{12}, or Φ_{12}, indicates the number of dimensions or factors common to both sets. The nature of these dimensions or factors can be clarified in at least two ways. One approach is given by Rao ([20], p.496): he shows that the minimum number of common factors needed to explain $\underline{Z}_1$ and $\underline{Z}_2$ separately (the same common factors being used in both representations) by the usual factor analysis model (see Rao ([20], p.498)) is just r. In Section 1.6, each canonical pair $\underline{Z}_{(i)}$ is expressed in terms of a one or two principal component factor model. It is only the first r pairs which admit interesting representations of this type. With respect to the single principal component factor model, the number of non-trivial factors, i.e., the number of $\underline{Z}_{(i)}$ with non-trivial representations in terms of one factor, is again r.

An important feature of a canonical analysis is that the end products, namely the canonical variates and correlations, are invariant under nonsingular transformations of either set; this point was made by Hotelling [12] and has been used advantageously by Horst ([8], [9], and [10]), Steel [21], and others. It is, therefore, legitimate to make preliminary transformations from $\underline{X}_1$ and $\underline{X}_2$ to $\underline{Y}_1$ and $\underline{Y}_2$ in order to achieve orthonormal bases for the two sets of variables. In terms of $\underline{Y}$, the canonical analysis is exposed as nothing more than the selection of orthogonal transformations (B_1^* and B_2^*) to new orthonormal bases, $\underline{Z}_1$ and $\underline{Z}_2$, which have a particularly simple correlational structure.

1.4 The Basic Theorem of Canonical Analysis[3]

Theorem 1.4.1. can be fairly described as the basic theorem of canonical analysis because, together with its corollary, it provides the essential information for carrying out the canonical analysis calculations. The theorem establishes the existence of a transformation D_B or D_{B*} which induces a unique Φ matrix with the property that the elements of Φ_{12} are all zero except along its main diagonal where they are non-negative, less than one, and in descending order of magnitude. The underlying $\underline{Z}$ must be a canonical vector as is demonstrated in Theorem 1.4.2.

THEOREM 1.4.1. Suppose Σ is positive definite. Then there exists a matrix

$$(1.4.1) \qquad D_B = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}$$

and a unique matrix

$$(1.4.2) \qquad \Phi = D_B \Sigma D_B'$$

such that

$$(1.4.3) \qquad \Phi = \begin{pmatrix} I & \Phi_{12} \\ \Phi_{21} & I \end{pmatrix},$$

with

$$(1.4.4) \qquad \Phi_{12} = (\underset{p_1}{D} \quad \underset{p_2-p_1}{0}\ )\ p_1 = \Phi_{21}'$$

and

[3] The material in this section is included mainly for completeness. The proof of Theorem 1.4.2, however, has a new twist which is useful for the considerations in Section 1.5.

(1.4.5) $$D = \text{diag}(\rho_1, \rho_2, \ldots, \rho_{p_1}),$$
with
$$1 > \rho_1 \geq \rho_2 \geq \ldots \geq \rho_{p_1} \geq 0.$$

($\rho_1^2, \rho_2^2, \ldots, \rho_{p_1}^2$ must be the eigenvalues of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.) Also $B_1 = B_1^* T_1^{-1}$ where B_1^* is any orthogonal matrix such that

(1.4.6) $$D^2 = B_1^* T_1^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}T_1^{-1\prime}B_1^{*\prime},$$

and $B_2 = B_2^* T_2^{-1}$, where B_2^* is any orthogonal matrix satisfying

(1.4.7) $$\Phi_{12}B_2^* = B_1^* T_1^{-1}\Sigma_{12}T_2^{-1\prime}.$$

(1.4.7) determines the first r rows of B_2^*, r being the number of non-zero eigenvalues ρ_j^2; the remaining $(p_2 - r)$ rows are any orthogonal completion of B_2^*.

Proof.
$$R = D_T^{-1}\Sigma D_T^{-1\prime} = \begin{pmatrix} I & R_{12} \\ R_{21} & I \end{pmatrix} \quad \text{where} \quad R_{12} = T_1^{-1}\Sigma_{12}T_2^{-1\prime} = R_{21}'.$$
Using Lemma A2,[4] there exist orthogonal matrices B_1^* and B_2^* such that $B_1^* R_{12} B_2^{*\prime} = \Phi_{12}$ with Φ_{12} as defined in (1.4.4) and (1.4.5). The ρ_j^2, according to Lemma A2, are the eigenvalues of $R_{12}R_{21}$. Moreover, B_1^* may be any orthogonal matrix such that

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

Note that the eigenvalues of $R_{12}R_{21}$ are the same as those of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Let $D_B = D_{B*}D_T^{-1}$. This definition of D_B yields a Φ matrix of the type specified in (1.4.2) - (1.4.5). Suppose now that $D_{\hat{B}}$ is any other matrix of the form given in (1.4.1) which yields a matrix $\hat{\Phi}$ of the same type. Then

[4] The "A" indicates that the lemma is in the appendix.

$$\hat{\Phi}_{12} = \hat{B}_1 \Sigma_{12} \hat{B}_2'$$
$$= (\hat{B}_1 T_1) T_1^{-1} \Sigma_{12} T_2^{-1}{}' (\hat{B}_2 T_2)'.$$
$$= \hat{B}_1^* R_{12} \hat{B}_2^*{}'.$$

Therefore

$$D^2 = \hat{B}_1^* R_{12} R_{21} \hat{B}_1^*{}',$$

where $\hat{B}_1^*$ ($= \hat{B}_1 T_1$) is an orthogonal matrix. Hence the ρ_j^2 must be the eigenvalues of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and there is one and only one Φ matrix of the prescribed form. The properties of B_2^* follow from the two equations

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

and

$$B_2^* R_{21} R_{12} B_2^*{}' = \begin{pmatrix} D^2 & 0 \\ 0 & 0 \end{pmatrix} \begin{matrix} p_1 \\ p_2 \end{matrix} .$$
$$\qquad\qquad\qquad\qquad\quad p_1 \quad p_2$$

The first corollary contains alternative expressions, in terms of the original $\underline{X}$ variables, for parts of the previous theorem.

COROLLARY 1.4.1.1. B_1 is any matrix such that

(1.48) $$D^2 = B_1\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}B_1' \quad \text{and} \quad B_1\Sigma_{11}B_1' = I.$$

Alternatively, the rows of B_1 are any complete set of eigenvectors of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ such that the i-th row, ${}_1\underline{b}_i$, of B_1 is associated with the i-th largest eigenvalue, ρ_i^2, and $B_1\Sigma_{11}B_1' = I$. The first r rows of B_2 are determined by the equation

(1.4.9) $$B_1\Sigma_{12}\Sigma_{22}^{-1} = \Phi_{12}B_2,$$

and the remaining $(p_2 - r)$ rows are chosen so that

(1.4.10) $$B_2 \Sigma_{22} B_2' = I.$$

Proof. Suppose (1.4.8) holds. Then

$$D^2 = (B_1 T_1)(T_1^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} T_1^{-1\prime})(B_1 T_1)' .$$

$B_1 T_1$ is orthogonal and thus may be taken as B_1^* in the theorem. Suppose now that

(1.4.11) $$\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} B_1' = B_1' D^2 \quad \text{and} \quad B_1 \Sigma_{11} B_1' = I.$$

But (1.4.11) holds if and only if

$$T_1^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} T_1^{-1\prime} (B_1 T_1)' = (B_1 T_1)' D^2 \quad \text{and} \quad (B_1 T_1)(B_1 T_1)' = I.$$

Again $B_1 T_1$ is orthogonal and, because of (1.4.6), may be taken as B_1^* in the theorem. The statement involving (1.4.9) and (1.4.10) is an immediate consequence of the theorem.

THEOREM 1.4.2. Any $\underline{Z}$ with covariance matrix equal to the unique Φ of Theorem 1.4.1 is a vector of canonical variates.

Proof. That $({}_1Z_1, {}_2Z_1)$ may be taken as the first pair of canonical variates follows from Lancaster [15] or Anderson ([1], p.295). (Both use an argument like the one used below for the other pairs.) Now for $i = 2, 3, \ldots, r$ in turn select $({}_1\tilde{Z}_i, {}_2\tilde{Z}_i)$ to maximize $\mathrm{corr}({}_1\tilde{Z}_i, {}_2\tilde{Z}_i)$ subject to (1.3.1) where ${}_1\tilde{Z}_i = \underline{\alpha}'\underline{Z}_1$, $\underline{\alpha}' = (\alpha_1, \alpha_2, \ldots, \alpha_{p_1})$, $||\underline{\alpha}|| = 1$, ${}_2\tilde{Z}_i = \underline{\beta}'\underline{Z}_2$, $\underline{\beta}' = (\beta_1, \beta_2, \ldots, \beta_{p_2})$, and $||\underline{\beta}|| = 1$. In other words,

$\sum_{j=i}^{r} \alpha_j\beta_j\rho_j$ is to be maximized subject to $\alpha_j = 0$, $j = 1, 2, \ldots, i-1$.

Arguing as in Anderson ([1], p.295), let $(\gamma_i, \gamma_{i+1}, \ldots, \gamma_{p_2}) = (\alpha_i\rho_i, \alpha_{i+1}\rho_{i+1}, \ldots, \alpha_r\rho_r, 0, \ldots, 0)$. Using the Cauchy-Schwarz inequality, $\left|\sum_{j=i}^{r} \alpha_j\beta_j\rho_j\right| = \left|\sum_{j=i}^{p_2} \gamma_j\beta_j\right| \le (\sum_{j=i}^{p_2} \gamma_j^2)^{\frac{1}{2}} (\sum_{j=i}^{p_2} \beta_j^2)^{\frac{1}{2}}$ with equality if and only if $(\gamma_i, \gamma_{i+1}, \ldots, \gamma_{p_2}) \propto (\beta_i, \beta_{i+1}, \ldots, \beta_{p_2})$. Assuming this equality condition, $\beta_j = 0$, $j = r+1, r+2, \ldots, p_2$. Changing the sign of all the α's, if necessary, one is left with

(1.4.12)
$$\sum_{j=i}^{r} \alpha_j\beta_j\rho_j = C \sum_{j=i}^{p_1} \alpha_j^2\rho_j^2 = C \{ \sum_{j=i}^{p_1} (\rho_j^2 - \rho_i^2)\, \alpha_j^2 + \alpha_i^2\} \text{ for some } C > 0 .$$

If $\rho_i = \rho_{i+1} = \cdots = \rho_k$ (and $\rho_k > \rho_{k+1}$ if $k < p_1$), then to maximize (1.4.12) it is necessary that $\alpha_j = 0$, $j = k+1, k+2, \ldots, p_1$ and hence $\beta_j = 0$, $j = k+1, k+2, \ldots, r$. Clearly $\beta_j = 0$, $j = 1, 2, \ldots, i-1$, in addition, so that (1.3.2) holds. Evidently

$$(1.4.13)\quad \begin{cases} {}_1\tilde{Z}_i = \sum_{j=i}^{k} \alpha_j\, {}_1Z_j, & \sum_{j=i}^{k} \alpha_j^2 = 1, \\ {}_2\tilde{Z}_i = \sum_{j=i}^{k} \beta_j\, {}_2Z_j, & \sum_{j=i}^{k} \beta_j^2 = 1, \\ \alpha_j = \beta_j, \quad j = i, i+1, \ldots, k , & \\ \text{and} & \\ \operatorname{corr}({}_1\tilde{Z}_i, {}_2\tilde{Z}_i) = \rho_i. & \end{cases}$$

Take $\alpha_i = \beta_i = 1$ and note that conditions (1.3.1) - (1.3.4) are satisfied so that $\underline{Z}_{(i)}$ is the i-th pair of canonical variates. (A similar proof holds starting from (1.3.2) instead of (1.3.1).) For $r < p_1$,

$({}_1Z_{r+1}, {}_1Z_{r+2}, \ldots, {}_1Z_{p_1})$ is uncorrelated with $\underline{Z}_2$ which means that the (r+1)-st canonical correlation is zero. In view of the structure of Φ, this is enough for $\underline{Z}$ to be a canonical vector.

The composition of an arbitrary i-th pair of canonical variates, $i \le r$, is spelled out in the corollary.

COROLLARY 1.4.2.1. Suppose $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(i-1)}$ are fixed pairs of canonical variates for $i \le r$. Suppose $\underline{Z}^*_{(i)}, Z^*_{(i+1)}, \ldots, \underline{Z}^*_{(k)}$ are also pairs of canonical variates with $\rho_i = \rho_{i+1} = \ldots = \rho_k$ and $\rho_k > \rho_{k+1}$ if $k < p_1$. Then any i-th pair of canonical variates must be of the form

$$\underline{Z}_{(i)} = \sum_{j=i}^{k} \alpha_j \underline{Z}^*_{(j)}, \qquad \sum_{j=i}^{k} \alpha_j^2 = 1 .$$

Proof. The proof for $i = 2, 3, \ldots, r$ follows from (1.4.13). The case $i = 1$ can be handled by a similar argument.

1.5 Minimal Conditions for Canonical Analysis

This section contains results pertaining to the effects of restrictions (1.3.1) - (1.3.4) on the generation of higher order canonical variates. The effects of two alternative constraints are also considered.

THEOREM 1.5.1. Let $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(i-1)}$ be the first (i-1) pairs of canonical variates with $i \le r+1$. Then restrictions (1.3.1) and (1.3.4) on the i-th pair of variates are equivalent, as are the restrictions (1.3.2) and (1.3.3). Furthermore, for $i \le r$, variables ${}_1\tilde{Z}_i$ and ${}_2\tilde{Z}_i$ having the maximum correlation, subject to only one of the constraints (1.3.1) - (1.3.4), must in fact satisfy all four and thus may be identified as the i-th pair of canonical variates.

Proof. Let $\underline{Z}_1$ and $\underline{Z}_2$ contain the elements of $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(i-1)}$ in their proper places. Form ${}_1\tilde{Z}_i = \underline{\alpha}'\underline{Z}_1$ and ${}_2\tilde{Z}_i = \underline{\beta}'\underline{Z}_2$ as in the proof of Theorem 1.4.2. Requiring that ${}_1\tilde{Z}_i$ satisfy (1.3.1) is equivalent to having $\alpha_j = 0$, $j = 1, 2, \ldots, i-1$. Requiring that ${}_1\tilde{Z}_i$ satisfy (1.3.4) is equivalent to having $\alpha_j\rho_j = 0$, $j = 1, 2, \ldots, i-1$ which is the same as $\alpha_j = 0$, $j = 1, 2, \ldots, i-1$ since $i \le r+1$. This demonstrates that (1.3.1) and (1.3.4) are equivalent for $i \le r+1$ and a similar argument proves the equivalence of (1.3.2) and (1.3.3), again for $i \le r+1$. The proof of the second part of the theorem is contained in the proof of Theorem 1.4.2.

One constraint of the type (1.3.1) - (1.3.4) will not suffice to determine canonical variates in the case $r < i \le p_1$. Suppose, as one possibility, that ${}_1\tilde{Z}_i$ and ${}_2\tilde{Z}_i$ have the maximum correlation subject only to (1.3.1). Then an admissible solution is $({}_1\tilde{Z}_i, {}_2\tilde{Z}_i) = ({}_1Z_i, {}_2Z_1)$ where ${}_1Z_i$ and ${}_2Z_1$ are canonical variates. But this is clearly not the i-th canonical pair. Perhaps the most natural way of dealing with this case is to impose the two constraints (1.3.1) and (1.3.2). One then has the following well known result.

THEOREM 1.5.2. Let $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(i-1)}$ be the first (i-1) pairs of canonical variates with $r < i \le p_1$. Then any $\underline{Z}_{(i)}$ may be taken as the i-th pair of canonical variates (in other words, any $\underline{\tilde{Z}}_{(i)}$ subject to (1.3.1) and (1.3.2)).

Proof. This follows from Theorems 1.4.1 and 1.4.2 noting that there exists a $\underline{Z}$ with covariance matrix Φ as in (1.4.2) - (1.4.5) and containing $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(i)}$ in the proper locations.

Another type of constraint can be employed to generate $\min(p_1, r+1)$ canonical pairs.

THEOREM 1.5.3. Suppose $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(i-1)}$ are the first (i-1) pairs of canonical variates. Then the variables ${}_1\tilde{Z}_i$ and ${}_2\tilde{Z}_i$ with the maximum correlation subject to the constraint

$$\underline{1}'\operatorname{corr}(\underline{Z}_{(j)}, \underline{\tilde{Z}}_{(i)})\underline{1} = 0, \quad j = 1, 2, \ldots, i-1 \tag{1.5.1}$$

must be the i-th canonical pair for $i \le \min(p_1, r+1)$ (but need not be in case $r + 1 < i \le p_1$).

Proof. Start with ${}_1\tilde{Z}_i = \underline{\alpha}'\underline{Z}_1$ and ${}_2\tilde{Z}_i = \underline{\beta}'\underline{Z}_2$ exactly as in the proof of Theorem 1.4.2.

$$\underline{1}'\operatorname{corr}(\underline{Z}_{(j)}, \underline{\tilde{Z}}_{(i)})\underline{1} = (\alpha_j + \beta_j)(1 + \rho_j). \tag{1.5.2}$$

Thus (1.5.1) holds if and only if $\alpha_j = -\beta_j$, $j = 1, 2, \ldots, i-1$. Then

$$\operatorname{corr}({}_1\tilde{Z}_i, {}_2\tilde{Z}_i) = -\sum_{j=1}^{i-1} \alpha_j^2\rho_j + \sum_{j=i}^{p_1} \alpha_j\beta_j\rho_j . \tag{1.5.3}$$

If $i \le r + 1$, $\alpha_j = \beta_j = 0$, $j = 1, 2, \ldots, i-1$ for a maximum. Using Theorems 1.5.1 and 1.5.2, it may now be inferred that $\underline{\tilde{Z}}_{(i)}$ must be the i-th pair of canonical variates. If $i > r + 1$, then $\rho_{i-1} = 0$ and the maximum value of (1.5.3), namely zero, may be obtained even with $\alpha_{i-1} \neq 0$. Consequently, $\underline{\tilde{Z}}_{(i)}$ need not be the i-th pair of canonical variates in this case.

A constraint similar to (1.5.1) can be used with the same effect.

THEOREM 1.5.4. Theorem 1.5.3 remains correct if (1.5.1) is replaced by

$$(1.5.4) \qquad (1, -1)\operatorname{corr}(\underline{Z}_{(j)}, \tilde{\underline{Z}}_{(i)})\begin{pmatrix}1\\-1\end{pmatrix} = 0, \quad j = 1, 2, \ldots, i-1.$$

Proof. The proof is the same as that for Theorem 1.5.3 except that (1.5.2) is replaced by

$$(1, -1)\operatorname{corr}(\underline{Z}_{(j)}, \tilde{\underline{Z}}_{(i)})\begin{pmatrix}1\\-1\end{pmatrix} = (\alpha_j + \beta_j)(1 - \rho_j).$$

1.6. Principal Component Models for Pairs of Canonical Variates

The concepts of canonical analysis and principal component analysis can be brought together in a way which gives new insight into the structure of the canonical variates. This approach is extremely useful in developing and relating methods for the canonical analysis of several sets of variables which is the subject of Chapter III.

The first theorem of this section and its corollary link the canonical variates to the principal components of $\underline{Y}$.

THEOREM 1.6.1. The eigenvalues of R are

$$c_j(R) = \begin{cases} 1 + \rho_j, & j = 1, 2, \ldots, p_1 \\ 1, & j = p_1+1, p_1+2, \ldots, p_2 \\ 1 - \rho_{p+1-j}, & j = p_2+1, p_2+2, \ldots, p \end{cases}.$$

If $\underline{v}_j$, $j = 1, 2, \ldots, p$, are corresponding orthonormal eigenvectors, then $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_r, \underline{v}_{p-r+1}, \underline{v}_{p-r+2}, \ldots, \underline{v}_p$ must be of the form

$$\underline{v}_j' = \begin{cases} 2^{-\frac{1}{2}}({}_1\underline{b}_j^{*\prime},\ {}_2\underline{b}_j^{*\prime}), & j = 1, 2, \ldots, r \\ 2^{-\frac{1}{2}}(-\ {}_1\underline{b}_j^{*\prime},\ {}_2\underline{b}_j^{*\prime}), & j = p-r+1, p-r+2, \ldots, p\ ; \end{cases}$$

and $\underline{v}_{r+1}, \underline{v}_{r+2}, \ldots, \underline{v}_{p-r}$ can (but need not) be of the form

$$\underline{v}_j' = \begin{cases} 2^{-\frac{1}{2}}({}_1\underline{b}_j^{*\prime},\ {}_2\underline{b}_j^{*\prime}), & j = r+1, r+2, \ldots, p_1 \\ (\underline{0}',\ {}_2\underline{b}_j^{*\prime}), & j = p_1+1, p_1+2, \ldots, p_2 \\ 2^{-\frac{1}{2}}(-\ {}_1\underline{b}_j^{*\prime},\ {}_2\underline{b}_j^{*\prime}), & j = p_2+1, p_2+2, \ldots, p-r\ . \end{cases}$$

(The ${}_1\underline{b}_j^*$ and ${}_2\underline{b}_j^*$ are the rows of B_1^* and B_2^* which define the canonical variates $\underline{Z}_1 = B_1^*\underline{Y}_1$ and $\underline{Z}_2 = B_2^*\underline{Y}_2$.)

Proof. Let $V = (\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_p)$ where the $\underline{v}_j$ are as defined above, and let $\Lambda = \mathrm{diag}(I + D, I, I - D)$ where D is as in (1.4.5). Now

(1.6.1) $$RV = V\Lambda \iff B_1^* R_{12} B_2^{*\prime} = (D\ 0).$$

The second equation in (1.6.1) holds by virtue of (1.4.7) so that the eigenvalues of R are as claimed, and the $\underline{v}_j$ form a complete set of orthonormal eigenvectors. The proof that the first r of the $\underline{v}_j$ and the last r of the $\underline{v}_j$ must be of the prescribed form follows readily from Corollary 1.4.2.1.

COROLLARY 1.6.1.1. The first r principal components of $\underline{Y}$ must be of the form

$$\underline{v}_j'\underline{Y} = 2^{-\frac{1}{2}}({}_1Z_j + {}_2Z_j), \quad j = 1, 2, \ldots, r\ ;$$

and the last r principal components must be of the form

$$\underline{v}_j'\underline{Y} = 2^{-\frac{1}{2}}(-{}_1Z_j + {}_2Z_j), \quad j = p-r+1, p-r+2, \ldots, p\ .$$

The corresponding variances are the respective $c_j(R)$. (The ${}_iZ_j$ are canonical variates.)

Okamoto and Kanazawa [18] present a development of principal component theory which is more general in scope than any previous approach. Their main result is recorded in Lemma A3. It can be used to further develop the work of Carroll [3]. Paraphrasing Carroll's central idea, the sum of squared sample correlations between each of the ${}_1Z_1$ and ${}_2Z_1$ sample vectors, standardized to zero sample mean and unit sample variance, and a third vector, at one's disposal, is a maximum with respect to the third vector when this third vector is proportional to the sum of the other two and overall when the sample vectors are canonical. Suppose fixed ${}_1Z_1$ and ${}_2Z_1$ sample vectors and an optimal third vector are used as the coordinates of three points in Euclidean space. Then the line passing through the origin and the third point is the first principal component line (cf. Gower [7]) for the other two points (i.e., those corresponding to the sample vectors). A random variable development of Carroll's work, in the context of principal components, is possible with the aid of Lemma A3.

Suppose $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(i-1)}$ are the first (i-1) pairs of canonical variates. Assume, for convenience, that $E(\underline{Z}_{(i)}) = \underline{0}$ and $\mathrm{corr}({}_1Z_i, {}_2Z_i) = \phi_i \geq 0$. Consider as a representation of an arbitrary $\underline{Z}_{(i)}$,

$$\underline{Z}_{(1)} = \underline{\ell}_i F_i + \underline{E}_i , \qquad (1.6.2)$$

where $\underline{\ell}_i$ is a non-null vector, F_i is a standardized (zero mean and unit variance) random variable, and $\underline{E}_i$ is an error vector. $\underline{\ell}_i$ and F_i

are to be chosen to produce the best fitting model for $\underline{Z}_{(i)}$ according to the following criterion:

(1.6.3) $$\text{minimize } f(\text{var}(\underline{E}_i)) = g(\theta_1, \theta_2),$$

where g is any strictly increasing function in θ_1 and θ_2, the eigenvalues of the non-negative definite matrix in the argument of f. Two possiblities for f are the trace and the Euclidean norm functions.

According to Lemma A3, optimal choices (for all f satisfying the above condition) are

(1.6.4) $$\begin{cases} \underline{\ell}_i = {}_1\lambda_i^{\frac{1}{2}}\ {}_1\underline{e}_i = \begin{cases} \{(1+\phi_i)/2\}^{\frac{1}{2}}\ \underline{1} & \text{if } \phi_i > 0 \\ {}_1\underline{e}_i & \text{if } \phi_i = 0 \end{cases} \\ \text{and} \\ F_i = {}_1\lambda_i^{-\frac{1}{2}}\ {}_1\underline{e}_i'\underline{Z}_{(i)} = \begin{cases} \{2(1+\phi_i)\}^{-\frac{1}{2}}\ \underline{1}'\underline{Z}_{(i)} & \text{if } \phi_i > 0 \\ {}_1\underline{e}_i'\underline{Z}_{(i)} & \text{if } \phi_i = 0, \end{cases} \end{cases}$$

in which case

(1.6.5) $$g(\theta_1, \theta_2) = g(1 - \phi_i, 0) = g({}_2\lambda_i, 0) = g(2 - {}_1\lambda_i, 0)\ .$$

The $\underline{Z}_{(i)}$ admitting the best fit of (1.6.2) in the sense of (1.6.3) must be the i-th pair of canonical variates since $(1 - \phi_i)$ is a minimum when ϕ_i is a maximum. This maximum is, of course, ρ_i.

Consider next a two factor model for $\underline{Z}_{(i)}$

(1.6.6) $$\underline{Z}_{(i)} = {}_1\underline{\ell}_i\ {}_1F_i + {}_2\underline{\ell}_i\ {}_2F_i + \underline{E}_i\ ,$$

where ${}_1\underline{\ell}_i$ and ${}_2\underline{\ell}_i$ are non-null vectors, ${}_1F_i$ and ${}_2F_i$ are standardized random variables, and $\underline{E}_i$ is an error vector. Clearly the $\underline{\ell}$'s and F's can be chosen so that $\underline{E}_i = \underline{0}$. This however is to be done in a special way: first, pick ${}_1\underline{\ell}_i$ and ${}_1F_i$ to

$$\text{(1.6.7)} \qquad \text{minimize } f(\text{var}(\underline{Z}_{(i)} - {}_1\underline{\ell}_i\, {}_1F_i)) = g(\theta_1, \theta_2);$$

second, pick ${}_2\underline{\ell}_i$ and ${}_2F_i$ to

$$\text{(1.6.8)} \qquad \text{minimize } f(\text{var}(\underline{Z}_{(i)} - {}_1\underline{\ell}_i\, {}_1F_i - {}_2\underline{\ell}_i\, {}_2F_i)) = g(\theta_1, \theta_2).$$

In both instances, g is the function defined in (1.6.3).

Applying Lemma A3, the choices are

$$\text{(1.6.9)} \begin{cases} {}_1\underline{\ell}_i' = {}_1\lambda_i^{\frac{1}{2}}\, {}_1\underline{e}_i' = \begin{cases} \{(1+\phi_i)/2\}^{\frac{1}{2}}(1,\ 1) & \text{if } \phi_i > 0 \\ {}_1\underline{e}_i' & \text{if } \phi_i = 0 \end{cases}; \\ {}_2\underline{\ell}_i = {}_2\lambda_i^{\frac{1}{2}}\, {}_2\underline{e}_i' = \begin{cases} \{(1-\phi_i)/2\}^{\frac{1}{2}}(1,\ -1) & \text{if } \phi_i > 0 \\ {}_2\underline{e}_i' & \text{if } \phi_i = 0 \end{cases}; \\ {}_1F_i = {}_1\lambda_i^{-\frac{1}{2}}\, {}_1\underline{e}_i'\underline{Z}_{(i)} = \begin{cases} \{2(1+\phi_i)\}^{-\frac{1}{2}}(1,\ 1)\underline{Z}_{(i)} & \text{if } \phi_i > 0 \\ {}_1\underline{e}_i'\underline{Z}_{(i)} & \text{if } \phi_i = 0 \end{cases}; \\ {}_2F_i = {}_2\lambda_i^{-\frac{1}{2}}\, {}_2\underline{e}_i'\underline{Z}_{(i)} = \begin{cases} \{2(1-\phi_i)\}^{-\frac{1}{2}}(1,\ -1)\underline{Z}_{(i)} & \text{if } \phi_i > 0 \\ {}_2\underline{e}_i'\underline{Z}_{(i)} & \text{if } \phi_i = 0 \end{cases}. \end{cases}$$

An external criterion is needed for the "optimal" selection of $\underline{Z}_{(i)}$. If f is the trace function, $g = 1 - \phi_i$ after the first fit (1.6.7) and $g = 0$ after the second fit (1.6.8). Translated into words, the first factor accounts for an amount ${}_1\lambda_i = 1 + \phi_i$ of the variability in $\underline{Z}_{(i)}$, and the second factor accounts for the remaining amount ${}_2\lambda_i = 1 - \phi_i$. A sensible approach is to choose $\underline{Z}_{(i)}$ to make ${}_1\lambda_i$ large and ${}_2\lambda_i$ small. This, after all, is the effect of model (1.6.2) as can be seen from (1.6.5). It will prove beneficial for future work to state the external criterion in the following two equivalent ways: choose $\underline{Z}_{(i)}$ to

(1.6.10) $$\text{maximize } {}_1\lambda_i^2 + {}_2\lambda_i^2 = 2(1 + \phi_i^2)$$

or

(1.6.11) $$\text{minimize } {}_1\lambda_i \, {}_2\lambda_i = 1 - \phi_i^2 .$$

Each leads to $\underline{Z}_{(i)}$ being the i-th pair of canonical variates.

These results are summarized below.

THEOREM 1.6.2. Let $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(i-1)}$ be the first (i-1) pairs of canonical variates. (Assume $E(\underline{Z}_{(i)}) = \underline{0}$ and $\phi_i \geq 0$.) The best fit of model (1.6.2), for arbitrary $\underline{Z}_{(i)}$, as measured by (1.6.3), is given by (1.6.4), and the $\underline{Z}_{(i)}$ yielding the best fit is the i-th canonical pair. The best fit of model (1.6.6), for arbitrary $\underline{Z}_{(i)}$, as measured by (1.6.7) and (1.6.8), is given by (1.6.9). The $\underline{Z}_{(i)}$ admitting the best fit using (1.6.10) or (1.6.11) is again the i-th canonical pair.

COROLLARY 1.6.2.1. $$\max_{\underline{Z}_{(i)}} {}_1\lambda_i = 1 + \rho_i .$$

$$\min_{\underline{Z}_{(i)}} {}_2\lambda_i = 1 - \rho_i .$$

COROLLARY 1.6.2.2. Let $\underline{\tilde{Z}}_{(i)}$ replace $\underline{Z}_{(i)}$ in model (1.6.2). (Assume $E(\underline{\tilde{Z}}_{(i)}) = \underline{0}$ and $\text{corr}({}_1\tilde{Z}_i, {}_2\tilde{Z}_i) = \tilde{\phi}_i \geq 0$.) Requiring that the best (cf. (1.6.3)) fitting factor F_i satisfy

$$\text{corr}(F_j, F_i) = 0, \quad j = 1, 2, \ldots, i-1$$

is equivalent to requiring that $\underline{\tilde{Z}}_{(i)}$ satisfy (1.5.1) if $\tilde{\phi}_i > 0$. This is also true for $\tilde{\phi}_i = 0$ if (and only if) ${}_1\underline{e}_i$ is chosen proportional to $\underline{1}$.

Proof. $\text{corr}(F_j, F_i) = 0 \iff \text{cov}(\underline{1}'\underline{Z}_{(j)}, \underline{1}'\underline{\tilde{Z}}_{(i)}) = 0$ $\iff \underline{1}'\text{corr}(\underline{Z}_{(j)}, \underline{\tilde{Z}}_{(i)})\underline{1} = 0$, $j = 1, 2, \ldots, i-1$. Note that $F_i \propto \underline{1}'\underline{\tilde{Z}}_{(i)}$ only when ${}_1\underline{e}_i \propto \underline{1}$.

COROLLARY 1.6.2.3. Let $\underline{\tilde{Z}}_{(i)}$ and $\tilde{\phi}_i$ be as in Corollary 1.6.2.2. Requiring that the best (cf. (1.6.3) fitting factor F_i be such that

$$\text{corr}(\underline{E}_j, \underline{E}_i) = 0, \quad j = 1, 2, \ldots, i-1$$

is equivalent to requiring that $\underline{\tilde{Z}}_{(i)}$ satisfy (1.5.4) if $\tilde{\phi}_i > 0$. This is also true for $\tilde{\phi}_i = 0$ if (and only if) ${}_1\underline{e}_i \propto \underline{1}$ (which implies ${}_2\underline{e}_i' \propto (1, -1)$).

Proof. $\text{corr}(\underline{E}_j, \underline{E}_i) = 0 \iff \text{cov}((1, -1)\underline{Z}_{(j)}, (1, -1)\underline{\tilde{Z}}_{(i)}) = 0$ $\iff (1, -1)\text{corr}(\underline{Z}_{(j)}, \underline{\tilde{Z}}_{(i)})(1, -1)' = 0$, $j = 1, 2, \ldots, i-1$. Note that $\underline{E}_i' \propto (1, -1)(1, -1)\underline{Z}_{(i)}$ if and only if ${}_1\underline{e}_i \propto \underline{1}$.

COROLLARY 1.6.2.4. Let $\underline{\tilde{Z}}_{(i)}$ replace $\underline{Z}_{(i)}$ in model (1.6.6). (Assume $E(\underline{\tilde{Z}}_{(i)}) = \underline{0}$ and $\text{corr}({}_1\tilde{Z}_i, {}_2\tilde{Z}_i) \geq 0$.) Requiring that the best (cf. (1.6.7) and (1.6.8)) fitting factors ${}_1F_i$ and ${}_2F_i$ satisfy

$$\text{corr}({}_uF_j, {}_vF_i) = 0, \quad u = 1, 2; \quad v = 1, 2; \quad j = 1, 2, \ldots, i-1$$

is equivalent to requiring that $\underline{\tilde{Z}}_{(i)}$ satisfy

$$\text{corr}(\underline{Z}_{(j)}, \underline{\tilde{Z}}_{(i)}) = 0, \quad j = 1, 2, \ldots, i-1$$

(i.e., conditions (1.3.1) - (1.3.4)). This is true for $i = 2,3,\ldots,p_1$.

Proof.

$$\begin{pmatrix} {}_1F_j \\ {}_2F_j \end{pmatrix} = A_j\underline{Z}_{(j)}, \quad j = 1, 2, \ldots, i-1, \quad \text{and} \quad \begin{pmatrix} {}_1F_i \\ {}_2F_i \end{pmatrix} = A_i\underline{\tilde{Z}}_{(i)}$$

for some nonsingular $A_1, A_2, \ldots, A_i$. Now

$\text{corr}(A_j\underline{Z}_{(j)}, A_i\underline{\tilde{Z}}_{(i)}) = 0 \iff \text{corr}(\underline{Z}_{(j)}, \underline{\tilde{Z}}_{(i)}) = 0,$

$j = 1, 2, \ldots, i-1; \quad i = 2, 3, \ldots, p_1.$

1.7 Some Invariants under Transformations of One Set

The next theorem is the main force behind certain iterative procedures developed in Chapter IV, but it is also of interest here for the further insight it gives into the relations between two sets of variates. In essence, one set of variables is held fixed and attention is directed to some special functions which are invariant under internal nonsingular transformations of the other set.

Let the second set $\underline{X}_2$ be fixed and consider the following functions of $\underline{Z}_1$:

$$(1.7.1) \quad \begin{cases} \alpha_i = \sum_{j=1}^{p_2} \{\text{corr}({}_1Z_i, {}_2X_j)\}^2, & i = 1, 2, \ldots, p_1, \\ \beta_i = \{\sum_{j=1}^{p_2} \text{corr}({}_1Z_i, {}_2X_j)\}^2, & i = 1, 2, \ldots, p_1, \\ \text{and} & \\ \gamma_i = \left|\text{var}\begin{pmatrix} {}_1Z_i \\ \underline{X}_2 \end{pmatrix}\right|, & i = 1, 2, \ldots, p_1. \end{cases}$$

The goal, for $i = 1, 2, \ldots, p_1$ in turn, is to select different ${}_1Z_i$ to (i) maximize α_i, (ii) maximize β_i, and (iii) minimize γ_i. The technique for doing this and the values attained are presented in Theorem 1.7.1. (Part (i) of this theorem was discussed by Fortier [6] and earlier by Rao [19].)

<u>THEOREM 1.7.1.</u> The optimal values for α_i, β_i, and γ_i defined in (1.7.1) are

(i) $\alpha_i = c_i(\Sigma_{11}^{-1}\Sigma_{12}D_{22}^{-1}\Sigma_{21}), \quad i = 1, 2, \ldots, p_1$;

(ii) $\beta_1 = \underline{1}' D_{22}^{-\frac{1}{2}}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}D_{22}^{-\frac{1}{2}}\underline{1}$

and

$\beta_i = 0, \qquad i = 2, 3, \ldots, p_1$;

(iii) $\gamma_i = |\Sigma_{22}|(1 - c_i(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})), \quad i = 1, 2, \ldots, p_1$;

where

$$D_{22} = \mathrm{diag}(\sigma_{p_1+1,p_1+1}^{\frac{1}{2}}, \sigma_{p_1+2,p_1+2}^{\frac{1}{2}}, \ldots, \sigma_{pp}^{\frac{1}{2}}).$$

These values are attained when the ${}_1\underline{b}_i$ defining the ${}_1Z_i$ are the eigenvectors, subject to $B_1\Sigma_{11}B_1' = I$, corresponding to the i-th largest eigenvalues of $\Sigma_{11}^{-1}\Sigma_{12}D_{22}^{-1}\Sigma_{21}$, $\Sigma_{11}^{-1}\Sigma_{12}D_{22}^{-\frac{1}{2}}\underline{1}\ \underline{1}'D_{22}^{-\frac{1}{2}}\Sigma_{21}$, and $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ respectively. For $\beta_1 > 0$, the corresponding eigenvector is $\beta_1^{-1}\Sigma_{11}^{-1}\Sigma_{12}D_{22}^{-\frac{1}{2}}\underline{1}$.

<u>Proof</u>. Repeated use will be made of Lemma A1. The maximum value of α_1 is

$$\sup_{{}_1\underline{b}_1} {}_1\underline{b}_1' \Sigma_{12}D_{22}^{-1}\Sigma_{21}\, {}_1\underline{b}_1 = c_1(\Sigma_{11}^{-1}\Sigma_{12}D_{22}^{-1}\Sigma_{21}) ,$$

and occurs when ${}_1\underline{b}_1$ is a corresponding eigenvector of $\Sigma_{11}^{-1}\Sigma_{12}D_{22}^{-1}\Sigma_{21}$. The maximum value of α_2 is

$$\sup_{\substack{{}_1\underline{b}_2 \\ {}_1\underline{b}_1' \Sigma_{11}\, {}_1\underline{b}_2 = 0}} {}_1\underline{b}_2' \Sigma_{12}D_{22}^{-1}\Sigma_{21}\, {}_1\underline{b}_2 = c_2(\Sigma_{11}^{-1}\Sigma_{12}D_{22}^{-1}\Sigma_{21})$$

and occurs when ${}_1\underline{b}_2$ is a corresponding eigenvector satisfying ${}_1\underline{b}_1' \Sigma_{11}\, {}_1\underline{b}_2 = 0$. Continuing in this way, the maximum value of α_{p_1} is

$$\sup_{\substack{{}_1\underline{b}_{p_1} \\ {}_1\underline{b}_j' \Sigma_{11}\, {}_1\underline{b}_{p_1} = 0 \\ j = 1, 2, \ldots, p_1-1}} {}_1\underline{b}_{p_1}' \Sigma_{12}D_{22}^{-1}\Sigma_{21}\, {}_1\underline{b}_{p_1} = c_{p_1}(\Sigma_{11}^{-1}\Sigma_{12}D_{22}^{-1}\Sigma_{21})$$

and occurs when ${}_1\underline{b}_{p_1}$ is a corresponding eigenvector satisfying ${}_1\underline{b}_j' \Sigma_{11}\, {}_1\underline{b}_{p_1} = 0$, $j = 1, 2, \ldots, p_1-1$. The results for the β_i and γ_i follow in a similar manner. To obtain the ${}_1Z_i$ yielding the maximal β_i, one needs to consider the variation of

$${}_1\underline{b}_i' \Sigma_{12}D_{22}^{-\frac{1}{2}}\underline{1}\,\underline{1}'D_{22}^{-\frac{1}{2}}\Sigma_{21}\, {}_1\underline{b}_i \,.$$

And for the ${}_1Z_i$ giving the minimal γ_i, it is the variation of

$$\begin{vmatrix} 1 & {}_1\underline{b}_i' \Sigma_{12} \\ \Sigma_{21}\, {}_1\underline{b}_i & \Sigma_{22} \end{vmatrix} = |\Sigma_{22}| - |\Sigma_{22}|\, {}_1\underline{b}_i' \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\, {}_1\underline{b}_i$$

which is of interest. Minimizing this last expression is equivalent to maximizing

$${}_1\underline{b}_i' \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\, {}_1\underline{b}_i$$

and this may be carried out using Lemma A1.

The minimal γ_i are especially noteworthy because of their intimate relation to the canonical correlations. Moreover, the corresponding ${}_1Z_i$ are canonical variables for the first set. Part (iii) of the theorem is in essence an application of the Cauchy-Schwarz inequality.

1.8 Modified Pairs of Canonical Variates

The restrictions (1.3.1) - (1.3.4) are quite stringent in that they leave the experimenter with little control over the extracted summary variables which he must analyze. Typically, it is very difficult, if not impossible, to attach meaning to the resulting higher order canonical pairs (and sometimes to the first pair as well). Another awkward point is that canonical variables derived from an estimated covariance matrix will usually be correlated in the population, when not members of the same pair, even though uncorrelated in the sample.

The purpose of this section is to consider one particular way of adding flexibility to the types of higher order summary variables at one's disposal. The resulting "modified" canonical variates could be used to supplement the first pair as the usual higher order pairs do.

The i-th pair of modified canonical variates consists of variables ${}_1\tilde{Z}_i$ and ${}_2\tilde{Z}_i$ which

(1.8.1) $$\text{maximize } \operatorname{corr}({}_1\tilde{Z}_i,\ {}_2\tilde{Z}_i)$$

subject to

(1.8.2) $$|\operatorname{corr}({}_kZ_j, {}_k\tilde{Z}_i)| \le \theta_{ji}(\ge 0), \quad k = 1,2;\ j = 1,2,\ldots,i-1;\ \sum_{j=1}^{i} \theta_{ji}^2 = 1$$ [5].

[5] This constraint on the θ_{ji} is given for convenience only and can be easily removed.

The maximum correlation is the i-th modified canonical correlation, designated $\tilde{\rho}_i$.

THEOREM 1.8.1. Suppose $({}_1Z_j, {}_2Z_j) = ({}_1\underline{b}_j^{*\prime}\, \underline{Y}_1, {}_2\underline{b}_j^{*\prime}\, \underline{Y}_2)$, $j = 1, 2, \ldots, i$, are the first i-pairs of canonical variates. Assume the corresponding canonical correlations are distinct. Let $({}_1\tilde{Z}_j, {}_2\tilde{Z}_j)$, $j = 1, 2, \ldots, i$, designate the modified pairs of canonical variates. Then the i-th pair of modified canonical variates, constructed to maximize (1.8.1) subject to (1.8.2), may be defined as

$${}_1\tilde{Z}_i = {}_1\tilde{\underline{b}}_i^{*\prime}\, \underline{Y}_1 \quad \text{and} \quad {}_2\tilde{Z}_i = {}_2\tilde{\underline{b}}_i^{*\prime}\, \underline{Y}_2$$

where

$${}_1\tilde{\underline{b}}_i^{*} = \sum_{j=1}^{i} \theta_{ji}\, {}_1\underline{b}_j^{*} \quad \text{and} \quad {}_2\tilde{\underline{b}}_i^{*} = \sum_{j=1}^{i} \theta_{ji}\, {}_2\underline{b}_j^{*}$$

in which case the inequalities in (1.8.2) are all equalities and the corresponding i-th modified canonical correlation is

$$\tilde{\rho}_i = \rho_i + \sum_{j=1}^{i-1} \theta_{ji}^2(\rho_j - \rho_i) .$$

Also, for $h < i$,

$$\operatorname{corr}({}_1Z_h, {}_2\tilde{Z}_i) = \operatorname{corr}({}_1\tilde{Z}_i, {}_2Z_h) = \theta_{hi}\rho_h$$

$$\operatorname{corr}({}_1\tilde{Z}_h, {}_1\tilde{Z}_i) = \operatorname{corr}({}_2\tilde{Z}_h, {}_2\tilde{Z}_i) = \sum_{j=1}^{h} \theta_{jh}\theta_{ji}$$

$$\operatorname{corr}({}_1\tilde{Z}_h, {}_2\tilde{Z}_i) = \operatorname{corr}({}_1\tilde{Z}_i, {}_2\tilde{Z}_h) = \sum_{j=1}^{h} \theta_{jh}\theta_{ji}\rho_j .$$

All of this holds for $i = 2, 3, \ldots, p_1$.

Proof. If $\theta_{ji} = 0$, $j = 1, 2, \ldots, i-1$, the results follow from Theorems 1.4.1 and 1.4.2. In general,

$$_1\underline{\tilde{b}}_i^* = \sum_{j=1}^{p_1} \alpha_j \, {}_1\underline{b}_j^*, \quad \sum_{j=1}^{p_1} \alpha_j^2 = 1;$$

$$_2\underline{\tilde{b}}_i^* = \sum_{j=1}^{p_2} \beta_j \, {}_2\underline{b}_j^*, \quad \sum_{j=1}^{p_2} \beta_j^2 = 1;$$

$$|\mathrm{corr}({}_1Z_j, {}_1\tilde{Z}_i)| = |\alpha_j| \le \theta_{ji}, \quad j = 1, 2, \ldots, i-1 \;;$$

$$|\mathrm{corr}({}_2Z_j, {}_2\tilde{Z}_i)| = |\beta_j| \le \theta_{ji}, \quad j = 1, 2, \ldots, i-1 \;;$$

$$\mathrm{corr}({}_1\tilde{Z}_i, {}_2\tilde{Z}_i) = \sum_{j=1}^{p_1} \alpha_j \beta_j \rho_j .$$

The α's and β's may be assumed non-negative without any loss. Consider α_j and β_j fixed for $j = 1, 2, \ldots, i-1$ and define

$$\gamma_j = \begin{cases} \alpha_j \rho_j, & j = i, i+1, \ldots, p_1 \\ 0, & j = p_1+1, p_1+2, \ldots, p_2. \end{cases}$$

The argument proceeds along the lines of Anderson ([1], p.295). $\sum_{j=i}^{p_1} \alpha_j\beta_j\rho_j = \sum_{j=i}^{p_2} \beta_j\gamma_j \le (\sum_{j=i}^{p_2} \beta_j^2)^{\frac{1}{2}} (\sum_{j=i}^{p_2} \gamma_j^2)^{\frac{1}{2}}$ with equality if and only if $(\gamma_i, \gamma_{i+1}, \ldots, \gamma_{p_2}) \propto (\beta_i, \beta_{i+1}, \ldots, \beta_{p_2})$. Thus, for a maximum, $\beta_j = 0$, $j = p_1+1, p_1+2, \ldots, p_2$ and $\sum_{j=i}^{p_2} \beta_j\gamma_j = K \sum_{j=i}^{p_2} \gamma_j^2 = K \sum_{j=i}^{p_1} \alpha_j^2\rho_j^2 = K\left\{ \sum_{j=i}^{p_1} (\rho_j^2 - \rho_i^2)\alpha_j^2 + (1 - \sum_{j=1}^{i-1} \alpha_j^2)\rho_i^2 \right\}$ for some $K > 0$. Maximizing over the free α's yields $\alpha_j = 0$, $j = i+1, i+2, \ldots, p_1$, $\alpha_i = (1 - \sum_{j=1}^{i-1} \alpha_j^2)^{\frac{1}{2}}$, $\beta_j = 0$, $j = i+1, i+2, \ldots, p_1$, and

$\beta_i = (1 - \sum_{j=1}^{i-1} \beta_j^2)^{\frac{1}{2}}$. What remains is to maximize

$f(\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_{i-1}, \beta_{i-1}) = \sum_{j=1}^{i-1} \alpha_j\beta_j\rho_j + (1 - \sum_{j=1}^{i-1} \alpha_j^2)^{\frac{1}{2}}$.

$(1 - \sum_{j=1}^{i-1} \beta_j^2)^{\frac{1}{2}}\rho_i$ in the region $0 \le \alpha_j \le \theta_{ji}$, $0 \le \beta_j \le \theta_{ji}$, $j = 1,2,..,i-1$.

If $\rho_i = 0$, it is easily seen that the maximum occurs when $\alpha_j = \beta_j = \theta_{ji}$, $j = 1, 2, \ldots, i-1$ so that $f = \sum_{j=1}^{i-1} \theta_{ji}^2\rho_j$. Hence suppose $\rho_i > 0$.

Since $\theta_{ji} = 0$ implies $\alpha_j = \beta_j = 0$, it is only necessary to consider $\theta_{ji} > 0$. For any such j,

$$\frac{\partial f}{\partial \alpha_j} = \beta_j\rho_j - \alpha_j\left(\frac{1 - \sum_{j=1}^{i-1} \beta_j^2}{1 - \sum_{j=1}^{i-1} \alpha_j^2}\right)^{\frac{1}{2}}\rho_i, \quad 0 \le \alpha_j < \theta_{ji}; \quad 0 \le \beta_j \le \theta_{ji}$$

and

$$\frac{\partial f}{\partial \beta_j} = \alpha_j\rho_j - \beta_j\left(\frac{1 - \sum_{j=1}^{i-1} \alpha_j^2}{1 - \sum_{j=1}^{i-1} \beta_j^2}\right)^{\frac{1}{2}}\rho_i, \quad 0 \le \alpha_j \le \theta_{ji}; \quad 0 \le \beta_j < \theta_{ji}.$$

Except for the solution $\alpha_j = \beta_j = 0$,

$$\frac{\partial f}{\partial \alpha_j} = 0 \Rightarrow \frac{\beta_j}{\left(1 - \sum_{j=1}^{i-1} \beta_j^2\right)^{\frac{1}{2}}} > \frac{\alpha_j}{\left(1 - \sum_{j=1}^{i-1} \alpha_j^2\right)^{\frac{1}{2}}}$$

and

$$\frac{\partial f}{\partial \beta_j} = 0 \Rightarrow \frac{\beta_j}{\left(1 - \sum_{j=1}^{i-1} \beta_j^2\right)^{\frac{1}{2}}} < \frac{\alpha_j}{\left(1 - \sum_{j=1}^{i-1} \alpha_j^2\right)^{\frac{1}{2}}} .$$

Consequently, there is only one solution and the maximum f is attained when, for each j, either $\alpha_j = 0$ or θ_{ji} or $\beta_j = 0$ or θ_{ji}. But, if $\alpha_j = 0$, then $\frac{\partial f}{\partial \beta_j} < 0$, $0 < \beta_j < \theta_{ji}$; if $\beta_j = 0$, then $\frac{\partial f}{\partial \alpha_j} < 0$, $0 < \alpha_j < \theta_{ji}$. This means there are four possibilities for each j $(\theta_{ji} > 0)$;

(i) $\alpha_j = \beta_j = \theta_{ji}$;

(ii) $\alpha_j = \theta_{ji}$, $0 < \beta_j < \theta_{ji}$;

(iii) $\beta_j = \theta_{ji}$, $0 < \alpha_j < \theta_{ji}$,

(iv) $\alpha_j = \beta_j = 0$.

Not all j's can be associated with (iv) for then f is less than that when all j's are associated with (i). Suppose j is classified with (ii) and j' with (iii). Then $\frac{\partial f}{\partial \beta_j} = \frac{\partial f}{\partial \alpha_{j'}} = 0 \Rightarrow \theta_{ji}\theta_{j'i}\rho_j\rho_{j'} = \beta_j\alpha_{j'}\rho_i^2 \Rightarrow \beta_j\alpha_{j'} > \theta_{ji}\theta_{j'i}$ which is impossible. This means that either (ii) or (iii) can be ruled out. Now, for a j connected with (ii), $\frac{\partial f}{\partial \beta_j} > \theta_{ji}(\rho_j - \rho_i) > 0$ which means that $\beta_j = \theta_{ji}$. Similarly, if the j were linked with (iii) instead of (ii), one would infer that $\alpha_j = \theta_{ji}$. Thus the only remaining possibilities are (i) and (iv). Finally, (iv) is ruled out by noting that, for any j connected with it, the corresponding value of f must be less than the value when

$\alpha_j = \beta_j = \theta_{ji}$. Evidently $\alpha_j = \beta_j = \theta_{ji}$, $j = 1, 2, \ldots, i-1$, and $f = \rho_i + \sum_{j=1}^{i-1} \theta_{ji}^2(\rho_j - \rho_i)$. The remainder of the theorem follows readily.

CHAPTER II

CANONICAL ANALYSIS IN THE PRESENCE OF SINGULARITIES

2.1 Introduction

The theory of canonical analysis is extended in this chapter to cover the situation where one or both of the sets may be singular. The key matrix in the canonical analysis of two nonsingular sets was shown in Theorem 1.4.1 to be $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Here, in the less restricted setting, the key matrix turns out to be $\Sigma_{11}^{-}\Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}$ ("-" indicates a generalized inverse). The main result is recorded in Theorem 2.3.1.

Three applications of the theorem are developed. In Section 2.4, the question of independence in a two way contingency table is considered as one of independence of two singular sets of variables. In Section 2.5, a scoring system which yields the maximum or minimum F statistic is constructed for qualitative responses in a one way analysis of variance. The corresponding most significant contrast in the treatment effects is obtained at the same time. In Section 2.6, a simple procedure is outlined for the appropriate canonical analysis when linear restrictions have been imposed on the coefficient vectors which define the canonical variates. The analysis is arrived at by first translating the problem into terms of two singular sets of variables.

2.2 Singular Sets of Variates

In many applications of canonical analysis, one must be prepared to handle singularities in one or both of the sets. From each singular set, it is customary to derive two subsets by an appropriate nonsingular transformation: one is nonsingular and the other is degenerate in that its covariance matrix is null. The degenerate subset is then eliminated on the grounds that it can contribute only a constant term to the canonical variates. The problem is thus reduced to the familiar analysis of two nonsingular sets of variates.

The theory presented in Section 2.3 demonstrates that the preliminary transformation and reduction are in fact unnecessary. In some situations, for example those of Sections 2.4, 2.5, and 2.6, considerable saving may be realized by not making the reduction. More generally, the theory yields a unified approach to the subject which covers the situation where both sets are nonsingular as a special case.

The degenerate subset mentioned above is formed from a particular basis of the null space of the covariance matrix of the singular set. The null space, in some cases, may contain useful statistical insights and thus be worthy of special attention. It is possible to obtain a basis for the null space as a by-product of the calculation of the generalized inverse of the singular set covariance matrix.

This generalized inverse plays an integral role in the ensuing development. A generalized inverse (g-inverse) of an arbitrary matrix A is any matrix A^- such that $AA^-A = A$. This definition has been popularized by Rao [20]; he has emphasized that it is often easier to compute than other (more restrictive) generalized inverses and that it is particularly relevant to statistical problems. For matrices with simple structures, like those in Sections 2.4, 2.5, and 2.6, one may be able to determine a g-inverse by inspection or by trial and error.

McKeon [16] appears to have been the first to employ generalized inverses in the context of canonical analysis. Confining attention to square matrices, he defines a generalized inverse A^+ of A as a matrix satisfying (i) $AA^+ = A^+A = E$ and (ii) $AE = EA = A$ for some idempotent E. Such a matrix A^+ exists whenever A is symmetric, but may not otherwise (e.g., when $A = \begin{pmatrix} 01 \\ 00 \end{pmatrix}$). With $A = A'$, $AA^+A = A$ so that A^+ is a g-inverse.

The theory presented here may be taken as justification of McKeon's development which he gives without proof. It is more general, however, in that it allows one to choose from the larger class of g-inverses where particularly simple and natural matrices, inadmissible in McKeon's setting, may be found.

2.3 The Extension Theorem

The ranks of Σ_{11} and Σ_{22} are denoted by q_1 and q_2. Because $q_1 \leq p_1$ and $q_2 \leq p_2$ there are only $(q_1 + q_2)$ canonical variates to be found instead of $(p_1 + p_2)$. Consequently, the dimensions of certain quantities defined in Section 1.2 need to be adjusted to reflect the correct number of canonical variates:

$$B_1' = ({}_1\underline{b}_1, {}_1\underline{b}_2, \ldots, {}_1\underline{b}_{q_1}) \quad \text{and} \quad B_2' = ({}_2\underline{b}_1, {}_2\underline{b}_2, \ldots, {}_2\underline{b}_{q_2});$$

$$\underline{Z}_1' = ({}_1Z_1, {}_1Z_2, \ldots, {}_1Z_{q_1}) \quad \text{and} \quad \underline{Z}_2' = ({}_2Z_1, {}_2Z_2, \ldots, {}_2Z_{q_2});$$

$$\Phi_{12} = B_1\Sigma_{12}B_2', \quad \text{a} \quad (q_1 \times q_2) \quad \text{matrix.}$$

THEOREM 2.3.1. Let q_1, q_2, and r be the ranks of Σ_{11}, Σ_{22}, and Σ_{12} respectively. Let Σ_{11}^- and Σ_{22}^- be g-inverses of Σ_{11} and Σ_{22}. Suppose $q_1 \le q_2$. The eigenvalues of $\Sigma_{11}^-\Sigma_{12}\Sigma_{22}^-\Sigma_{21}$ are $\rho_1^2, \rho_2^2, \ldots, \rho_{q_1}^2, 0, \ldots, 0$ where $1 > \rho_1^2 \ge \rho_2^2 \ge \ldots \ge \rho_{q_1}^2$ are the ordered squared canonical correlations. Eigenvectors $_1\underline{b}_i$ associated with ρ_i^2 may be chosen so that $B_1\Sigma_{11}B_1' = I$ and used to define canonical variates $\underline{Z}_1 = B_1\underline{X}_1$ for the first set. Any matrix B_2 satisfying

(2.3.1) $$\Sigma_{22}^-\Sigma_{21}B_1' = B_2'\Phi_{12}' \quad \text{and} \quad B_2\Sigma_{22}B_2' = I ,$$

where

(2.3.2) $$\Phi_{12} = \begin{pmatrix} \underset{q_1}{D} & \underset{q_2-q_1}{0} \end{pmatrix} q_1 \quad \text{and} \quad D = \operatorname{diag}(\rho_1, \rho_2, \ldots, \rho_{q_1}),$$

may be used to define corresponding canonical variates $\underline{Z}_2 = B_2\underline{X}_2$ for the second set.

Proof.[1] Use will be made of the existence of matrices M_{12} and M_{21} such that $\Sigma_{12} = M_{12}\Sigma_{22}$ and $\Sigma_{21} = M_{21}\Sigma_{11}$ (cf. Lemma A4). $\Sigma_{11}^-\Sigma_{11}$ is idempotent since

$$(\Sigma_{11}^-\Sigma_{11})(\Sigma_{11}^-\Sigma_{11}) = \Sigma_{11}^-(\Sigma_{11}\Sigma_{11}^-\Sigma_{11}) = \Sigma_{11}^-\Sigma_{11}$$

and $N(\Sigma_{11}^-\Sigma_{11}) = N(\Sigma_{11})$ since

$$\Sigma_{11}\underline{u} = \underline{0} \Rightarrow \Sigma_{11}^-\Sigma_{11}\underline{u} = \underline{0}$$

and

---

[1] This method of proof was suggested to me by Professor R. I. Jennrich.

$$\Sigma_{11}^{-}\Sigma_{11}\underline{u} = \underline{0} \Rightarrow \Sigma_{11}\Sigma_{11}^{-}\Sigma_{11}\underline{u} = \underline{0} \Rightarrow \Sigma_{11}\underline{u} = \underline{0}.$$

Thus, by the matrix analogue of Lemma A5, the direct sum decomposition, $V(I_{p_1}) = N(\Sigma_{11}) \oplus V(\Sigma_{11}^{-}\Sigma_{11})$, can be made. Similarly, $V(I_{p_2}) = N(\Sigma_{22}) \oplus V(\Sigma_{22}^{-}\Sigma_{22})$. Also note that $\dim[V(\Sigma_{11}^{-}\Sigma_{11})] = q_1$ and $\dim[V(\Sigma_{22}^{-}\Sigma_{22})] = q_2$. From Theorems 1.4.1 and 1.4.2, it follows that there exist vectors ${}_1\underline{b}_i$ and ${}_2\underline{b}_j$ defining canonical variates ${}_1Z_i = {}_1\underline{b}_i' \underline{X}_1$ and ${}_2Z_j = {}_2\underline{b}_j' \underline{X}_2$ with canonical correlation $\rho_i = \text{corr}({}_1Z_i, {}_2Z_i)$, $i = 1, 2, \ldots, q_1$; $j = 1, 2, \ldots, q_2$. The component of ${}_1\underline{b}_i$ in $N(\Sigma_{11})$ adds only a constant to ${}_1Z_i$ so that without loss it may be assumed that ${}_1\underline{b}_i \in V(\Sigma_{11}^{-}\Sigma_{11})$ and likewise that ${}_2\underline{b}_j \in V(\Sigma_{22}^{-}\Sigma_{22})$.

$$\begin{aligned} 0 &= \text{cov}({}_1Z_i, \rho_j\, {}_1Z_j - {}_2Z_j) \\ &= {}_1\underline{b}_i'(\rho_j\Sigma_{11}\, {}_1\underline{b}_j - \Sigma_{12}\, {}_2\underline{b}_j), \quad i = 1, 2, \ldots, q_1 . \end{aligned}$$

Therefore

$$0 = B_1\Sigma_{11}(\rho_j\, {}_1\underline{b}_j - M_{21}'\, {}_2\underline{b}_j)$$

which implies that

$$0 = \Sigma_{11}(\rho_j\, {}_1\underline{b}_j - M_{21}'\, {}_2\underline{b}_j).$$

Arguing similarly on $\text{cov}({}_2Z_i, \rho_j\, {}_2Z_j - {}_1Z_j)$, one arrives at

$$\left.\begin{aligned} \rho_j\Sigma_{11}\, {}_1\underline{b}_j &= \Sigma_{12}\, {}_2\underline{b}_j \\ \rho_j\Sigma_{22}\, {}_2\underline{b}_j &= \Sigma_{21}\, {}_1\underline{b}_j \end{aligned}\right\} \quad j = 1, 2, \ldots, q_1.$$

Now

$$\Sigma_{11}^{-}\Sigma_{11}\, {}_1\underline{b}_j = {}_1\underline{b}_j \quad \text{and} \quad \Sigma_{22}^{-}\Sigma_{22}\, {}_2\underline{b}_j = {}_2\underline{b}_j$$

so that

$$\rho_j \; {}_1\underline{b}_j = \Sigma_{11}^{-}\Sigma_{12} \; {}_2\underline{b}_j$$

and

$$\rho_j \; {}_2\underline{b}_j = \Sigma_{22}^{-}\Sigma_{21} \; {}_1\underline{b}_j .$$

Therefore

$$\rho_j^2 \; {}_1\underline{b}_j = \Sigma_{11}^{-}\Sigma_{12}\Sigma_{22}^{-}\Sigma_{21} \; {}_1\underline{b}_j, \quad j = 1, 2, \ldots, q_1 .$$

Every ρ_j^2 is an eigenvalue of $\Sigma_{11}^{-}\Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}$ and the rows of B_1 form a linearly independent set of eigenvectors with each ${}_1\underline{b}_j \in V(\Sigma_{11}^{-}\Sigma_{11})$. Since $\dim[N(\Sigma_{11})] = p_1 - q_1$ and $\Sigma_{11}^{-}\Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}\underline{b} = \Sigma_{11}^{-}\Sigma_{12}\Sigma_{22}^{-}M_{21}\Sigma_{11}\underline{b} = \underline{0}$ for any $\underline{b} \in N(\Sigma_{11})$, it follows that the remaining $(p_1 - q_1)$ eigenvalues are all zero. Now let B_1 be any matrix such that $B_1\Sigma_{11}B_1' = I$ and $\Sigma_{11}^{-}\Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}B_1' = B_1'D^2$ with D as in (2.3.2). Then $B_1\Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}B_1' = D^2$. Let the first r rows of B_2 be defined by $\Sigma_{22}^{-}\Sigma_{21}B_1' = B_2'\Phi_{12}'$ with Φ_{12} as in (2.3.2) so that $D^2 = \Phi_{12}B_2\Sigma_{22}B_2'\Phi_{12}'$. Thus the matrix B_2 may be completed so that $B_2\Sigma_{22}B_2' = I$. Now $B_1\Sigma_{12}B_2' = B_1\Sigma_{12}\Sigma_{22}^{-}\Sigma_{22}B_2' = \Phi_{12}B_2\Sigma_{22}B_2' = \Phi_{12}$. $\underline{Z}_1$ and $\underline{Z}_2$, therefore, are vectors of canonical variates.

2.4 Canonical Analysis of a Two Way Contingency Table

The analysis of association in a two way contingency table is a particular application of canonical analysis. This is well documented (Fisher [5], Williams [24], Kendall and Stuart [14], and McKeon [16]).

Let f_{ij}, $f_{i\cdot}$, and $f_{\cdot j}$ be the probabilities of an observation falling in the (i,j) cell, the i-th row, and the j-th column, respectively. Write

$$F = ((f_{ij})),$$

$$\underline{f}_1' = (f_{1\cdot}, f_{2\cdot}, \ldots, f_{p_1\cdot}) ,$$

$$\underline{f}_2' = (f_{\cdot 1}, f_{\cdot 2}, \ldots, f_{\cdot p_2}) ,$$

$$D_1 = \text{diag}(f_{1\cdot}, f_{2\cdot}, \ldots, f_{p_1\cdot}) ,$$

and

$$D_2 = \text{diag}(f_{\cdot 1}, f_{\cdot 2}, \ldots, f_{\cdot p_2}) .$$

To make the connection with canonical analysis, it is necessary to define the vector variates $\underline{X}_1$ and $\underline{X}_2$. They will be used as indicators of row and column membership: for an observation in the i-th row and j-th column, $\underline{X}_1$ would have a one as the i-th element with zeros elsewhere and $\underline{X}_2$ would have a one as the j-th element with zeros elsewhere. Then

$$\Sigma_{11} = E(\underline{X}_1\underline{X}_1') - E(\underline{X}_1)E(\underline{X}_1)' = D_1 - \underline{f}_1\underline{f}_1',$$

$$\Sigma_{22} = E(\underline{X}_2\underline{X}_2') - E(\underline{X}_2)E(\underline{X}_2)' = D_2 - \underline{f}_2\underline{f}_2',$$

and

$$\Sigma_{12} = E(\underline{X}_1\underline{X}_2') - E(\underline{X}_1)E(\underline{X}_2)' = F - \underline{f}_1\underline{f}_2' .$$

It is easy to show that $q_1 = p_1 - 1$ and $q_2 = p_2 - 1$. Natural choices for the g-inverses are $\Sigma_{11}^- = D_1^{-1}$ and $\Sigma_{22}^- = D_2^{-1}$. (McKeon used $\Sigma_{11}^- = E_1 D_1^{-1} E_1$ and $\Sigma_{22}^- = E_2 D_2^{-1} E_2$ where $E_1 = (I - p_1^{-1}\underline{1}\,\underline{1}')$ and

$E_2 = (I - p_2^{-1}\underline{1}\,\underline{1}').)$ With $q_1 \le q_2$ there are q_1 (squared) canonical correlations which may be obtained as the q_1 largest eigenvalues of

$$\Sigma_{11}^{-}\Sigma_{12}\Sigma_{22}^{-}\Sigma_{21} = D_1^{-1}FD_2^{-1}F' - \underline{1}\,\underline{f}_1'.$$

One eigenvalue is always zero and $\underline{1} \in N(\Sigma_{11})$ an associated eigenvector. This reflects the known constraint $\underline{f}_1'\underline{1} = 1$. The sum of the eigenvalues is

$$\operatorname{tr}(\Sigma_{11}^{-}\Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}) = \sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\frac{f_{ij}^2}{f_{i\cdot}f_{\cdot j}} - 1.$$

It is easy to show that $\Sigma_{12} = 0$ if and only if $\underline{X}_1$ and $\underline{X}_2$ are independent, as in the case of the multivariate normal distribution. It is for this reason that the canonical correlations are directly related to independence statements or hypotheses.

The sample covariance matrix S, based on n observations, has the same structure as Σ. One need only replace f_{ij} by n_{ij}/n, $f_{i\cdot}$ by $n_{i\cdot}/n$, and $f_{\cdot j}$ by $n_{\cdot j}/n$; n_{ij}, $n_{i\cdot}$, and $n_{\cdot j}$ are the cell, row, and column totals, respectively. Then, partitioning S like Σ, it follows that

$$\operatorname{tr}(S_{11}^{-}S_{12}S_{22}^{-}S_{21}) = \sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} - 1 = \frac{1}{n}\chi^2 ,$$

where χ^2 is the "square contingency" statistic for measuring association in a two way table. A test based on χ^2 is, therefore, a test based on the sum of squared sample canonical correlations.

2.5 Scoring Qualitative Responses in a One Way Analysis of Variance

A problem concerning qualitative responses in a one way analysis of variance (ANOVA) is mathematically equivalent to the contingency table analysis of Section 2.4.[2] The ANOVA problem has been included to provide one example from the multitude of rotational type problems which arise frequently and, even though not reducible to the study of two sets of random variables, can be translated into a form to which the usual canonical analysis techniques, or hybrids of them, are applicable. (Other examples are discriminant analysis, as discussed by Bartlett [2] and McKeon [16], and the problem of orthogonal rotation to congruence in factor analysis. Some references to the latter are Wrigley and Neuhaus [25], Horst [8], and Cliff [4].)

Imagine a one way ANOVA with n observations representing qualitative responses at any of p_1 possible levels on a total of p_2 treatments. The object is to affix scores $\underline{s}' = (s_1, s_2, \ldots, s_{p_1})$ to the response levels such that the resulting F statistic is a maximum (or a minimum). The corresponding most significant contrast $\underline{c}' = (c_1, c_2, \ldots, c_{p_1})$ is also to be determined. The sampling can be summarized in n pairs of vectors which are realizations of indicator variables $(\underline{X}_1, \underline{X}_2)$ like those used in the previous section. The observation on $\underline{X}_1$ indicates the level of response while the $\underline{X}_2$ value indicates which treatment was used. One can formally calculate a sample covariance matrix S using these n pairs as suggested at the end of Section 2.4. Making the substitutions specified there, one arrives at

---

2 This connection became apparent during discussions with Dr. Peter Nemenyi.

$$S_{11} = \hat{D}_1 - \hat{\underline{f}}_1 \hat{\underline{f}}_1' ,$$

$$S_{22} = \hat{D}_2 - \hat{\underline{f}}_2 \hat{\underline{f}}_2' ,$$

and

$$S_{12} = \hat{F} - \hat{\underline{f}}_1 \hat{\underline{f}}_2'$$

where

$$\hat{\underline{f}}_1' = n^{-1}(n_{1\cdot}, n_{2\cdot}, \ldots, n_{p_1\cdot}),$$

$$\hat{\underline{f}}_2' = n^{-1}(n_{\cdot 1}, n_{\cdot 2}, \ldots, n_{\cdot p_2}),$$

$$\hat{D}_1 = n^{-1}\mathrm{diag}(n_{1\cdot}, n_{2\cdot}, \ldots, n_{p_1\cdot}),$$

$$\hat{D}_2 = n^{-1}\mathrm{diag}(n_{\cdot 1}, n_{\cdot 2}, \ldots, n_{\cdot p_2}),$$

and

$$\hat{F} = n^{-1}((n_{ij})) .$$

The F statistic, for fixed $\underline{s}$, is a function of the most significant contrast; that is,

(2.5.1) $$F = \sup_{\underline{1}'\underline{c} = 0} \left(\frac{n-p_2}{p_2-1}\right) \frac{(\underline{s}\hat{F}\hat{D}_2^{-1}\underline{c})^2}{(\underline{c}'\hat{D}_2^{-1}\underline{c})\underline{s}'(\hat{D}_1 - \hat{F}\hat{D}_2^{-1}\hat{F}')\underline{s}}$$

$$= \sup_{\hat{\underline{f}}_2'\underline{d} = 0} \left(\frac{n-p_2}{p_2-1}\right) \frac{(\underline{s}'\hat{F}\underline{d})^2}{(\underline{d}'\hat{D}_2\underline{d})\underline{s}'(\hat{D}_1 - \hat{F}\hat{D}_2^{-1}\hat{F}')\underline{s}} ,$$

after making the substitution

(2.5.2) $$\underline{d} = \hat{D}_2^{-1}\underline{c}.$$

The expression for F in terms of the usual sums of squares is

$$(2.5.3) \qquad F = \left(\frac{n-p_2}{p_2-1}\right) \frac{\underline{s}'(\hat{F}\hat{D}_2^{-1}\hat{F}' - \hat{\underline{f}}_1\hat{\underline{f}}_1')\underline{s}}{\underline{s}'(\hat{D}_1 - \hat{F}\hat{D}_2^{-1}\hat{F}')\underline{s}} .$$

Define

$$F^* = \sup_{\underline{s} \in V(S_{11}^- S_{11})} F \quad \text{and} \quad F^{**} = \inf_{\underline{s} \in V(S_{11}^- S_{11})} F .$$

F^* is a monotonic function of

$$\frac{(p_2-1)F^*}{(n-p_2) + (p_2-1)F^*} = \sup_{\underline{s} \in V(S_{11}^- S_{11})} \frac{(p_2-1)F}{(n-p_2) + (p_2-1)F} .$$

Substituting (2.5.1) in the numerator and (2.5.3) in the denominator leads to

$$(2.5.4) \qquad \frac{(p_2-1)F^*}{(n-p_2) + (p_2-1)F^*} = \sup_{\underline{s} \in V(S_{11}^- S_{11})} \; \sup_{\underline{d} \in V(S_{22}^- S_{22})} \frac{(\underline{s}'\hat{F}\underline{d})^2}{(\underline{s}'S_{11}\underline{s})(\underline{d}'\hat{D}_2\underline{d})}$$

$$= \sup_{\underline{s} \in V(S_{11}^- S_{11})} \; \sup_{\underline{d} \in V(S_{22}^- S_{22})} \frac{(\underline{s}'S_{12}\underline{d})^2}{(\underline{s}'S_{11}\underline{s})(\underline{d}'S_{22}\underline{d})} .$$

Thus the optimal $\underline{s}$ and $\underline{d}$ are consistent with the optimal ${}_1\underline{b}_1$ and ${}_2\underline{b}_1$ for the sample two way contingency analysis. Introducing $\hat{\rho}_1^2$ for (2.5.4), i.e., the squared first sample canonical correlation and solving for F^* yields

$$F^* = \left(\frac{n-p_2}{p_2-1}\right)\left(\frac{\hat{\rho}_1^2}{1 - \hat{\rho}_1^2}\right) .$$

Having found an optimal $\underline{s}$ in accordance with Theorem 2.3.1 and assuming $\hat{\rho}_1^2 > 0$, it follows from (2.3.1) and (2.5.2) that the optimal $\underline{c}$ is proportional to

$$S_{21}\underline{s} \ . \tag{2.5.5}$$

The foregoing argument can be repeated in substantially the same way to find an $\underline{s}$ yielding F** and the corresponding most significant $\underline{c}$: one need only replace "$\sup_{\underline{s} \in V(S_{11}^- S_{11})}$" by "$\inf_{\underline{s} \in V(S_{11}^- S_{11})}$" throughout. Then

$$F^{**} = \left(\frac{n-p_2}{p_2-1}\right)\left(\frac{\hat{\rho}^2_{p_1-1}}{1-\hat{\rho}^2_{p_1-1}}\right)$$

where $\hat{\rho}^2_{p_1-1}$ is the smallest of the squared sample canonical correlations associated with S. If $\hat{\rho}^2_{p_1-1} > 0$, the most significant contrast $\underline{c}$ is proportional to (2.5.5).

2.6 Canonical Analysis with Linear Restrictions

A thorough study of the relations between two sets of variates may include the systematic placement of linear restrictions on the coefficient vectors which define the canonical variates. Such restrictions may be used, for example to force particular coefficients to add to zero, be equal, or occur in specified proportions; they may also assure that the canonical variates are uncorrelated with some specified set of variables.

The restrictions are embedded in matrices M_1 and M_2; it is required that ${}_1\underline{b}_i \in V(M_1)$, $i = 1, 2, \ldots, \tilde{q}_1$, and ${}_2\underline{b}_j \in V(M_2)$, $j = 1, 2, \ldots, \tilde{q}_2$. $\tilde{q}_1 = \text{rank}(M_1'\Sigma_{11}M_1)$,

$\tilde{q}_2 = \text{rank}(M_2'\Sigma_{22}M_2)$, and $\tilde{q}_1 \le \tilde{q}_2$. The effect of M_1 and M_2 is to induce a new covariance matrix

$$\begin{pmatrix} M_1'\Sigma_{11}M_1 & M_1'\Sigma_{12}M_2 \\ M_2'\Sigma_{21}M_1 & M_2'\Sigma_{22}M_{22} \end{pmatrix},$$

which can be formally treated according to the theorem. The coefficient vectors $\tilde{B}_1$ and $\tilde{B}_2$ (say) obtained in this way are related to the desired B_1 and B_2 by

$$B_1 = \tilde{B}_1M_1' \quad \text{and} \quad B_2 = \tilde{B}_2M_2'.$$

It may be most natural to formulate the restrictions in terms of matrices G_1 and G_2[3] and then, as a second step, to define M_1 and M_2 so that $V(M_1) \perp V(\Sigma_{11}G_1)$ and $V(M_2) \perp V(\Sigma_{22}G_2)$. Thus let $M_1 = \Sigma_{11}^{-1} - G_1(G_1'\Sigma_{11}G_1)^-G_1'$ and $M_2 = \Sigma_{22}^{-1} - G_2(G_2'\Sigma_{22}G_2)^-G_2'$, assuming now that Σ is positive definite. [Note: g-inverses of M_1 and M_2 are Σ_{11} and Σ_{22}, respectively.]

In compliance with the theorem, one then forms

$$M_1^-(M_1\Sigma_{12}M_2)M_2^-(M_2\Sigma_{21}M_1) = M_1^-M_1\Sigma_{12}M_2\Sigma_{21}M_1$$

---

[3] For example, if the canonical variables for the first set are to be uncorrelated with ${}_1X_1$ and ${}_2X_1$, then

$$G_1 = \left(\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; \Sigma_{11}^{-1}\Sigma_{12}\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right).$$

and works from there. The eigenvalues are the same as those of $M_1\Sigma_{12}M_2\Sigma_{21}$. This latter matrix, when expanded, is

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + G_1(G_1'\Sigma_{11}G_1)^{-}G_1'\Sigma_{12}G_2(G_2'\Sigma_{22}G_2)^{-}G_2'\Sigma_{21}$$

$$- \Sigma_{11}^{-1}\Sigma_{12}G_2(G_2'\Sigma_{22}G_2)^{-}G_2'\Sigma_{21} - G_1(G_1'\Sigma_{11}G_1)^{-}G_1'\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Taking the trace, one finds that the sum of squared canonical correlations between the constrained variables $(M_1\underline{X}_1,\ M_2\underline{X}_2)$ is equal to the sum for $(\underline{X}_1,\ \underline{X}_2)$ plus the sum for $(G_1'\underline{X}_1,\ G_2'\underline{X}_2)$ minus the sum for $(\underline{X}_1,\ G_2'\underline{X}_2)$ minus the sum for $(G_1'\underline{X}_1,\ \underline{X}_2)$.

CHAPTER III

MODELS FOR THE CANONICAL ANALYSIS OF SEVERAL SETS OF VARIABLES

3.1 Introduction

Hotelling [12] recognized the need of extending his theory to deal with more than two sets of random variables. He did not solve this problem, however, and it has only been since 1950 that much attention has been given to it.

In the canonical analysis of several sets of variables, one would like, as for two sets, to identify "canonical variates" which isolate and summarize existing linear relations among the sets in a convenient and insightful manner. One would at the same time like to measure the intensity of the relationship through some sort of generalization of the canonical correlation coefficient.

Five different techniques for the canonical analysis of several sets will be presented here. They are

(i) the sum of correlations method (SUMCOR);

(ii) the maximum variance method (MAXVAR);

(iii) the sum of squared correlations method (SSQCOR);

(iv) the minimum variance method (MINVAR);

(v) the generalized variance method (GENVAR).

Of these, the SSQCOR and MINVAR procedures are new. Each is a generalization of the two set analysis in that it reduces to Hotelling's classical method whenever the number of sets is only two. Each is designed to detect a different form of linear relationship among the sets. It is, therefore, often advisable, especially in exploratory studies, to employ more than one and perhaps all of these methods.

The notation relevant to the several set problem is contained in Section 3.2. The criteria for picking first stage canonical variates (analogous to the first pair of canonical variates for two sets) according to the different methods are set down in Section 3.3. The main purpose of this chapter is to develop models of the general principal component type, extensions of those in Section 1.6, for the various methods. This is done in Sections 3.4 - 3.8 in terms of first stage canonical variates. The models are important in that they reveal the types of effects which the methods can uncover. They also help to motivate and to interrelate the methods. The question of how to construct higher order canonical variates is postponed to Section 3.9.

3.2 Notation

The main notation for the next three chapters is recorded here for future reference. Essentially the notation is the logical extension of that given in Section 1.2 for two sets so it will not be necessary to explain the symbols in detail.

The number of sets is designated by m. Thus

$$\underline{X}_j(p_j \times 1), \quad j = 1, 2, \ldots, m, \quad \text{with} \quad p_1 \le p_2 \le \ldots \le p_m$$

$$\text{and} \quad p = \sum_{i=1}^{m} p_i$$

are the original variables. The other random variables of interest are

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

$$\underline{Y}' = (\underline{Y}_1', \underline{Y}_2', \ldots, \underline{Y}_m'), \quad \text{with} \quad \underline{Y}_j = T_j^{-1}\underline{X}_j,$$

$$\underline{Z}' = (\underline{Z}_1', \underline{Z}_2', \ldots, \underline{Z}_m'), \quad \text{with} \quad \underline{Z}_j = B_j^*\underline{Y}_j = B_j\underline{X}_j,$$

and

$$\underline{Z}_{(j)}' = ({}_1Z_j, {}_2Z_j, \ldots, {}_mZ_j), \quad \text{with} \quad {}_iZ_j = {}_i\underline{b}_j^{*\prime}\underline{Y}_j = {}_i\underline{b}_j'\,\underline{X}_j.$$

Although the elements of $\underline{Z}$ are still required to have unit variance, they no longer need be uncorrelated within sets as in Chapter I.

The corresponding covariance matrices are

$$\Sigma_{jj},$$

$$\Sigma = ((\Sigma_{ij})),$$

$$R = ((R_{ij})) \quad \text{with} \quad R_{ii} = I,$$

$$\Phi = ((\Phi_{ij})),$$

and

$$\Phi_{(j)}.$$

Introducing block diagonal matrices

$$D_T = \text{diag}(T_1, T_2, \ldots, T_m),$$

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

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

$$D_{B(j)} = \text{diag}({}_1\underline{b}_j', {}_2\underline{b}_j', \ldots, {}_m\underline{b}_j'),$$

and

$$D_{B^*(j)} = \text{diag}({}_1b_j^{*\prime}, {}_2\underline{b}_j^*, \ldots, {}_m\underline{b}_j^{*\prime}),$$

it follows that

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

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

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

and

$$\Phi_{(j)} = D_{B*(j)} R D'_{B*(j)} = D_{B(j)} \Sigma D'_{B(j)}.$$

Additionally, let $\underline{v}_j$ be orthonormal eigenvectors of R corresponding to the eigenvalues $c_j(R)$, and let ${}_j\underline{e}_i$ be orthonormal eigenvectors of $\Phi_{(i)}$ corresponding to the eigenvalues ${}_1\lambda_i \geq {}_2\lambda_i \geq \ldots \geq {}_m\lambda_i > 0$, with $\sum_{j=1}^{m} {}_j\lambda_i = m$. It is also convenient to define

$$E_{(i)} = ({}_1\underline{e}_i, {}_2\underline{e}_i, \ldots, {}_m\underline{e}_i)$$

and

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

Finally, let ϕ_{max} be the largest canonical correlation obtainable between any two of the m sets.

3.3 Criteria for the Selection of First Stage Canonical Variables

The methods of canonical analysis involve a number of stages. The goal at the s-th stage is to find $\underline{Z}_{(s)}$ which optimizes some criterion function. All of the criteria considered here involve the optimization of some function, f, of $\Phi_{(s)}$. The chosen ${}_jZ_s$ are called s-th stage canonical variates. The same criterion function is used at each stage although certain restrictions on the admissible $\underline{Z}_{(s)}$ are added as one progresses beyond the first stage. The nature of these restrictions and the number of possible stages are matters which will be dealt with in Section 3.9. It will suffice here to consider the construction of the optimal $\underline{Z}_{(1)}$.

The SUMCOR method was introduced by Horst in [8]. It is briefly mentioned in [9] and again, more extensively, in [10] under the name of the maximum correlation method. The material in [10], however, is essentially the same as in [8].

The SUMCOR criterion is to select $\underline{Z}_{(1)}$ to

$$(3.3.1)\quad \text{maximize } f(\Phi_{(1)}) = \sum_{i=1}^{m} \sum_{j=1}^{m} \text{corr}({}_iZ_1, {}_jZ_1) = \underline{1}'\Phi_{(1)}\underline{1} .$$

An equivalent criterion is to

$$(3.3.2)\quad \text{minimize } f(\Phi_{(1)}) = \sum_{i=1}^{m} \sum_{j=1}^{m} \text{var}({}_iZ_1 - {}_jZ_1).$$

It is worth remarking that the f in (3.3.1) or (3.3.2) is sensitive to sign changes in the ${}_jZ_1$ whereas the other four criteria to be considered are not.

Horst also originated the MAXVAR method (see [9] or [10]). His motivation was to find a $\underline{Z}_{(1)}$ such that the associated $\Phi_{(1)}$ gives the best least squares or Euclidean norm approximation to a rank one matrix; and, consequently, Horst refers to this in [10] as the "rank one approximation method." He shows that the procedure[1] is equivalent to picking $\underline{Z}_{(1)}$ to

$$(3.3.3)\quad \text{maximize } f(\Phi_{(1)}) = {}_1\lambda_1 .$$

---

[1] See Corollary 1.6.2.1 which deals with the special case $m = 2$.

For any $\underline{Z}_{(1)}$, the corresponding ${}_1\lambda_1$ is the variance of the first principal component, ${}_1\underline{e}_1'\,\underline{Z}_{(1)}$. Thus the first stage canonical variates have the property that their first principal component has a variance at least as large as that for any other choice of $\underline{Z}_{(1)}$.

McKeon [16] and Carroll [3] have arrived at this same procedure using different approaches. Carroll's is particularly interesting. It is the logical extension of his formulation for two sets which was described in Section 1.6. Working in terms of a sample space, he finds (sample) canonical variates and an auxiliary (sample) variate which are most closely related in the sense that the sum of squared correlations of the auxiliary with each of the canonical variates is a maximum. The intimate connection between the auxiliary vector and the first principal component line that was stated in Section 1.6 is also valid in this context. More will be said about Carroll's work in Section 3.5.

The least informative $\Phi_{(1)}$ is the identity matrix in that there is no correlation between any two of the ${}_jZ_1$. It seems, therefore, that a logical criterion for picking $\underline{Z}_{(1)}$ would be to make $\Phi_{(1)}$ as "distant" from the identity as possible. This idea can be expressed mathematically in terms of the Euclidean norm: pick $\underline{Z}_{(1)}$ to maximize $||I - \Phi_{(1)}||$. Or, equivalently,

$$\text{(3.3.4)}\quad \text{maximize } f(\Phi_{(1)}) = \sum_{i=1}^{m} \sum_{j=1}^{m} \{\text{corr}({}_iZ_1, {}_jZ_1)\}^2 = \sum_{j=1}^{m} {}_j\lambda_1^2 \,.$$

At first glance, (3.3.4) appears to be only a variation of (3.3.1), but in effect it is a much different criterion. (3.3.4) is the SSQCOR criterion for determining $\underline{Z}_{(1)}$. It is the obvious generalization of (1.6.10). Loosely speaking, (3.3.4) calls for maximizing the variance of the ${}_j\lambda_1$.

The MINVAR criterion is, to some extent, the antithesis of the MAXVAR criterion. The MINVAR $\underline{Z}_{(1)}$ has the property that the variance of its m-th principal component is a minimum. In other words, $\underline{Z}_{(1)}$ is chosen to

(3.3.5) $$\text{minimize } f(\Phi_{(1)}) = {}_m\lambda_1 .$$

That this criterion is a generalization of the two set procedure follows from Corollary 1.6.2.1.

The GENVAR procedure was proposed by Steel [21] in 1951 and is the oldest of the five methods. As its name suggests, the criterion is to find $\underline{Z}_{(1)}$ with the smallest possible generalized variance. That is, one wants to

(3.3.6) $$\text{minimize } f(\Phi_{(1)}) = |\Phi_{(1)}| = \prod_{j=1}^{m} {}_j\lambda_1,$$

which is the natural generalization of (1.6.11).

3.4 The Sum of Correlations Model

Further insight into the statistical nature of the SUMCOR criterion can be obtained by considering the following model for $\underline{Z}_{(1)}$[2]:

(3.4.1) $$\underline{Z}_{(1)} = \underline{\ell}_1 F_1 + \underline{E}_1,$$

where $\underline{\ell}_1$ is a known non-null vector, F_1 is a standardized random variable, and $\underline{E}_1$ is a vector of error variables.

Suppose that F_1 is chosen so as to minimize the sum of residual variances, i.e.,

---

[2] The ${}_jZ_1$ will be assumed to have zero means in Sections 3.4 - 3.8 for convenience.

(3.4.2) $$\text{minimize } \operatorname{tr}\{\operatorname{var}(\underline{E}_1)\} .$$

Now

(3.4.3) $$\operatorname{tr}\{\operatorname{var}(\underline{E}_1)\} = \operatorname{tr}\{\Phi_{(1)} + \underline{\ell}_1\underline{\ell}_1' - \underline{\ell}_1\underline{\alpha}_1' - \underline{\alpha}_1\underline{\ell}_1'\}$$

$$= m + \underline{\ell}_1'\underline{\ell}_1 - 2\underline{\ell}_1'\underline{\alpha}_1$$

where

$$\underline{\alpha}_1 = \operatorname{cov}(\underline{Z}_{(1)}, F_1) .$$

The problem, therefore, is to choose F_1 so as to maximize $\underline{\ell}_1'\underline{\alpha}_1$. Adjusting the sign of F_1, if necessary, to make $\underline{\ell}_1'\underline{\alpha}_1$ non-negative, it follows that

$$0 \le \underline{\ell}_1'\underline{\alpha}_1 = \operatorname{cov}(\underline{\ell}_1'\underline{Z}_{(1)}, F_1) \le (\underline{\ell}_1'\Phi_{(1)}\underline{\ell}_1)^{\frac{1}{2}} ,$$

by the Cauchy-Schwarz inequality. The second inequality becomes an equality if and only if $\underline{\ell}_1'\underline{Z}_{(1)}$ and F_1 are linearly related (with probability one). Since the ${}_jZ_1$ and F_1 are standardized, $\underline{\ell}_1'\underline{\alpha}_1$ is maximized when

(3.4.4) $$F_1 = (\underline{\ell}_1'\Phi_{(1)}\underline{\ell}_1)^{-\frac{1}{2}}\, \underline{\ell}_1'\underline{Z}_{(1)}$$

and, for this choice of F_1,

$$\operatorname{tr}\{\operatorname{var}(\underline{E}_1)\} = m + \underline{\ell}_1'\underline{\ell}_1 - 2(\underline{\ell}_1'\Phi_{(1)}\underline{\ell}_1)^{\frac{1}{2}}.$$

Choosing canonical variates to give the best fit to (3.4.1), as measured by (3.4.2), is thus equivalent to finding ${}_jZ_1$ which maximize $\underline{\ell}_1'\Phi_{(1)}\underline{\ell}_1$. If $\underline{\ell}_1 \propto \underline{1}$, then this is just the SUMCOR procedure. In other words, Horst's method generates a $\underline{Z}_{(1)}$ having the "best" fitting common factor, assuming the factor contributes with the same potency to each of the ${}_jZ_1$.

3.5 The Maximum Variance Model

Carroll's development, mentioned in Section 3.3 and earlier in Section 1.6, can be reformulated in terms of random variables and the nature of the auxiliary variable revealed. A model for the MAXVAR procedure will be arrived at in the process. Returning to the model (3.4.1), suppose that $\underline{\ell}_1$ is also allowed to vary so that both $\underline{\ell}_1$ and F_1 are to be found according to the criterion (3.4.2).

Equation (3.4.3) can be rewritten as

$$\text{(3.5.1)} \qquad \text{tr}\{\text{var}(\underline{E}_1)\} = m + (\underline{\ell}_1 - \underline{\alpha}_1)'(\underline{\ell}_1 - \underline{\alpha}_1) - \underline{\alpha}_1'\underline{\alpha}_1.$$

The second term on the right hand side is a positive definite quadratic form in $(\underline{\ell}_1 - \underline{\alpha}_1)$; consequently, the minimum of (3.5.1) with respect to $\underline{\ell}_1$ occurs when $\underline{\ell}_1 = \underline{\alpha}_1$. With this choice of $\underline{\ell}_1$,

$$\text{tr}\{\text{var}(\underline{E}_1)\} = m - \underline{\alpha}_1'\underline{\alpha}_1$$

$$= m - \sum_{j=1}^{m} \{\text{corr}({}_jZ_1, F_1)\}^2 .$$

The last sum corresponds to Carroll's criterion with F_1 in the role of the auxiliary variate.

The overall minimum of (3.5.1), for given $\underline{Z}_{(1)}$, is obtained when

$$\text{(3.5.2)} \qquad \underline{\ell}_1 = {}_1\lambda_1^{\frac{1}{2}}\ {}_1\underline{e}_1 \quad \text{and} \quad F_1 = {}_1\lambda_1^{-\frac{1}{2}}\ {}_1\underline{e}_1'\ \underline{Z}_{(1)}.$$

Using (3.5.2),

$$\text{(3.5.3)} \qquad \mathrm{tr}\{\mathrm{var}(\underline{E}_1)\} = m - {}_1\lambda_1 .$$

In the limiting case given by ${}_1\lambda_1 = m$, the fit is perfect, each element of ${}_1\underline{e}_1$ is $\pm m^{-\frac{1}{2}}$, and $|\mathrm{corr}({}_jZ_1, F_1)| = 1$ for $j = 1, 2, \ldots, m$.

The $\underline{\ell}_1$ and F_1 in (3.5.2) are, in fact, the optimal choices for a general class of criterion functions, one of which is the trace function (3.4.2). This should be clear from the discussion, in Section 1.6, of the work of Okamoto and Kanazawa [18] (see also Lemma A3). Thus (3.4.2) can be replaced, in this section, by

$$\text{(3.5.4)} \qquad \text{minimize } f(\mathrm{var}(\underline{E}_1)) = g(\theta_1, \theta_2, \ldots, \theta_m) ,$$

where g is any strictly increasing function in each θ_j, the eigenvalues of the non-negative definite matrix in the argument of f.

The goodness of fit obtainable depends, of course, on the choice of $\underline{Z}_{(1)}$. The MAXVAR method is equivalent to finding $\underline{Z}_{(1)}$ which admits the best fit of (3.4.1) as measured by (3.5.4), with both $\underline{\ell}_1$ and F_1 at one's disposal. This follows from (3.3.3) and (3.5.3) or, more generally, Lemma A3.

3.6 The Sum of Squared Correlations Model

The model appropriate for the sum of squared correlations method requires m factors: for a given $\underline{Z}_{(1)}$,

$$\text{(3.6.1)} \qquad \underline{Z}_{(1)} = \sum_{j=1}^{m} {}_j\underline{\ell}_1 \, {}_jF_1 + \underline{E}_1 ,$$

where the ${}_j\underline{\ell}_1$ are arbitrary non-null vectors, the ${}_jF_1$ are standardized random variables, and $\underline{E}_1$ is a vector of error variables. (Of course, with m factors, it is possible to choose the ${}_j\underline{\ell}_1$ and ${}_jF_1$ such that $\underline{E}_1 = \underline{0}$.)

The factors are determined so that ${}_1F_1$ is the most important, ${}_2F_1$ the second most important, and so on. More precisely, the ${}_j\underline{\ell}_1$ and ${}_jF_1$ are chosen to

$$\text{(3.6.2)} \quad \text{minimize } f(\text{var}(\underline{Z}_{(1)} - \sum_{i=1}^{j} {}_i\underline{\ell}_1 \, {}_iF_1)) = g(\theta_1, \theta_2, \ldots, \theta_m), \quad j = 1, 2, \ldots, m,$$

where g is the function in (3.5.4). The correct choice for ${}_1\underline{\ell}_1$ and ${}_1F_1$ follows from (3.5.2). To obtain ${}_2\underline{\ell}_1$ and ${}_2F_1$, subtract ${}_1\underline{\ell}_1 \, {}_1F_1$ from both sides of (3.6.1). Then (3.5.2) may be used to determine the optimal ${}_2\underline{\ell}_1$ and ${}_2F_1$, remembering that the largest eigenvalue of $(\Phi_{(1)} - {}_1\lambda_1 \, {}_1\underline{e}_1 \, {}_1\underline{e}_1')$ is ${}_2\lambda_1$ with associated eigenvector ${}_2\underline{e}_1$. Continuing in this way through m steps, it is seen that

$$\text{(3.6.3)} \quad {}_j\underline{\ell}_1 = {}_j\lambda_1^{\frac{1}{2}} \, {}_j\underline{e}_1 \quad \text{and} \quad {}_jF_1 = {}_j\lambda_1^{-\frac{1}{2}} \, {}_j\underline{e}_1' \, \underline{Z}_{(1)}, \quad j = 1, 2, \ldots, m.$$

The factors are just the standardized principal components of $\underline{Z}_{(1)}$ and hence are uncorrelated. With f as the trace function in (3.6.2), the minimum sum of residual variances with j factors is $m - \sum_{i=1}^{j} {}_i\lambda_1$. Thus, by adding on the j-th factor, a reduction of ${}_j\lambda_1$ from the minimum sum with (j-1) factors may be attained.

An external criterion is needed as a basis for choosing the optimal $\underline{Z}_{(1)}$, and (3.3.4) is the one which will be used here.

The maximum of the sum of squares of m numbers which add to m and lie in the range zero to one is m^2, and it occurs when one number is m and the rest are zero. The minimum sum is m and results when each number is equal to one. This suggests that the MAXVAR and SSQCOR methods will yield similar $\underline{Z}_{(1)}$'s whenever most of the variability can be accounted for by a single factor. It further suggests that the effect of the criterion (3.3.4) will be to produce a $\underline{Z}_{(1)}$ such that its first "few" factors account for most of the variability and the last "few" very little of the variability. That is, the tendency should be to spread out the ${}_j\lambda_1$ as much as possible. This is in contrast with each factor being equally important $({}_j\lambda_1 = 1, \; j = 1, 2, \ldots, m)$ which is the situation when and only when $\Phi_{(1)}$ is the identity matrix.

3.7 The Minimum Variance Model

MINVAR canonical variates admit the best possible representation in terms of (m-1) factors. To see in what sense this is true, consider an (m-1) factor model,

$$\underline{Z}_{(1)} = \sum_{j=1}^{m-1} {}_j\underline{\ell}_1 \, {}_jF_1 + \underline{E}_1, \tag{3.7.1}$$

for an arbitrary $\underline{Z}_{(1)}$. The ${}_j\underline{\ell}_1$, ${}_jF_1$, and $\underline{E}_1$ are as in (3.6.1). If the ${}_j\underline{\ell}_1$ and ${}_jF_1$ are chosen to satisfy (3.5.4), then the minimum is reached when

$${}_j\underline{\ell}_1 = {}_j\lambda_1^{\frac{1}{2}} \, {}_j\underline{e}_1 \quad \text{and} \quad {}_jF_1 = {}_j\lambda_1^{-\frac{1}{2}} \, {}_j\underline{e}_1' \, \underline{Z}_{(1)}, \quad j = 1, 2, \ldots, m-1. \tag{3.7.2}$$

This follows from Lemma A3. Alternatively the ${}_j\underline{\ell}_1$ and ${}_jF_1$ can be picked sequentially using (3.6.2) for $j = 1, 2, \ldots, m-1$. The optimal choices are, again, as recorded in (3.7.2). Operating with criterion (3.5.4) has the advantage that orthogonal rotations on the factors (and simultaneously on the ${}_j\underline{\ell}_1$) can be made without disturbing the measure of fit. These rotated factors, however, would not be expected to possess the pleasing properties of those in (3.7.2).

If the ${}_j\underline{\ell}_1$ and ${}_jF_1$ specified in (3.7.2) are substituted into (3.7.1), then (3.5.4) becomes

$$f(\mathrm{var}(\underline{E}_1)) = g({}_m\lambda_1, 0, \ldots, 0) .$$

From this point, it is clear why MINVAR canonical variates can be interpreted as those which admit the best fit in terms of (m-1) factors.

Evidently the optimal (m-1) factors in (3.7.2) are such that the single remaining (or m-th) factor needed to obtain a perfect fit in (3.7.1) contributes the smallest possible amount to the residual variances. The m-th factor is ${}_mF_1 = {}_m\lambda_1^{-\frac{1}{2}}\, {}_m\underline{e}_1'\, \underline{Z}_{(1)}$, the corresponding vector of weights is ${}_m\underline{\ell}_1 = {}_m\lambda_1^{\frac{1}{2}}\, {}_m\underline{e}_1$, and the error vector is $\underline{E}_1 = {}_m\underline{\ell}_1\, {}_mF_1$.

3.8 <u>The Generalized Variance Model</u>

The GENVAR method can be motivated in terms of the m factor model in (3.6.1), using the criterion (3.6.2) and the results in (3.6.3). Having done this for a fixed $\underline{Z}_{(1)}$, it remains to define an effective external criterion, in this case (3.3.6), for the optimal selection of $\underline{Z}_{(1)}$.

The SSQCOR criterion places main emphasis on the first "few" factors or ${}_j\lambda_1$ and less on the last "few." The GENVAR criterion has the opposite emphasis: it is directed at diminishing the contribution of the last "few" factors or ${}_j\lambda_1$ with less weight given to the effect of the first "few."

One might expect that the MINVAR and GENVAR methods would yield similar $\underline{Z}_{(1)}$'s whenever it is possible to account for most of the variability (in some $\underline{Z}_{(1)}$) with (m-1) factors. A note of caution should be added, however, as the multiplicative type function in (3.3.6) is especially sensitive to the values of the "smaller" ${}_j\lambda_1$ and, consequently, it is difficult to predict exactly how it will behave.

3.9 Higher Stage Canonical Variables

The study of relations among the sets can be continued beyond the first stage by considering higher stage or higher order canonical variates $\underline{Z}_{(2)}, \underline{Z}_{(3)}, \cdots$ to supplement the optimal $\underline{Z}_{(1)}$. The same criterion function is used at each stage, but restrictions are added to assure that the canonical variables for a particular stage differ from those for the previous stages. The restrictions are such that if the number of sets were only two, the selected variates would be the usual canonical pair as defined in Section 1.3.

It has already been demonstrated that there are a number of workable criterion functions for the several set situation which provide generalizations of the two set criterion. It is also true that for each method there exist several different kinds of restrictions which can be used in the construction of higher stage canonical variables, each leading to a different choice of variables and each being a generalization of the restrictions used for two sets.

The models introduced in Sections 3.4 - 3.8 (or, in two instances to be specified later, generalizations of them) can be adapted to a general s-th stage by expressing each in terms of an arbitrary member of the class of admissible $\underline{Z}_{(s)}$'s. One then seeks that $\underline{Z}_{(s)}$ which admits the best fit according to the appropriate standard.

A simple but useful type of restriction on $\underline{Z}_{(s)}$ is

$$(3.9.1) \qquad \mathrm{corr}({}_jZ_i, {}_jZ_s) = 0, \quad i = 1, 2, \ldots, s-1; \; j = 1, 2, \ldots, m.$$

($\underline{Z}_{(j)}$, $j = 1, 2, \ldots, s-1$, are, of course, the canonical variates for the preceding stages.) In other words, the canonical variates are required to be uncorrelated within sets. (3.9.1) has been used by Horst ([8], [9], and [10]) in connection with the SUMCOR and MAXVAR procedures. When $m = 2$, (3.9.1) is equivalent to (1.3.1) and (1.3.2).

A more flexible class of constraints is obtained by requiring the canonical variables to be uncorrelated in some but not necessarily all of the sets:

$$(3.9.2) \qquad \mathrm{corr}({}_jZ_i, {}_jZ_s) = 0, \quad i = 1, 2, \ldots, s-1; \; j \in I_k$$

where I_k is a non-empty subset of k of the first m integers, specifically, $i_1 < i_2 < \ldots < i_k$. When $k = m$, (3.9.2) is the same as (3.9.1).

The most rigid restriction to be considered is

$$(3.9.3) \qquad \mathrm{corr}(\underline{Z}_{(j)}, \underline{Z}_{(s)}) = 0, \quad j = 1, 2, \ldots, s-1.$$

That is, the canonical variables at the s-th stage must be uncorrelated with all the lower order canonical variables. When $m = 2$, (3.9.3) is equivalent to (1.3.1) - (1.3.4). If ${}_jF_s$, $j = 1, 2, \ldots, m$, are the best fitting factors associated with the s-th stage fit of an m factor model like (3.6.1) using a criterion like (3.6.2), and if the admissible $\underline{Z}_{(s)}$'s are those for which

$$(3.9.4) \quad \text{corr}({}_i F_j, {}_k F_s) = 0, \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, s-1; \quad k = 1, 2, \ldots, m,$$

then (3.9.4) is equivalent to (3.9.3). This was proved in Corollary 1.6.2.4 for the case $m = 2$ and a similar proof applies here.

The final type of constraint to be considered is relevant to the MAXVAR or MINVAR method. This constraint is important because it leads to canonical variables which correspond to Carroll's higher stage variables.

Extensions of the MAXVAR and MINVAR first stage models are required in this connection. For the MAXVAR method, the model is

$$(3.9.5) \qquad \underline{Z}_{(s)} = \sum_{j=1}^{k_s} {}_j\underline{\ell}_s \, {}_jF_s + \underline{E}_s, \quad 1 \le k_s < m ,$$

where the ${}_j\underline{\ell}_s$ and ${}_jF_s$ are chosen sequentially in accordance with the s-th stage equivalent of (3.6.2) for $j = 1, 2, \ldots, k_s$. The constraint is in terms of the "last" factors, namely,

$$(3.9.6) \qquad \text{corr}({}_{k_j}F_j, {}_{k_s}F_s) = 0, \quad j = 1, 2, \ldots, s-1,$$

and the criterion is to

$$(3.9.7) \qquad \text{maximize } {}_{k_s}\lambda_s$$

with respect to admissible k_s and $\underline{Z}_{(s)}$. It will be shown in Section 4.5 that $k_1 = 1$ always and that the maximum of (3.9.7) is $c_s(R)$.

The revised model for the MINVAR method is

$$(3.9.8)\qquad \underline{Z}_{(s)} = \sum_{j=1}^{k_s-1} {}_j\underline{\ell}_s \; {}_jF_s + \underline{E}_s, \quad 1 < k_s \le m,$$

where the ${}_j\underline{\ell}_s$ and ${}_jF_s$ are selected in accordance with the s-th stage equivalent of (3.5.4) as prescribed by Lemma A3. The constraint on the higher stages is again (3.9.6) where ${}_{k_s}F_s$ is now the next factor which would be added if an additional term were allowed in (3.9.8). (This new factor is just ${}_{k_s}F_s$ in (3.9.5).) The criterion is to

$$(3.9.9)\qquad \text{minimize } {}_{k_s}\lambda_s$$

with respect to admissible k_s and $\underline{Z}_{(s)}$. It will be seen in Section 4.6 that $k_1 = m$ of necessity and that the minimum in (3.9.9) is $c_{p-s+1}(R)$.

In either situation, it is easy to prove that (3.9.6) can be rewritten as

$$(3.9.10)\qquad {}_{k_j}\underline{e}_j' \operatorname{corr}(\underline{Z}_{(j)}, \underline{Z}_{(s)}) \; {}_{k_s}\underline{e}_s = 0, \quad j = 1, 2, \ldots, s-1.$$

(See Corollaries 1.6.2.2 and 1.6.2.3 for the special cases with $m = 2$.)

When restrictions (3.9.1) or (3.9.3) are used in place of (3.9.10) with the criteria (3.9.7) and (3.9.9), $k_j = 1$ at all stages of the MAXVAR procedure and $k_j = m$ at all stages of the MINVAR procedure. This, too, follows from Sections 4.5 and 4.6.

Nothing has been said so far about the number of stages that can be usefully employed in conjunction with any one of the indicated restrictions. This number, s_o say, specifies how many non-redundant $\underline{Z}_{(s)}$'s can be generated in a manner which is consistent with the procedure when $m = 2$.

Restrictions (3.9.1), (3.9.2), and (3.9.3) are applicable to any of the five methods. For (3.9.1), $s_o = p_1$, the number of variables in the least numerous set. For (3.9.2) with $k > 1$, $s_o = p_{i_1}$. The case $k = 1$ is more complex. Simply stated s_o is the greater of one and n, where n is the number of stages for which there exist canonical variables with $\Phi_{(s)} \neq I$ (cf. Section 1.5). [It is easy to prove that n is the greatest s such that the value of (4.5.8) is greater than one -- if $R = I$, $n = 0$.] As for (3.9.3), s_o is the largest value of s for which (3.9.3) can be satisfied. [In terms of the matrices ${}_jC^*_s$ defined in (4.2.8), and interpreting ${}_jC^*_1$ as the zero matrix, s_o is the maximum s such that $(p_j - \text{rank}({}_jC^*_s)) > 0$ for $j = 1, 2, \ldots, m$.]

One can deduce from Theorem 1.6.1 and Corollary 1.6.2.2 (1.6.2.3) that the number of stages for the MAXVAR (MINVAR) method using restriction (3.9.6) when $m = 2$ is the number of eigenvalues of R greater (less) than one. The situation for $m > 2$ is somewhat arbitrary. The approach taken here is that the number of MAXVAR (MINVAR) stages should be at least equal to the number of $c_j(R)$ greater (less) than one. The ambiguity arises in connection with the unit eigenvalues of R. For each $c_j(R) = 1$, it is possible that the associated variables, $\underline{Z}_{(s)}$ say, have $\Phi_{(s)} = I$ in which case they should not be considered as a separate stage for the same reasons as when $m = 2$. If $\Phi_{(s)} \neq I$, $\underline{Z}_{(s)}$ can reasonably be considered as part of either the MAXVAR or MINVAR system.

Horst introduced a technique in [9], later labeled the "oblique maximum variance method" in [10], which is mathematically equivalent to p_1 stages of the MAXVAR procedure using restriction (3.9.6). His "rank one approximation method" ([9] and [10]) is equivalent to p_1 stages of the MAXVAR procedure using (3.9.1). From the present point of view, the only difference between these two methods of Horst's is in the type of restriction used at the higher stages. However, it should be clear from Theorem 1.6.1 that the first of these is not strictly within the province of canonical analysis when $m = 2$, unless R_{12} is of full rank $(r = p_1)$.

Carroll used his technique to generate what amounts to p stages of the MAXVAR (or MINVAR) method using (3.9.6). Thus his derived variates are a mixture of MAXVAR and MINVAR canonical variates plus, possibly, some others corresponding to eigenvalues of R equal to one.

CHAPTER IV

PROCEDURES FOR THE CANONICAL ANALYSIS OF SEVERAL SETS OF VARIABLES

4.1 Introduction

Procedures are developed in this chapter for finding the various stages of canonical variates and the corresponding criterion values.

The SUMCOR, SSQCOR, and GENVAR problems, considered in Sections 4.2 - 4.4, require iterative procedures for generating the optimal $\underline{Z}_{(s)}$. Fortunately, each procedure is of the same basic type so that one general computer program can be used for all of them. The procedures are such that convergence is guaranteed; and, given that the starting points are appropriately chosen, the end products will be coefficient vectors which define the desired canonical variates.

Horst proposed a different type of iterative procedure for finding SUMCOR first stage and higher order canonical variates in accordance with restriction (3.9.1). The convergence properties of his procedure, however, are not known.

The MAXVAR and MINVAR procedures are presented in Sections 4.5 and 4.6. The associated canonical variates and criterion values are related to the eigenvectors and eigenvalues of certain known matrices and hence are easy to determine. The MAXVAR procedure, using (3.9.1) as the constraint on the higher stages, is due to Horst [9].

4.2 The Sum of Correlations Procedure

The following special notation will be used:

$$f_s = \underline{1}' D_{B*(s)} R D'_{B*(s)} \underline{1} = \underline{1}' \Phi_{(s)} \underline{1},$$

(4.2.1) $${}_j N_s = (\Sigma_{j1}\, {}_1\underline{b}_s, \ldots, \Sigma_{jj-1}\, {}_{j-1}\underline{b}_s, \Sigma_{jj+1}\, {}_{j+1}\underline{b}_s, \ldots, \Sigma_{jm}\, {}_m\underline{b}_s),$$

(4.2.2) $${}_j N^*_s = (R_{j1}\, {}_1\underline{b}^*_s, \ldots, R_{jj-1}\, {}_{j-1}\underline{b}^*_s, R_{jj+1}\, {}_{j+1}\underline{b}^*_s, \ldots, R_{jm}\, {}_m\underline{b}^*_s),$$

$${}_j\underline{h}_s = {}_j N_s \underline{1},$$

and

$${}_j\underline{h}^*_s = {}_j N^*_s \underline{1}.$$

The s-th stage of the SUMCOR method involves the maximization of f_s subject to certain restrictions on $D_{B*(s)}$. For the first stage, start from the Lagrangian equation

$$g_1 = f_1 + \underline{1}'(I - D_{B*(1)} D'_{B*(1)})\underline{\theta}_1 ,$$

where $\underline{\theta}'_1 = ({}_1\theta_1, {}_2\theta_1, \ldots, {}_m\theta_1)$ is a vector of Lagrange multipliers. Differentiating with respect to $D'_{B*(1)}\underline{1}$ gives

$$\frac{\partial g_1}{\partial D'_{B*(1)}\underline{1}} = 2RD'_{B*(1)}\underline{1} - 2D'_{B*(1)}\underline{\theta}_1 .$$

Equating the derivative to zero leads to

(4.2.3) $$RD'_{B*(1)}\underline{1} = D'_{B*(1)}\underline{\theta}_1$$

or

$$(4.2.4) \qquad {}_j\underline{h}^*_1 = ({}_j\theta_1 - 1)\,{}_j\underline{b}^*_1, \quad j = 1, 2, \ldots, m.$$

Premultiply (4.2.3) by $D_{B^*(1)}$ or (4.2.4) by ${}_j\underline{b}^{*\prime}_1$; then

$$\underline{\theta}_1 = D_{B^*(1)} R D'_{B^*(1)} \underline{1} = \Phi_{(1)} \underline{1}$$

which means that

$$(4.2.5) \qquad {}_j\theta_1 = \sum_{i=1}^{m} \operatorname{corr}({}_jZ_1, {}_iZ_1), \quad j = 1, 2, \ldots, m.$$

The relation between f_1 and the Lagrange multipliers is

$$(4.2.6) \qquad f_1 = \underline{1}'\underline{\theta}_1 .$$

Suppose that the restrictions imposed at the s-th stage $(s > 1)$ can be expressed in terms of matrices ${}_jC^*_s$ in the following manner:

$$(4.2.7) \qquad {}_j\underline{b}^{*\prime}_s \, {}_jC^*_s = \underline{0}', \quad j = 1, 2, \ldots, m.$$

Restrictions (3.9.1), (3.9.2), and (3.9.3), in particular, can be expressed in this form. For (3.9.2) (which includes (3.9.1))

$${}_jC^*_s = \begin{cases} ({}_j\underline{b}^*_1, {}_j\underline{b}^*_2, \ldots, {}_j\underline{b}^*_{s-1}), & j \in I_k \\ 0 & , \text{ otherwise .} \end{cases}$$

Alternatively, if the restrictions are like those in (3.9.3), then

$$(4.2.8) \qquad {}_jC^*_s = ({}_j\underline{b}^*_1, {}_j\underline{b}^*_2, \ldots, {}_j\underline{b}^*_{s-1}, {}_jN^*_1, {}_jN^*_2, \ldots, {}_jN^*_{s-1}).$$

The appropriate s-th stage Lagrangian equation is

$$(4.2.9) \qquad g_s = f_s + \sum_{j=1}^{m} (1 - {}_j\underline{b}^{*\prime}_s \, {}_j\underline{b}^*_s)\, {}_j\theta_s - 2 \sum_{j=1}^{m} {}_j\underline{b}^{*\prime}_s \, {}_jC^*_s \, {}_j\underline{\gamma}_s ,$$

where the ${}_j\theta_s$ and ${}_j\underline{\gamma}_s$ are Lagrange multipliers with $\underline{\theta}_s' = ({}_1\theta_s, {}_2\theta_s, \ldots, {}_m\theta_s)$.

$$\frac{\partial g_s}{\partial {}_j\underline{b}_s^*} = 2{}_j\underline{h}_s^* - 2({}_j\theta_s - 1){}_j\underline{b}_s^* - 2{}_jC_s^*\,{}_j\underline{\gamma}_s \,.$$

Setting this equal to zero yields

$$\text{(4.2.10)} \qquad \underline{0} = {}_j\underline{h}_s^* - ({}_j\theta_s - 1){}_j\underline{b}_s^* - {}_jC_s^*\,{}_j\underline{\gamma}_s \,.$$

Multiplying by ${}_jC_s^{*\prime}$ and solving for ${}_j\underline{\gamma}_s$, one obtains the general solution (cf. Rao [20], p.26)

$$\text{(4.2.11)} \quad {}_j\underline{\gamma}_s = ({}_jC_s^{*\prime}\,{}_jC_s^*)^-\,{}_jC_s^{*\prime}\,{}_j\underline{h}_s^* + \{I - ({}_jC_s^{*\prime}\,{}_jC_s^*)^-\,({}_jC_s^{*\prime}\,{}_jC_s^*)\}\underline{u} \,,$$

where $\underline{u}$ is arbitrary. Substituting (4.2.11) into (4.2.10) gives

$$\text{(4.2.12)} \quad \underline{0} = \{I - {}_jC_s^*({}_jC_s^{*\prime}\,{}_jC_s^*)^-\,{}_jC_s^{*\prime}\}\,{}_j\underline{h}_s^* - ({}_j\theta_s - 1)\,{}_j\underline{b}_s^* \,, \qquad j = 1, 2, \ldots, m.$$

It may be inferred from (4.2.12) that

$$\text{(4.2.13)} \qquad \begin{cases} {}_j\theta_s = \sum_{i=1}^{m} \operatorname{corr}({}_jZ_s, {}_iZ_s) \\ \text{subject to} \\ {}_j\underline{b}_s^* \in V(I - {}_jC_s^*({}_jC_s^{*\prime}\,{}_jC_s^*)^-\,{}_jC_s^{*\prime}) \end{cases}$$

and

$$\text{(4.2.14)} \qquad f_s = \underline{1}'\underline{\theta}_s.$$

If ${}_jC_1^*$ is defined to be the zero matrix, then the above results apply for $s = 1$ as well: (4.2.4), (4.2.5), and (4.2.6) are just special cases of (4.2.12), (4.2.13), and (4.2.14).

In terms of the original $\underline{X}$ variables, (4.2.12) becomes

$$(4.2.15)\quad \underline{0} = \{\Sigma_{jj}^{-1} - \Sigma_{jj}^{-1}\, {}_jC_s({}_jC_s'\, \Sigma_{jj}^{-1}\, {}_jC_s)^{-}\, {}_jC_s'\, \Sigma_{jj}^{-1}\}\, {}_j\underline{h}_s - ({}_j\theta_s - 1)\, {}_j\underline{b}_s,$$
$$j = 1, 2, \ldots, m,$$

where

$${}_jC_s = \Sigma_{jj}^{\frac{1}{2}}\, {}_jC_s^*,$$

and (4.2.7) becomes

$${}_j\underline{b}_s'\, {}_jC_s = \underline{0}', \quad j = 1, 2, \ldots, m.$$

The form of (4.2.12) or (4.2.15) suggests the following iterative procedure (described in terms of the $\underline{Y}$ variables) for generating the optimal s-th stage $\underline{Z}_{(s)}$:

(i) specify initial variables ${}_jZ_s^{(o)}$ by vectors ${}_j\underline{b}_s^{*(o)}$ which satisfy (4.2.7);

(ii) for $n = 1, 2, \ldots$ and with ${}_j\underline{h}_s^{*(n)}$ fixed, solve

$$(4.2.16)\quad \underline{0} = \{I - {}_jC_s^*({}_jC_s^{*\prime}\, {}_jC_s^*)^{-}\, {}_jC_s^{*\prime}\}\, {}_j\underline{h}_s^{*(n)} - ({}_j\theta_s^{(n)} - 1)\, {}_j\underline{b}_s^{*(n)},$$
$$j = 1, 2, \ldots, m,$$

obtaining the maximum ${}_j\theta_s^{(n)}$ and associated ${}_j\underline{b}_s^{*(n)}$.

The vector ${}_j\underline{h}_s^{*(n)}$ is computed like ${}_j\underline{h}_s^*$ using

$$(4.2.17)\quad {}_1\underline{b}_s^{*(n)}, \ldots, {}_{j-1}\underline{b}_s^{*(n)}, {}_{j+1}\underline{b}_s^{*(n-1)}, \ldots, {}_m\underline{b}_s^{*(n-1)}$$

so that

$$(4.2.18)\quad ({}_j\theta_s^{(n)} - 1) = \sum_{i=1}^{j-1} \mathrm{corr}({}_jZ_s^{(n)}, {}_iZ_s^{(n)}) + \sum_{i=j+1}^{m} \mathrm{corr}({}_jZ_s^{(n)}, {}_iZ_s^{(n-1)}) .$$

This procedure, as will be shown, must converge monotonically in ${}_j f_s^{(n)}$, the value of f_s using ${}_1Z_s^{(n)}, \ldots, {}_jZ_s^{(n)}, {}_{j+1}Z_s^{(n-1)}, \ldots, {}_mZ_s^{(n-1)}$, to a solution of (4.2.12). The solution will be optimal if the initial variables are appropriately chosen. A sufficient condition for the solution to be optimal is for $\sum_{i=1}^{m} \sum_{j=1}^{m} \operatorname{corr}({}_iZ_s^{(o)}, {}_jZ_s^{(o)})$ to be greater than for any non-optimal $\underline{Z}_{(s)}$ which comprises a solution of (4.2.12).

Observe that for $s = 1$ the calculation which is needed at each step of the iteration procedure is equivalent to the calculation of β_1 in Theorem 1.7.1 (ii). For $s > 1$, the calculation is of the same type but with linear restrictions on the admissible coefficient vectors. In either case, the calculation is quick and easy to execute.

The nature of the calculation is such that

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

Thus, since ${}_jf_s^{(n)}$ is bounded above by m^2,

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

where $\tilde{f}_s$ is a positive finite number.

Define ${}_j\hat{\theta}_s^{(n)}$ by the equation

(4.2.21)
$$({}_j\hat{\theta}_s^{(n)} - 1) = \sum_{i=1}^{j-1} \operatorname{corr}({}_jZ_s^{(n-1)}, {}_iZ_s^{(n)}) + \sum_{i=j+1}^{m} \operatorname{corr}({}_jZ_s^{(n-1)}, {}_iZ_s^{(n-1)}),$$

and let

(4.2.22) $$ {}_of_s^{(n)} = {}_mf_s^{(n-1)} .$$

Note that

$$(4.2.23) \qquad 1 \le {}_j\hat{\theta}_s^{(n)} \le {}_j\theta_s^{(n)} < m$$

and

$$(4.2.24) \qquad {}_jf_s^{(n)} - {}_{j-1}f_s^{(n)} = 2({}_j\theta_s^{(n)} - {}_j\hat{\theta}_s^{(n)}), \quad j = 1, 2, \ldots, m.$$

The limit of the left hand side of (4.2.24) is zero as $n \to \infty$ by virtue of (4.2.19) and (4.2.20) so that

$$(4.2.25) \qquad \lim_{n \to \infty} ({}_j\theta_s^{(n)} - {}_j\hat{\theta}_s^{(n)}) = 0.$$

This implies that each element of $({}_j\underline{b}_s^{*(n)} - {}_j\underline{b}_s^{*(n-1)})$ can be made arbitrarily small in magnitude by taking n sufficiently large, provided that ${}_j\theta_s^{(n)} > 0$. (For ${}_j\theta_s^{(n)} = 0$, ${}_j\underline{b}_s^{*(n)}$ is arbitrary.)

Suppose $\underline{\tilde{Z}}_{(s)}$ is a particular choice of s-th stage variables with associated ${}_j\tilde{\underline{b}}_s^*$, ${}_j\tilde{\theta}_s$, ${}_j\tilde{\underline{h}}_s^*$, and $\tilde{\Phi}_{(s)}$ such that

$$\tilde{f}_s = \underline{1}'\tilde{\Phi}_{(s)}\underline{1}$$

where $\tilde{f}_s$ is the limiting value in (4.2.20). Then the ${}_j\tilde{\underline{b}}_s^*$ and ${}_j\tilde{\theta}_s$ comprise a solution of (4.2.12) since $\tilde{f}_s$ can not be increased by changing any one of the ${}_j\tilde{\underline{b}}_s^*$.

In practice, where the criterion values connected with different solutions of (4.2.12) will be distinct and where the ${}_j\theta_s^{(n)}$ will always be positive, the situation is such that

(4.2.26) $$\lim_{n\to\infty} ({}_j\underline{b}_s^{*(n)} - {}_j\underline{b}_s^{*(n-1)}) = \underline{0},$$

(4.2.27) $$\lim_{n\to\infty} {}_j\underline{b}_s^{*(n)} = {}_j\tilde{\underline{b}}_s^{*},$$

(4.2.28) $$\lim_{n\to\infty} {}_j\underline{h}_s^{*(n)} = {}_j\tilde{\underline{h}}_s^{*},$$ $$\lim_{n\to\infty} {}_j\theta_s^{(n)} = {}_j\tilde{\theta}_s,$$

(4.2.20) holds, and the ${}_j\tilde{\underline{b}}_s^*$ and ${}_j\tilde{\theta}_s$ form a solution of (4.2.12).

4.3 The Sum of Squared Correlations Procedure

The SSQCOR procedure is developed in much the same way as the SUMCOR procedure. Define

(4.3.1) $${}_jP_s = {}_jN_s\,{}_jN_s' \quad \text{and} \quad {}_jP_s^* = {}_jN_s^*\,{}_jN_s^{*\prime}$$

where ${}_jN_s$ and ${}_jN_s^*$ are as in (4.2.1) and (4.2.2). The criterion function at the s-th stage is

$$f_s = \sum_{j=1}^{m} ({}_j\underline{b}_s^{*\prime}\,{}_j\underline{b}_s^*)^2 + \sum_{\substack{i=1\\i\neq j}}^{m}\sum_{j=1}^{m} ({}_i\underline{b}_s^{*\prime}\,R_{ij}\,{}_j\underline{b}_s^*)^2,$$

which is to be maximized subject to certain restrictions on $D_{B^*(s)}$.

The restrictions which are relevant to the higher stages of the SSQCOR procedure are all of the form (4.2.7). If ${}_jC_1^*$ is again used to represent the zero matrix, one need only consider the general Lagrangian equation

$$g_s = f_s + 2\sum_{j=1}^{m} (1 - {}_j\underline{b}_s^{*\prime}\,{}_j\underline{b}_s^*)\,{}_j\theta_s - 4\sum_{j=1}^{m} {}_j\underline{b}_s^{*\prime}\,{}_jC_s^*\,{}_j\underline{\gamma}_s$$

where the ${}_j\theta_s$ and ${}_j\underline{\gamma}_s$ are Lagrange multipliers as in (4.2.9).

$$\frac{\partial g_s}{\partial\, {}_j\underline{b}^*_s} = 4\,{}_jP^*_s\,{}_j\underline{b}^*_s - 4({}_j\theta_s - 1)\,{}_j\underline{b}^*_s - 4\,{}_jC^*_s\,{}_j\underline{\gamma}_s$$

which, equating to zero, gives

(4.3.2) $$\underline{0} = {}_jP^*_s\,{}_j\underline{b}^*_s - ({}_j\theta_s - 1)\,{}_j\underline{b}^*_s - {}_jC^*_s\,{}_j\underline{\gamma}_s \,.$$

Multiplying by ${}_jC^{*\prime}_s$ and solving for ${}_j\underline{\gamma}_s$ yields

(4.3.3) $${}_j\underline{\gamma}_s = ({}_jC^{*\prime}_s\,{}_jC^*_s)^-\,{}_jC^{*\prime}_s\,{}_jP^*_s\,{}_j\underline{b}^*_s + \{I - ({}_jC^{*\prime}_s\,{}_jC^*_s)^-\,({}_jC^{*\prime}_s\,{}_jC^*_s)\}\underline{u},$$

where $\underline{u}$ is arbitrary. Substituting (4.3.3) into (4.3.2), one arrives at

(4.3.4) $$\underline{0} = [\{I - {}_jC^*_s({}_jC^{*\prime}_s\,{}_jC^*_s)^-\,{}_jC^{*\prime}_s\}\,{}_jP^*_s - ({}_j\theta_s - 1)\,I]\,{}_j\underline{b}^*_s,$$
$$j = 1, 2, \ldots, m,$$

which implies that

$$\begin{cases} {}_j\theta_s = \sum_{i=1}^{m} \{\mathrm{corr}({}_jZ_s, {}_iZ_s)\}^2 \\ \text{subject to} \\ {}_j\underline{b}^*_s \in V(I - {}_jC^*_s({}_jC^{*\prime}_s\,{}_jC^*_s)^-\,{}_jC^{*\prime}_s) \end{cases}$$

and

$$f_s = \underline{1}'\underline{\theta}_s.$$

In terms of the original $\underline{X}$ variables, (4.3.4) becomes

(4.3.5) $$\underline{0} = [\{\Sigma_{jj}^{-1} - \Sigma_{jj}^{-1}\,{}_jC_s({}_jC'_s\,\Sigma_{jj}^{-1}\,{}_jC_s)^-\,{}_jC'_s\,\Sigma_{jj}^{-1}\}\,{}_jP_s - ({}_j\theta_s - 1)I]\,{}_j\underline{b}_s,$$
$$j = 1, 2, \ldots, m.$$

Equation (4.3.4) or (4.3.5) can be solved with the help of an iterative procedure like the one described in the last section. The principal change is that ${}_j\underline{h}^{*(n)}_s$ is replaced by ${}_jP^{*(n)}_s \; {}_j\underline{b}^{*(n)}_s$ in (4.2.16). ${}_jP^{*(n)}_s$ is computed like ${}_jP^*_s$ in (4.3.1) using (4.2.17). In place of (4.2.18), one has

$$({}_j\theta^{(n)}_s - 1) = \sum_{i=1}^{j-1} \{\mathrm{corr}({}_jZ^{(n)}_s, {}_iZ^{(n)}_s)\}^2 + \sum_{i=j+1}^{m} \{\mathrm{corr}({}_jZ^{(n)}_s, {}_iZ^{(n-1)}_s)\}^2 .$$

Again, the procedure must converge monotonically in ${}_jf^{(n)}_s$ -- i.e., (4.2.19) and (4.2.20) hold -- to a solution of the derivative equations, (4.3.4) in this case. Thus the solution is optimal provided the initial variables are appropriately chosen.

For $s = 1$, each iteration involves a calculation like the one used to find α_1 and ${}_1\underline{b}^*_1$ in Theorem 1.7.1 (i). For $s > 1$, the only difference is that linear restrictions on the admissible coefficient vectors are imposed. In both instances, the calculation involves the determination of the largest eigenvalue and associated eigenvector of a certain matrix.

Define ${}_j\hat{\theta}^{(n)}_s$ by the equation

$$({}_j\hat{\theta}^{(n)}_s - 1) = \sum_{i=1}^{j-1} \{\mathrm{corr}({}_jZ^{(n-1)}_s, {}_iZ^{(n)}_s)\}^2 + \sum_{i=j+1}^{m} \{\mathrm{corr}({}_jZ^{(n-1)}_s, {}_iZ^{(n-1)}_s)\}^2,$$

analogous to (4.2.21), and let ${}_of^{(n)}_s$ be as in (4.2.22). It is apparent that equations (4.2.23), (4.2.24), and (4.2.25) are valid here, although their interpretation is somewhat different in this context.

The most general conclusion is that ${}_j\underline{b}^{*(n-1)}_s$ and ${}_j\underline{b}^{*(n)}_s$ are, in the limit as $n \to \infty$, in the same eigenspace -- the space corresponding to the largest eigenvalue ${}_j\theta_s^{(n)}$ of ${}_jP^{*(n)}_s$.

Consider $\tilde{\underline{Z}}_{(s)}$, like in Section 4.2, as well as the associated ${}_j\tilde{\underline{b}}^*_s$, ${}_j\tilde{\theta}_s$, ${}_j\tilde{P}^*_s$, and $\tilde{\Phi}_{(s)}$. But here

$$\tilde{f}_s = tr\{\tilde{\Phi}^2_{(s)}\},$$

where $\tilde{f}_s$ is the limiting value in (4.2.20) for the SSQCOR iterative procedure. As before, the ${}_j\tilde{\underline{b}}^*_s$ and ${}_j\tilde{\theta}_s$ must form a solution of the current derivative equations, namely (4.3.4).

In practice, the criterion values corresponding to different solutions of (4.3.4) will be distinct, the ${}_j\theta_s^{(n)}$ will be of multiplicity one, and ${}_j\underline{b}^{*(n-1)}_s$ will equal ${}_j\underline{b}^{*(n)}_s$, apart from a possible scalar factor of (-1), in the limit as $n \to \infty$. With no loss in generality, the ${}_j\underline{b}^{*(n)}_s$ may be adjusted in sign so that

$$\lim_{n \to \infty} {}_jP^{*(n)}_s = {}_j\tilde{P}^*_s ;$$

equations (4.2.26), (4.2.27), and (4.2.28) hold for the SSQCOR procedure; and the ${}_j\tilde{\underline{b}}^*_s$ and ${}_j\tilde{\theta}_s$ form a solution of (4.3.4).

4.4 The Generalized Variance Procedure

Steel [21] developed a system of non-linear equations for finding first stage GENVAR canonical variates. The crux of his argument is that, for any orthogonal D_{B*}, $|\Phi_{(1)}|$ appears as a diagonal element of the m-th 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^*, $B_2^*, \ldots, B_m^*$. Although some solution to these equations will minimize $|\Phi_{(1)}|$, the equations appear difficult to solve except in special cases.

It is possible to obtain appropriate derivative equations without recourse to compound matrices by making repeated use of a simple determinantal expansion. These equations can be put in a form which is amenable to solution.

Let ${}_jM_s$ be the matrix obtained from $\Phi_{(s)}$ by deleting the j-th row and j-th column. Define

$${}_jQ_s = {}_jN_s \, {}_jM_s^{-1} \, {}_jN_s' \quad \text{and} \quad {}_jQ_s^* = {}_jN_s^* \, {}_jM_s^{-1} \, {}_jN_s^{*\prime} .$$

The criterion function at the s-th stage is

$$\begin{aligned} f_s &= |\Phi_{(s)}| \\ &= |{}_jM_s| \, {}_j\underline{b}_s^{*\prime}(I - {}_jQ_s^*) \, {}_j\underline{b}_s^*, \quad j = 1, 2, \ldots, m . \end{aligned}$$

The crucial point here is that ${}_jM_s$ and ${}_jQ_s^*$ do not involve ${}_j\underline{b}_s^*$.

Considering restrictions like (3.9.1), (3.9.2), and (3.9.3), which can be phrased in the form of (4.2.7), one has the general s-th stage Lagrangian equation

$$g_s = f_s - \sum_{j=1}^{m} (1 - {}_j\underline{b}_s^{*\prime} \, {}_j\underline{b}_s^*) |{}_jM_s| \, {}_j\theta_s + 2 \sum_{j=1}^{m} |{}_jM_s| \, {}_j\underline{b}_s^{*\prime} \, {}_jC_s^* \, {}_j\underline{\gamma}_s$$

where the ${}_j\theta_s$ and ${}_j\underline{\gamma}_s$ are Lagrange multipliers and ${}_jC_1^*$ is the zero matrix as before.

$$\frac{\partial g_s}{\partial\, {}_j\underline{b}^*_s} = -2|{}_jM_s|\,{}_jQ^*_s\,{}_j\underline{b}^*_s + 2|{}_jM_s|({}_j\theta_s - 1)\,{}_j\underline{b}^*_s + 2|{}_jM_s|\,{}_jC^*_s\,{}_j\underline{\gamma}_s$$

and equating this to zero gives

$$(4.4.1) \qquad \underline{0} = {}_jQ^*_s\,{}_j\underline{b}^*_s - ({}_j\theta_s - 1)\,{}_j\underline{b}^*_s - {}_jC^*_s\,{}_j\underline{\gamma}_s \,.$$

The general solution for ${}_j\underline{\gamma}_s$ is

$$(4.4.2) \quad {}_j\underline{\gamma}_s = ({}_jC^{*\prime}_s\,{}_jC^*_s)^-\,{}_jQ^*_s\,{}_j\underline{b}^*_s + \{I - ({}_jC^{*\prime}_s\,{}_jC^*_s)^-({}_jC^{*\prime}_s\,{}_jC^*_s)\}\underline{u},$$

with $\underline{u}$ arbitrary. Inserting (4.4.2) into (4.4.1) results in

$$(4.4.3) \quad \underline{0} = [\{I - {}_jC^*_s({}_jC^{*\prime}_s\,{}_jC^*_s)^-\,{}_jC^{*\prime}_s\}\,{}_jQ^*_s - ({}_j\theta_s - 1)I]\,{}_j\underline{b}^*_s\,, \qquad j = 1, 2, \ldots, m.$$

It follows from (4.4.3) that

$$\begin{cases} {}_j\theta_s = 1 + {}_j\underline{b}^{*\prime}_s\,{}_jQ^*_s\,{}_j\underline{b}^*_s \\ \text{subject to} \\ {}_j\underline{b}^*_s \in \mathcal{V}(I - {}_jC^*_s({}_jC^{*\prime}_s\,{}_jC^*_s)^-\,{}_jC^{*\prime}_s) \end{cases}$$

and

$$f_s = |{}_jM_s|(2 - {}_j\theta_s), \qquad j = 1, 2, \ldots, m.$$

In other words, $({}_j\theta_s - 1)$ is the squared multiple correlation between ${}_jZ_s$ and $({}_1Z_s, \ldots, {}_{j-1}Z_s, {}_{j+1}Z_s, \ldots, {}_mZ_s)$. The companion expression of (4.4.3), in terms of the original $\underline{X}$ variables, is

$$(4.4.4)\quad \underline{0} = [\{\Sigma_{jj}^{-1} - \Sigma_{jj}^{-1}\, {}_jC_s({}_jC_s'\, \Sigma_{jj}^{-1}\, {}_jC_s)^-\, {}_jC_s'\, \Sigma_{jj}^{-1}\}{}_jQ_s - ({}_j\theta_s - 1)I]\, {}_j\underline{b}_s,$$

$$j = 1, 2, \ldots, m.$$

Equation (4.4.3) or (4.4.4) can be solved using an iterative procedure which is basically the same as the ones suggested for (4.2.12) and (4.3.4). One now requires ${}_jQ_s^{*(n)}$ which is computed like ${}_jQ_s^*$ using (4.2.17). The quantity ${}_jQ_s^{*(n)}\ {}_j\underline{b}_s^{*(n)}$ is substituted in (4.2.16) for ${}_j\underline{h}_s^{*(n)}$. In this case, $({}_j\theta_s^{(n)} - 1)$ is the square of the first canonical correlation between $[I - {}_jC_s^*({}_jC_s^{*\prime}\ {}_jC_s^*)^-\ {}_jC_s^{*\prime}]\underline{Y}_j$ and $({}_1Z_s^{(n)}, \ldots, {}_{j-1}Z_s^{(n)}, {}_{j+1}Z_s^{(n-1)}, \ldots, {}_mZ_s^{(n-1)})$, and ${}_j\underline{b}_s^{*(n)}$ is the vector which defines the corresponding first canonical variate for the first of these sets (cf. Theorem 1.7.1 (iii)). For $s > 1$, this is an example of a canonical analysis with linear restrictions (on the set of variables $\underline{Y}_j$), as discussed in Section 2.5.

In contrast to (4.2.19), the iterative procedure is such that

$$(4.4.5)\qquad {}_1f_s^{(1)} \geq \ldots \geq {}_mf_s^{(1)} \geq {}_1f_s^{(2)} \geq \ldots \geq {}_mf_s^{(2)} \geq \ldots\ .$$

Furthermore,

$$(4.4.6)\qquad 0 < {}_jf_s^{(n)} \leq 1\ ,$$

and (4.2.20) holds for the GENVAR sequence ${}_jf_s^{(n)}$.

Now let

$$({}_j\hat{\theta}_s^{(n)} - 1) = {}_j\underline{b}_s^{*(n-1)\prime}\ {}_jQ_s^{*(n)}\ {}_j\underline{b}_s^{*(n-1)},$$

that is, the squared multiple correlation coefficient between ${}_jZ_s^{(n-1)}$ and $({}_1Z_s^{(n)}, \ldots, {}_{j-1}Z_s^{(n)}, {}_{j+1}Z_s^{(n-1)}, \ldots, {}_mZ_s^{(n-1)})$. Instead of (4.2.23), one now has

$$(4.4.7) \qquad 1 \le {}_j\hat{\theta}_s^{(n)} \le {}_j\theta_s^{(n)} < 2.$$

Consider the ratio

$$(4.4.8) \qquad \frac{{}_{j-1}f_s^{(n)}}{{}_jf_s^{(n)}} = \frac{|{}_jM_s^{(n)}|(2 - {}_j\hat{\theta}_s^{(n)})}{|{}_jM_s^{(n)}|(2 - {}_j\theta_s^{(n)})},$$

where ${}_of_s^{(n)}$ is defined as in (4.2.22) and ${}_jM_s^{(n)}$ is computed like ${}_jM_s$ using (4.2.17). The limit of the left hand side as $n \to \infty$ is one because of (4.4.5) and (4.2.20). This, together with (4.4.7), assures that (4.2.25) remains correct for the GENVAR sequence $({}_j\theta_s^{(n)} - {}_j\hat{\theta}_s^{(n)})$.

From this point, the development proceeds exactly like the last part of the previous section: one need only read ${}_jQ_s^{*(n)}$ for ${}_jP_s^{*(n)}$ and GENVAR for SSQCOR and redefine $\tilde{f}_s$ to be $|\tilde{\Phi}_{(s)}|$.

4.5 The Maximum Variance Procedure

The criterion function at the s-th stage of the MAXVAR procedure is

$$(4.5.1) \qquad f_s = {}_k\underline{e}_s' D_{B^*(s)} R D_{B^*(s)}' {}_k\underline{e}_s = {}_k\underline{e}_s' \Phi_{(s)} {}_k\underline{e}_s$$

which is to be maximized with respect to choice of $D_{B^*(s)}$ and k_s, $1 \le k_s < m$, subject to restrictions (3.9.10) or (4.2.7). Note that $D_{B^*(s)}' {}_k\underline{e}_s$ is a length one vector.

The optimal criterion value and the corresponding canonical variables at the first stage can be easily found with the aid of Lemma A1.

Let $\underline{v}$ be an arbitrary unit length vector. Then

$$\sup_{\underline{v}} \underline{v}'R\underline{v} = \underline{v}_1'R\underline{v}_1 = c_1(R) .$$

Suppose $D_{B*(1)}$ and ${}_{k_1}\underline{e}_1$ are required to satisfy

(4.5.2) $$D'_{B*(1)}\ {}_{k_1}\underline{e}_1 = \underline{v}_1 .$$

(4.5.2) implies that[1]

(4.5.3) $${}_j\underline{b}_1^* = \frac{\pm\ {}_j\underline{v}_1}{||{}_j\underline{v}_1||} , \quad j = 1, 2, \ldots, m .$$

Next observe that ${}_{k_1}\underline{e}_1'\ \Phi_{(1)}\ {}_{k_1}\underline{e}_1 = \underline{v}_1'R\underline{v}_1 = c_1(R)$. There cannot exist an $\underline{f}$, $||\underline{f}|| = 1$, such that $\underline{f}'\Phi_{(1)}\underline{f} = \underline{f}'D_{B*(1)}RD'_{B*(1)}\underline{f} = \underline{v}'R\underline{v} > c_1(R)$, since $c_1(R)$ is the largest eigenvalue of R. This implies that

$$\sup_{\underline{Z}_{(1)}} f_1 = c_1(R)$$

and occurs when $\underline{Z}_{(1)}$ is defined by (4.5.3). It also implies that $c_1(R)$ is the largest eigenvalue of $\Phi_{(1)}$ with corresponding eigenvector ${}_{k_1}\underline{e}_1$. Thus $k_1 = 1$.

Suppose that the restrictions imposed at the s-th stage are of the type (4.2.7) with

(4.5.4) $$D_{C_s^*} = \text{diag}({}_1C_s^*, {}_2C_s^*, \ldots, {}_mC_s^*) .$$

Let $\underline{v} = D'_{B*(s)}\ {}_{k_s}\underline{e}_s$. It is desired to find $\underline{v}$ with the appropriate properties which maximizes (4.5.1). Form the Lagrangian equation

---

[1] If $||{}_j\underline{v}_1|| = 0$, ${}_j\underline{b}_1^*$ is arbitrary. A similar comment applies to the other expressions of this type found in this and the next section.

$$g_s = f_s + (1 - \underline{v}'v)\theta_s - 2\underline{v}'D_{C^*_s}\underline{\gamma}_s ,$$

where θ_s and $\underline{\gamma}_s$ are Lagrange multipliers. The derivative with respect to $\underline{v}$ is

$$\frac{\partial g_s}{\partial \underline{v}} = 2R\underline{v} - 2\,\theta_s\underline{v} - 2\,D_{C^*_s}\underline{\gamma}_s$$

which, when equated to zero, gives

(4.5.5) $$\underline{0} = R\underline{v} - \theta_s\underline{v} - D_{C^*_s}\underline{\gamma}_s .$$

Multiplying (4.5.5) by $\underline{v}'$ results in

(4.5.6) $$f_s = \theta_s .$$

Solving for $\underline{\gamma}_s$ and substituting the general expression for it into (4.5.5) leads to

(4.5.7) $$\underline{0} = [\{I - D_{C^*_s}(D'_{C^*_s}D_{C^*_s})^{-} D'_{C^*_s}\} R - \theta_s I]\,\underline{v}.$$

Thus

(4.5.8) $$\sup_{\underline{Z}(s)} f_s = c_1(\{I - D_{C^*_s}(D'_{C^*_s}D_{C^*_s})^{-} D'_{C^*_s}\}R)$$

and $D_{B^*(s)}$ and $k_s\underline{e}_s$ are such that

(4.5.9) $$D'_{B^*(s)}\,k_s\underline{e}_s = \underline{\tilde{v}}_s ,$$

where

$$\underline{\tilde{v}}'_s = ({}_1\underline{\tilde{v}}'_s,\ {}_2\underline{\tilde{v}}'_s,\ \ldots,\ {}_m\underline{\tilde{v}}'_s)$$

is a unit length eigenvector associated with the largest eigenvalue in (4.5.7). (4.5.9) implies

$$(4.5.10) \qquad {}_j\underline{b}^*_s = \pm \frac{{}_j\tilde{\underline{v}}_s}{||{}_j\tilde{\underline{v}}_s||} , \quad j = 1, 2, \ldots, m .$$

It remains to verify that ${}_{k_s}\underline{e}_s$ is an eigenvector of the induced $\Phi_{(s)}$. To see this, first note that

$\Phi_{(s)} = D_{B^*(s)} R D'_{B^*(s)} = D_{B^*(s)}\{I - D_{C^*_s}(D'_{C^*_s} D_{C^*_s})^- D'_{C^*_s}\} R D'_{B^*(s)}$. Now arguing as in the case $s = 1$, it follows not only that ${}_{k_s}\underline{e}_s$ is an eigenvector but also that $k_s = 1$ and ${}_1\lambda_s = \sup_{\underline{Z}_{(s)}} f_s$, the value of which is recorded in (4.5.8).

Restriction (3.9.10) is also easy to handle. At the second stage, and using Lemma A1 once more,

$$\sup_{\substack{\underline{Z}_{(2)} \\ \underline{v}_1'R\underline{v} = 0}} f_2 = \sup_{\substack{\underline{Z}_{(2)} \\ \underline{v}_1'\underline{v} = 0}} \underline{v}'R\underline{v} = \underline{v}_2'R\underline{v}_2 = c_2(R).$$

Suppose $D_{B^*(2)}$ and ${}_{k_2}\underline{e}_2$ are required to satisfy

$$D'_{B^*(2)}\, {}_{k_2}\underline{e}_2 = \underline{v}_2$$

which implies that

$${}_j\underline{b}^*_2 = \pm \frac{{}_j\underline{v}_2}{||{}_j\underline{v}_2||} , \quad j = 1, 2, \ldots, m.$$

Then, if ${}_{k_2}\underline{e}_2$ can be shown to be an eigenvector of the associated $\Phi_{(2)}$, $\underline{Z}_{(2)}$ must be the required vector of canonical variables.

Let $R^{(2)} = (I - \underline{v}_1\underline{v}_1')R(I - \underline{v}_1\underline{v}_1')$ and form $\Phi^{(2)} = D_{B*(2)}R^{(2)}D'_{B*(2)} = \Phi_{(2)} - c_1(R)D_{B*(2)}\underline{v}_1\underline{v}_1'D'_{B*(2)}$. Now $k_2\underline{e}_2$ must be an eigenvector of $\Phi^{(2)}$ because $\Phi^{(2)}\ k_2\underline{e}_2 = D_{B*(2)}R^{(2)}\underline{v}_2 = D_{B*(2)}(c_2(R)\underline{v}_2) = c_2(R)\ k_2\underline{e}_2$. But $\Phi^{(2)}\ k_2\underline{e}_2 = \Phi(2)\ k_2\underline{e}_2 - c_1(R)D_{B*(2)}\underline{v}_1\underline{v}_1'\underline{v}_2 = \Phi(2)\ k_2\underline{e}_2$ which proves that $k_2\underline{e}_2$ is an eigenvector of $\Phi_{(2)}$ corresponding to the eigenvalue $c_2(R)$. It is not true, however, that k_2 must be equal to one in this situation; that is, $_{k_2}\lambda_2$ need not be the largest eigenvalue of $\Phi_{(2)}$. Examples are given in Sections 5.4 - 5.6 which demonstrate that, at least for some $s > 1$, one can have $k_s > 1$.

Continuing in this way, one has at the s-th stage

$$\sup_{\substack{\underline{Z}_{(s)} \\ \underline{v}_j'R\underline{v} = 0 \\ j = 1, 2, \ldots, s-1}} f_s = c_s(R), \tag{4.5.11}$$

with

$$D'_{B*(s)}\ k_s\underline{e}_s = \underline{v}_s$$

and

$$_j\underline{b}_s^* = \pm \frac{_j\underline{v}_s}{||_j\underline{v}_s||}, \quad j = 1, 2, \ldots, m. \tag{4.5.12}$$

One can easily show, as for $s = 2$, that $_{k_s}\underline{e}_s$ is an eigenvector of $\Phi_{(s)}$ corresponding to the eigenvalue $c_s(R)$.

As with the other procedures, one does not in practice need to make the transformation from $\underline{X}$ to $\underline{Y}$. The equation analogous to (4.5.10) which defines the s-th stage canonical variate coefficient vectors with respect to $\underline{X}$ is

$$(4.5.13) \qquad {}_j\underline{b}_s = \pm \frac{\Sigma_{jj}^{-1}\, {}_j\tilde{\underline{w}}_s}{||\Sigma_{jj}^{-\frac{1}{2}}\, {}_j\tilde{\underline{w}}_s||} , \quad j = 1, 2, \ldots, m ,$$

where

$$\tilde{\underline{w}}_s' = ({}_1\tilde{\underline{w}}_s', {}_2\tilde{\underline{w}}_s', \ldots, {}_m\tilde{\underline{w}}_s')$$

is an eigenvector corresponding to the largest eigenvalue of the matrix

$$(4.5.14) \qquad \{I - D_{C_s}(D_{C_s}' D_\Sigma^{-1} D_{C_s})^{-} D_{C_s}' D_\Sigma^{-1}\}\Sigma D_\Sigma^{-1}$$

with

$$D_{C_s} = D_\Sigma^{\frac{1}{2}} D_{C_s^*}.$$

The largest eigenvalue of (4.5.14) is the maximum value of f_s, as given in (4.5.8). The companion expression of (4.5.12), in terms of the original variables, is

$${}_j\underline{b}_s = \pm \frac{\Sigma_{jj}^{-1}\, {}_j\underline{w}_s}{||\Sigma_{jj}^{-\frac{1}{2}}\, {}_j\underline{w}_s||} , \quad j = 1, 2, \ldots, m ,$$

where

$$(4.5.15) \qquad \underline{w}_s' = ({}_1\underline{w}_s', {}_2\underline{w}_s', \ldots, {}_m\underline{w}_s')$$

is an eigenvector of ΣD_Σ^{-1} corresponding to the s-th largest eigenvalue of this matrix subject to the restriction

$$\underline{w}_j' D_\Sigma^{-1} \Sigma D_\Sigma^{-1} \underline{w}_s = 0, \quad j = 1, 2, \ldots, s-1.$$

Note that $c_s(\Sigma D_\Sigma^{-1}) = c_s(R)$, the maximum value of f_s as recorded in (4.5.11).

4.6 The Minimum Variance Procedure

The mathematical developments of the MAXVAR and MINVAR procedures are so much alike that many of the details for the latter procedure can be safely omitted. The s-th stage criterion function is the same as (4.5.1). The object now is to minimize (4.5.1) with respect to choice of $D_{B*(s)}$ and k_s, $1 < k_s \le m$, subject to restrictions (3.9.10) or (4.2.7).

First of all, using (1f.2.1) of Rao [20], one has

$$\inf_{\underline{v}} \underline{v}'R\underline{v} = \underline{v}_p'R\underline{v}_p = c_p(R) .$$

Suppose $D_{B*(1)}$ and $_{k_1}\underline{e}_1$ are required to satisfy

$$D'_{B*(1)}\ {}_{k_1}\underline{e}_1 = \underline{v}_p .$$

It follows then that

$$\text{(4.6.1)} \qquad {}_j\underline{b}_1^* = \frac{\pm\ {}_j\underline{v}_p}{||{}_j\underline{v}_p||} , \qquad j = 1, 2, \ldots, m.$$

Arguing in a manner similar to that in the previous section, one concludes that

$$\inf_{\underline{Z}_{(1)}} f_1 = c_p(R)$$

and occurs when $\underline{Z}_{(1)}$ is defined by (4.6.1), $c_p(R)$ is the smallest eigenvalue of the associated $\Phi_{(1)}$, and $k_1 = m$.

For restrictions of the type (4.2.7) on the higher stage canonical variables, the key equation is again (4.5.7). The relationship between the present θ_s and f_s is the same as in (4.5.6). The eigenvalues of the 'adjusted' R matrix in (4.5.7) are the same as those of $\{I - D_{C_s^*}(D'_{C_s^*}D_{C_s^*})^- D'_{C_s^*}\}^2 R$ because of the idempotency of the matrix in braces. These in turn are the same as those of

$$(4.6.2) \qquad \{I - D_{C_s^*}(D'_{C_s^*}D_{C_s^*})^- D'_{C_s^*}\}R\{I - D_{C_s^*}(D'_{C_s^*}D_{C_s^*})^- D'_{C_s^*}\}.$$

If the rank of $D_{C_s^*}$ is denoted by q, then q of the eigenvalues of (4.6.2) are zero, and associated eigenvectors are q independent columns of $D_{C_s^*}$. Since the matrix in (4.6.2) is symmetric, the eigenvectors associated with its non-zero eigenvalues must be orthogonal to the columns of $D_{C_s^*}$ -- i.e., each eigenvector is in $V(I - D_{C_s^*}(D'_{C_s^*}D_{C_s^*})^- D'_{C_s^*})$. One can infer from this fact that any eigenvector associated with a non-zero eigenvalue of (4.6.2) is also an eigenvector corresponding to this same eigenvalue for the 'adjusted' R matrix in (4.5.7). This, together with (1f.2.3) of Rao [20], shows that the smallest admissible θ_s is the (p-q)-th largest eigenvalue of the 'adjusted' R matrix. [In either the MAXVAR or MINVAR procedure, one can work with the symmetric matrix (4.6.2) in place of $\{I - D_{C_s^*}(D'_{C_s^*}D_{C_s^*})^- D'_{C_s^*}\}R$.]

If

$$\underline{\tilde{v}}'_{p-s+1} = ({}_1\underline{\tilde{v}}'_{p-s+1}, {}_2\underline{\tilde{v}}'_{p-s+1}, \ldots, {}_m\underline{\tilde{v}}'_{p-s+1})$$

is the eigenvector associated with the smallest non-zero eigenvalue, and if $D_{B*(s)}$ and $k_s\underline{e}_s$ satisfy

$$D'_{B*(s)}\, k_s\underline{e}_s = \underline{\tilde{v}}_{p-s+1} ,$$

which implies that

$$(4.6.3) \qquad {}_j\underline{b}^*_s = \pm \frac{{}_j\tilde{\underline{v}}_{p-s+1}}{||{}_j\tilde{\underline{v}}_{p-s+1}||}, \quad j = 1, 2, \ldots, m,$$

then it follows readily that (i) ${}_{k_s}\underline{e}_s$ is an eigenvector of the associated $\Phi_{(s)}$, (ii) $k_s = m$, and

$$\text{(iii)} \quad {}_m\lambda_s = \inf_{\underline{Z}(s)} f_s = c_{p-q}(\{I - D_{C^*_s}(D'_{C^*_s}D_{C^*_s})^- D'_{C^*_s}\}R).$$

Turning to restriction (3.9.10), the main results are

$$\inf_{\substack{\underline{Z}(s) \\ \underline{v}'_j R\underline{v} = 0 \\ j = 1, 2, \ldots, s-1}} f_s = c_{p-s+1}(R)$$

with the optimal $D'_{B^*(s)}$ and associated ${}_{k_s}\underline{e}_s$ satisfying

$$D'_{B^*(s)}\, {}_{k_s}\underline{e}_s = \underline{v}_{p-s+1}$$

which implies that

$$(4.6.4) \qquad {}_j\underline{b}^*_s = \pm \frac{{}_j\underline{v}_{p-s+1}}{||{}_j\underline{v}_{p-s+1}||}, \quad j = 1, 2, \ldots, m.$$

${}_{k_s}\underline{e}_s$ is indeed the eigenvector of $\Phi_{(s)}$ associated with the eigenvalue $c_{p-s+1}(R)$.

The counterpart of (4.6.3) in terms of the $\underline{X}$ variables is

$${}_j\underline{b}_s = \pm \frac{\Sigma^{-1}_{jj}\, {}_j\tilde{\underline{w}}_{p-s+1}}{||\Sigma^{-\frac{1}{2}}_{jj}\, {}_j\tilde{\underline{w}}_{p-s+1}||}, \quad j = 1, 2, \ldots, m,$$

where

$$\tilde{\underline{w}}'_{p-s+1} = ({}_1\tilde{\underline{w}}'_{p-s+1},\ {}_2\tilde{\underline{w}}'_{p-s+1},\ \ldots,\ {}_m\tilde{\underline{w}}'_{p-s+1})$$

is an eigenvector corresponding to the smallest non-zero eigenvalue of the matrix in (4.5.14) using the current D_{C_s}. The counterpart of (4.6.4) is

$$_j\underline{b}_s = \pm \frac{\Sigma_{jj}^{-1}\ {}_j\tilde{\underline{w}}_{p-s+1}}{||\Sigma_{jj}^{-\frac{1}{2}}\ {}_j\underline{w}_{p-s+1}||}, \quad j = 1, 2, \ldots, m,$$

where

$$\underline{w}'_{p-s+1} = ({}_1\underline{w}'_{p-s+1},\ {}_2\underline{w}'_{p-s+1},\ \ldots,\ {}_m\underline{w}'_{p-s+1})$$

is like $\underline{w}_s$ in (4.5.15) but subject to

$$\underline{w}'_{p-j+1}\ D_\Sigma^{-1}\Sigma D_\Sigma^{-1}\ \underline{w}_{p-s+1} = 0, \quad j = 1, 2, \ldots, s-1.$$

CHAPTER V

PRACTICAL CONSIDERATIONS AND EXAMPLES

5.1 Introduction

The final chapter is devoted to practical considerations and examples. Suggestions for the interpretation of the results of a canonical analysis are made in Section 5.2. The third section contains specific recommendations on starting points for the iterative procedures. Selected calculations and comments for three examples are presented in Sections 5.4 - 5.6.

The examples are based on data taken from Thurstone and Thurstone [22]. They study in their monograph a number of different variables, each of which measures one of seven "primary abilities." Three variables are associated with each ability. The seven abilities are: verbal comprehension (V), word fluency (W), number (N), space (S), rote memory (M), reasoning (R), and perception (P). The sample correlations, based on 437 observations, are displayed in Table 5.4.1.

The SUMCOR, SSQCOR, and GENVAR iterative procedures were programmed to terminate as soon as the condition

$$\sum_{j=1}^{m} \left| {}_j\theta_s^{(n)} - {}_j\theta_s^{(n-1)} \right| < 0.0001$$

holds. This condition has proved to be an effective test for convergence of the criterion values ${}_j f_s^{(n)}$. (See Sections 4.2 - 4.4 for the definitions of ${}_j\theta_s^{(n)}$ and ${}_j f_s^{(n)}$.)

5.2 Interpretation of the Results

The analysis should include more than an examination of the canonical variates and their correlations. Given the canonical vector $\underline{Z}_{(s)}$, one should also study the basic ingredients of the associated best fitting model.

Taking the MAXVAR $\underline{Z}_{(1)}$ as one example, attention should be directed to ${}_1\lambda_1$, ${}_1\underline{e}_1$, and F_1. Hopefully, some meaning can be attached to the "common factor" F_1. The elements of ${}_1\underline{e}_1$, together with ${}_1\lambda_1$, are indicators of the potency of the factor in the different sets. For the MINVAR $\underline{Z}_{(1)}$, as a second example, the most important quantities are ${}_m\lambda_1$, ${}_m\underline{e}_1$, and ${}_mF_1$. If $c_p(R)$ is near zero, the canonical variables form a "nearly" singular set; that is, the elements of $\underline{Z}_{(1)}$ are "nearly" linearly related. The relevant linear compound is, of course, ${}_mF_1$, which should be interpreted if possible. In the situation where only two elements of ${}_m\underline{e}_1$ are non-zero (it will often happen that this is <u>approximately</u> true), the interpretation is simplified in that ${}_mF_1$, apart from a constant multiplier, is the difference between the first pair of canonical variates for the sets corresponding to the non-zero elements of ${}_m\underline{e}_1$, and the accompanying first canonical correlation is just $(1 - c_p(R))$. Furthermore, this correlation must be ϕ_{max}, the largest possible first canonical correlation, for otherwise $c_p(R)$ would not be the minimum eigenvalue of R as claimed (cf. Theorem 1.6.1).

Sometimes the correlation ϕ_{max} is of interest in itself. One would like, especially when m is large, to obtain some information about it without separately studying all $\binom{m}{2}$ possible pairs of sets. The off-diagonal element of any $\Phi_{(1)}$ matrix which is largest in magnitude provides a simple lower bound for ϕ_{max}. A non-trivial upper bound can be found in terms of the extreme eigenvalues of R. It follows from Theorem 1.6.1 that

$$(1 + \phi_{max}) \leq c_1(R) \quad \text{and} \quad (1 - \phi_{max}) \geq c_p(R).$$

Combining these inequalities, one has

$$\phi_{max} \leq \min\{c_1(R) - 1, \ 1 - c_p(R)\}. \tag{5.2.1}$$

Since $c_1(R) \geq 1$ and $0 < c_p(R) \leq 1$, by virtue of R being a positive definite correlation matrix, the right hand side of (5.2.1) is always non-negative and less than one.

If the upper bound is small, that is, if $c_1(R)$ or $c_p(R)$ is close to one, there would be little reason to further pursue the search for relations between two (or more) of the sets. Otherwise, it is reasonable to consider the sets corresponding to the dominant elements (assuming there are some) of the MAXVAR ${}_1\underline{e}_1$, if $(c_1(R) - 1) < (1 - c_p(R))$, or the MINVAR ${}_m\underline{e}_1$, if $(c_1(R) - 1) > (1 - c_p(R))$, as the ones most likely to produce canonical variates with correlation equal to ϕ_{max}. The heuristic justification for this notion, in the case where $(c_1(R) - 1)$ is the minimum term in (5.2.1), is that ${}_1\underline{e}_1'\Phi_{(1)}\,{}_1\underline{e}_1 = c_1(R)$ is large (but less than two) by assumption which suggests that the off-diagonal elements of $\Phi_{(1)}$ corresponding to the dominant elements of ${}_1\underline{e}_1$ are also large. A similar argument can be made when $(c_1(R) - 1) > (1 - c_p(R))$.

5.3 Starting Points for Iterative Procedures

Iterative procedures were proposed in the last chapter for determining the coefficient vectors which define the SUMCOR, SSQCOR, and GENVAR canonical variates. These procedures can be used not only at the first stage but also at the higher stages, so long as the associated constraints have the effect of confining the coefficient vectors to some specified vector space.

There are two primary objectives to consider in the selection of starting points for the iterative procedures: first, limiting the number of iterations needed to achieve convergence and, second, assuring that convergence is to an optimal solution.

An "all purpose" starting point, which has been used successfully with each of the procedures, begins with either ${}_j\underline{b}_s \propto \underline{1}$ or ${}_j\underline{b}^*_s \propto \underline{1}$, $j = 1, 2, \ldots, m$, depending on whether one is working with $\underline{X}$ or $\underline{Y}$. This equal weight starting point is most appropriate when one does not want to bother to make the calculations needed to determine a more refined starting point. It can also be used to generate variables to compare with those obtained using other starting points.

A better starting point for the first stage of the SUMCOR procedure can usually be obtained by selecting the MAXVAR $\underline{Z}_{(1)}$ for the initial variables. Recall from Sections 3.4 and 3.5 that the MAXVAR and SUMCOR $\underline{Z}_{(1)}$ vectors will be the same when the MAXVAR ${}_1\underline{e}_1$ is proportional to $\underline{1}$. Thus one would expect that the MAXVAR $\underline{Z}_{(1)}$ will be a particularly good starting point whenever the associated ${}_1\underline{e}_1$ is nearly proportional to a unit vector. A similar type of starting point can be found for the higher stages: choose that $\underline{Z}_{(s)}$ which results when the MAXVAR procedure is applied subject to the constraint (4.2.7) but with $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(s-1)}$ being the previously selected SUMCOR variables. Again, if ${}_1\underline{e}_s$ is nearly proportional to $\underline{1}$, convergence should be quick.

It was argued in Section 3.6 that the MAXVAR and SSQCOR procedures tend to produce similar $\underline{Z}_{(1)}$'s whenever there exist first stage variables which can be well explained by a single factor. Accordingly, the MAXVAR $\underline{Z}_{(1)}$ can often serve as a useful starting point for the SSQCOR procedure. One can argue on similar grounds that a reasonable starting point for the s-th stage of the SSQCOR procedure would be $\underline{Z}_{(s)}$ constructed by the MAXVAR procedure subject to the constraint (4.2.7), in which the SSQCOR $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(s-1)}$ are utilized instead of the usual MAXVAR canonical variables.

Finally, the MINVAR $\underline{Z}_{(1)}$ is often a satisfactory starting point for the GENVAR first stage, especially when the number of sets is large. (For a small number of sets, the MAXVAR $\underline{Z}_{(1)}$ seems frequently to be a better choice.) And, at the higher stages, one can begin with $\underline{Z}_{(s)}$ constructed by the MINVAR procedure subject to (4.2.7), in which the GENVAR $\underline{Z}_{(1)}, \underline{Z}_{(2)}, \ldots, \underline{Z}_{(s-1)}$ are inserted in place of the usual MINVAR canonical variables.

For all of the procedures requiring iterative techniques, it would be advisable to try more than one starting point. Experience has shown that convergence is usually obtained rapidly enough to permit this kind of check without excessive expense.

5.4 Example Number One

The first example is the one which Horst dealt with in [8], [9], and [10]. In this example, $m = 3$, $p_1 = 3$, $p_2 = 3$, $p_3 = 3$, and $p = 9$. The first set contains variables 5, 6, and 7 (cf. Table 5.4.1); the second contains variables 12, 13, and 14; and the third contains variables 19, 20, and 21. Horst made a preliminary transformation from the original variables with (sample) covariance matrix Σ_1 (the subscript indicates the example number) to new variables $\underline{Y}$ with (sample) correlation matrix R_1. The analysis given here starts from this latter matrix as found in [8] or [9] and displayed below:

$$R_1 = \begin{pmatrix} 1.000 & 0.000 & 0.000 & 0.636 & 0.126 & 0.059 & 0.626 & 0.195 & 0.059 \\ & 1.000 & 0.000 & -0.021 & 0.633 & 0.049 & 0.035 & 0.459 & 0.129 \\ & & 1.000 & 0.016 & 0.157 & 0.521 & 0.048 & 0.238 & 0.426 \\ & & & 1.000 & 0.000 & 0.000 & 0.709 & 0.050 & -0.002 \\ & & & & 1.000 & 0.000 & 0.039 & 0.532 & 0.190 \\ & & & & & 1.000 & 0.067 & 0.258 & 0.299 \\ & & & & & & 1.000 & 0.000 & 0.000 \\ & & & & & & & 1.000 & 0.000 \\ & & & & & & & & 1.000 \end{pmatrix}$$

TABLE 5.4.1

Correlation Coefficients for the Twenty-one Variables

	2(M)	3(V)	4(W)	5(S)	6(N)	7(R)	8(P)	9(M)	10(V)	11(W)	12(S)	13(N)	14(R)	15(P)	16(M)	17(V)	18(W)	19(S)	20(N)	21(R)
1(P)	.129	.261	.297	.181	.434	.292	.461	.156	.247	.239	.143	.497	.259	.430	.129	.204	.231	.221	.369	.348
2(M)		.295	.286	.103	.196	.320	.209	.280	.364	.299	.046	.268	.381	.263	.478	.282	.311	.061	.261	.285
3(V)			.419	.108	.298	.492	.297	.151	.829	.356	.033	.309	.555	.343	.234	.768	.407	.108	.351	.425
4(W)				.176	.293	.391	.256	.243	.472	.654	.092	.348	.355	.447	.240	.428	.557	.127	.319	.398
5(S)					.249	.271	.416	.227	.115	.192	.636	.183	.185	.279	.100	.272	.100	.626	.369	.279
6(N)						.399	.339	.121	.323	.261	.138	.654	.262	.317	.216	.296	.233	.190	.527	.356
7(R)							.378	.290	.468	.367	.180	.407	.613	.418	.249	.446	.305	.225	.471	.610
8(P)								.338	.264	.221	.402	.334	.398	.505	.175	.400	.183	.424	.433	.401
9(M)									.242	.184	.183	.040	.254	.224	.292	.234	.122	.252	.155	.217
10(V)										.415	.061	.347	.525	.349	.260	.775	.482	.125	.369	.381
11(W)											.165	.308	.323	.372	.217	.354	.514	.144	.334	.381
12(S)												.091	.147	.291	.078	.205	.009	.709	.254	.191
13(N)													.296	.354	.203	.271	.254	.103	.541	.394
14(R)														.385	.254	.523	.319	.179	.437	.496
15(P)															.209	.332	.350	.298	.347	.460
16(M)																.251	.236	.166	.140	.192
17(V)																	.433	.238	.385	.396
18(W)																		.066	.293	.303
19(S)																			.291	.245
20(N)																				.429

P = perception
M = rote memory
V = verbal comprehension
W = word fluency

S = space
N = number
R = reasoning

The SSQCOR calculations for R_1 using restriction (3.9.1) give

$$\Phi_{(1)} = \begin{bmatrix} 1.000 & 0.735 & 0.756 \\ & 1.000 & 0.743 \\ & & 1.000 \end{bmatrix},$$

$$\Phi_{(2)} = \begin{bmatrix} 1.000 & 0.603 & 0.504 \\ & 1.000 & 0.635 \\ & & 1.000 \end{bmatrix},$$

$$\Phi_{(3)} = \begin{bmatrix} 1.000 & 0.464 & 0.268 \\ & 1.000 & 0.165 \\ & & 1.000 \end{bmatrix},$$

$$E_{(1)} = \begin{bmatrix} 0.578 & -0.533 & -0.619 \\ 0.574 & 0.804 & -0.155 \\ 0.580 & -0.265 & 0.770 \end{bmatrix},$$

$$E_{(2)} = \begin{bmatrix} 0.559 & -0.753 & -0.348 \\ 0.602 & 0.080 & 0.794 \\ 0.571 & 0.653 & -0.498 \end{bmatrix},$$

$$E_{(3)} = \begin{bmatrix} 0.653 & -0.176 & -0.736 \\ 0.611 & -0.451 & 0.650 \\ 0.447 & 0.875 & 0.187 \end{bmatrix},$$

$$\underline{\lambda}'_{(1)} = (\ 2.490 \quad 0.267 \quad 0.243\ ),$$

$$\underline{\lambda}'_{(2)} = (\ 2.163 \quad 0.498 \quad 0.338\ ),$$

and

$$\underline{\lambda}'_{(3)} = (\ 1.617 \quad 0.861 \quad 0.522\ ).$$

The GENVAR, MAXVAR, and SUMCOR results using restriction (3.9.1) are virtually the same in each case as those for the SSQCOR procedure.

The SUMCOR calculations also have been carried out using restriction (3.9.2), first with $I_k = \{1\}$ and then again with $I_k = \{1, 2\}$. The second and third stage correlation matrices in the first instance are

$$\Phi_{(2)} = \begin{bmatrix} 1.000 & 0.601 & 0.505 \\ & 1.000 & 0.638 \\ & & 1.000 \end{bmatrix}$$

and

$$\Phi_{(3)} = \begin{bmatrix} 1.000 & 0.315 & 0.153 \\ & 1.000 & 0.526 \\ & & 1.000 \end{bmatrix},$$

and in the second instance are

$$\Phi_{(2)} = \begin{bmatrix} 1.000 & 0.603 & 0.505 \\ & 1.000 & 0.635 \\ & & 1.000 \end{bmatrix}$$

and

$$\Phi_{(3)} = \begin{bmatrix} 1.000 & 0.465 & 0.269 \\ & 1.000 & 0.166 \\ & & 1.000 \end{bmatrix}.$$

As one would expect, higher criterion values are attained as the number of sets upon which the restrictions are imposed is reduced. The results are summarized in Table 5.4.2.

TABLE 5.4.2

Values of the SUMCOR Criterion Function for R_1 Using Different Restrictions from the Class (3.9.2)

I_k:	{1}	{1, 2}	{1, 2, 3}
Stage 2	6.489	6.487	6.484
Stage 3	4.988	4.800	4.794

The eigenvalues which are relevant to the MAXVAR/MINVAR procedures using restriction (3.9.10) along with the values of k_s are displayed in Table 5.4.3.

TABLE 5.4.3

Values of $c_j(R_1)$ and Related MAXVAR/MINVAR Quantities

		MAXVAR			
j	$c_j(R_1)$	k_j	$\underline{\lambda}'_{(j)}$		
1	2.490	1	2.490	0.267	0.243
2	2.164	1	2.164	0.497	0.338
3	1.620	1	1.620	0.860	0.520
		MINVAR			
		k_{10-j}	$\underline{\lambda}'_{(10-j)}$		
4	0.867	2	1.517	0.867	0.616
5	0.541	3	1.455	1.004	0.541
6	0.487	3	1.757	0.756	0.487
7	0.337	3	1.954	0.709	0.337
8	0.259	3	2.365	0.376	0.259
9	0.235	3	2.082	0.683	0.235

There are evidently three MAXVAR and six MINVAR stages. Each MAXVAR $\underline{Z}_{(s)}$ has associated with it a representation of the type (3.9.5) involving a single factor while each MINVAR $\underline{Z}_{(s)}$, except $\underline{Z}_{(6)}$, has associated with it a representation of the type (3.9.8) using two factors.

The close resemblance of the MAXVAR $\underline{\lambda}_{(j)}$ in Table 5.4.3 to those previously listed for the SSQCOR procedure suggests that the corresponding $\Phi_{(j)}$ and $E_{(j)}$ will be virtually the same in both cases, as they indeed turn out to be. (The same resemblance holds if (3.9.10) is replaced by (3.9.1).)

The $\Phi_{(j)}$ and $E_{(j)}$ matrices for the first two MINVAR stages (using restriction (3.9.10)) are as follows:

$$\Phi_{(1)} = \begin{bmatrix} 1.000 & 0.345 & -0.736 \\ & 1.000 & -0.517 \\ & & 1.000 \end{bmatrix},$$

$$\Phi_{(2)} = \begin{bmatrix} 1.000 & -0.725 & 0.626 \\ & 1.000 & -0.695 \\ & & 1.000 \end{bmatrix},$$

$$E_{(1)} = \begin{bmatrix} 0.591 & -0.516 & 0.620 \\ 0.493 & 0.839 & 0.228 \\ -0.638 & 0.171 & 0.751 \end{bmatrix},$$

and

$$E_{(2)} = \begin{bmatrix} 0.574 & -0.632 & 0.520 \\ -0.593 & 0.118 & 0.797 \\ 0.565 & 0.766 & 0.307 \end{bmatrix}.$$

Taking the element 0.756 from the SSQCOR $\Phi_{(1)}$ matrix and using (5.2.1), one has

$$0.756 \leq \phi_{max} \leq 0.765 .$$

The dominant elements of the MINVAR ${}_3\underline{e}_1$ vector are the first and the third which suggests that the corresponding sets are the ones which yield ϕ_{max}. In fact, $\rho_1 = 0.742$ for the first and second sets, $\rho_1 = 0.757$ for the first and the third sets, and $\rho_1 = 0.750$ for the second and third sets.

As a rough approximation, the canonical variates ${}_jZ_1$, ${}_jZ_2$, and ${}_jZ_3$, other than the MINVAR ones, are proportional to $\underline{1}'\underline{X}_j$, $(2, -1, -1)\underline{X}_j$, and $(0, 1, -1)\underline{X}_j$, $j = 1, 2, 3$. $\underline{Z}_{(1)}$ is well explained by a single factor accounting for 83.0 per cent of its variability and contributing with equal weight to each of the ${}_jZ_1$. The factor is approximately proportional to $\underline{1}'\underline{X}$. For $\underline{Z}_{(2)}$, the first factor accounts for 72.1 per cent and the second factor 16.6 per cent of the variability. The second factor is present mainly in ${}_1Z_2$ and ${}_3Z_2$. Two factors account for 82.6 per cent of the variability in $\underline{Z}_{(3)}$, about as much as one factor does for $\underline{Z}_{(1)}$.

Table 5.4.4[1] shows the first stage SSQCOR criterion values at various steps of the iteration procedure. The following starting vectors ${}_j\underline{b}_1^{*(o)}$, $j = 1, 2, 3$, were used:

(i) (0.574 0.574 0.574);

(ii) (1.000 0.000 0.000);

(iii) (0.000 0.000 1.000);

---

[1] The figures in Table 5.4.4 are subject to error in the sixth decimal place.

(iv) the SUMCOR ${}_{j}\underline{b}_2^*$ under restriction (3.9.1);

(v) the SUMCOR ${}_{j}\underline{b}_3^*$ under restriction (3.9.1);

(vi) the MAXVAR ${}_{j}\underline{b}_1^*$.

Starting point (vi) gave the quickest convergence which is not surprising in view of the MAXVAR ${}_{1}\underline{e}_1$ being nearly proportional to $\underline{1}$.

TABLE 5.4.4

Values of SSQCOR ${}_{3}f_1^{(n)}$ on R_1 Using Different Starting Points

	Starting Points					
n	(i)	(ii)	(iii)	(iv)	(v)	(vi)
1	6.303480	5.799538	5.118418	5.046118	3.635250	6.329807
2	6.324824	6.051978	6.044878	5.047182	3.686164	6.329809
3	6.328192	6.209252	6.235992	5.049262	4.686334	
4	6.329200	6.281344	6.293624	5.054866	6.182884	
5	6.329578	6.311030	6.315908	5.069552	6.323596	
6	6.329720	6.322640	6.324514	5.107198	6.329160	
7	6.329778	6.327086	6.327802	5.197832	6.329620	
8	6.329796	6.328780	6.329044	5.385994	6.329738	
9	6.329804	6.329422	6.329522	5.677556	6.329786	
10	6.329810	6.329666	6.329700	5.972386	6.329798	
11	6.329810	6.329752	6.329770	6.167168	6.329802	
12		6.329792	6.329792	6.263168	6.329808	
13		6.329802	6.329804	6.303786	6.329810	
14		6.329808	6.329808	6.319844		
15		6.329808	6.329808	6.326020		
16		6.329806		6.328374		
17				6.329264		
18				6.329602		
19				6.329734		
20				6.329780		
21				6.329800		
22				6.329808		
23				6.329804		
24				6.329806		

5.5 Example Number Two

Σ_2 is the correlation matrix of the twenty-one variables in Table 5.4.1. The first set consists of the first seven variables, the second set consists of the middle seven variables, and the third set consists of the last seven variables.

The main features of the data can be extracted by the MAXVAR and MINVAR procedures. The eigenvalues of R_2 are given in Table 5.5.1 along with the number of factors for the models (3.9.5) and (3.9.8) using restriction (3.9.10).

The other key quantities for the first two MAXVAR stages are

$$\Phi_{(1)} = \begin{bmatrix} 1.000 & 0.889 & 0.869 \\ & 1.000 & 0.871 \\ & & 1.000 \end{bmatrix},$$

$$\Phi_{(2)} = \begin{bmatrix} 1.000 & 0.710 & 0.636 \\ & 1.000 & 0.689 \\ & & 1.000 \end{bmatrix},$$

$$E_{(1)} = \begin{bmatrix} 0.578 & -0.449 & -0.681 \\ 0.579 & -0.362 & 0.730 \\ 0.575 & 0.817 & -0.050 \end{bmatrix},$$

$$E_{(2)} = \begin{bmatrix} 0.575 & -0.640 & -0.510 \\ 0.589 & -0.109 & 0.801 \\ 0.568 & 0.760 & -0.314 \end{bmatrix},$$

$$\underline{\lambda}'_{(1)} = (\ 2.753 \quad 0.136 \quad 0.111\ ),$$

$$\underline{\lambda}'_{(2)} = (\ 2.357 \quad 0.366 \quad 0.277\ ),$$

$$_1\underline{b}'_1 = (\; 0.114 \quad 0.149 \quad 0.532 \quad 0.216 \quad 0.190 \quad 0.124 \quad 0.190 \;),$$

$$_2\underline{b}'_1 = (\; 0.208 \quad 0.031 \quad 0.538 \quad 0.179 \quad 0.097 \quad 0.184 \quad 0.224 \;),$$

$$_3\underline{b}'_1 = (\; 0.193 \quad 0.109 \quad 0.534 \quad 0.134 \quad 0.028 \quad 0.223 \quad 0.206 \;),$$

$$_1\underline{b}'_2 = (\; -0.284 \quad 0.034 \quad 0.618 \quad 0.135 \quad -0.759 \quad -0.096 \quad -0.166 \;),$$

$$_2\underline{b}'_2 = (\; -0.191 \quad -0.149 \quad 0.605 \quad 0.041 \quad -0.723 \quad -0.291 \quad 0.014 \;),$$

and

$$_3\underline{b}'_2 = (\; -0.162 \quad 0.035 \quad 0.530 \quad 0.205 \quad -0.819 \quad -0.291 \quad -0.051 \;).$$

It appears that the best fitting factor for $\underline{Z}_{(1)}$ is largely associated with the verbal primary ability while the best fitting factor for $\underline{Z}_{(2)}$ is, for the most part, a weighted difference of verbal and spatial abilities.

For the MINVAR first stage

$$\Phi_{(1)} = \begin{bmatrix} 1.000 & -0.894 & -0.022 \\ & 1.000 & -0.073 \\ & & 1.000 \end{bmatrix},$$

$$E_{(1)} = \begin{bmatrix} 0.705 & -0.081 & 0.704 \\ -0.708 & -0.024 & 0.706 \\ 0.040 & 0.996 & 0.074 \end{bmatrix},$$

$$\underline{\lambda}'_{(1)} = (\; 1.896 \quad 1.003 \quad 0.101 \;),$$

$$_1\underline{b}'_1 = (\; -0.120 \quad -0.188 \quad -0.547 \quad -0.219 \quad -0.024 \quad -0.137 \quad -0.203 \;),$$

$$_2\underline{b}'_1 = (\; 0.140 \quad 0.018 \quad 0.607 \quad 0.171 \quad -0.025 \quad 0.208 \quad 0.221 \;),$$

and

$$_3\underline{b}'_1 = (\; 0.074 \quad 0.553 \quad -0.645 \quad -0.084 \quad -0.014 \quad -0.358 \quad 0.820 \;).$$

TABLE 5.5.1

Values of $c_j(R_2)$ and $c_j(R_3)$ Plus Related k_j and k_{22-j}

j	$c_j(R_2)$	k_j/k_{22-j}	$c_j(R_3)$	k_j/k_{22-j}
1	2.752	1	4.406	1
2	2.357	1	2.650	1
3	2.148	1	2.107	1
4	1.808	1	1.815	1
5	1.606	1	1.519	2
6	1.482	1	1.321	2
7	1.356	1	0.892	4
8	0.961	3	0.840	5
9	0.888	2	0.742	4
10	0.838	3	0.645	5
11	0.800	3	0.625	5
12	0.666	3	0.573	5
13	0.641	3	0.506	6
14	0.586	3	0.480	5
15	0.478	3	0.428	6
16	0.434	3	0.390	6
17	0.404	3	0.307	6
18	0.310	3	0.239	6
19	0.259	3	0.228	6
20	0.123	3	0.161	6
21	0.101	3	0.124	6

Using the MINVAR $\Phi_{(1)}$ and (5.2.1), one obtains

(5.5.1) $$0.894 \le \phi_{max} \le 0.899 .$$

The elements of ${}_3\underline{e}_1$ suggest that ϕ_{max} occurs between the first and second sets.

Note the close resemblance between the MAXVAR ${}_1\underline{b}_1$ and ${}_2\underline{b}_1$ and the corresponding MINVAR ${}_1\underline{b}_1$ (apart from a factor of (-1)) and ${}_2\underline{b}_1$. Without separately analyzing the first and second sets, one can surmise, in light of (5.5.1) and the MINVAR $\Phi_{(1)}$, that the MINVAR residual factor ${}_3F_1$ is, in essence, the difference between the first canonical variates for these sets.

5.6 Example Number Three

Σ_3 is the same matrix as Σ_2, but the breakdown of variables into sets is different (and hence R_2 and R_3 are different). Now there are six sets: the first, third, and fifth sets consist of the first four variables in the three sets used in Example Two; the second, fourth, and sixth sets consist of the last three variables from the sets in Example Two.

The eigenvalues of R_3 and the number of factors for models (3.9.5) and (3.9.8), using restriction (3.9.10), may be found in Table 5.5.1. The detailed calculations are given only for the first stages of the SSQCOR and GENVAR procedures. The SSQCOR results are given first:

$$\Phi_{(1)} = \begin{bmatrix} 1.000 & 0.598 & 0.822 & 0.662 & 0.791 & 0.591 \\ & 1.000 & 0.619 & 0.730 & 0.590 & 0.739 \\ & & 1.000 & 0.677 & 0.823 & 0.628 \\ & & & 1.000 & 0.636 & 0.712 \\ & & & & 1.000 & 0.592 \\ & & & & & 1.000 \end{bmatrix},$$

$$E_{(1)} = \begin{bmatrix} 0.415 & -0.398 & 0.089 & -0.026 & -0.715 & 0.387 \\ 0.394 & 0.482 & 0.018 & -0.781 & -0.027 & -0.033 \\ 0.425 & -0.361 & -0.068 & 0.025 & -0.008 & -0.827 \\ 0.409 & 0.292 & 0.749 & 0.396 & 0.171 & 0.032 \\ 0.412 & -0.421 & -0.181 & -0.095 & 0.671 & 0.401 \\ 0.393 & 0.465 & -0.628 & 0.472 & -0.091 & 0.065 \end{bmatrix},$$

$$\underline{\lambda}'_{(1)} = (\ 4.406 \quad 0.683 \quad 0.284 \quad 0.255 \quad 0.206 \quad 0.167\ ),$$

$${}_1\underline{b}'_1 = (\ 0.322 \quad 0.229 \quad 0.562 \quad 0.307\ ),$$

$${}_3\underline{b}'_1 = (\ 0.463 \quad 0.024 \quad 0.592 \quad 0.279\ ),$$

$${}_5\underline{b}'_1 = (\ 0.449 \quad 0.133 \quad 0.608 \quad 0.154\ ),$$

$${}_2\underline{b}'_1 = (\ 0.239 \quad 0.369 \quad 0.686\ ),$$

$${}_4\underline{b}'_1 = (\ 0.254 \quad 0.498 \quad 0.651\ ),$$

and

$${}_6\underline{b}'_1 = (\ 0.181 \quad 0.520 \quad 0.578\ ).$$

The GENVAR results are

$$\Phi_{(1)} = \begin{bmatrix} 1.000 & 0.548 & 0.858 & 0.604 & 0.814 & 0.538 \\ & 1.000 & 0.573 & 0.736 & 0.565 & 0.751 \\ & & 1.000 & 0.636 & 0.840 & 0.575 \\ & & & 1.000 & 0.607 & 0.727 \\ & & & & 1.000 & 0.567 \\ & & & & & 1.000 \end{bmatrix},$$

$$E_{(1)} = \begin{pmatrix} 0.415 & -0.423 & -0.020 & 0.064 & -0.590 & -0.545 \\ 0.392 & 0.460 & 0.192 & 0.772 & -0.008 & 0.034 \\ 0.427 & -0.380 & -0.004 & -0.026 & -0.171 & 0.802 \\ 0.406 & 0.332 & -0.821 & -0.197 & 0.099 & -0.049 \\ 0.418 & -0.378 & 0.164 & -0.009 & 0.775 & -0.235 \\ 0.391 & 0.461 & 0.512 & -0.600 & -0.110 & -0.030 \end{pmatrix},$$

$$\underline{\lambda}'_{(1)} = (\ 4.317 \quad 0.842 \quad 0.271 \quad 0.248 \quad 0.186 \quad 0.136\ ),$$

$${}_1\underline{b}'_1 = (\ 0.149 \quad 0.166 \quad 0.706 \quad 0.299\ ),$$

$${}_3\underline{b}'_1 = (\ 0.298 \quad -0.006 \quad 0.754 \quad 0.230\ ),$$

$${}_5\underline{b}'_1 = (\ 0.293 \quad 0.094 \quad 0.708 \quad 0.207\ ),$$

$${}_2\underline{b}'_1 = (\ 0.364 \quad 0.370 \quad 0.604\ ),$$

$${}_4\underline{b}'_1 = (\ 0.355 \quad 0.515 \quad 0.581\ ),$$

and

$${}_6\underline{b}'_1 = (\ 0.288 \quad 0.535 \quad 0.494\ ).$$

The largest off-diagonal element of the MINVAR $\Phi_{(1)}$ matrix, in absolute value, turns out to be 0.864 and appears in the first row and third column. Thus, with the aid of (5.2.1), it follows that

$$0.864 \leq \phi_{max} \leq 0.876 .$$

The dominant elements of the MINVAR ${}_6\underline{e}_1$ are the first and the third which suggests that it is the corresponding sets which achieve ϕ_{max}.

The main difference between the two $\Phi_{(1)}$ matrices is that the off-diagonal elements of the SSQCOR matrix which are less (greater) than 0.7 are greater (less) than the corresponding elements in the GENVAR matrix. Looking at the two $\underline{\lambda}_{(1)}$ vectors, one sees that ${}_1\lambda_1$, ${}_3\lambda_1$, ${}_4\lambda_1$, ${}_5\lambda_1$, and ${}_6\lambda_1$ are less for the GENVAR procedure than for the SSQCOR procedure while only ${}_2\lambda_1$ is greater. Passing from the SSQCOR to the GENVAR ${}_j\underline{b}_1$, it is apparent that the most important differences are a uniform increase in the verbal comprehension and space coefficients and a uniform decrease in the perception and reasoning coefficients.

Focusing on the SSQCOR results, the first two factors account for 84.8 per cent of the variability in $\underline{Z}_{(1)}$. The elements of ${}_1\underline{e}_1$ indicate that the first factor contributes with approximately the same weight to each ${}_jZ_1$. The second factor, however, has positive weight in the representation of ${}_2Z_1$, ${}_4Z_1$, and ${}_6Z_1$ and negative weight in the representation of ${}_1Z_1$, ${}_3Z_1$, and ${}_5Z_1$. Each of the remaining four factors appears to be common to at most three of the ${}_jZ_1$.

BIBLIOGRAPHY

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

[2] Bartlett, M.S. (1951). "The goodness of fit of a single hypothetical discriminant function in the case of several groups." Annals of Eugenics, 16, 199-214.

[3] Carroll, J. Douglas (1968). "A generalization of canonical correlation analysis to three or more sets of variables." To appear in Proceedings of American Psychological Association.

[4] Cliff, Norman (1966). "Orthogonal rotation to congruence." Psychometrika, 31, 33-42.

[5] Fisher, R.A. (1940). "The precision of discriminant functions." Annals of Eugenics, 10, 422-429.

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

[7] Gower, J.C. (1967). "Multivariate analysis and multidimensional geometry." The Statistician, 17, 13-28.

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

[9] Horst, Paul (1961). "Generalized canonical correlations and their applications to experimental data." Journal of Clinical Psychology (monograph supplement), 14, 331-347.

[10] Horst, Paul (1965). Factor Analysis of Data Matrices. Holt, Rinehart, and Winston, New York.

[11] Hotelling, Harold (1935). "The most predictable criterion." Journal of Educational Psychology, 26, 139-142.

[12] Hotelling, Harold (1936). "Relations between two sets of variates." Biometrika, 28, 321-377.

[13] Hsu, P.L. (1946). "On a factorization of pseudo-orthogonal matrices." Quarterly Journal of Mathematics, 17, 162-165.

[14] Kendall, Maurice G. and Stuart, Alan (1961). The Advanced Theory of Statistics, Vol. 2. Hafner, New York.

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

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

[17] Mostow, George D., Sampson, Joseph H., and Meyer, Jean-Pierre (1963). Fundamental Structures of Algebra. McGraw-Hill, New York.

[18] Okamoto, Masashi and Kanazawa, Mitsuyo (1968). "Minimization of eigenvalues of a matrix and optimality of principal components." Annals of Mathematical Statistics, 39, 859-863.

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

[20] Rao, C. Radhakrishna (1965). Linear Statistical Inference and its Applications. Wiley, New York.

[21] Steel, Robert G.D. (1951). "Minimum generalized variance for a set of linear functions." Annals of Mathematical Statistics, 22, 456-460.

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

[23] Vinograde, Bernard (1950). "Canonical positive definite matrices under internal linear transformations." Proceedings of the American Mathematical Society, 1, 159-161.

[24] Williams, E.J. (1952). "Use of scores for the analysis of association in contingency tables." Biometrika, 39, 274-289.

[25] Wrigley, Charles and Neuhaus, Jack O. (1955). "The matching of two sets of factors." American Psychologist, 10, 418-419.

APPENDIX

LEMMA A1. Let A and B be $(p \times p)$ matrices with A symmetric and B positive definite. Let $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_p$ be the eigenvalues of $B^{-1}A$ and $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_p$ a corresponding set of eigenvectors such that $\underline{f}_i'B\underline{f}_j = \delta_{ij}$ (Kronecker delta). Then

$$\sup_{\underline{y}} \frac{\underline{y}'A\underline{y}}{\underline{y}'B\underline{y}} = \frac{\underline{f}_1'A\underline{f}_1}{\underline{f}_1'B\underline{f}_1} = \lambda_1$$

and

$$\sup_{\substack{\underline{y} \\ \underline{f}_i'B\underline{y} = 0 \\ i = 1,2,\ldots,k}} \frac{\underline{y}'A\underline{y}}{\underline{y}'B\underline{y}} = \frac{\underline{f}_{k+1}'A\underline{f}_{k+1}}{\underline{f}_{k+1}'B\underline{f}_{k+1}} = \lambda_{k+1}, \quad k = 1, 2, \ldots, p-1.$$

Proof. Let $\underline{x} = B^{\frac{1}{2}}\underline{y}$ and apply (1f.2.1) - (1f.2.3) of Rao [20].

LEMMA A2. Let M be a $(p \times q)$ matrix with $p \leq q$. Then there exist orthogonal matrices Γ and Δ such that

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

where

$$D = \text{diag}(\theta_1, \theta_2, \ldots, \theta_p)$$

and $\theta_1^2, \theta_2^2, \ldots, \theta_p^2$ are the eigenvalues of MM'.

Proof. See Hsu [13] or Vinograde [23].

LEMMA A3. Let $\underline{X}$ be a $(p \times 1)$ random vector with $E(\underline{X}) = \underline{0}$ and $\text{var}(\underline{X}) = \Sigma$. Let M be any $(p \times q)$ matrix and $\underline{Y}$ any $(q \times 1)$ random vector. Let f be any real-valued function defined on the set of all non-negative definite matrices of order p which is identical to some function $g(\theta_1, \theta_2, \ldots, \theta_p)$ where $\theta_1, \theta_2, \ldots, \theta_p$ are the eigenvalues of the matrix in the argument of f and g is strictly increasing in each θ_j. Then

$$f(E\{(\underline{X} - M\underline{Y})(\underline{X} - M\underline{Y})'\})$$

is minimized with respect to M and $\underline{Y}$ when and only when

$$M\underline{Y} = \sum_{j=1}^{q} \underline{v}_j \underline{v}_j' \underline{X} ,$$

the $\underline{v}_j$ being orthonormal eigenvectors of Σ corresponding to the θ_j. The minimum value of f is

$$g(\theta_{q+1}, \theta_{q+2}, \ldots, \theta_p, 0, \ldots, 0).$$

Proof. See Okamoto and Kanazawa [18].

The next lemma is based on an assertion by Rao ([20], p.441).

LEMMA A4. Suppose

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

is non-negative definite. Then there exist matrices M_{12} and M_{21} such that

$$\Sigma_{12} = M_{12}\Sigma_{22} \quad \text{and} \quad \Sigma_{21} = M_{21}\Sigma_{11}.$$

Proof. $\Sigma_{ij} = T_iT_j'$ for some T_i and T_j, $i = 1,2$; $j = 1,2$. $V(T_1T_2') \subset V(T_1) = V(T_1T_1')$. Therefore, $\Sigma_{21} = M_{21}\Sigma_{11}$ for some M_{21}. Similarly, $\Sigma_{12} = M_{12}\Sigma_{22}$.

LEMMA A5. Let $T : U \to U$ be an idempotent (i.e., $T^2 = T$) endomorphism of a vector space. Then U is the direct sum of the kernel of T and the image of T.

Proof. See Mostow, Sampson, and Meyer ([17], p.384).