Understanding Canonical Correlation Through the General Linear Model and Principal Components

KEITH E. MULLER*

Canonical correlation has been little used and little understood, even by otherwise sophisticated analysts. An alternative approach to canonical correlation, based on a general linear multivariate model, is presented. Properties of principal component analysis are used to help explain the method. Standard computational methods for full rank canonical correlation, techniques for canonical correlation on component scores, and canonical correlation with less than full rank are discussed. They are seen to be essentially equivalent when the model equation for canonical correlation on component scores is presented. The two approaches to less than full rank situations are equivalent in some senses, but quite different in usefulness, depending on the application. An example dataset is analyzed in detail to help demonstrate the conclusions.

KEY WORDS: Canonical correlation; General linear model; Less than full rank regression; Component matching; Regression on component scores.

1. INTRODUCTION

Canonical correlation was introduced by Hotelling (1935, 1936) as the answer to a simple problem with vast implications: find that linear combination of a set of variables which is most highly correlated with any linear combination of a second set of variables. Like most multivariate techniques, it received more attention from theoreticians than it did practical application until powerful digital computers became available in the 1960's. That attention centered mostly on either answering the difficult distributional problems associated with the technique or dealing with the many ways that more common techniques are either kin to or special cases of canonical correlation. Bartlett (1948) and McKeon (1964) provided useful reviews of that issue.

This article centers on understanding canonical correlation as a method of estimation of weights used to predict each set from the other. The general linear multivariate model underlying canonical correlation is introduced and used, along with standard results on principal components, to help explain the method. An example is analyzed in detail to demonstrate many of the conclusions.

By exploring the general linear multivariate model behind canonical correlation, a number of useful insights may be had. First, the nature of canonical correlation is made clear. Second, simpler derivations are available. Third, computational methods for canonical correlation, canonical correlation on component scores, and the less than full rank case can be seen to be different interpretations of the same situation (the collinearity problem in regression is essentially the same problem as the less than full rank problem). Canonical correlation has never been among the commonly used multivariate techniques. It is hoped that the reader will be better able to judge its uses and limitations having read this article.

This article will ignore questions associated with significance testing. As mentioned earlier, canonical correlation contains as special cases many multivariate techniques for linear models. Suffice it to say that, given conventional distributional assumptions, the canonical solution provides statistics that are optimal in many useful senses. Much progress has been made towards solving the various distributional questions but work remains to be done. Anderson (1958) presented a good deal of information on significance testing. For purposes of estimation, the focus of this article, only the usual least squares assumptions will be required: linearity of relationships, independence of observations, and homoscedasticity.

2. THE STANDARD STATEMENT OF CANONICAL CORRELATION

Many multivariate texts treat canonical correlation. See, for example, Rao (1973), Morrison (1967), Finn (1974), or Tatsuoka (1971). Gittens (1979) provided one of the more detailed presentations. This section first briefly presents the standard statement of canonical correlation. This will introduce the notation used throughout. A number of well-known properties are also mentioned.

In general, Greek letters will denote population parameters and Roman-alphabet letters will denote sample statistics. Capital letters will be used for matrices, small letters for scalars, and small bold letters for column vectors. With that in hand, consider n observations on two sets of variables with

*Keith E. Muller is Research Assistant Professor, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27514. This research was supported in part by U.S. Environmental Protection Agency contract 68-02-3402. Portions of this paper were presented at the 1980 American Statistical Association Meetings in Houston. The author gratefully acknowledges corrections and changes suggested by Professor N.L. Johnson for an earlier version. The manuscript was also improved by revisions suggested by an anonymous associate editor. The data for the example were used with the kind permission of Dr. E.S. Johnson.

$$X = \{x_{ij}\} \quad i = 1, \ldots n \quad j = 1, \ldots p \quad (2.1)$$

$$Y = \{y_{ij}\} \quad i = 1, \ldots n \quad j = 1, \ldots q \quad (2.2)$$

where $n > p + q$.

Consider the parallel matrices Z_x and Z_y with zero means and unit variance. Using sample standardized values will help simplify expressions although it implicitly causes some technical problems. All such technical problems can be resolved, usually by working with the "raw" (unstandardized) data. Ignoring the problems allows greatly simplifying the exposition.

The matrices of correlations among the X variables, among the Y variables, and between the two sets are

$$R_{xx} = (1/n)Z_x{}'Z_x, \quad (2.3)$$

$$R_{yy} = (1/n)Z_y{}'Z_y, \quad (2.4)$$

$$R_{xy} = (1/n)Z_x{}'Z_y. \quad (2.5)$$

Sets of coefficients, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, are sought which maximize

$$\rho(Z_x\boldsymbol{\alpha},\ Z_y\boldsymbol{\beta}) = (1/n)\boldsymbol{\alpha}Z_x{}'Z_y\boldsymbol{\beta} = \boldsymbol{\alpha}'R_{xy}\boldsymbol{\beta}, \quad (2.6)$$

under the constraints

$$(1/n)(Z_x\boldsymbol{\alpha})'(Z_x\boldsymbol{\alpha}) = \boldsymbol{\alpha}'R_{xx}\boldsymbol{\alpha} = 1, \quad (2.7)$$

and

$$(1/n)(Z_y\boldsymbol{\beta})'(Z_y\boldsymbol{\beta}) = \boldsymbol{\beta}'R_{yy}\boldsymbol{\beta} = 1. \quad (2.8)$$

The second and third conditions require the resultant combinations, namely $Z_x\boldsymbol{\alpha}$ and $Z_y\boldsymbol{\beta}$, to have unit variance. In order to derive the canonical solution, LaGrangian multipliers are introduced. Then matrix derivatives are computed, set to zero, and simplified. The results are what may be thought of as the normal equations for canonical correlation:

$$\rho R_{xx}\boldsymbol{\alpha} = R_{xy}\boldsymbol{\beta} \quad (2.9)$$

and

$$\rho R_{yy}\boldsymbol{\beta} = R_{yx}\boldsymbol{\alpha}. \quad (2.10)$$

To continue the standard proof, at this point one must assume R_{xx} and R_{yy} to be full rank and the canonical correlation to be nonzero. With these restrictions, (2.8) implies

$$\boldsymbol{\beta} = (1/\rho)R_{yy}{}^{-1}R_{yx}\boldsymbol{\alpha}. \quad (2.11)$$

It is then easy to generate the solution equation

$$\mathbf{0} = (R_{xy}R_{yy}{}^{-1}R_{yx} - \rho^2R_{xx})\boldsymbol{\alpha}, \quad (2.12)$$

and

$$\mathbf{0} = (R_{xx}{}^{-1}R_{xy}R_{yy}{}^{-1}R_{yx} - \rho^2I)\boldsymbol{\alpha}. \quad (2.13)$$

Consequently, the weights for the X set are the right eigenvectors of the matrix

$$M_x = R_{xx}{}^{-1}R_{xy}R_{yy}{}^{-1}R_{yx} \quad (2.14)$$

and its eigenvalues are the squared canonical correlations.

At most $d = \min(p,q)$ distinct solutions may exist. The canonical variates are uncorrelated (orthogonal) within set, as required. Furthermore, the ith variate is uncorrelated with any variate from the other set, except for the ith (the property of bi-orthogonality). Each successive canonical variate pair achieves the maximum relationship orthogonal to the preceding pair. An equivalent set of equations, parallel to (2.11) through (2.16), exists based on the Y set. M_x, in general, is not symmetric. This lack of symmetry implies that the canonical weights themselves are also generally not orthogonal. The canonical solution may be said to characterize completely all joint linear relationships between two sets of variables. The canonical correlations remain invariant under any full-rank transformation of either or both sets, for example. Consequently, the canonical solution may be computed on the sums of squares and cross-products matrices, covariance, or correlation matrices. A simple transformation allows conversion to the preferred metric.

A number of special cases of canonical correlation are important and more commonly used techniques. If $q = 1$ and $p > 1$ then canonical correlation becomes multiple correlation. If $q = 1$ and $p = 1$ then canonical correlation becomes univariate correlation. If one set of variables is a set of dummy variables indicating group membership, then canonical correlation becomes discriminant analysis and, equivalently, multivariate analysis of variance. Bartlett (1948) and McKeon (1964) discussed in detail how various techniques are special cases of canonical correlation.

3. STANDARD GENERAL LINEAR MULTIVARIATE MODEL

It is worth noting that the standard derivations of canonical correlation and multiple regression bear little relation to each other. For multivariate multiple regression a standard approach may be sketched as follows. Given n observations on each of two sets of variables with

$$X = \{x_{ij}\} \quad i = 1, \ldots n, \quad j = 1, \ldots p \quad (3.1)$$

and

$$Y = \{y_{ij}\} \quad i = 1, \ldots n, \quad j = 1, \ldots q, \quad (3.2)$$

assume that the linear combinations of X that best predict Y are sought. This amounts to finding an estimate of Ξ in the following model equation:

$$Y = X\Xi + E. \quad (3.3)$$

Here E is an $n \times q$ matrix of errors with zero column means. It is often convenient to consider the model for standardized scores:

$$Z_y = Z\Xi + E. \quad (3.4)$$

The method of moments and the least squares method both provide the same estimate, based on the equation

$$X'Y = X'X\hat{\Xi}. \quad (3.5)$$

This is usually referred to as the system of normal equations for the model. For standard scores, the normal equations may be expressed as

$$R_{xy} = R_{xx}\hat{\Xi}. \quad (3.6)$$

If X is full rank, then it follows that

$$\hat{\Xi} = (X'X)^{-1}X'Y \quad (3.7)$$

or, for standard scores,

$$\hat{\Xi} = R_{xx}{}^{-1}R_{xy}. \quad (3.8)$$

The least squares method treats the residual covariance matrix

$$V_y = \frac{1}{n}E'E = (Y - X\Xi)'(Y - X\Xi)/n \quad (3.9)$$

$$= (Y'Y - Y'X\Xi - \Xi'X'Y + \Xi'X'X\Xi)/n. \quad (3.10)$$

The same solution follows from minimizing (via matrix differentiation techniques) either the generalized variance, which is the determinant of V_y, or the trace of V_y (Press 1972). In turn, the estimator is the same as the one just discussed.

4. GENERAL LINEAR MULTIVARIATE MODEL FOR CANONICAL CORRELATION

Canonical correlation may also be cast in a regression model. The emphasis shifts to estimation of the weights to predict the linear combinations of one set from the linear combinations of the other set. This section introduces a model for canonical correlation. It will be used in the remainder of the article to provide a novel and, one hopes, clearer approach.

Consider the model equation

$$Z_y B = Z_x A D(\rho_k) + E. \quad (4.1)$$

B is a $q \times d$ matrix, with the kth column being the canonical weights for the Y set for the kth canonical variate pair. $D(\rho_k)$ is a $d \times d$ diagonal matrix of canonical correlations. A is a $p \times d$ matrix, with the kth column being the canonical weights for the X set for the kth canonical variate. The matrices B, A, and $D(\rho_k)$ must correspond in the sense that the kth columns of B and A provide the linear combinations that are correlated ρ_k, which is the (k,k) element of $D(\rho_k)$. It is most natural, although not necessary, for the matrices to be ordered such that the largest canonical correlation is the (1,1) element of $D(\rho_k)$, the second largest the (2,2) element, and so forth. As earlier, d is at most $\min(p,q)$. It will often be convenient to drop any zero canonical correlations and associated vectors from the equation. E here is $n \times d$.

Equation (4.1) may be treated as a special case of (3.3) with Y identified as $Z_y B$, X as Z_x, and Ξ as $AD(\rho_k)$. This special structure arises from the orthogonality and bi-orthogonality of the canonical variates. Consider the equation treating just the kth canonical correlation:

$$Z_y \boldsymbol{\beta}_k = Z_x \boldsymbol{\alpha}_k \rho_k + \mathbf{e}_k. \quad (4.2)$$

This is a univariate regression equation relating a single canonical "Y" variate to its corresponding single X variate. The standard statement of canonical correlation has more in common with this univariate equation than with the multivariate statement of (4.1).

The same derivational approach that is used in the standard proof may, of course, be applied to a multivariate statement of the problem. With appropriate extra care to deal with the greater complication, the proof follows in a parallel fashion. The solution equations become

$$R_{xy} B = R_{xx} A D(\rho_k) \quad (4.3)$$

and

$$R_{yx} A = R_{yy} B D(\rho_k). \quad (4.4)$$

Two advantages accrue from the extra work. First, it makes obvious the fact that the resultant variates are required to be orthogonal within each set, a crucial characteristic that is only implicit in a univariate approach. Second, the multivariate approach helps make more explicit the simultaneous factorizations at the heart of canonical correlation. That observation will be explored later in this article.

New derivations of the canonical solution follow naturally from (4.1). Straightforward method-of-moments and least squares proofs are both available.

5. PROPERTIES OF THE RESIDUAL COVARIANCE MATRIX

A number of elegant properties of the model become apparent upon examining the residual covariance matrix. Equation (4.1) gives

$$V_y = \frac{1}{n}E'E = (Z_y B - Z_x A D(\rho_k))'(Z_y B - Z_x A D(\rho_k))/n \quad (5.1)$$

$$= B'R_{yy}B - B'R_{yx}AD(\rho_k) - D(\rho_k)A'R_{xy}B + D(\rho_k)A'R_{xx}AD(\rho_k). \quad (5.2)$$

Substituting $D(\rho_k)$ for $A'R_{xy}B$ and $B'R_{yx}A$ and I for $A'R_{xx}A$ and $B'R_{yy}B$ in (5.2) gives

$$V_y = I - D(\rho_k)D(\rho_k) - D(\rho_k)D(\rho_k) + D(\rho_k)ID(\rho_k) \quad (5.3)$$

$$= I - D(\rho_k^2) \quad (5.4)$$

$$= D(1 - \rho_k^2). \quad (5.5)$$

It is easy to show that the same result holds for V_x, and hence that $V_x = V_y$. Since the residual covariance matrix is diagonal the residuals are uncorrelated. The residual associated with the kth canonical variate pair has variance $1 - \rho_k^2$. In simple linear regression $1 - r^2$ is the proportion of variance in Y not accounted for in X.

The diagonality of V may be contrasted to the situation in multivariate multiple regression. From (3.4) and (3.8) it is easily seen that for standard scores

$$V_y = (Z_y - Z_x\hat{\Xi})'(Z_y - Z_x\hat{\Xi})(1/n) \quad (5.6)$$

$$= R_{yy} - R_{yx}R_{xx}^{-1}R_{xy}. \quad (5.7)$$

The off-diagonal elements, the partial covariances, are not, in general, zero. The diagonal elements are $1 - R_k^2$, where R_k^2 is the squared multiple correlation for Y_k given the X set. Either X, Y, or both being orthogonal does not guarantee that the matrix is diagonal. One of the defining characteristics of the canonical solution is the required orthogonality.

6. COMPUTATIONAL METHODS FOR CANONICAL CORRELATION

The canonical "solution" equation actually leaves the analyst a few steps away from the calculations for a data set. The steps needed help explain the technique. Furthermore, the solution methods will be reinterpreted later in this article in order to demonstrate certain relationships to component analysis.

The methods commonly employed to compute the canonical solution depend on factoring R_{xx}. Finn (1974), among others, outlined the most common approach. Equations (2.15) and (2.16) indicate that the canonical weights may be found as right eigenvectors of the matrix M_x. Since M_x is not, in general, symmetric, its right and left eigenvectors are distinct and not orthogonal. To deal with this, begin by finding a factor of R_{xx}, F_x (such that $F_xF_x' = R_{xx}$). It is easy to see that we may define $F_x^{-t} \equiv (F_x^{-1})' \equiv (F_x')^{-1}$. The following simple manipulations of (2.12) lead to a solution.

$$(R_{xy}R_{yy}^{-1}R_{yx} - \rho_k^2 R_{xx})\boldsymbol{\alpha} = \mathbf{0}; \quad (2.12)$$

$$F_x^{-1}(R_{xy}R_{yy}^{-1}R_{yx} - \rho_k^2 F_xF_x')F_x^{-t}F_x'\boldsymbol{\alpha} = F^{-1}\mathbf{0}, \quad (6.1)$$

$$(F_x^{-1}R_{xy}R_{yy}^{-1}R_{yx}F_x^{-t} - \rho_k^2 I)F_x'\boldsymbol{\alpha} = \mathbf{0}, \quad (6.2)$$

and

$$(M_x^* - \rho_k^2 I)\boldsymbol{\alpha}^* = \mathbf{0}. \quad (6.3)$$

M_x^* is clearly symmetric with orthogonal eigenvectors $\boldsymbol{\alpha}^*$ and eigenvalues ρ_k^2. Equation (6.3) may be solved with readily available computer programs. Having solved for $\boldsymbol{\alpha}^*$, it follows that

$$\boldsymbol{\alpha} = F^{-t}\boldsymbol{\alpha}^*. \quad (6.4)$$

Equation (2.11) gives $\boldsymbol{\beta} = (1/\rho)R_{yy}^{-1}R_{yx}\boldsymbol{\alpha}$ to complete the solution.

7. CANONICAL CORRELATION ON COMPONENT SCORES

Principal components may be defined in many ways. Assume that new variates, which are linear combinations of an existing set, are sought such that the new variates are successively maximum variance and mutually uncorrelated (orthogonal). Operationally, they may be computed from the spectral decomposition of either the covariance or correlation matrices. Note that distinct sets arise according to which matrix is factored.

In a multiple regression context, it may happen that the set of predictors, say X, may be collinear. This is equivalent to saying that the correlation matrix, R_{xx} (or the corresponding covariance matrix, C_{xx}), is rank $p^* < p$. A common technique is to replace the original deficient rank scores with a full rank set of component scores. Such an approach may be exploited here.

Define a factor of R_{xx} as F_x such that

$$R_{xx} = F_xF_x', \quad (7.1)$$

where F_x is $p \times p^*$, $p^* \leq p$, and $p^* = rk(R_{xx})$. The spectral decomposition of R_{xx} is

$$\underset{(p \times p^*)(p^* \times p^*)(p^* \times p)}{R_{xx} = E_xD(\lambda_x)E_x'}. \quad (7.2)$$

Here E_x is a matrix whose columns are the eigenvectors of R_{xx} and $D(\lambda_x)$ is the diagonal matrix of eigenvalues of R_{xx}. It is easy to see that choosing

$$F_x = E_xD(\lambda_x^{1/2}) \quad (7.3)$$

is one choice meeting the requirement (7.1).

If uncorrelated (orthogonal), unit-variance new variates, Z_{x^*}, are sought such that

$$Z_x = Z_{x^*}F'_x \quad (7.4)$$

then

$$Z_{x^*} = Z_xF_x(F_x'F_x)^{-1}. \quad (7.5)$$

Even if $p^* < rk(X)$, (7.5) may be a useful approximation. The covariance matrix among the new variates is

$$C_{x^*x^*} = \frac{1}{n}Z_{x^*}'Z_{x^*} = (F_x'F_x)^{-1}F_x'R_{xx}F_x(F_x'F_x)^{-1} \quad (7.6)$$

$$= I = R_{x^*x^*} \quad (7.7)$$

If $rk(X) = p$ then F_x is $p \times p$, full rank, and has an inverse. This leads to useful simplifications of (7.5):

$$Z_{x^*} = Z_xF_x(F_x^{-1}F_x^{-t}) \quad (7.8)$$

$$= Z_xF_x^{-t} = Z_xE_xD(\lambda_x^{-1/2}). \quad (7.9)$$

If Z_{y^*} are computed for the Y set in a parallel manner, it is easy to see that

$$R_{x^*y^*} = (F_y'F_y)^{-1}F_y'R_{yx}F_x(F_x'F_x)^{-1}. \quad (7.10)$$

If $rk(X) = p$ then (7.9) implies

$$R_{x^*y^*} = F_x^{-1}R_{xy}F_y^{-t}. \quad (7.11)$$

The new variates are principal component scores normalized to have unit variance. Most computers packages will provide them, labeled as "factor scores." Component scores, defined by

$$Z_xE_x = Z_{xc}, \quad (7.12)$$

have covariance matrix

$$\frac{1}{n}(Z_xE_x)'(Z_xE_x) = E_x'R_{xx}E_x = D(\lambda_x).$$

These results may be used to show that the canonical solution is equivalent to a three-step process. First, compute factor matrices for R_{xx} and R_{yy}. Second, compute the intercorrelation matrix for the associated component scores using (7.10). Third, decompose that matrix. If required, the results may be stated in terms of the original variables by using the transformations in (6.4) and (2.11).

In order to see this, consider the solution equation for component scores:

$$(R_{x^*y^*}R_{y^*y^*}^{-1}R_{y^*x^*} - \rho_k^2 I)\boldsymbol{\alpha}^* = \mathbf{0}. \quad (7.14)$$

Since $R_{y^*y^*} = I$, $R_{y^*y^*}^{-1} = I$, (7.14) becomes (7.15)

$$(R_{x^*y^*}R_{y^*x^*} - \rho_k^2 I)\boldsymbol{\alpha}^* = \mathbf{0}. \quad (7.16)$$

Using (7.11) in (7.16) gives

$$((F_x^{-1}R_{xy}F_y^{-t})(F_y^{-1}R_{yx}F_x^{-t}) - \rho_k^2 I)\boldsymbol{\alpha}^* = \mathbf{0}. \quad (7.17)$$

Since $R_{yy} = F_yF_y'$ and $R_{yy}^{-1} = F_y^{-t}F_y^{-1}$, (7.17), (7.16), and (7.14) are seen to be identical to (6.3), the most commonly used solution equation. Since $\boldsymbol{\alpha}^*$ must be an eigenvector of

$$R_{x^*y^*}R_{y^*x^*} = M_{x^*}, \quad (7.18)$$

it is appropriate to use the same symbols as in (6.3).

The results presented show that the canonical solution provides a means of comparing the component structures of two sets of variables. The weights A^* and

B^* are the rotations needed to produce maximum congruence. The patterns of their values show which factors of each set are related. The canonical correlations indicate the amount of overlap of the components.

The general problem of comparing two factor structures is known as the factor matching or Procrustes problem (Harman 1967). An extensive literature exists, treating many different cases. Levine (1977) provided an introduction to the topic as it relates to canonical correlation. Browne (1979) derived a method for maximum likelihood interbattery factor analysis. The interbattery factor loadings are rescaled canonical factors. Note that Browne's work centers on a genuine common-factor model rather than the components approach studied here.

The ideas in this section are not new. Hotelling (1936) discussed canonical correlation as a process of diagonalizing the correlation matrices R_{xx}, R_{yy}, and R_{xy}. Bartlett (1948) and McKeon (1964), among others, directly suggested that canonical correlation be used to evaluate factor matching. Rao (1973) discussed testing the canonical correlations to test how many factors are shared. In contrast to most factor-analytic methods, the technique can have appropriate significance tests. Furthermore, other techniques of factor matching do not have measures of congruence as well defined as the canonical correlation. The technique depends on having both sets of variables measured on the same subjects since it exploits the inter-set correlations.

The reader should keep in mind certain distinctions. First, A and B are, in general, nonorthogonal factors of R_{xx}^{-1} and R_{yy}^{-1}, respectively. Second, the initial factor matrix F_x, with $F_xF_x' = R_{xx}$, is orthogonal. Third, the matrix A^*, the canonical weights for the component scores, is orthogonal. It is the rotation that transforms the original components into the canonical factors. The canonical factors are usually computed as $R_{xx}A$. Since $A = F^{-\prime}A^*$ it follows that

$$R_{xx}A = (F_xF_x')F_x^{-\prime}A^* = F_xA^*. \qquad (7.19)$$

Since this holds for any factoring, A^* depends on the choice of factoring, and hence A^* is usually ignored. The next two sections will discuss a case in which A^* may be of more interest than A.

8. MODEL FOR CANONICAL CORRELATION ON COMPONENT SCORES

For the purpose of factor matching, a special case of the canonical model provides a model for canonical correlation on component scores:

$$Z_xF_y^{-\prime}B^* = Z_xF_x^{-\prime}A^*D_+(\rho_k) + E. \qquad (8.1)$$

Here the kth column of A^* is α^*_k, the kth eigenvector of M_x^*, and the kth column of B^* is β^*_k, the kth eigenvector of M_x^*. If $q < p$, $D_+(\rho_k)$ is $D(\rho_k)$ with p-q zero rows adjoined at the bottom, while if $q > p$, then $D_+(\rho_k)$ is $D(\rho_k)$ with $q - p$ zero columns adjoined on the right. If $p = q$ then $D_+(\rho_k) = D(\rho_k)$.

The model brings together results presented earlier and makes them more accessible. The canonical model requires two separate transformations of both sets of variables. The roles of the original variables and the factoring process are clarified by this model equation. Simple derivations of the canonical solution follow from (8.1). Equation (8.1) provides a model for component matching, since it may be identified as a model for canonical correlation in which the scores of interest are component scores on two different sets of variables.

9. LESS THAN FULL RANK CANONICAL CORRELATION

Less than full rank correlation matrices, particularly in a numerical sense, may arise in many ways. It seems a natural event in what is implicitly a component analysis. Despite that, for most of this article it has been assumed that $rk(X) = p$ and $rk(Y) = q$. This assumption is convenient although not necessary. Relaxing the assumption requires more caution both in the mathematics and in the application of canonical correlation to data. This section will review earlier work in the area. The next section will propose an alternative method, then relate the two to each other and to an important special case, multiple regression.

Recall that no assumption of rank was needed in the derivation of the normal equations for canonical analysis:

$$\rho R_{xx}\alpha = R_{xy}\beta \qquad (2.4)$$

and

$$\rho R_{yy}\beta = R_{yx}\alpha. \qquad (2.10)$$

Searle (1971), p. 26, theorem 8) gives a theorem, in the context of multiple regression, that allows proceeding even though R_{yy} is not full rank. In the notation used here, it says that, for all generalized inverses R_{yy}^-,

$$\tilde{\beta} = \frac{1}{\rho}R_{yy}^-\alpha \qquad (9.1)$$

generates all solutions to (2.10). Following the technique of the standard proof, this is substituted in (2.9) giving

$$(R_{xy}R_{yy}^-R_{yx} - \rho^2R_{xx})\tilde{\alpha} = \mathbf{0}. \qquad (9.2)$$

It is important to realize that for $rk(X) = p^* < p$ infinitely many R_{xx}- exist, therefore infinitely many versions of (9.2) exist and so infinitely many $\tilde{\alpha}$ exist. Khatri (1976) showed that the canonical correlations are unique even though the weights are not. This follows from

$$R_{xy}R_{yy}^-R_{yx} = R_{xy}R_{yy}^+R_{yx} \qquad (9.3)$$

and the fact that R_{xx}^+ is unique. Here R_{xx}^+ is the unique Moore-Penrose inverse. If one set of variables is full rank then the weights for that set are unique but not the weights for the other (not of full rank) set. The use of the unique Moore-Penrose inverse defines unique weights. In any case, for any less than full rank set the weights are not estimable (in the statistical sense of an unbiased estimator existing).

An alternative approach to dealing with singular correlation matrices follows from the material in this article. A common technique in multiple regression with an X set of less than full rank is to replace the X scores with the (full rank) set of principal components scores. No information is lost, the resultant parameter matrix of regression weights is estimable, and the process may reveal a more accurate understanding of the structure of the variables. In some cases the investigator has a strong commitment to the original variable space, and therefore the generalized inverse approach has merit. Harris (1975) briefly discussed the approach of using factor scores in canonical correlation.

Results in (7.5) may be applied to (8.1) to provide a model for a singular X and Y:

$$Z_yF_y(F_y'F_y)^{-1}B^* = Z_xF_x(F_x'F_x)^{-1}A^*D_+(\rho_k) + E. \quad (9.4)$$

As before, any complete orthogonal factoring will do. However, in this context the factoring may be exploited as part of the interpretation since it amounts to a reparameterization of a less than full rank model to a full rank model. The use of square-root (Cholesky) factors, the usual method for computer programs, is commendable when just a numerical way station. It would be recommended by few analysts over a principal-components analysis if the factors themselves are to see the light of day, as they should here.

In the full-rank factor space, the generalized inverse approach is equivalent to the factor-score approach. This depends on two basic ideas. Khatri (1976) indicated one convenient method for computing R_{yy}^+ (for a square symmetric matrix). Find F_y such that $R_{yy} = F_yF_y'$ where F_y is $q \times q^*$, where $q^* = rk(Y)$. Here, of course, the concern is with $q^* < q$. Then R_{yy}^+ may be computed as

$$R_{yy}^+ = F_y(F_y'F_y)^{-2}F_y. \quad (9.5)$$

Using this and other results mentioned earlier, it can be shown that the solution equation for both methods is

$$[(F_x'F_x)^{-1}F_x'R_{xy}F_y(F_y'F_y)^{-1}(F_y'F_y)^{-1} \times F_y'R_{yx}F_x(F_x'F_x)^{-1} - \rho^2I]\boldsymbol{\alpha}^* = \mathbf{0}. \quad (9.6)$$

This shows the equivalence of the generalized inverse approach and the computation of the solution on factor scores. More often than not, much is lost by projecting the solution back into the rank-deficient original variables' space, which is done with the generalized inverse approach. Either provides a completely general approach to canonical correlation.

10. UNDERSTANDING MANOVA VIA THE CANONICAL MODEL

An important application of canonical correlation is in understanding a multivariate analysis of variance (MANOVA). A true multivariate analysis, as distinguished from a collection of univariate analyses, uses multivariate test statistics. All of the standard multivariate test statistics are simple functions of the canonical correlations.

Recall that the residual covariance matrix for the canonical model is

$$V = D(1 - \rho_k^2). \quad (5.5)$$

Hence the trace of V is

$$\mathrm{tr}(V) = \sum_k (1 - \rho_k^2) = d - \sum_k \rho_k^2. \quad (10.1)$$

Since V is diagonal, its determinant is

$$V = \prod_k (1 - \rho_k^2) = \Lambda. \quad (10.2)$$

Here Λ is Wilk's Lambda likelihood ration criterion (Tatsuoka 1971). The test statistic for Roy's largest root test is ρ_m^2, where ρ_m^2 is the largest canonical correlation. Bartlett's smallest root test considers the smallest canonical correlation. The Hotelling-Lawley trace statistic is $\sum_k (\rho_k^2/(1 - \rho_k^2))$. The Pillai-Bartlett trace is simply $\sum \rho_k^2$. Therefore all of these standard statistics are simple functions of V. The expressions for the statistics are from Olsen (1976). Cramer and Nicewander (1979) reviewed measures of multivariate association. All six measures they treated may be expressed as simple functions of the canonical correlations, and so of V.

Implicitly, then, one is testing canonical correlations in a multivariate analysis of variance. By not considering the canonical variates, one ignores exactly those linear combinations judged significant. Some MANOVA programs do give at least one set of the "discriminant weights," the linear combinations of the dependent variables. The canonical correlations may be useful measures of the strength of a relationship, something too often ignored in the MANOVA setting. It may also be useful to study the weights on the predictors. Canonical correlation may provide a useful tool in decomposing a complex multivariate-multivariate relationship space. The general linear multivariate model approach to computing a MANOVA now dominates. The general linear multivariate model approach to canonical analysis presented here provides a natural framework for interpreting a MANOVA.

11. EXAMPLE ANALYSIS

This section summarizes a canonical correlation analysis for a particular set of data. The goal is to demonstrate the concepts discussed earlier in the article. Canonical correlation analysis involves evaluating relationships within sets of variables and between sets of variables. It is most natural to first consider summary statistics for the data, then the within-set characteristics, and finally the between-set characteristics.

The data of this example are observations on 200 college students from introductory psychology classes, each solving four different problems. Details of the problem-solving task are documented in Johnson (1971). Of the 200 subjects, 88 were female and 112 were males. These data came from a study involving a treatment that was found not to affect problem-solving performance. Furthermore, no differences were seen between male and female performance on this task. If any treatment effects were present, it would be appropriate to conduct canonical analysis on the residual correlation matrix rather than the unadjusted correlated matrix. Muller, Hosking, and Helms (1979) discussed this problem in detail. They demonstrated that the presence of treatment effects leads to biased estimation of the covariance matrix, and usually to distorted factor structure.

Table 1 provides summary statistics on the 16 variables included in the canonical analysis. Four measures are available for each subject for each of the four problems solved. For every problem, each response of a subject, a "trial," may be classified as either a "hypothesis" or a "pattern" trial. Each hypothesis trial may be classified as either "valid," useful, or "invalid," useless, for solving the problem. Each pattern response may be classified as either "informative," useful, or "redundant," useless, for solving the problem. In a weak sense, valid hypotheses and informative patterns are incorrect responses. The sum of the four numbers, valid hypotheses plus invalid hypotheses plus informative patterns plus redundant patterns, is equal to the number of trials taken to solve a particular problem. In all cases the value measured is the simple count for an entire problem for a particular subject. As counts, all are bounded below by zero. Table 1 indicates that on the average a subject presented approximately five valid hypotheses for each problem, from two to five invalid hypotheses, approximately six informative patterns, and from two to four redundant patterns.

The point of this canonical analysis is to investigate the similarity of structure of the hypothesis variables and the pattern variables. Consequently, one set of variables studied included the four valid hypothesis responses and the four invalid hypothesis responses, while the other set included the four informative pattern variables and the four redundant pattern variables.

Since canonical correlation is essentially invariant under a linear transformation of either or both sets of variables, it is most appropriate to treat the correlation matrices rather than the covariance or sums of squares matrices. Tables 2, 3, and 4 report the relevant correlation matrices. Table 2 gives the correlations among the hypothesis variables. Table 3 provides the correlations among the pattern variables. Table 4 lists the correlations between hypothesis and pattern variables.

The first step in a canonical analysis is the decomposition of each of the two sets of variables. For this example, principal component analysis was chosen. Tables 5 and 6 summarize the eigen analysis for the hypothesis and pattern variables, respectively. In a factor analysis, a commonly used rule of thumb is that any factor that does not account for one variable's worth of information is not included. In this case eight variables are present and so one variable's share of variance is .125. Therefore approximately two or three factors would be retained in each case. The canonical correlation analysis uses all eight components from both sets.

Tables 7 and 8 present the factor matrices for both the hypothesis and pattern variables. These matrices are factors in the sense of (7.1) for a correlation matrix among either the hypothesis or pattern variables. For a correlation matrix its factor matrix is also a matrix of correlations. The correlations are between the original variables, the row labels, and the components, the new variates, which are the column labels. The elements of a factor of a covariance matrix are covariances. As mentioned earlier, the factor matrix for principal-components analysis equals the matrix of eigenvectors multiplied by the diagonal matrix of the square roots of the eigenvalues.

It is common to plot the factor matrix of a correlation

Table 1. Problem-Solving Performance for Four Problems (n = 200)

Variable	*Mean*	*Standard Deviation*	*Minimum*	*Maximum*
Valid Hypotheses 1	5.20	6.66	1	30
Valid Hypotheses 2	5.74	7.14	1	30
Valid Hypotheses 3	5.15	7.41	1	30
Valid Hypotheses 4	4.93	7.37	1	30
Invalid Hypotheses 1	4.52	6.53	0	32
Invalid Hypotheses 2	3.66	6.45	0	48
Invalid Hypotheses 3	2.74	5.95	0	53
Invalid Hypotheses 4	1.77	4.30	0	30
Informative Patterns 1	6.49	3.05	0	19
Informative Patterns 2	6.11	3.43	0	19
Informative Patterns 3	6.38	3.39	0	19
Informative Patterns 4	6.24	3.35	0	19
Redundant Patterns 1	3.80	5.04	0	33
Redundant Patterns 2	2.31	3.03	0	17
Redundant Patterns 3	2.08	3.09	0	23
Redundant Patterns 4	1.87	3.94	0	41

Table 2. Correlations Among Hypothesis Variables (n = 200)

	Valid Hypothesis				*Invalid Hypothesis*			
	1	*2*	*3*	*4*	*1*	*2*	*3*	*4*
Valid 1	1.00							
Valid 2	.70	1.00						
Valid 3	.69	.80	1.00					
Valid 4	.64	.77	.85	1.00				
Invalid 1	.12	.24	.21	.28	1.00			
Invalid 2	.14	.08	.11	.21	.27	1.00		
Invalid 3	.03	.13	.06	.14	.29	.43	1.00	
Invalid 4	.13	.06	.15	.11	.14	.37	.37	1.00

Table 3. Correlations Among Pattern Variables (n = 200)

	Informative Patterns				*Redundant Patterns*			
	1	*2*	*3*	*4*	*1*	*2*	*3*	*4*
Informative 1	1.00							
Informative 2	.57	1.00						
Informative 3	.55	.70	1.00					
Informative 4	.54	.73	.83	1.00				
Redundant 1	.32	.28	.30	.32	1.00			
Redundant 2	.01	.20	.17	.17	.27	1.00		
Redundant 3	.08	.11	.23	.22	.22	.40	1.00	
Redundant 4	.07	.14	.13	.14	.08	.21	.31	1.00

Table 4. Correlations Between Hypothesis and Pattern Variables (n = 200)

	Informative Patterns				Redundant Patterns			
	1	2	3	4	1	2	3	4
Valid 1	−.64	−.43	−.43	−.42	−.19	−.14	−.15	−.10
Valid 2	−.45	−.65	−.54	−.53	−.04	−.15	−.17	−.15
Valid 3	−.44	−.45	−.64	−.55	−.06	−.12	−.21	−.11
Valid 4	−.41	−.46	−.55	−.06	−.12	−.13	−.21	−.13
Invalid 1	−.13	−.04	−.05	−.07	.43	.23	.14	−.00
Invalid 2	−.20	−.13	−.18	−.20	.05	.40	.16	.01
Invalid 3	−.14	−.21	−.17	−.13	−.01	.12	.15	.09
Invalid 4	−.12	−.13	−.13	−.17	−.07	.03	.15	.37

matrix as in Figures 1 and 2. Such a plot displays the component loadings, which are the correlations between the variables and the components. The structures for the two sets of variables look quite similar. The first component may be labeled a "type of response" dimension. Both invalid and valid hypothesis variables load positively on the first component. The second dimension may be labeled an "error" dimension. Invalid hypotheses load positively while valids load negatively. From problem 1 to 4 average performance improved from 20.0 trials to 14.8 trials to solution. Despite that, no learning or trial dimension, on which variables would be ordered by problem number, was apparent.

The next step in the analysis involves evaluating relationships between the sets of component variables. A transformation as indicated in (7.5) from the raw variables to the new components has been completed. The next step is to compute the correlation matrix between the sets of component scores. This, of course, could be done by first computing the component scores on the raw data and then computing the resultant correlation matrix. Alternately, one may simply use (7.10) or (7.11). Table 9 reports the correlation matrix between the principal component scores for the two sets of variables. With a small exception in row five, the largest value within each row and column always occurs on the diagonal. Hence the largest correlation for the ith component in the hypothesis set is with the ith component in the pattern set. These correlations are approximately .5 to .6 in absolute value. Most of the off-diagonal correlations are much smaller. Even though no attempt has been made to simplify the cross-relationship pattern, it has been simplified by orthogonalizing the two sets individually.

Table 5. Eigen Analysis Summary for Hypothesis Variables

Component	Eigenvalues	Proportion of Variance	Cumulative Proportion
1	3.44	.43	.43
2	1.78	.22	.65
3	.87	.11	.76
4	.61	.08	.84
5	.59	.07	.91
6	.37	.05	.96
7	.20	.02	.98
8	.13	.02	1.00

Table 6. Eigen Analysis Summary for Pattern Variables

Component	Eigenvalues	Proportion of Variance	Cumulative Proportion
1	3.33	.42	.42
2	1.49	.19	.60
3	.93	.11	.72
4	.73	.09	.81
5	.61	.08	.88
6	.48	.06	.94
7	.29	.04	.98
8	.17	.02	1.00

The next step is to decompose the between-component set correlation matrix, as is done in (7.14) and (7.16), to find weights for the hypothesis set and weights for the pattern set. The first analysis will include all components for both variable sets since this corresponds to the usual canonical analysis. Subsequent analysis will use only two components.

Since eight variables are in each set, eight canonical variate pairs may exist. All eight canonical correlations are numerically nonzero. Table 10 contains information that may be misinterpreted. The first canonical correlation is .72 and its square is .52. Approximately half of the variance in some linear combination of the eight hypothesis variables is accounted for by some linear combination of the pattern variables. The second canonical correlation is .68 and its square is .46. This

Table 7. Component Loadings for Hypothesis Variables

	Component								
Variable	1	2	3	4	5	6	7	8	Plot Symbol
Valid 1	.80	−.24	−.19	−.07	.09	.49	.13	−.01	A
Valid 2	.88	−.24	.04	.01	−.15	.02	.37	−.06	B
Valid 3	.90	−.24	−.06	.06	.00	−.20	−.07	.28	C
Valid 4	.90	−.14	.04	−.06	−.01	−.28	−.19	−.22	D
Invalid 1	.39	.40	.75	.28	.21	.09	−.07	.01	E
Invalid 2	.32	.70	−.07	−.52	.34	−.05	.07	.03	F
Invalid 3	.27	.74	.04	.08	−.61	.07	−.06	.03	G
Invalid 4	.28	.63	−.51	.49	.17	−.03	.03	−.04	H

Table 8. Component Loadings for Pattern Variables

	Component								
Variable	1	2	3	4	5	6	7	8	Plot Symbol
Informative 1	.71	−.33	−.01	.26	−.10	.55	.10	−.02	A
Informative 2	.83	−.22	.09	−.12	.21	.00	−.43	.05	B
Informative 3	.87	−.18	.09	−.18	−.07	−.18	.23	.27	C
Informative 4	.88	−.19	.09	−.17	−.03	−.22	.13	−.30	D
Redundant 1	.51	.20	−.62	.52	−.01	−.22	−.03	.01	E
Redundant 2	.35	.67	−.30	−.33	.42	.20	.10	−.01	F
Redundant 3	.38	.70	.04	−.16	−.57	.07	−.12	−.00	G
Redundant 4	.27	.54	.64	.43	.20	−.06	.04	.00	H

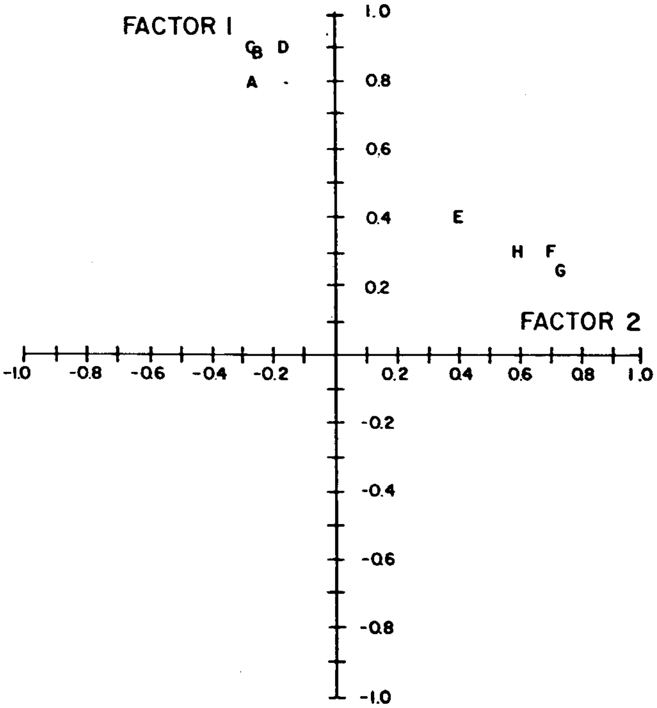

Figure 1. Principal Components Loadings for Hypothesis Variables' Correlation Matrix

indicates that, of the variance left after the first canonical variate has been removed, approximately half is predictable. The successive variate pairs are orthogonal. One computes the first variate pair, then finds residuals from the pair. In turn a subsequent analysis is done within that residual space, and so forth for each successive variate pair. Consequently, the second and later canonical correlations are correlations between residual spaces.

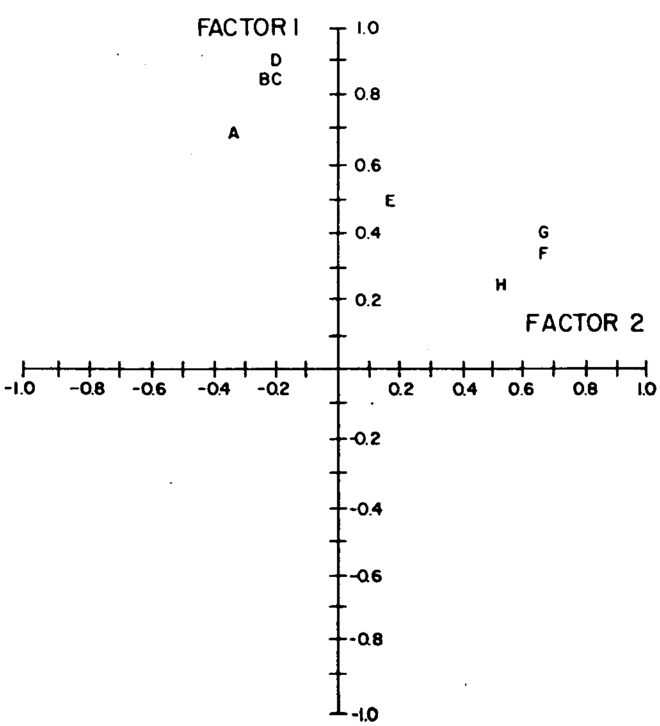

Figure 2. Principal Components Loadings for Pattern Variables' Correlation Matrix

Table 9. Correlation Matrix Between Principal Component Scores

Hypothesis Components	Pattern Components 1	2	3	4	5	6	7	8
1	.60	−.22	.19	.15	.06	.13	−.06	−.02
2	−.16	−.36	.01	−.01	.04	−.08	.03	−.03
3	.20	.02	−.47	−.07	−.00	−.09	.01	.02
4	.10	−.01	.21	−.33	.11	−.13	.01	.03
5	−.14	−.13	.11	−.05	.15	.07	−.18	−.16
6	−.02	−.06	−.09	−.21	−.06	.50	.11	.07
7	.02	−.10	.13	.10	−.15	−.07	.51	.20
8	−.04	−.09	.11	−.05	.14	−.01	−.23	−.58

Since the nature of the example data precludes the use of normal-theory statistics, the usual significance tests have not been reported. From a descriptive perspective, seven canonical variate pairs may be useful. Considering the above discussion concerning the inflation of the apparent importance of the later variates, Wilk's Likelihood Ratio indicates that two or perhaps three canonical variate pairs should be retained. Recall that this number is bounded by the dimensions of the original spaces.

Tables 11 and 12 provide the canonical weights for the component scores. These two matrices are the orthogonal transformation matrices necessary to transform the components into the canonical variates. The transformations are A^* and B^*, given in (8.1). They are the weights necessary to transform the component scores into the new canonical variates. Each may be thought of as an orthogonal rotation, in eight-space, of the original orthogonal structure.

Tables 13 and 14 report the canonical factor loadings for the hypothesis variables and the pattern variables, respectively. The values are also the correlations between the original variables and the canonical variates. The first two canonical factors are plotted for the hypothesis variables in Figure 3 and for the pattern variables in Figure 4. The structure in Figures 3 and 4 is similar to that in Figures 1 and 2 but has some differences. The differences stem from the fact that rigid rotation in eight-dimensional space is not necessarily a rigid rotation in any two-dimensional subspace. The canonical solution works simultaneously with all eight component variables. The figures only present the first two components. Given the original component struc-

Table 10. Canonical Correlation in Summary for Eight Components

Variate Pair	Canonical Correlation	Canonical Squared	Wilks' Λ
1	.72	.52	.04
2	.68	.46	.10
3	.65	.42	.18
4	.60	.36	.31
5	.56	.31	.48
6	.41	.17	.70
7	.40	.16	.84
8	.06	.00	1.00

Table 11. Orthogonal Transformation for Hypothesis Variables for Eight Components

Hypothesis Component	Canonical Variate 1	2	3	4	5	6	7	8
1	.85	.37	−.34	−.02	.06	.13	.05	−.02
2	.00	−.16	.09	−.15	.26	.68	−.50	−.40
3	−.13	.33	−.12	.26	−.65	.21	−.49	.29
4	.22	−.25	−.32	−.05	.12	−.64	−.69	−.05
5	−.06	−.25	.32	.05	.41	.20	−.10	.78
6	.06	−.03	.13	.94	.31	.03	−.01	.06
7	.25	.21	.86	−.13	.07	.05	−.06	.37
8	.38	−.76	.10	.09	−.47	.14	.13	.06

tures and the magnitude of the canonical correlations, the two canonical structures look quite similar. Note that the interpretation would potentially be different from the original factors. Certainly the first component may still be interpreted as a type-of-response dimension. However, the second component is not as clearly defined as it was originally. As with any factor of a correlation matrix, Tables 13 and 14 have elements that are correlation coefficients between the original variables and the components.

Tables 15 and 16 report what is more commonly shown for canonical analysis. This provides the weights for the standardized original hypothesis and pattern variables necessary to produce the new canonical variates. These tables may be computed using the transformations indicated in (6.4) and (2.11). Two transformations are combined, one being from the components to the canonical variates. Note that the weights depend upon the scale of the original variables, if standardized variables are not treated. It must be emphasized that the results in Tables 10, 15, and 16 are exactly what is reported by direct calculation of the canonical solution. Although they are only a small fraction of the information, they are usually a major portion of most computer programs' output and written reports.

Different results follow if only the first two components are retained for each set. For this approach, use the first two columns of Tables 7 and 8 as F_x and F_y, respectively. Then $R_{x \cdot y \cdot}$ becomes the first two rows of the first two columns of Table 9. Calculation of the canonical solution on that 2×2 matrix gives the rotations, which are the canonical weights. Tables 17 and 18

Table 12. Orthogonal Transformation for Pattern Variables for Eight Components

Pattern Component	Canonical Variate 1	2	3	4	5	6	7	8
1	.70	.54	−.28	.07	−.29	−.21	−.12	−.03
2	−.35	.09	−.03	.03	−.29	−.75	.44	.18
3	.46	−.31	.09	−.38	.57	−.39	.20	−.10
4	.07	.27	.03	−.45	−.11	.46	.68	.19
5	.11	−.27	−.29	−.08	.00	.02	−.21	.88
6	.14	.03	−.08	.78	.38	.12	.44	.13
7	.01	.43	.78	.04	.18	−.09	−.19	.35
8	.38	.53	.45	.18	−.56	.05	.15	−.01

Table 13. Canonical Factor Loadings for Hypothesis Variables for Eight Components

Variable	Canonical Factor 1	2	3	4	5	6	7	8	Plot Symbol
Valid 1	.66	−.33	.39	−.54	.06	.04	−.02	.12	A
Valid 2	.83	−.52	−.05	.01	.02	.07	−.18	.08	B
Valid 3	.82	−.19	.39	.12	−.25	−.04	−.23	.11	C
Valid 4	.64	−.47	.53	.27	.05	.08	−.08	−.05	D
Invalid 1	.36	.17	.07	.17	.69	−.39	−.35	−.22	E
Invalid 2	.38	.17	.12	.19	.34	−.27	.79	.01	F
Invalid 3	.29	−.07	−.13	.03	−.00	−.36	.32	−.82	G
Invalid 4	.04	−.27	.12	−.01	−.13	−.91	.24	.06	H

Table 14. Canonical Factor Loadings for Pattern Variables for Eight Components

Variable	Canonical Factor 1	2	3	4	5	6	7	8	Plot Symbol
Informative 1	.66	−.24	−.35	.57	−.12	.05	−.18	−.00	A
Informative 2	.71	−.65	.21	.01	.08	−.03	−.13	.09	B
Informative 3	.81	−.21	−.28	−.16	.43	.01	−.06	.08	C
Informative 4	.59	−.56	−.45	−.28	.10	−.03	−.18	.08	D
Redundant 1	−.05	−.34	−.18	.34	.58	−.28	−.55	.03	E
Redundant 2	−.12	−.50	−.12	.19	.44	−.29	.61	.19	F
Redundant 3	.12	−.19	−.20	−.03	.26	−.54	.27	−.70	G
Redundant 4	.24	.09	−.01	−.05	−.15	−.91	.07	.26	H

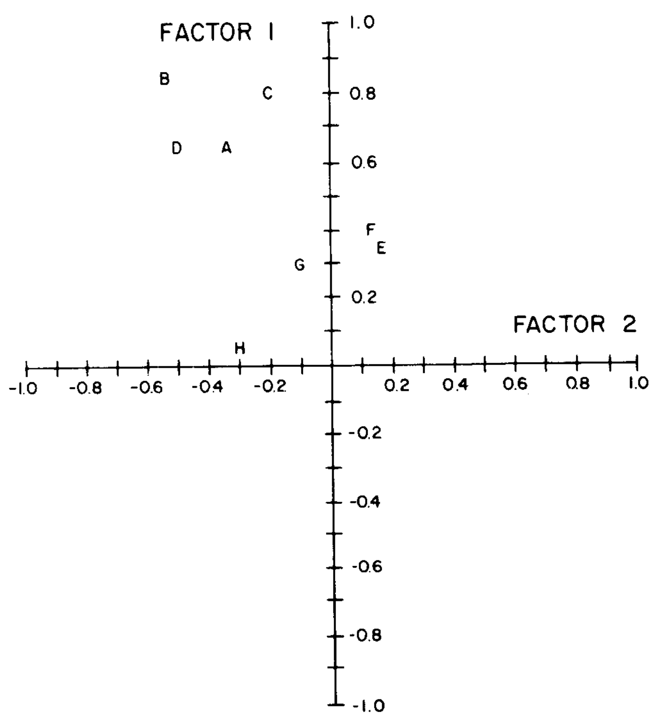

Figure 3. Canonical Factor Loadings for Hypothesis Variables' Correlation Matrix

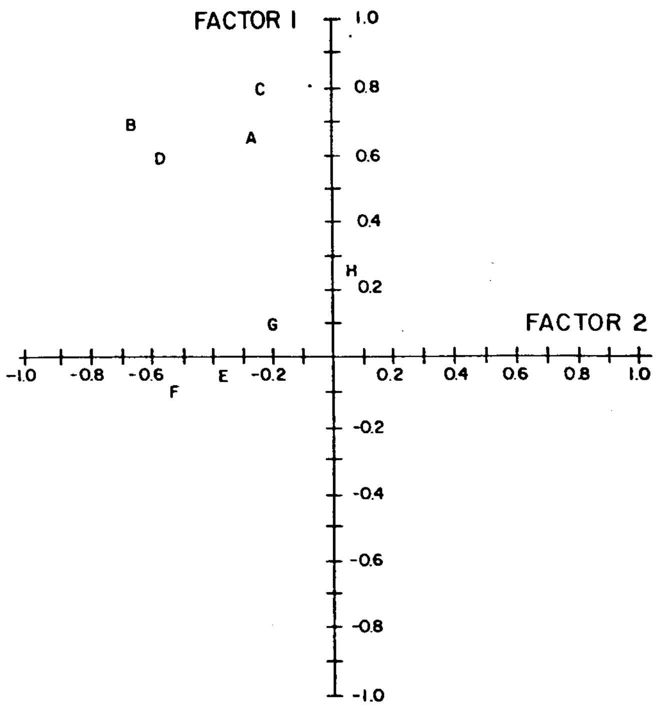

Figure 4. Canonical Factor Loadings for Pattern Variables' Correlation Matrix

Table 15. Canonical Weights for Standardized Original Hypothesis Variables for Eight Components

	Canonical Variate							
Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Valid 1	.07	−.31	.56	−1.35	.31	.07	.10	−.10
Valid 2	.51	−.90	−1.51	.21	.28	.09	.03	.47
Valid 3	.89	1.49	.24	.24	−1.26	−.29	−.42	−.12
Valid 4	−.65	−1.17	1.15	.74	.50	.40	.14	−.40
Invalid 1	.11	.35	.05	.02	.75	−.37	−.60	−.08
Invalid 2	.38	.40	−.09	.26	.29	.10	.88	.42
Invalid 3	.17	.05	−.11	−.20	−.33	−.01	.12	−1.11
Invalid 4	−.28	−.53	.04	.01	−.15	−.91	−.02	.34

Table 16. Canonical Weights for Standardized Original Pattern Variables for Eight Components

	Canonical Variate							
Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Informative 1	.32	.09	−.52	1.02	−.48	.11	.14	.08
Informative 2	.37	−.71	1.34	.04	−.13	−.04	−.14	−.20
Informative 3	.86	1.07	.08	−.17	1.21	.22	.29	.23
Informative 4	−.45	−.90	−1.18	−.83	−.75	.00	−.30	.15
Redundant 1	−.34	−.11	.01	.25	.58	−.25	−.80	.06
Redundant 2	−.24	−.39	−.19	.30	.27	.04	.81	.49
Redundant 3	.07	−.08	−.00	−.04	.04	−.30	.14	−1.11
Redundant 4	.17	.28	.02	−.06	−.27	−.84	−.06	.48

Table 17. Orthogonal Transformation for Hypothesis Variables for Two Components

	Canonical Variate	
Hypothesis Component	*1*	*2*
1	−.9987	.0504
2	−.0504	−.9987

Table 18. Orthogonal Transformation for Pattern Variables for Two Components

	Canonical Variate	
Pattern Component	*1*	*2*
1	.9471	.3210
2	.3410	−.9471

report these for this solution. The associated canonical correlations are .64 and .39. Tables 19 and 20 give the correlations between the original sets of variables and the two corresponding canonical variates. These may be thought of as the canonical factor loadings. Tables 21 and 22 give the weights necessary to compute the two canonical variates directly from the original variable sets.

The canonical factors are orthogonal rotations of the first two factors in Tables 7 and 8, as well as what are plotted in Figures 1 and 2. Interpreting Tables 17 and 18 as geometric rotation matrices indicates that finding this canonical solution amounts to rotating the hypothesis components approximately 3 degrees and rotating the pattern components approximately 19 degrees. Hence canonical factor plots should display the same relationships as in Figures 1 and 2.

The weights are simple and easy to understand with this approach. This follows from the correspondence with the component analysis. The canonical analysis weights for all eight components are less appealing. For example, the second canonical variates put heavy weight on the eighth components, which are essentially noise.

Table 19. Canonical Factor Loadings for Hypothesis Variables for Two Components

	Canonical Factor	
Variable	*1*	*2*
Valid 1	.81	−.20
Valid 2	.89	−.20
Valid 3	.91	−.20
Valid 4	.91	−.10
Invalid 1	.37	.42
Invalid 2	.29	.71
Invalid 3	.23	.75
Invalid 4	.24	.64

Table 20. Canonical Factor Loadings for Pattern Variables for Two Components

Variable	1	2
Informative 1	.76	−.09
Informative 2	.86	.06
Informative 3	.89	.11
Informative 4	.89	.11
Redundant 1	.42	.35
Redundant 2	.11	.75
Redundant 3	.13	.78
Redundant 4	.08	.59

12. CONCLUSIONS

Choosing to retain only meaningful components, as in the example, leads to simpler interpretation, one more consonant with the within-set structure. The analytic results and the example analysis provided a sequential construction of the model for canonical correlation based on component scores given in (8.1). First, the sets were decomposed (one at a time) into principal components. Next, the between-component set correlation matrix was computed. Finally, that matrix was decomposed, providing maximally correlated linear combinations of the two sets of components. Reporting any canonical analysis accurately will require some detail since so much is implicitly being done. Making the implicit properties explicit should help understanding and exposition.

The example analysis has certain limitations. First, no form of significance testing was presented. The stepwise nature of the analysis makes the usual test statistics of questionable value. Any sequential descriptive analysis such as a stepwise regression, factor analysis, or the canonical analysis reported here, demands some form of replication. Another set of data on the same variables is available. If the purpose of this analysis were psychological modeling rather than statistical demonstration, it would be necessary to analyze the second data set to demonstrate the replicability of the conclusions. Many other split-sample techniques are available. Since canonical correlation usually involves stepwise data analysis, it is important to remember the general applicability and need for such an approach.

Table 21. Canonical Weights for Standardized Original Hypothesis Variables for Two Components

	Canonical Variate	
Variable	1	2
Valid 1	.22	−.14
Valid 2	.25	−.15
Valid 3	.25	−.15
Valid 4	.26	−.09
Invalid 1	.12	.22
Invalid 2	.11	.39
Invalid 3	.10	.41
Invalid 4	.10	.35

Table 22. Canonical Weights for Standardized Original Pattern Variables for Two Components

	Canonical Variate	
Variable	1	2
Informative 1	.27	−.14
Informative 2	.28	−.06
Informative 3	.29	−.03
Informative 4	.29	−.03
Redundant 1	.10	.17
Redundant 2	−.05	.46
Redundant 3	−.04	.48
Redundant 4	−.04	.37

Canonical correlation stands outside the set of commonly used linear-models analyses. Despite that, understanding canonical correlation helps in understanding many commonly used linear-models analyses. As mentioned earlier, most of the more common techniques are special cases of canonical correlation. McKeon (1964) and Bartlett (1948) each discussed a number of (distinct) examples. An important feature of canonical correlation is the symmetry as to which set is "X" and which is "Y." In contrast, multivariate multiple regression (including its special cases univariate and multiple regression), MANOVA (and its special cases), and discriminant analysis all are not symmetric. Yet the canonical solution encompasses them all.

[Received January 1981. Revised November 1981.]

REFERENCES

ANDERSON, T.W. (1958), *An Introduction to Multivariate Statistical Analysis,* New York: John Wiley and Sons.

BARTLETT, M.S. (1948), "Internal and External Factor Analysis," *British Journal of Psychology, Statistics Section,* 1, 73–81.

BROWNE, M.W. (1979), "The Maximum-likelihood Solution in Inter-battery Factory Analysis," *British Journal of Mathematical and Statistical Psychology,* 32, 75–86.

CRAMER, E.M., and NICEWANDER, W.A. (1979), "Some Symmetric Invariant Measures of Multivariate Association," *Psychometrika,* 44, 43–54.

FINN, J.D. (1974), *A General Model for Multivariate Analysis,* New York: Holt, Rinehart and Winston.

GITTENS, R. (1979), "Ecological Applications of Canonical Analysis," in *Multivariate Methods in Ecological Work,* eds. L. Orloci and C.R. Rao, Fairland, Md.: International Co-op Publication.

GOLUB, G.H. (1969), "Matrix Decompositions and Statistical Calculations," in *Statistical Computation,* eds. R.C. Milton and J.A. Nelder, New York: Academic Press.

HARMAN, J. (1967), *Modern Factor Analysis,* Chicago: University of Chicago Press.

HARRIS, R.J. (1975), *A Primer of Multivariate Statistics,* New York: Academic Press.

HOTELLING, H. (1935), "The Most Predictable Criterion," *Journal of Education Psychology,* 26, 139–142.

——— (1936), "Relations Between Two Sets of Variates," *Biometrika,* 28, 321–377.

JOHNSON, E.S. (1971), "Objective Identification of Strategy on a Selection Concept Learning Task," *Journal of Experimental Psychology,* 90, 167–196 (monograph).

KHATRI, C.G. (1976), "A Note on Multiple and Canonical Correlation for a Singular Covariance Matrix," *Psychometrika,* 41, 465–470.

LEVINE, M.S. (1977), *Canonical Analysis and Factor Comparison,* Beverly Hills, Calif.: Sage Publications (monograph in Qualitative Applications in the Social Sciences series).

McKEON, J.J. (1964), "Canonical Analysis: Some Relations Between Canonical Correlation, Factor Analysis, Discriminant Function Analysis, and Scaling Theory," *Psychometric Monograph,* 13.

MORRISON, D.F. (1967), *Multivariate Statistical Methods,* New York: McGraw-Hill.

MULLER, K.E.; HOSKING, J.D.; and HELMS, R.W. (1979), "Using LINMOD to Adjust for Treatment Effects When Analyzing the Covariance Matrix," *Proceedings of the Statistical Computing Section, American Statistical Association,* 136–140.

OLSON, C.L. (1976), "On Choosing a Test Statistic in Multivariate Analysis," *Psychological Bulletin,* 83, 579–586.

PRESS, S.J. (1972), *Applied Multivariate Analysis,* New York: Holt, Rinehart, and Winston.

RAO, C.R. (1973), *Linear Statistical Inference and Its Applications* (2nd ed.), New York: John Wiley.

SEARLE, S.R. (1971), *Linear Models,* New York: John Wiley.

TATSUOKA, M.M. (1971), *Multivariate Analysis: Techniques for Educational and Psychological Research,* New York: John Wiley.