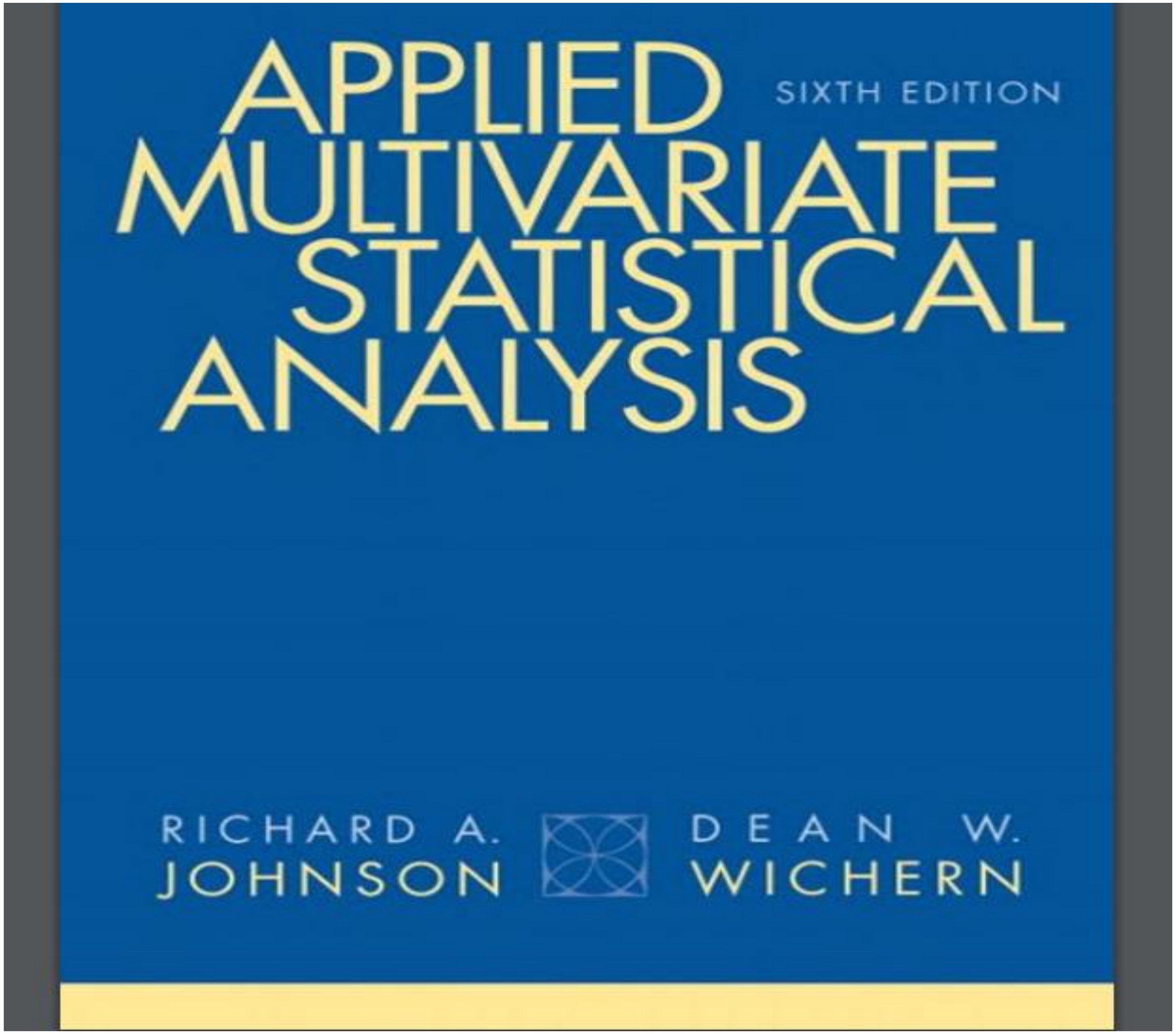
Advanced Multivariate Methods



Canonical Correlation Analysis

10.1 Introduction

Canonical correlation analysis seeks to identify and quantify the associations between *two sets* of variables. H. Hotelling ([5], [6]), who initially developed the technique, provided the example of relating arithmetic speed and arithmetic power to reading speed and reading power. (See Exercise 10.9.) Other examples include relating governmental policy variables with economic goal variables and relating college "performance" variables with precollege "achievement" variables.

Canonical correlation analysis focuses on the correlation between a *linear combination* of the variables in one set and a *linear combination* of the variables in another set. The idea is first to determine the pair of linear combinations having the largest correlation. Next, we determine the pair of linear combinations having the largest correlation among all pairs uncorrelated with the initially selected pair, and so on. The pairs of linear combinations are called the *canonical variables*, and their correlations are called *canonical correlations*.

The canonical correlations measure the strength of association between the two sets of variables. The maximization aspect of the technique represents an attempt to concentrate a high-dimensional relationship between two sets of variables into a few pairs of canonical variables.

Canonical Variates and Canonical Correlations: Measures of Association Between Two Groups of Variables

10.2 Canonical Variates and Canonical Correlations

We shall be interested in measures of association between two groups of variables. The first group, of p variables, is represented by the $(p \times 1)$ random vector $\mathbf{X}^{(1)}$. The second group, of q variables, is represented by the $(q \times 1)$ random vector $\mathbf{X}^{(2)}$. We assume, in the theoretical development, that $\mathbf{X}^{(1)}$ represents the *smaller* set, so that $p \leq q$.

For the random vectors $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, let

$$\begin{aligned} E(\mathbf{X}^{(1)}) &= \boldsymbol{\mu}^{(1)}; \qquad \operatorname{Cov}(\mathbf{X}^{(1)}) = \boldsymbol{\Sigma}_{11} \\ E(\mathbf{X}^{(2)}) &= \boldsymbol{\mu}^{(2)}; \qquad \operatorname{Cov}(\mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{22} \\ \operatorname{Cov}(\mathbf{X}^{(1)}, &\mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' \end{aligned} \tag{10-1}$$

It will be convenient to consider $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ jointly, so, using results (2-38) through (2-40) and (10-1), we find that the random vector

$$\underset{((p+q)\times 1)}{\mathbf{X}} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \hline \mathbf{X}^{(2)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \hline X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{bmatrix} \tag{10-2}$$

Measures of Association Between Two Groups of Variables: Mean Vector and Covariance Matrix

$$\underset{((p+q)\times 1)}{\mathbf{X}} = \left[\begin{array}{c}\mathbf{X}^{(1)} \\ \hdashline \mathbf{X}^{(2)}\end{array}\right] = \left[\begin{array}{c} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \hdashline X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{array}\right] \tag{10-2}$$

has mean vector

$$\underset{((p+q)\times 1)}{\boldsymbol{\mu}} = E(\mathbf{X}) = \left[\begin{array}{c} E(\mathbf{X}^{(1)}) \\ \hdashline E(\mathbf{X}^{(2)}) \end{array}\right] = \left[\begin{array}{c} \boldsymbol{\mu}^{(1)} \\ \hdashline \boldsymbol{\mu}^{(2)} \end{array}\right] \tag{10-3}$$

and covariance matrix

$$\begin{aligned}
\underset{(p+q)\times(p+q)}{\boldsymbol{\Sigma}} &= E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' \\
&= \left[\begin{array}{c:c} E(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})' & E(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})' \\ \hdashline E(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})' & E(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})' \end{array}\right] \\
&= \left[\begin{array}{c:c} \underset{(p\times p)}{\boldsymbol{\Sigma}_{11}} & \underset{(p\times q)}{\boldsymbol{\Sigma}_{12}} \\ \hdashline \underset{(q\times p)}{\boldsymbol{\Sigma}_{21}} & \underset{(q\times q)}{\boldsymbol{\Sigma}_{22}} \end{array}\right]
\end{aligned} \tag{10-4}$$

Covariance Between Pairs of Variables From Different Sets Combined in ∑

The covariances between pairs of variables from different sets—one variable from $\mathbf{X}^{(1)}$, one variable from $\mathbf{X}^{(2)}$—are contained in $\boldsymbol{\Sigma}_{12}$ or, equivalently, in $\boldsymbol{\Sigma}_{21}$. That is, the pq elements of $\boldsymbol{\Sigma}_{12}$ measure the association between the two sets. When p and q are relatively large, interpreting the elements of $\boldsymbol{\Sigma}_{12}$ collectively is ordinarily hopeless. Moreover, it is often linear combinations of variables that are interesting and useful for predictive or comparative purposes. The main task of canonical correlation analysis is to summarize the associations between the $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ sets in terms of a *few* carefully chosen covariances (or correlations) rather than the pq covariances in $\boldsymbol{\Sigma}_{12}$.

Linear combinations provide simple summary measures of a set of variables. Set

$$\begin{aligned} U &= \mathbf{a}'\mathbf{X}^{(1)} \\ V &= \mathbf{b}'\mathbf{X}^{(2)} \end{aligned} \tag{10-5}$$

for some pair of coefficient vectors $\mathbf{a}$ and $\mathbf{b}$. Then, using (10-5) and (2–45), we obtain

$$\begin{aligned} \operatorname{Var}(U) &= \mathbf{a}' \operatorname{Cov}(\mathbf{X}^{(1)})\,\mathbf{a} = \mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a} \\ \operatorname{Var}(V) &= \mathbf{b}' \operatorname{Cov}(\mathbf{X}^{(2)})\,\mathbf{b} = \mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b} \\ \operatorname{Cov}(U,V) &= \mathbf{a}' \operatorname{Cov}(\mathbf{X}^{(1)},\mathbf{X}^{(2)})\,\mathbf{b} = \mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b} \end{aligned} \tag{10-6}$$

We shall seek coefficient vectors $\mathbf{a}$ and $\mathbf{b}$ such that

$$\operatorname{Corr}(U,V) = \frac{\mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a}}\,\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b}}} \tag{10-7}$$

is as large as possible.

Correlation Between the kth Pair of Canonical Variables is Called the kth Canonical Correlation

We define the following:

The *first pair of canonical variables*, or *first canonical variate pair*, is the pair of linear combinations U_1, V_1 having unit variances, which maximize the correlation (10-7);

The *second pair of canonical variables*, or *second canonical variate pair*, is the pair of linear combinations U_2, V_2 having unit variances, which maximize the correlation (10-7) among all choices that are uncorrelated with the first pair of canonical variables.

At the kth step,

The *kth pair of canonical variables*, or *kth canonical variate pair*, is the pair of linear combinations U_k, V_k having unit variances, which maximize the correlation (10-7) among all choices uncorrelated with the previous $k - 1$ canonical variable pairs.

The correlation between the kth pair of canonical variables is called the *kth canonical correlation*.

The following result gives the necessary details for obtaining the canonical variables and their correlations.

Maximizing Correlation Between Two Sets of Variables

Result 10.1. Suppose $p \leq q$ and let the random vectors $\underset{(p\times 1)}{\mathbf{X}^{(1)}}$ and $\underset{(q\times 1)}{\mathbf{X}^{(2)}}$ have $\text{Cov}(\mathbf{X}^{(1)}) = \underset{(p\times p)}{\mathbf{\Sigma}_{11}}$, $\text{Cov}(\mathbf{X}^{(2)}) = \underset{(q\times q)}{\mathbf{\Sigma}_{22}}$ and $\text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \underset{(p\times q)}{\mathbf{\Sigma}_{12}}$, where $\mathbf{\Sigma}$ has full rank. For coefficient vectors $\underset{(p\times 1)}{\mathbf{a}}$ and $\underset{(q\times 1)}{\mathbf{b}}$, form the linear combinations $U = \mathbf{a}'\mathbf{X}^{(1)}$ and $V = \mathbf{b}'\mathbf{X}^{(2)}$. Then

$$\max_{\mathbf{a},\mathbf{b}} \text{Corr}(U, V) = \rho_1^*$$

attained by the linear combinations (first canonical variate pair)

$$U_1 = \underbrace{\mathbf{e}_1'\mathbf{\Sigma}_{11}^{-1/2}}_{\mathbf{a}_1'}\mathbf{X}^{(1)} \quad \text{and} \quad V_1 = \underbrace{\mathbf{f}_1'\mathbf{\Sigma}_{22}^{-1/2}}_{\mathbf{b}_1'}\mathbf{X}^{(2)}$$

Pair of Linear Combinations Premise: Variance = 1; Covariance and Correlation = 0

The kth pair of canonical variates, $k = 2, 3, \ldots, p$,

$$U_k = \mathbf{e}_k' \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{X}^{(1)} \qquad V_k = \mathbf{f}_k' \boldsymbol{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}$$

maximizes

$$\operatorname{Corr}(U_k, V_k) = \rho_k^*$$

among those linear combinations uncorrelated with the preceding $1, 2, \ldots, k-1$ canonical variables.

Here $\rho_1^{*2} \geq \rho_2^{*2} \geq \cdots \geq \rho_p^{*2}$ are the eigenvalues of $\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}$, and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$ are the associated $(p \times 1)$ eigenvectors. [The quantities $\rho_1^{*2}, \rho_2^{*2}, \ldots, \rho_p^{*2}$ are also the p largest eigenvalues of the matrix $\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2}$ with corresponding $(q \times 1)$ eigenvectors $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_p$. Each $\mathbf{f}_i$ is proportional to $\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{e}_i$.]

The canonical variates have the properties

$$\begin{aligned}
\operatorname{Var}(U_k) &= \operatorname{Var}(V_k) = 1 \\
\operatorname{Cov}(U_k, U_\ell) &= \operatorname{Corr}(U_k, U_\ell) = 0 \quad k \neq \ell \\
\operatorname{Cov}(V_k, V_\ell) &= \operatorname{Corr}(V_k, V_\ell) = 0 \quad k \neq \ell \\
\operatorname{Cov}(U_k, V_\ell) &= \operatorname{Corr}(U_k, V_\ell) = 0 \quad k \neq \ell
\end{aligned}$$

for $k, \ell = 1, 2, \ldots, p$.

Calculating Canonical Variables and Canonical Correlations

Example 10.1 (Calculating canonical variates and canonical correlations for standardized variables) Suppose $\mathbf{Z}^{(1)} = [Z_1^{(1)}, Z_2^{(1)}]'$ are standardized variables and $\mathbf{Z}^{(2)} = [Z_1^{(2)}, Z_2^{(2)}]'$ are also standardized variables. Let $\mathbf{Z} = [\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}]'$ and

$$\text{Cov}(\mathbf{Z}) = \left[\begin{array}{c|c} \boldsymbol{\rho}_{11} & \boldsymbol{\rho}_{12} \\ \hline \boldsymbol{\rho}_{21} & \boldsymbol{\rho}_{22} \end{array}\right] = \left[\begin{array}{cc|cc} 1.0 & .4 & .5 & .6 \\ .4 & 1.0 & .3 & .4 \\ \hline .5 & .3 & 1.0 & .2 \\ .6 & .4 & .2 & 1.0 \end{array}\right]$$

Then

$$\boldsymbol{\rho}_{11}^{-1/2} = \begin{bmatrix} 1.0681 & -.2229 \\ -.2229 & 1.0681 \end{bmatrix}$$

$$\boldsymbol{\rho}_{22}^{-1} = \begin{bmatrix} 1.0417 & -.2083 \\ -.2083 & 1.0417 \end{bmatrix}$$

and

$$\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2} = \begin{bmatrix} .4371 & .2178 \\ .2178 & .1096 \end{bmatrix}$$

Calculating Canonical Variables and Canonical Correlations

The eigenvalues, ρ_1^{*2}, ρ_2^{*2}, of $\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}$ are obtained from

$$0 = \begin{vmatrix} .4371 - \lambda & .2178 \\ .2178 & .1096 - \lambda \end{vmatrix} = (.4371 - \lambda)(.1096 - \lambda) - (2.178)^2$$

$$= \lambda^2 - .5467\lambda + .0005$$

yielding $\rho_1^{*2} = .5458$ and $\rho_2^{*2} = .0009$. The eigenvector $\mathbf{e}_1$ follows from the vector equation

$$\begin{bmatrix} .4371 & .2178 \\ .2178 & .1096 \end{bmatrix} \mathbf{e}_1 = (.5458)\mathbf{e}_1$$

Thus, $\mathbf{e}_1' = [.8947, .4466]$ and

$$\mathbf{a}_1 = \boldsymbol{\rho}_{11}^{-1/2}\mathbf{e}_1 = \begin{bmatrix} .8561 \\ .2776 \end{bmatrix}$$

From Result 10.1, $\mathbf{f}_1 \propto \boldsymbol{\rho}_{22}^{-1/2}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}\mathbf{e}_1$ and $\mathbf{b}_1 = \boldsymbol{\rho}_{22}^{-1/2}\mathbf{f}_1$. Consequently,

$$\mathbf{b}_1 \propto \boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\mathbf{a}_1 = \begin{bmatrix} .3959 & .2292 \\ .5209 & .3542 \end{bmatrix} \begin{bmatrix} .8561 \\ .2776 \end{bmatrix} = \begin{bmatrix} .4026 \\ .5443 \end{bmatrix}$$

Note: First Pair of Canonical Variables and Their Canonical Correlation

We must scale $\mathbf{b}_1$ so that

$$\operatorname{Var}(V_1) = \operatorname{Var}(\mathbf{b}_1'\mathbf{Z}^{(2)}) = \mathbf{b}_1'\boldsymbol{\rho}_{22}\mathbf{b}_1 = 1$$

The vector $[.4026, .5443]'$ gives

$$[.4026, .5443]\begin{bmatrix} 1.0 & .2 \\ .2 & 1.0 \end{bmatrix}\begin{bmatrix} .4026 \\ .5443 \end{bmatrix} = .5460$$

Using $\sqrt{.5460} = .7389$, we take

$$\mathbf{b}_1 = \frac{1}{.7389}\begin{bmatrix} .4026 \\ .5443 \end{bmatrix} = \begin{bmatrix} .5448 \\ .7366 \end{bmatrix}$$

The first pair of canonical variates is

$$U_1 = \mathbf{a}_1'\mathbf{Z}^{(1)} = .86Z_1^{(1)} + .28Z_2^{(1)}$$
$$V_1 = \mathbf{b}_1'\mathbf{Z}^{(2)} = .54Z_1^{(2)} + .74Z_2^{(2)}$$

and their canonical correlation is

$$\rho_1^* = \sqrt{\rho_1^{*2}} = \sqrt{.5458} = .74$$

This is the largest correlation possible between linear combinations of variables from the $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ sets.

Note: Second Pair of Canonical Variables and Their Canonical Correlation

The second canonical correlation, $\rho_2^* = \sqrt{.0009} = .03$, is very small, and consequently, the second pair of canonical variates, although uncorrelated with members of the first pair, conveys very little information about the association between sets. (The calculation of the second pair of canonical variates is considered in Exercise 10.5.)

We note that U_1 and V_1, apart from a scale change, are not much different from the pair

$$\widetilde{U}_1 = \mathbf{a}'\mathbf{Z}^{(1)} = [3, 1]\begin{bmatrix} Z_1^{(1)} \\ Z_2^{(1)} \end{bmatrix} = 3Z_1^{(1)} + Z_2^{(1)}$$

$$\widetilde{V}_1 = \mathbf{b}'\mathbf{Z}^{(2)} = [1, 1]\begin{bmatrix} Z_1^{(2)} \\ Z_2^{(2)} \end{bmatrix} = Z_1^{(2)} + Z_2^{(2)}$$

For these variates,

$$\mathrm{Var}(\widetilde{U}_1) = \mathbf{a}'\boldsymbol{\rho}_{11}\mathbf{a} = 12.4$$

$$\mathrm{Var}(\widetilde{V}_1) = \mathbf{b}'\boldsymbol{\rho}_{22}\mathbf{b} = 2.4$$

$$\mathrm{Cov}(\widetilde{U}_1, \widetilde{V}_1) = \mathbf{a}'\boldsymbol{\rho}_{12}\mathbf{b} = 4.0$$

and

$$\mathrm{Corr}(\widetilde{U}_1, \widetilde{V}_1) = \frac{4.0}{\sqrt{12.4}\,\sqrt{2.4}} = .73$$

The correlation between the rather simple and, perhaps, easily interpretable linear combinations $\widetilde{U}_1, \widetilde{V}_1$ is almost the maximum value $\rho_1^* = .74$. ∎

Conditions: Symmetric Matrices, Eigenvalues Determine Canonical Coefficients

The procedure for obtaining the canonical variates presented in Result 10.1 has certain advantages. The symmetric matrices, whose eigenvectors determine the canonical coefficients, are readily handled by computer routines. Moreover, writing the coefficient vectors as $\mathbf{a}_k = \mathbf{\Sigma}_{11}^{-1/2}\mathbf{e}_k$ and $\mathbf{b}_k = \mathbf{\Sigma}_{22}^{-1/2}\mathbf{f}_k$ facilitates analytic descriptions and their geometric interpretations. To ease the computational burden, many people prefer to get the canonical correlations from the eigenvalue equation

$$|\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} - \rho^{*2}\mathbf{I}| = 0 \qquad (10\text{-}10)$$

The coefficient vectors **a** and **b** follow directly from the eigenvector equations

$$\begin{aligned} \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{a} &= \rho^{*2}\mathbf{a} \\ \mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{b} &= \rho^{*2}\mathbf{b} \end{aligned} \qquad (10\text{-}11)$$

The matrices $\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$ and $\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$ are, in general, not symmetric. (See Exercise 10.4 for more details.)

Interpreting Canonical Variables

10.3 Interpreting the Population Canonical Variables

Canonical variables are, in general, artificial. That is, they have no physical meaning. If the original variables $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are used, the canonical coefficients $\mathbf{a}$ and $\mathbf{b}$ have units proportional to those of the $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ sets. If the original variables are standardized to have zero means and unit variances, the canonical coefficients have no units of measurement, and they must be interpreted in terms of the standardized variables.

Result 10.1 gives the technical definitions of the canonical variables and canonical correlations. In this section, we concentrate on interpreting these quantities.

Identifying the Canonical Variables

Even though the canonical variables are artificial, they can often be "identified" in terms of the subject-matter variables. Many times this identification is aided by computing the correlations between the canonical variates and the original variables. These correlations, however, must be interpreted with caution. They provide only univariate information, in the sense that they do not indicate how the original variables contribute *jointly* to the canonical analyses. (See, for example, [11].)

Strategy: Assess Contributions Original Variables Directly from Standardized Coefficients

For this reason, many investigators prefer to assess the contributions of the original variables directly from the standardized coefficients (10-8).

Let $\underset{(p\times p)}{\mathbf{A}} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p]'$ and $\underset{(q\times q)}{\mathbf{B}} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_q]'$, so that the vectors of canonical variables are

$$\underset{(p\times 1)}{\mathbf{U}} = \mathbf{A}\mathbf{X}^{(1)} \qquad \underset{(q\times 1)}{\mathbf{V}} = \mathbf{B}\mathbf{X}^{(2)} \tag{10-12}$$

where we are primarily interested in the first p canonical variables in $\mathbf{V}$. Then

$$\operatorname{Cov}(\mathbf{U}, \mathbf{X}^{(1)}) = \operatorname{Cov}(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{X}^{(1)}) = \mathbf{A}\boldsymbol{\Sigma}_{11} \tag{10-13}$$

Because $\operatorname{Var}(U_i) = 1$, $\operatorname{Corr}(U_i, X_k^{(1)})$ is obtained by dividing $\operatorname{Cov}(U_i, X_k^{(1)})$ by $\sqrt{\operatorname{Var}(X_k^{(1)})} = \sigma_{kk}^{1/2}$. Equivalently, $\operatorname{Corr}(U_i, X_k^{(1)}) = \operatorname{Cov}(U_i, \sigma_{kk}^{-1/2} X_k^{(1)})$. Introducing the $(p \times p)$ diagonal matrix $\mathbf{V}_{11}^{-1/2}$ with kth diagonal element $\sigma_{kk}^{-1/2}$, we have, in matrix terms,

$$\begin{aligned}\underset{(p\times p)}{\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(1)}}} = \operatorname{Corr}(\mathbf{U}, \mathbf{X}^{(1)}) = \operatorname{Cov}(\mathbf{U}, \mathbf{V}_{11}^{-1/2}\mathbf{X}^{(1)}) &= \operatorname{Cov}(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{V}_{11}^{-1/2}\mathbf{X}^{(1)}) \\ &= \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2}\end{aligned}$$

Canonical Variables Derived from Standardized Variables Interpreted by Computing the Correlations

Similar calculations for the pairs $(\mathbf{U}, \mathbf{X}^{(2)})$, $(\mathbf{V}, \mathbf{X}^{(2)})$ and $(\mathbf{V}, \mathbf{X}^{(1)})$ yield

$$\underset{(p\times p)}{\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(1)}}} = \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2} \qquad \underset{(q\times q)}{\boldsymbol{\rho}_{\mathbf{V},\mathbf{X}^{(2)}}} = \mathbf{B}\boldsymbol{\Sigma}_{22}\mathbf{V}_{22}^{-1/2}$$
$$\underset{(p\times q)}{\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(2)}}} = \mathbf{A}\boldsymbol{\Sigma}_{12}\mathbf{V}_{22}^{-1/2} \qquad \underset{(q\times p)}{\boldsymbol{\rho}_{\mathbf{V},\mathbf{X}^{(1)}}} = \mathbf{B}\boldsymbol{\Sigma}_{21}\mathbf{V}_{11}^{-1/2} \tag{10-14}$$

where $\mathbf{V}_{22}^{-1/2}$ is the $(q \times q)$ diagonal matrix with ith diagonal element $[\mathrm{Var}(X_i^{(2)})]$.

Canonical variables derived from standardized variables are sometimes interpreted by computing the correlations. Thus,

$$\boldsymbol{\rho}_{\mathbf{U},\mathbf{Z}^{(1)}} = \mathbf{A}_{\mathbf{z}}\boldsymbol{\rho}_{11} \qquad \boldsymbol{\rho}_{\mathbf{V},\mathbf{Z}^{(2)}} = \mathbf{B}_{\mathbf{z}}\boldsymbol{\rho}_{22}$$
$$\boldsymbol{\rho}_{\mathbf{U},\mathbf{Z}^{(2)}} = \mathbf{A}_{\mathbf{z}}\boldsymbol{\rho}_{12} \qquad \boldsymbol{\rho}_{\mathbf{V},\mathbf{Z}^{(1)}} = \mathbf{B}_{\mathbf{z}}\boldsymbol{\rho}_{21} \tag{10-15}$$

where $\underset{(p\times p)}{\mathbf{A}_{\mathbf{z}}}$ and $\underset{(q\times q)}{\mathbf{B}_{\mathbf{z}}}$ are the matrices whose rows contain the canonical coefficients for the $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ sets, respectively. The correlations in the matrices displayed in (10–15) have the *same* numerical values as those appearing in (10–14); that is, $\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(1)}} = \boldsymbol{\rho}_{\mathbf{U},\mathbf{Z}^{(1)}}$, and so forth. This follows because, for example, $\boldsymbol{\rho}_{\mathbf{U},\mathbf{X}^{(1)}} = \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2} = \mathbf{A}\mathbf{V}_{11}^{1/2}\mathbf{V}_{11}^{-1/2}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2} = \mathbf{A}_{\mathbf{z}}\boldsymbol{\rho}_{11} = \boldsymbol{\rho}_{\mathbf{U},\mathbf{Z}^{(1)}}$. The correlations are unaffected by the standardization.

Computing Correlations Between Canonical Variates and Component Variables

Example 10.2 (Computing correlations between canonical variates and their component variables) Compute the correlations between the first pair of canonical variates and their component variables for the situation considered in Example 10.1.

The variables in Example 10.1 are already standardized, so equation (10–15) is applicable. For the standardized variables,

$$\boldsymbol{\rho}_{11} = \begin{bmatrix} 1.0 & .4 \\ .4 & 1.0 \end{bmatrix} \qquad \boldsymbol{\rho}_{22} = \begin{bmatrix} 1.0 & .2 \\ .2 & 1.0 \end{bmatrix}$$

and

$$\boldsymbol{\rho}_{12} = \begin{bmatrix} .5 & .6 \\ .3 & .4 \end{bmatrix}$$

With $p = 1$,

$$\mathbf{A_z} = [.86, .28] \qquad \mathbf{B_z} = [.54, .74]$$

so

$$\boldsymbol{\rho}_{U_1,\mathbf{z}^{(1)}} = \mathbf{A_z}\boldsymbol{\rho}_{11} = [.86, .28] \begin{bmatrix} 1.0 & .4 \\ .4 & 1.0 \end{bmatrix} = [.97, .62]$$

and

$$\boldsymbol{\rho}_{V_1,\mathbf{z}^{(2)}} = \mathbf{B_z}\boldsymbol{\rho}_{22} = [.54, .74] \begin{bmatrix} 1.0 & .2 \\ .2 & 1.0 \end{bmatrix} = [.69, .85]$$

Question Posed: Which of the Two Variables Most Closely Associated with Canonical Variate U_1

We conclude that, of the two variables in the set $\mathbf{Z}^{(1)}$, the first is most closely associated with the canonical variate U_1. Of the two variables in the set $\mathbf{Z}^{(2)}$, the second is most closely associated with V_1. In this case, the correlations reinforce the information supplied by the standardized coefficients $\mathbf{A}_\mathbf{z}$ and $\mathbf{B}_\mathbf{z}$. However, the correlations elevate the relative importance of $Z_2^{(1)}$ in the first set and $Z_1^{(2)}$ in the second set because they ignore the contribution of the remaining variable in each set.

From (10-15), we also obtain the correlations

$$\boldsymbol{\rho}_{U_1,\mathbf{Z}^{(2)}} = \mathbf{A}_\mathbf{z}\boldsymbol{\rho}_{12} = [.86, .28]\begin{bmatrix} .5 & .6 \\ .3 & .4 \end{bmatrix} = [.51, .63]$$

and

$$\boldsymbol{\rho}_{V_1,\mathbf{Z}^{(1)}} = \mathbf{B}_\mathbf{z}\boldsymbol{\rho}_{21} = \mathbf{B}_\mathbf{z}\boldsymbol{\rho}'_{12} = [.54, .74]\begin{bmatrix} .5 & .3 \\ .6 & .4 \end{bmatrix} = [.71, .46]$$

Later, in our discussion of the sample canonical variates, we shall comment on the interpretation of these last correlations. ∎

Correlations Help Supply Meanings for Canonical Variates Similar to Principal Component (PC) Analysis : Relevance Correlations Between the PC and Associated Variables

The correlations $\boldsymbol{\rho}_{U,\mathbf{X}^{(1)}}$ and $\boldsymbol{\rho}_{V,\mathbf{X}^{(2)}}$ can help supply meanings for the canonical variates. The spirit is the same as in principal component analysis when the correlations between the principal components and their associated variables may provide subject-matter interpretations for the components.

Canonical Correlations as Generalizations of Other Correlation Coefficients

First, the canonical correlation generalizes the correlation between two variables. When $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ each consist of a single variable, so that $p = q = 1$,

$$|\mathrm{Corr}(X_1^{(1)}, X_1^{(2)})| = |\mathrm{Corr}(aX_1^{(1)}, bX_1^{(2)})| \qquad \text{for all } a, b \neq 0$$

Therefore, the "canonical variates" $U_1 = X_1^{(1)}$ and $V_1 = X_1^{(2)}$ have correlation $\rho_1^* = |\mathrm{Corr}(X_1^{(1)}, X_1^{(2)})|$. When $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ have more components, setting $\mathbf{a}' = [0, \ldots, 0, 1, 0, \ldots, 0]$ with 1 in the ith position and $\mathbf{b}' = [0, \ldots, 0, 1, 0, \ldots, 0]$ with 1 in the kth position yields

$$\begin{aligned} |\mathrm{Corr}(X_i^{(1)}, X_k^{(2)})| &= |\mathrm{Corr}(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)})| \\ &\leq \max_{\mathbf{a},\mathbf{b}} \mathrm{Corr}(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)}) = \rho_1^* \end{aligned} \tag{10-16}$$

That is, the first canonical correlation is larger than the absolute value of any entry in $\boldsymbol{\rho}_{12} = \mathbf{V}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\mathbf{V}_{22}^{-1/2}$.

Second, the multiple correlation coefficient $\rho_{1(\mathbf{X}^{(2)})}$ [see (7-48)] is a special case of a canonical correlation when $\mathbf{X}^{(1)}$ has the single element $X_1^{(1)}$ $(p = 1)$. Recall that

$$\rho_{1(\mathbf{X}^{(2)})} = \max_{\mathbf{b}} \mathrm{Corr}(X_1^{(1)}, \mathbf{b}'\mathbf{X}^{(2)}) = \rho_1^* \qquad \text{for} \quad p = 1 \tag{10-17}$$

Kth Squared Canonical Correlation: Proportion Variance Canonical Variate V_k Explained by the Set X i.e. Shared Variance

When $p > 1$, ρ_1^* is larger than each of the multiple correlations of $X_i^{(1)}$ with $\mathbf{X}^{(2)}$ or the multiple correlations of $X_i^{(2)}$ with $\mathbf{X}^{(1)}$.

Finally, we note that

$$\rho_{U_k(\mathbf{X}^{(2)})} = \max_{\mathbf{b}} \operatorname{Corr}(U_k, \mathbf{b}'\mathbf{X}^{(2)}) = \operatorname{Corr}(U_k, V_k) = \rho_k^*, \qquad k = 1, 2, \ldots, p \tag{10-18}$$

from the proof of Result 10.1 (see website: www.prenhall.com/statistics). Similarly,

$$\rho_{V_k(\mathbf{X}^{(1)})} = \max_{\mathbf{a}} \operatorname{Corr}(\mathbf{a}'\mathbf{X}^{(1)}, V_k) = \operatorname{Corr}(U_k, V_k) = \rho_k^*, \qquad k = 1, 2, \ldots, p \tag{10-19}$$

That is, the canonical correlations are also the multiple correlation coefficients of U_k with $\mathbf{X}^{(2)}$ or the multiple correlation coefficients of V_k with $\mathbf{X}^{(1)}$.

Because of its multiple correlation coefficient interpretation, the kth *squared* canonical correlation ρ_k^{*2} is the proportion of the variance of canonical variate U_k "explained" by the set $\mathbf{X}^{(2)}$. It is also the proportion of the variance of canonical variate V_k "explained" by the set $\mathbf{X}^{(1)}$. Therefore, ρ_k^{*2} is often called the *shared variance* between the two sets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. The largest value, ρ_1^{*2}, is sometimes regarded as a measure of set "overlap."

First r Canonical Variables As Summary of Variability

The First r Canonical Variables as a Summary of Variability

The change of coordinates from $\mathbf{X}^{(1)}$ to $\mathbf{U} = \mathbf{A}\mathbf{X}^{(1)}$ and from $\mathbf{X}^{(2)}$ to $\mathbf{V} = \mathbf{B}\mathbf{X}^{(2)}$ is chosen to maximize Corr (U_1, V_1) and, successively, Corr (U_i, V_i), where (U_i, V_i) have zero correlation with the previous pairs $(U_1, V_1), (U_2, V_2), \ldots, (U_{i-1}, V_{i-1})$. Correlation between the sets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ has been isolated in the pairs of canonical variables

By design, the coefficient vectors $\mathbf{a}_i, \mathbf{b}_i$ are selected to maximize correlations, not necessarily to provide variables that (approximately) account for the subset covariances $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$. When the first few pairs of canonical variables provide poor summaries of the variability in $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$, it is not clear how a high canonical correlation should be interpreted.

Canonical Correlation as Poor Sumary of Variability

Example 10.3 (Canonical correlation as a poor summary of variability) Consider the covariance matrix

$$\text{Cov}\left(\begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \hdashline X_1^{(2)} \\ X_2^{(2)} \end{bmatrix}\right) = \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right] = \left[\begin{array}{cc:cc} 100 & 0 & 0 & 0 \\ 0 & 1 & .95 & 0 \\ \hdashline 0 & .95 & 1 & 0 \\ 0 & 0 & 0 & 100 \end{array}\right]$$

The reader may verify (see Exercise 10.1) that the first pair of canonical variates $U_1 = X_2^{(1)}$ and $V_1 = X_1^{(2)}$ has correlation

$$\rho_1^* = \text{Corr}(U_1, V_1) = .95$$

Yet $U_1 = X_2^{(1)}$ provides a very poor summary of the variability in the first set. Most of the variability in this set is in $X_1^{(1)}$, which is uncorrelated with U_1. The same situation is true for $V_1 = X_1^{(2)}$ in the second set. ■

Geometrical Interpretation Population Canonical Correlation Analysis

A Geometrical Interpretation of the Population Canonical Correlation Analysis

A geometrical interpretation of the procedure for selecting canonical variables provides some valuable insights into the nature of a canonical correlation analysis.

The transformation

$$\mathbf{U} = \mathbf{A}\mathbf{X}^{(1)}$$

from $\mathbf{X}^{(1)}$ to $\mathbf{U}$ gives

$$\mathrm{Cov}(\mathbf{U}) = \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{A}' = \mathbf{I}$$

From Result 10.1 and (2-22), $\mathbf{A} = \mathbf{E}'\boldsymbol{\Sigma}_{11}^{-1/2} = \mathbf{E}'\mathbf{P}_1\Lambda_1^{-1/2}\mathbf{P}_1'$ where $\mathbf{E}'$ is an orthogonal matrix with row $\mathbf{e}_i'$, and $\boldsymbol{\Sigma}_{11} = \mathbf{P}_1\Lambda_1\mathbf{P}_1'$. Now, $\mathbf{P}_1'\mathbf{X}^{(1)}$ is the set of principal components derived from $\mathbf{X}^{(1)}$ alone. The matrix $\Lambda_1^{-1/2}\mathbf{P}_1'\mathbf{X}^{(1)}$ has ith row $(1/\sqrt{\lambda_i})\,\mathbf{p}_i'\mathbf{X}^{(1)}$, which is the ith principal component scaled to have unit variance. That is,

Process: Transformation of $X^{(1)}$ to Uncorrelated Principal Components With Orthogonal Rotation Determined by Covariance Matrix of Pair 1, with Another Rotation from the Full Covariance Matrix

From Result 10.1 and (2-22), $\mathbf{A} = \mathbf{E}'\mathbf{\Sigma}_{11}^{-1/2} = \mathbf{E}'\mathbf{P}_1\Lambda_1^{-1/2}\mathbf{P}_1'$ where $\mathbf{E}'$ is an orthogonal matrix with row $\mathbf{e}_i'$, and $\mathbf{\Sigma}_{11} = \mathbf{P}_1\Lambda_1\mathbf{P}_1'$. Now, $\mathbf{P}_1'\mathbf{X}^{(1)}$ is the set of principal components derived from $\mathbf{X}^{(1)}$ alone. The matrix $\Lambda_1^{-1/2}\mathbf{P}_1'\mathbf{X}^{(1)}$ has ith row $(1/\sqrt{\lambda_i})\,\mathbf{p}_i'\mathbf{X}^{(1)}$, which is the ith principal component scaled to have unit variance. That is,

$$\begin{aligned}\mathrm{Cov}(\Lambda_1^{-1/2}\mathbf{P}_1'\mathbf{X}^{(1)}) &= \Lambda_1^{-1/2}\mathbf{P}_1'\mathbf{\Sigma}_{11}\mathbf{P}_1\Lambda_1^{-1/2} = \Lambda_1^{-1/2}\mathbf{P}_1'\mathbf{P}_1\Lambda_1\mathbf{P}_1'\mathbf{P}_1\Lambda_1^{-1/2}\\ &= \Lambda_1^{-1/2}\Lambda_1\Lambda_1^{-1/2} = \mathbf{I}\end{aligned}$$

Consequently, $\mathbf{U} = \mathbf{A}\mathbf{X}^{(1)} = \mathbf{E}'\mathbf{P}_1\Lambda_1^{-1/2}\mathbf{P}_1'\mathbf{X}^{(1)}$ can be interpreted as (1) a transformation of $\mathbf{X}^{(1)}$ to uncorrelated standardized principal components, followed by (2) a rigid (orthogonal) rotation $\mathbf{P}_1$ determined by $\mathbf{\Sigma}_{11}$ and then (3) another rotation $\mathbf{E}'$ determined from the full covariance matrix $\mathbf{\Sigma}$. A similar interpretation applies to $\mathbf{V} = \mathbf{B}\mathbf{X}^{(2)}$.

Sample Canonical Variates and Their Correlations

10.4 The Sample Canonical Variates and Sample Canonical Correlations

A random sample of n observations on each of the $(p+q)$ variables $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ can be assembled into the $n \times (p+q)$ data matrix

$$\mathbf{X} = \left[\mathbf{X}^{(1)} \mid \mathbf{X}^{(2)}\right]$$

$$= \left[\begin{array}{cccc|cccc} x_{11}^{(1)} & x_{12}^{(1)} & \cdots & x_{1p}^{(1)} & x_{11}^{(2)} & x_{12}^{(2)} & \cdots & x_{1q}^{(2)} \\ x_{21}^{(1)} & x_{22}^{(1)} & \cdots & x_{2p}^{(1)} & x_{21}^{(2)} & x_{22}^{(2)} & \cdots & x_{2q}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1}^{(1)} & x_{n2}^{(1)} & \cdots & x_{np}^{(1)} & x_{n1}^{(2)} & x_{n2}^{(2)} & \cdots & x_{nq}^{(2)} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{x}_1^{(1)\prime} & \mathbf{x}_1^{(2)\prime} \\ \vdots & \vdots \\ \mathbf{x}_n^{(1)\prime} & \mathbf{x}_n^{(2)\prime} \end{array}\right] \quad (10\text{-}20)$$

The vector of sample means can be organized as

$$\underset{(p+q)\times 1}{\bar{\mathbf{x}}} = \left[\begin{array}{c} \bar{\mathbf{x}}^{(1)} \\ \hline \bar{\mathbf{x}}^{(2)} \end{array}\right] \quad \text{where} \quad \bar{\mathbf{x}}^{(1)} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j^{(1)}$$

$$\bar{\mathbf{x}}^{(2)} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j^{(2)} \quad (10\text{-}21)$$

Calculating the kth Sample Canonical Correlation

Similarly, the sample covariance matrix can be arranged analogous to the representation (10-4). Thus,

$$\underset{(p+q)\times(p+q)}{\mathbf{S}} = \left[\begin{array}{c:c} \underset{(p\times p)}{\mathbf{S}_{11}} & \underset{(p\times q)}{\mathbf{S}_{12}} \\ \hdashline \underset{(q\times p)}{\mathbf{S}_{21}} & \underset{(q\times q)}{\mathbf{S}_{22}} \end{array}\right]$$

where

$$\mathbf{S}_{kl} = \frac{1}{n-1}\sum_{j=1}^{n} (\mathbf{x}_j^{(k)} - \bar{\mathbf{x}}^{(k)})(\mathbf{x}_j^{(l)} - \bar{\mathbf{x}}^{(l)})', \qquad k, l = 1, 2 \qquad (10\text{-}22)$$

The linear combinations

$$\hat{U} = \hat{\mathbf{a}}'\mathbf{x}^{(1)}; \qquad \hat{V} = \hat{\mathbf{b}}'\mathbf{x}^{(2)} \qquad (10\text{-}23)$$

have sample correlation [see (3-36)]

$$r_{\hat{U},\hat{V}} = \frac{\hat{\mathbf{a}}'\mathbf{S}_{12}\hat{\mathbf{b}}}{\sqrt{\hat{\mathbf{a}}'\mathbf{S}_{11}\hat{\mathbf{a}}}\sqrt{\hat{\mathbf{b}}'\mathbf{S}_{22}\hat{\mathbf{b}}}} \qquad (10\text{-}24)$$

The *first pair of sample canonical variates* is the pair of linear combinations $\hat{U}_1, \hat{V}_1$ having unit sample variances that maximize the ratio (10-24).

In general, the *kth pair of sample canonical variates* is the pair of linear combinations $\hat{U}_k, \hat{V}_k$ having unit sample variances that maximize the ratio (10-24) among those linear combinations uncorrelated with the previous $k-1$ sample canonical variates.

The sample correlation between $\hat{U}_k$ and $\hat{V}_k$ is called the *kth sample canonical correlation*.

The sample canonical variates and the sample canonical correlations can be obtained from the sample covariance matrices $\mathbf{S}_{11}$, $\mathbf{S}_{12} = \mathbf{S}_{21}'$, and $\mathbf{S}_{22}$ in a manner consistent with the population case described in Result 10.1.

For the kth Pair, Largest Possible Correlation Among Linear Combinations Uncorrelated with k-1 Sample Canonical Variates

Result 10.2. Let $\widehat{\rho_1^{*2}} \geq \widehat{\rho_2^{*2}} \geq \cdots \geq \widehat{\rho_p^{*2}}$ be the p ordered eigenvalues of $\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$ with corresponding eigenvectors $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \ldots, \hat{\mathbf{e}}_p$, where the $\mathbf{S}_{kl}$ are defined in (10-22) and $p \leq q$. Let $\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \ldots, \hat{\mathbf{f}}_p$ be the eigenvectors of $\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}$, where the first p $\hat{\mathbf{f}}$'s may be obtained from $\hat{\mathbf{f}}_k = (1/\widehat{\rho_k^*})\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\hat{\mathbf{e}}_k$, $k = 1, 2, \ldots, p$. Then the kth sample canonical variate pair[1] is

$$\hat{U}_k = \underbrace{\hat{\mathbf{e}}_k'\mathbf{S}_{11}^{-1/2}}_{\hat{\mathbf{a}}_k'}\mathbf{x}^{(1)} \qquad \hat{V}_k = \underbrace{\hat{\mathbf{f}}_k'\mathbf{S}_{22}^{-1/2}}_{\hat{\mathbf{b}}_k'}\mathbf{x}^{(2)}$$

where $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are the values of the variables $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ for a particular experimental unit. Also, the first sample canonical variate pair has the maximum sample correlation

$$r_{\hat{U}_1, \hat{V}_1} = \widehat{\rho_1^*}$$

and for the kth pair,

$$r_{\hat{U}_k, \hat{V}_k} = \widehat{\rho_k^*}$$

is the largest possible correlation among linear combinations uncorrelated with the preceding $k - 1$ sample canonical variates.

The quantities $\widehat{\rho_1^*}, \widehat{\rho_2^*}, \ldots, \widehat{\rho_p^*}$ are the sample canonical correlations.[2]

Proof. The proof of this result follows the proof of Result 10.1, with $\mathbf{S}_{kl}$ substituted for Σ_{kl}, $k, l = 1, 2$. ■

[1] When the distribution is normal, the maximum likelihood method can be employed using $\hat{\Sigma} = \mathbf{S}_n$ in place of $\mathbf{S}$. The sample canonical correlations $\widehat{\rho_k^*}$ are, therefore, the maximum likelihood estimates of ρ_k^* and $\sqrt{n/(n-1)}\,\hat{\mathbf{a}}_k$, $\sqrt{n/(n-1)}\,\hat{\mathbf{b}}_k$ are the maximum likelihood estimates of $\mathbf{a}_k$ and $\mathbf{b}_k$, respectively.

[2] If $p > \text{rank}(\mathbf{S}_{12}) = p_1$, the nonzero sample canonical correlations are $\widehat{\rho_1^*}, \ldots, \widehat{\rho_{p_1}^*}$.

[3] The vectors $\hat{\mathbf{b}}_{p_1+1} = \mathbf{S}_{22}^{-1/2}\hat{\mathbf{f}}_{p_1+1}, \hat{\mathbf{b}}_{p_1+2} = \mathbf{S}_{22}^{-1/2}\hat{\mathbf{f}}_{p_1+2}, \ldots, \hat{\mathbf{b}}_q = \mathbf{S}_{22}^{-1/2}\hat{\mathbf{f}}_q$ are determined from a choice of the last $q - p_1$ mutually orthogonal eigenvectors $\hat{\mathbf{f}}$ associated with the *zero* eigenvalue of $\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}$.

Sample Canonical Variates, With Variances and Sample Correlations

Proof. The proof of this result follows the proof of Result 10.1, with $\mathbf{S}_{kl}$ substituted for $\boldsymbol{\Sigma}_{kl}$, $k, l = 1, 2$. ∎

The sample canonical variates have unit sample variances

$$s_{\hat{U}_k, \hat{U}_k} = s_{\hat{V}_k, \hat{V}_k} = 1 \tag{10-25}$$

and their sample correlations are

$$\begin{aligned} r_{\hat{U}_k, \hat{U}_\ell} = r_{\hat{V}_k, \hat{V}_\ell} &= 0, \qquad k \neq \ell \\ r_{\hat{U}_k, \hat{V}_\ell} &= 0, \qquad k \neq \ell \end{aligned} \tag{10-26}$$

The interpretation of $\hat{U}_k$, $\hat{V}_k$ is often aided by computing the sample correlations between the canonical variates and the variables in the sets $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. We define the matrices

$$\underset{(p\times p)}{\hat{\mathbf{A}}} = [\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \ldots, \hat{\mathbf{a}}_p]' \qquad \underset{(q\times q)}{\hat{\mathbf{B}}} = [\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \ldots, \hat{\mathbf{b}}_q]' \tag{10-27}$$

whose *rows* are the coefficient vectors for the sample canonical variates.[3] Analogous to (10-12), we have

$$\underset{(p\times 1)}{\hat{\mathbf{U}}} = \hat{\mathbf{A}}\,\mathbf{x}^{(1)} \qquad \underset{(q\times 1)}{\hat{\mathbf{V}}} = \hat{\mathbf{B}}\,\mathbf{x}^{(2)} \tag{10-28}$$

[1] When the distribution is normal, the maximum likelihood method can be employed using $\hat{\boldsymbol{\Sigma}} = \mathbf{S}_n$ in place of $\mathbf{S}$. The sample canonical correlations $\widehat{\rho_k^*}$ are, therefore, the maximum likelihood estimates of ρ_k^* and $\sqrt{n/(n-1)}\,\hat{\mathbf{a}}_k$, $\sqrt{n/(n-1)}\,\hat{\mathbf{b}}_k$ are the maximum likelihood estimates of $\mathbf{a}_k$ and $\mathbf{b}_k$, respectively.

[2] If $p > \text{rank}(\mathbf{S}_{12}) = p_1$, the nonzero sample canonical correlations are $\widehat{\rho_1^*}, \ldots, \widehat{\rho_{p_1}^*}$.

[3] The vectors $\hat{\mathbf{b}}_{p_1+1} = \mathbf{S}_{22}^{-1/2}\hat{\mathbf{f}}_{p_1+1}$, $\hat{\mathbf{b}}_{p_1+2} = \mathbf{S}_{22}^{-1/2}\hat{\mathbf{f}}_{p_1+2}, \ldots, \hat{\mathbf{b}}_q = \mathbf{S}_{22}^{-1/2}\hat{\mathbf{f}}_q$ are determined from a choice of the last $q - p_1$ mutually orthogonal eigenvectors $\hat{\mathbf{f}}$ associated with the *zero* eigenvalue of $\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}$.

Matrix of Sample Correlations

and we can define

$$\begin{aligned}
\mathbf{R}_{\hat{\mathbf{U}},\mathbf{x}^{(1)}} &= \text{matrix of sample correlations of } \hat{\mathbf{U}} \text{ with } \mathbf{x}^{(1)} \\
\mathbf{R}_{\hat{\mathbf{V}},\mathbf{x}^{(2)}} &= \text{matrix of sample correlations of } \hat{\mathbf{V}} \text{ with } \mathbf{x}^{(2)} \\
\mathbf{R}_{\hat{\mathbf{U}},\mathbf{x}^{(2)}} &= \text{matrix of sample correlations of } \hat{\mathbf{U}} \text{ with } \mathbf{x}^{(2)} \\
\mathbf{R}_{\hat{\mathbf{V}},\mathbf{x}^{(1)}} &= \text{matrix of sample correlations of } \hat{\mathbf{V}} \text{ with } \mathbf{x}^{(1)}
\end{aligned}$$

Corresponding to (10-19), we have

$$\begin{aligned}
\mathbf{R}_{\hat{\mathbf{U}},\mathbf{x}^{(1)}} &= \hat{\mathbf{A}}\mathbf{S}_{11}\mathbf{D}_{11}^{-1/2} \\
\mathbf{R}_{\hat{\mathbf{V}},\mathbf{x}^{(2)}} &= \hat{\mathbf{B}}\mathbf{S}_{22}\mathbf{D}_{22}^{-1/2} \\
\mathbf{R}_{\hat{\mathbf{U}},\mathbf{x}^{(2)}} &= \hat{\mathbf{A}}\mathbf{S}_{12}\mathbf{D}_{22}^{-1/2} \\
\mathbf{R}_{\hat{\mathbf{V}},\mathbf{x}^{(1)}} &= \hat{\mathbf{B}}\mathbf{S}_{21}\mathbf{D}_{11}^{-1/2}
\end{aligned} \tag{10-29}$$

where $\mathbf{D}_{11}^{-1/2}$ is the $(p \times p)$ diagonal matrix with ith diagonal element (sample $\text{var}(x_i^{(1)}))^{-1/2}$ and $\mathbf{D}_{22}^{-1/2}$ is the $(q \times q)$ diagonal matrix with ith diagonal element (sample $\text{var}(x_i^{(2)}))^{-1/2}$.

Note: Sample Canonical Correlations Unaffected by Standardizations

Comment. If the observations are standardized [see (8-25)], the data matrix becomes

$$\mathbf{Z} = \left[\mathbf{Z}^{(1)} \;\vdots\; \mathbf{Z}^{(2)}\right] = \begin{bmatrix} \mathbf{z}_1^{(1)\prime} & \vdots & \mathbf{z}_1^{(2)\prime} \\ \vdots & \vdots & \vdots \\ \mathbf{z}_n^{(1)\prime} & \vdots & \mathbf{z}_n^{(2)\prime} \end{bmatrix}$$

and the sample canonical variates become

$$\underset{(p\times 1)}{\hat{\mathbf{U}}} = \hat{\mathbf{A}}_{\mathbf{z}}\mathbf{z}^{(1)} \qquad \underset{(q\times 1)}{\hat{\mathbf{V}}} = \hat{\mathbf{B}}_{\mathbf{z}}\mathbf{z}^{(2)} \tag{10-30}$$

where $\hat{\mathbf{A}}_{\mathbf{z}} = \hat{\mathbf{A}}\mathbf{D}_{11}^{1/2}$ and $\hat{\mathbf{B}}_{\mathbf{z}} = \hat{\mathbf{B}}\mathbf{D}_{22}^{1/2}$. The sample canonical correlations are unaffected by the standardization. The correlations displayed in (10–29) remain unchanged and may be calculated, for standardized observations, by substituting $\hat{\mathbf{A}}_{\mathbf{z}}$ for $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}_{\mathbf{z}}$ for $\hat{\mathbf{B}}$, and $\mathbf{R}$ for $\mathbf{S}$. Note that $\mathbf{D}_{11}^{-1/2} = \underset{(p\times p)}{\mathbf{I}}$ and $\mathbf{D}_{22}^{-1/2} = \underset{(q\times q)}{\mathbf{I}}$ for standardized observations.

Canonical Correlation Analysis of the Chicken-Bone Data

Example 10.4 (Canonical correlation analysis of the chicken-bone data) In Example 9.14, data consisting of bone and skull measurements of white leghorn fowl were described. From this example, the chicken-bone measurements for

$$\text{Head } (\mathbf{X}^{(1)}): \quad \begin{cases} X_1^{(1)} = \text{skull length} \\ X_2^{(1)} = \text{skull breadth} \end{cases}$$

$$\text{Leg } (\mathbf{X}^{(2)}): \quad \begin{cases} X_1^{(2)} = \text{femur length} \\ X_2^{(2)} = \text{tibia length} \end{cases}$$

Canonical Correlation Analysis of the Chicken-Bone Data

have the sample correlation matrix

$$\mathbf{R} = \left[\begin{array}{c|c} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \hline \mathbf{R}_{21} & \mathbf{R}_{22} \end{array}\right] = \left[\begin{array}{cc|cc} 1.0 & .505 & .569 & .602 \\ .505 & 1.0 & .422 & .467 \\ \hline .569 & .422 & 1.0 & .926 \\ .602 & .467 & .926 & 1.0 \end{array}\right]$$

A canonical correlation analysis of the head and leg sets of variables using $\mathbf{R}$ produces the two canonical correlations and corresponding pairs of variables

$$\widehat{\rho_1^*} = .631 \qquad \begin{aligned} \hat{U}_1 &= .781z_1^{(1)} + .345z_2^{(1)} \\ \hat{V}_1 &= .060z_1^{(2)} + .944z_2^{(2)} \end{aligned}$$

and

$$\widehat{\rho_2^*} = .057 \qquad \begin{aligned} \hat{U}_2 &= -.856z_1^{(1)} + 1.106z_2^{(1)} \\ \hat{V}_2 &= -2.648z_1^{(2)} + 2.475z_2^{(2)} \end{aligned}$$

Here $z_i^{(1)}, i = 1, 2$ and $z_i^{(2)}, i = 1, 2$ are the standardized data values for sets 1 and 2, respectively. The preceding results were taken from the SAS statistical software output shown in Panel 10.1. In addition, the correlations of the original variables with the canonical variables are highlighted in that panel. ■

Canonical Correlation Analysis of Job Satisfaction: SAS Output

Example 10.5 (Canonical correlation analysis of job satisfaction) As part of a larger study of the effects of organizational structure on "job satisfaction," Dunham [4] investigated the extent to which measures of job satisfaction are related to job characteristics. Using a survey instrument, Dunham obtained measurements of $p = 5$ job characteristics and $q = 7$ job satisfaction variables for $n = 784$ executives from the corporate branch of a large retail merchandising corporation. Are measures of job satisfaction associated with job characteristics? The answer may have implications for job design.

PANEL 10.1 SAS ANALYSIS FOR EXAMPLE 10.4 USING PROC CANCORR.

```
title 'Canonical Correlation Analysis';
data skull (type = corr);
_type_ = 'CORR';
input _name_$ x1 x2 x3 x4;
cards;
x1   1.0    .      .     .
x2    .505   1.0    .     .
x3    .569    .422  1.0   .
x4    .602    .467   .926  1.0
;
proc cancorr data = skull vprefix = head wprefix = leg;
   var x1 x2; with x3 x4;
```

PROGRAM COMMANDS

(continues on next page)

Canonical Correlation Analysis of the Chicken-Bone Data

PANEL 10.1 *(continued)*

Canonical Correlation Analysis

	Canonical Correlation	Adjusted Canonical Correlation	Approx Standard Error	Squared Canonical Correlation
1	0.631085	0.628291	0.036286	0.398268
2	0.056794		0.060108	0.003226

Raw Canonical Coefficient for the 'VAR' Variables

	HEAD1	HEAD2
X1	0.7807924389	−0.855973184
X2	0.3445068301	1.1061835145

OUTPUT

Raw Canonical Coefficient for the 'WITH' Variables

	LEG1	LEG2
X3	0.0602508775	−2.648156338
X4	0.943948961	2.4749388913

Canonical Structure

Correlations Between the 'VAR' Variables and Their Canonical Variables

	HEAD1	HEAD2
X1	0.9548	−0.2974
X2	0.7388	0.6739

(see 10-29)

Canonical Correlation Analysis of the Chicken-Bone Data

Correlations Between the 'WITH' Variables and Their Canonical Variables

	LEG1	LEG2	
X3	0.9343	−0.3564	(see 10-29)
X4	0.9997	0.0227	

Correlations Between the 'VAR' Variables and the Canonical Variables of the 'WITH' Variables

	LEG1	LEG2	
X1	0.6025	−0.0169	(see 10-29)
X2	0.4663	0.0383	

Correlations Between the 'WITH' Variables and the Canonical Variables of the 'VAR' Variables

	HEAD1	HEAD2	
X3	0.5897	−0.0202	(see 10-29)
X4	0.6309	0.0013	

Original Job Characteristic Variables and Job Satisfaction Variables

The original job characteristic variables, $\mathbf{X}^{(1)}$, and job satisfaction variables, $\mathbf{X}^{(2)}$, were respectively defined as

$$\mathbf{X}^{(1)} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ X_3^{(1)} \\ X_4^{(1)} \\ X_5^{(1)} \end{bmatrix} = \begin{bmatrix} \text{feedback} \\ \text{task significance} \\ \text{task variety} \\ \text{task identity} \\ \text{autonomy} \end{bmatrix}$$

$$\mathbf{X}^{(2)} = \begin{bmatrix} X_1^{(2)} \\ X_2^{(2)} \\ X_3^{(2)} \\ X_4^{(2)} \\ X_5^{(2)} \\ X_6^{(2)} \\ X_7^{(2)} \end{bmatrix} = \begin{bmatrix} \text{supervisor satisfaction} \\ \text{career-future satisfaction} \\ \text{financial satisfaction} \\ \text{workload satisfaction} \\ \text{company identification} \\ \text{kind-of-work-satisfaction} \\ \text{general satisfaction} \end{bmatrix}$$

Responses for variables $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ were recorded on a scale and then standardized. The sample correlation matrix based on 784 responses is

Sample Correlation Matrix n = 784

Responses for variables $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ were recorded on a scale and then standardized. The sample correlation matrix based on 784 responses is

$$
\mathbf{R} = \left[\begin{array}{c|c} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \hline \mathbf{R}_{21} & \mathbf{R}_{22} \end{array}\right]
$$

$$
= \left[\begin{array}{ccccc|ccccccc}
1.0 & & & & & .33 & .32 & .20 & .19 & .30 & .37 & .21 \\
.49 & 1.0 & & & & .30 & .21 & .16 & .08 & .27 & .35 & .20 \\
.53 & .57 & 1.0 & & & .31 & .23 & .14 & .07 & .24 & .37 & .18 \\
.49 & .46 & .48 & 1.0 & & .24 & .22 & .12 & .19 & .21 & .29 & .16 \\
.51 & .53 & .57 & .57 & 1.0 & .38 & .32 & .17 & .23 & .32 & .36 & .27 \\
\hline
.33 & .30 & .31 & .24 & .38 & 1.0 & & & & & & \\
.32 & .21 & .23 & .22 & .32 & .43 & 1.0 & & & & & \\
.20 & .16 & .14 & .12 & .17 & .27 & .33 & 1.0 & & & & \\
.19 & .08 & .07 & .19 & .23 & .24 & .26 & .25 & 1.0 & & & \\
.30 & .27 & .24 & .21 & .32 & .34 & .54 & .46 & .28 & 1.0 & & \\
.37 & .35 & .37 & .29 & .36 & .37 & .32 & .29 & .30 & .35 & 1.0 & \\
.21 & .20 & .18 & .16 & .27 & .40 & .58 & .45 & .27 & .59 & .31 & 1.0
\end{array}\right]
$$

The $\min(p, q) = \min(5, 7) = 5$ sample canonical correlations and the sample canonical variate coefficient vectors (from Dunham [4]) are displayed in the following table:

Example: Canonical Variate Coefficients and Correlations

Canonical Variate Coefficients and Canonical Correlations

	Standardized variables							Standardized variables						
	$z_1^{(1)}$	$z_2^{(1)}$	$z_3^{(1)}$	$z_4^{(1)}$	$z_5^{(1)}$	$\widehat{\rho_1^*}$		$z_1^{(2)}$	$z_2^{(2)}$	$z_3^{(2)}$	$z_4^{(2)}$	$z_5^{(2)}$	$z_6^{(2)}$	$z_7^{(2)}$
$\hat{\mathbf{a}}_1'$:	.42	.21	.17	−.02	.44	.55	$\hat{\mathbf{b}}_1'$:	.42	.22	−.03	.01	.29	.52	−.12
$\hat{\mathbf{a}}_2'$:	−.30	.65	.85	−.29	−.81	.23	$\hat{\mathbf{b}}_2'$:	.03	−.42	.08	−.91	.14	.59	−.02
$\hat{\mathbf{a}}_3'$:	−.86	.47	−.19	−.49	.95	.12	$\hat{\mathbf{b}}_3'$:	.58	−.76	−.41	−.07	.19	−.43	.92
$\hat{\mathbf{a}}_4'$:	.76	−.06	−.12	−1.14	−.25	.08	$\hat{\mathbf{b}}_4'$:	.23	.49	.52	−.47	.34	−.69	−.37
$\hat{\mathbf{a}}_5'$:	.27	1.01	−1.04	.16	.32	.05	$\hat{\mathbf{b}}_5'$:	−.52	−.63	.41	.21	.76	.02	.10

Job Characteristic and Job Satisfaction: First Sample Canonical Variat

For example, the first sample canonical variate pair is

$$\hat{U}_1 = .42z_1^{(1)} + .21z_2^{(1)} + .17z_3^{(1)} - .02z_4^{(1)} + .44z_5^{(1)}$$

$$\hat{V}_1 = .42z_1^{(2)} + .22z_2^{(2)} - .03z_3^{(2)} + .01z_4^{(2)} + .29z_5^{(2)} + .52z_6^{(2)} - .12z_7^{(2)}$$

with sample canonical correlation $\widehat{\rho_1^*} = .55$.

According to the coefficients, $\hat{U}_1$ is primarily a feedback and autonomy variable, while $\hat{V}_1$ represents supervisor, career-future, and kind-of-work satisfaction, along with company identification.

To provide interpretations for $\hat{U}_1$ and $\hat{V}_1$, the sample correlations between $\hat{U}_1$ and its component variables and between $\hat{V}_1$ and its component variables were computed. Also, the following table shows the sample correlations between variables in one set and the first sample canonical variate of the other set. These correlations can be calculated using (10-29).

Correlations Between Original Variables and Canonical Variables

Sample Correlations Between Original Variables and Canonical Variables

$\mathbf{X}^{(1)}$ variables	Sample canonical variates $\hat{U}_1$	$\hat{V}_1$	$\mathbf{X}^{(2)}$ variables	Sample canonical variates $\hat{U}_1$	$\hat{V}_1$
1. Feedback	.83	.46	1. Supervisor satisfaction	.42	.75
2. Task significance	.74	.41	2. Career-future satisfaction	.35	.65
3. Task variety	.75	.42	3. Financial satisfaction	.21	.39
4. Task identity	.62	.34	4. Workload satisfaction	.21	.37
5. Autonomy	.85	.48	5. Company identification	.36	.65
			6. Kind-of-work satisfaction	.44	.80
			7. General satisfaction	.28	.50

All five job characteristic variables have roughly the same correlations with the first canonical variate $\hat{U}_1$. From this standpoint, $\hat{U}_1$ might be interpreted as a job characteristic "index." This differs from the preferred interpretation, based on coefficients, where the task variables are not important.

The other member of the first canonical variate pair, $\hat{V}_1$, seems to be representing, primarily, supervisor satisfaction, career-future satisfaction, company identification, and kind-of-work satisfaction. As the variables suggest, $\hat{V}_1$ might be regarded as a job satisfaction–company identification index. This agrees with the preceding interpretation based on the canonical coefficients of the $z_i^{(2)}$'s. The sample correlation between the two indices $\hat{U}_1$ and $\hat{V}_1$ is $\widehat{\rho_1^*} = .55$. There appears to be some overlap between job characteristics and job satisfaction. We explore this issue further in Example 10.7. ■

Scatter Plots: Significant Canonical Variates Against Their Component Variables

Scatter plots of the first $(\hat{U}_1, \hat{V}_1)$ pair may reveal atypical observations $\mathbf{x}_j$ requiring further study. If the canonical correlations $\widehat{\rho_2^*}, \widehat{\rho_3^*}, \ldots$ are also moderately large,

scatter plots of the pairs $(\hat{U}_2, \hat{V}_2), (\hat{U}_3, \hat{V}_3), \ldots$ may also be helpful in this res[illegible] Many analysts suggest plotting "significant" canonical variates against their com[illegible] nent variables as an aid in subject-matter interpretation. These plots reinforce [illegible] correlation coefficients in (10-29).

If the sample size is large, it is often desirable to split the sample in half. [illegible] first half of the sample can be used to construct and evaluate the sample cano[illegible] cal variates and canonical correlations. The results can then be "validated" [illegible] the remaining observations. The change (if any) in the nature of the canon[illegible] analysis will provide an indication of the sampling variability and the stabili[illegible] the conclusions.

Additional Sample Descriptive Measures: Calculating Proportion of Variance in One Set of Variables Explained by the Canonical Variates of the Other Set

10.5 Additional Sample Descriptive Measures

If the canonical variates are "good" summaries of their respective sets of variables, then the associations between variables can be described in terms of the canonical variates and their correlations. It is useful to have summary measures of the extent to which the canonical variates account for the variation in their respective sets. It is also useful, on occasion, to calculate the proportion of variance in one set of variables explained by the canonical variates of the other set.

Matrices of Errors of Approximations

Given the matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ defined in (10-27), let $\hat{\mathbf{a}}^{(i)}$ and $\hat{\mathbf{b}}^{(i)}$ denote the ith column of $\hat{\mathbf{A}}^{-1}$ and $\hat{\mathbf{B}}^{-1}$, respectively. Since $\hat{\mathbf{U}} = \hat{\mathbf{A}}\mathbf{x}^{(1)}$ and $\hat{\mathbf{V}} = \hat{\mathbf{B}}\mathbf{x}^{(2)}$ we can write

$$\underset{(p\times 1)}{\mathbf{x}^{(1)}} = \underset{(p\times p)}{\hat{\mathbf{A}}^{-1}}\ \underset{(p\times 1)}{\hat{\mathbf{U}}} \qquad \underset{(q\times 1)}{\mathbf{x}^{(2)}} = \underset{(q\times q)}{\hat{\mathbf{B}}^{-1}}\ \underset{(q\times 1)}{\hat{\mathbf{V}}} \tag{10-31}$$

Sample Co-variances of the 1st r Canonical Variates With Component Variables

Because sample $\text{Cov}(\hat{\mathbf{U}}, \hat{\mathbf{V}}) = \hat{\mathbf{A}}\mathbf{S}_{12}\hat{\mathbf{B}}'$, sample $\text{Cov}(\hat{\mathbf{U}}) = \hat{\mathbf{A}}\mathbf{S}_{11}\hat{\mathbf{A}}' = \underset{(p \times p)}{\mathbf{I}}$, and sample $\text{Cov}(\hat{\mathbf{V}}) = \hat{\mathbf{B}}\mathbf{S}_{22}\hat{\mathbf{B}}' = \underset{(q \times q)}{\mathbf{I}}$,

$$\mathbf{S}_{12} = \hat{\mathbf{A}}^{-1} \left[\begin{array}{cccc:c} \widehat{\rho_1^*} & 0 & \cdots & 0 & \\ 0 & \widehat{\rho_2^*} & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & \mathbf{0} \\ 0 & 0 & \cdots & \widehat{\rho_p^*} & \end{array} \right] (\hat{\mathbf{B}}^{-1})' = \widehat{\rho_1^*}\hat{\mathbf{a}}^{(1)}\hat{\mathbf{b}}^{(1)\prime} + \widehat{\rho_2^*}\hat{\mathbf{a}}^{(2)}\hat{\mathbf{b}}^{(2)\prime} + \cdots + \widehat{\rho_p^*}\hat{\mathbf{a}}^{(p)}\hat{\mathbf{b}}^{(p)\prime} \quad (10\text{-}$$

$$\mathbf{S}_{11} = (\hat{\mathbf{A}}^{-1})(\hat{\mathbf{A}}^{-1})' = \hat{\mathbf{a}}^{(1)}\hat{\mathbf{a}}^{(1)\prime} + \hat{\mathbf{a}}^{(2)}\hat{\mathbf{a}}^{(2)\prime} + \cdots + \hat{\mathbf{a}}^{(p)}\hat{\mathbf{a}}^{(p)\prime}$$

$$\mathbf{S}_{22} = (\hat{\mathbf{B}}^{-1})(\hat{\mathbf{B}}^{-1})' = \hat{\mathbf{b}}^{(1)}\hat{\mathbf{b}}^{(1)\prime} + \hat{\mathbf{b}}^{(2)}\hat{\mathbf{b}}^{(2)\prime} + \cdots + \hat{\mathbf{b}}^{(q)}\hat{\mathbf{b}}^{(q)\prime}$$

Since $\mathbf{x}^{(1)} = \hat{\mathbf{A}}^{-1}\hat{\mathbf{U}}$ and $\hat{\mathbf{U}}$ has sample covariance $\mathbf{I}$, the first r columns o contain the sample covariances of the first r canonical variates $\hat{U}_1, \hat{U}_2, \ldots$ their component variables $X_1^{(1)}, X_2^{(1)}, \ldots, X_p^{(1)}$. Similarly, the first r column contain the sample covariances of $\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_r$ with their component variab

If Only 1st r Canonical Pairs Used, then S_{12} Approximated by Sample Covariance $(x^{(1)}, x^{(2)})$

If only the first r canonical pairs are used, so that for instance,

$$\widetilde{\mathbf{x}}^{(1)} = [\hat{\mathbf{a}}^{(1)} \mid \hat{\mathbf{a}}^{(2)} \mid \cdots \mid \hat{\mathbf{a}}^{(r)}] \begin{bmatrix} \hat{U}_1 \\ \hat{U}_2 \\ \vdots \\ \hat{U}_r \end{bmatrix}$$

and (10-33)

$$\widetilde{\mathbf{x}}^{(2)} = [\hat{\mathbf{b}}^{(1)} \mid \hat{\mathbf{b}}^{(2)} \mid \cdots \mid \hat{\mathbf{b}}^{(r)}] \begin{bmatrix} \hat{V}_1 \\ \hat{V}_2 \\ \vdots \\ \hat{V}_r \end{bmatrix}$$

then $\mathbf{S}_{12}$ is approximated by sample Cov $(\widetilde{\mathbf{x}}^{(1)}, \widetilde{\mathbf{x}}^{(2)})$.

Continuing, we see that the *matrices of errors of approximation* are

$$\mathbf{S}_{11} - (\hat{\mathbf{a}}^{(1)}\hat{\mathbf{a}}^{(1)\prime} + \hat{\mathbf{a}}^{(2)}\hat{\mathbf{a}}^{(2)\prime} + \cdots + \hat{\mathbf{a}}^{(r)}\hat{\mathbf{a}}^{(r)\prime}) = \hat{\mathbf{a}}^{(r+1)}\hat{\mathbf{a}}^{(r+1)\prime} + \cdots + \hat{\mathbf{a}}^{(p)}\hat{\mathbf{a}}^{(p)\prime}$$

$$\mathbf{S}_{22} - (\hat{\mathbf{b}}^{(1)}\hat{\mathbf{b}}^{(1)\prime} + \hat{\mathbf{b}}^{(2)}\hat{\mathbf{b}}^{(2)\prime} + \cdots + \hat{\mathbf{b}}^{(r)}\hat{\mathbf{b}}^{(r)\prime}) = \hat{\mathbf{b}}^{(r+1)}\hat{\mathbf{b}}^{(r+1)\prime} + \cdots + \hat{\mathbf{b}}^{(q)}\hat{\mathbf{b}}^{(q)\prime}$$

$$\mathbf{S}_{12} - (\widehat{\rho_1^*}\,\hat{\mathbf{a}}^{(1)}\hat{\mathbf{b}}^{(1)\prime} + \widehat{\rho_2^*}\,\hat{\mathbf{a}}^{(2)}\hat{\mathbf{b}}^{(2)\prime} + \cdots + \widehat{\rho_r^*}\,\hat{\mathbf{a}}^{(r)}\hat{\mathbf{b}}^{(r)\prime})$$

$$= \widehat{\rho_{r+1}^*}\,\hat{\mathbf{a}}^{(r+1)}\hat{\mathbf{b}}^{(r+1)\prime} + \cdots + \widehat{\rho_p^*}\,\hat{\mathbf{a}}^{(p)}\hat{\mathbf{b}}^{(p)\prime} \quad (10\text{-}34)$$

Approximation Error Matrices May Be Interpreted as Descriptive Summaries How Well 1st r Sample Canonical Variates Reproduce Sample Covariance Matrices

The approximation error matrices (10-34) may be interpreted as descriptive summaries of how well the first r sample canonical variates reproduce the sample covariance matrices. Patterns of large entries in the rows and/or columns of the approximation error matrices indicate a poor "fit" to the corresponding variable(s).

Ordinarily, the first r variates do a better job of reproducing the elements of $\mathbf{S}_{12} = \mathbf{S}'_{21}$ than the elements of $\mathbf{S}_{11}$ or $\mathbf{S}_{22}$. Mathematically, this occurs because the residual matrix in the former case is directly related to the smallest $p - r$ sample canonical correlations. These correlations are usually all close to zero. On the other hand, the residual matrices associated with the approximations to the matrices $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ depend only on the last $p - r$ and $q - r$ coefficient vectors. The elements in these vectors may be relatively large, and hence, the residual matrices can have "large" entries.

For standardized observations, $\mathbf{R}_{kl}$ replaces $\mathbf{S}_{kl}$ and $\hat{\mathbf{a}}_z^{(k)}$, $\hat{\mathbf{b}}_z^{(l)}$ replace $\hat{\mathbf{a}}^{(k)}$, $\hat{\mathbf{b}}^{(l)}$ in (10-34).

Calculating Matrices of Errors of Approximation

Example 10.6 (Calculating matrices of errors of approximation) In Example 10.4, we obtained the canonical correlations between the two head and the two leg variables for white leghorn fowl. Starting with the sample correlation matrix

$$\mathbf{R} = \left[\begin{array}{c|c} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \hline \mathbf{R}_{21} & \mathbf{R}_{22} \end{array}\right] = \left[\begin{array}{cc|cc} 1.0 & .505 & .569 & .602 \\ .505 & 1.0 & .422 & .467 \\ \hline .569 & .422 & 1.0 & .926 \\ .602 & .467 & .926 & 1.0 \end{array}\right]$$

we obtained the two sets of canonical correlations and variables

$$\widehat{\rho_1^*} = .631 \qquad \begin{aligned} \hat{U}_1 &= .781z_1^{(1)} + .345z_2^{(1)} \\ \hat{V}_1 &= .060z_1^{(2)} + .944z_2^{(2)} \end{aligned}$$

and

$$\widehat{\rho_2^*} = .057 \qquad \begin{aligned} \hat{U}_2 &= -.856z_1^{(1)} + 1.106z_2^{(1)} \\ \hat{V}_2 &= -2.648z_1^{(2)} + 2.475z_2^{(2)} \end{aligned}$$

where $z_i^{(1)}, i = 1, 2$ and $z_i^{(2)}, i = 1, 2$ are the standardized data values for sets 1 and 2, respectively.

We first calculate (see Panel 10.1)

$$\hat{\mathbf{A}}_\mathbf{z}^{-1} = \begin{bmatrix} .781 & .345 \\ -.856 & 1.106 \end{bmatrix}^{-1} = \begin{bmatrix} .9548 & -.2974 \\ .7388 & .6739 \end{bmatrix}$$

$$\hat{\mathbf{B}}_\mathbf{z}^{-1} = \begin{bmatrix} .9343 & -.3564 \\ .9997 & .0227 \end{bmatrix}$$

Calculating Matrices of Errors of Approximation

Consequently, the matrices of errors of approximation created by using only the first canonical pair are

$$\mathbf{R}_{12} - \text{sample}\,\text{Cov}(\tilde{\mathbf{z}}^{(1)}, \tilde{\mathbf{z}}^{(2)}) = (.057)\begin{bmatrix} -.2974 \\ .6739 \end{bmatrix}\begin{bmatrix} -.3564 & .0227 \end{bmatrix}$$

$$= \begin{bmatrix} .006 & -.000 \\ -.014 & .001 \end{bmatrix}$$

$$\mathbf{R}_{11} - \text{sample}\,\text{Cov}(\tilde{\mathbf{z}}^{(1)}) = \begin{bmatrix} -.2974 \\ .6739 \end{bmatrix}\begin{bmatrix} -.2974 & .6739 \end{bmatrix}$$

$$= \begin{bmatrix} .088 & -.200 \\ -.200 & .454 \end{bmatrix}$$

$$\mathbf{R}_{22} - \text{sample}\,\text{Cov}(\tilde{\mathbf{z}}^{(2)}) = \begin{bmatrix} -.3564 \\ .0227 \end{bmatrix}\begin{bmatrix} -.3564 & .0227 \end{bmatrix}$$

$$= \begin{bmatrix} .127 & -.008 \\ -.008 & .001 \end{bmatrix}$$

where $\tilde{\mathbf{z}}^{(1)}, \tilde{\mathbf{z}}^{(2)}$ are given by (10-33) with $r = 1$ and $\hat{\mathbf{a}}_z^{(1)}, \hat{\mathbf{b}}_z^{(1)}$ replace $\hat{\mathbf{a}}^{(1)}, \hat{\mathbf{b}}^{(1)}$, respectively.

Proportions of Explained Sample Variance

We see that the first pair of canonical variables effectively summarizes (reproduces) the intraset correlations in $\mathbf{R}_{12}$. However, the individual variates are not particularly effective summaries of the sampling variability in the original $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$ sets, respectively. This is especially true for $\hat{U}_1$. ∎

Proportions of Explained Sample Variance

When the observations are standardized, the sample covariance matrices $\mathbf{S}_{kl}$ are correlation matrices $\mathbf{R}_{kl}$. The canonical coefficient vectors are the *rows* of the matrices $\hat{\mathbf{A}}_{\mathbf{z}}$ and $\hat{\mathbf{B}}_{\mathbf{z}}$ and the *columns* of $\hat{\mathbf{A}}_{\mathbf{z}}^{-1}$ and $\hat{\mathbf{B}}_{\mathbf{z}}^{-1}$ are the sample correlations between the canonical variates and their component variables.

Specifically,

$$\text{sample Cov}(\mathbf{z}^{(1)}, \hat{\mathbf{U}}) = \text{sample Cov}(\hat{\mathbf{A}}_{\mathbf{z}}^{-1}\hat{\mathbf{U}}, \hat{\mathbf{U}}) = \hat{\mathbf{A}}_{\mathbf{z}}^{-1}$$

and

$$\text{sample Cov}(\mathbf{z}^{(2)}, \hat{\mathbf{V}}) = \text{sample Cov}(\hat{\mathbf{B}}_{\mathbf{z}}^{-1}\hat{\mathbf{V}}, \hat{\mathbf{V}}) = \hat{\mathbf{B}}_{\mathbf{z}}^{-1}$$

Total Standardized Variance in Each Set

so

$$\hat{\mathbf{A}}_{\mathbf{z}}^{-1} = [\hat{\mathbf{a}}_{\mathbf{z}}^{(1)}, \hat{\mathbf{a}}_{\mathbf{z}}^{(2)}, \ldots, \hat{\mathbf{a}}_{\mathbf{z}}^{(p)}] = \begin{bmatrix} r_{\hat{U}_1, z_1^{(1)}} & r_{\hat{U}_2, z_1^{(1)}} & \cdots & r_{\hat{U}_p, z_1^{(1)}} \\ r_{\hat{U}_1, z_2^{(1)}} & r_{\hat{U}_2, z_2^{(1)}} & \cdots & r_{\hat{U}_p, z_2^{(1)}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\hat{U}_1, z_p^{(1)}} & r_{\hat{U}_2, z_p^{(1)}} & \cdots & r_{\hat{U}_p, z_p^{(1)}} \end{bmatrix}$$

$$\hat{\mathbf{B}}_{\mathbf{z}}^{-1} = [\hat{\mathbf{b}}_{\mathbf{z}}^{(1)}, \hat{\mathbf{b}}_{\mathbf{z}}^{(2)}, \ldots, \hat{\mathbf{b}}_{\mathbf{z}}^{(q)}] = \begin{bmatrix} r_{\hat{V}_1, z_1^{(2)}} & r_{\hat{V}_2, z_1^{(2)}} & \cdots & r_{\hat{V}_q, z_1^{(2)}} \\ r_{\hat{V}_1, z_2^{(2)}} & r_{\hat{V}_2, z_2^{(2)}} & \cdots & r_{\hat{V}_q, z_2^{(2)}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\hat{V}_1, z_q^{(2)}} & r_{\hat{V}_2, z_q^{(2)}} & \cdots & r_{\hat{V}_q, z_q^{(2)}} \end{bmatrix} \quad (10\text{-}35)$$

where $r_{\hat{U}_i, z_k^{(1)}}$ and $r_{\hat{V}_i, z_k^{(2)}}$ are the sample correlation coefficients between the quantities with subscripts.

Using (10-32) with standardized observations, we obtain

Total (standardized) sample variance in first set

$$= \operatorname{tr}(\mathbf{R}_{11}) = \operatorname{tr}(\hat{\mathbf{a}}_{\mathbf{z}}^{(1)}\hat{\mathbf{a}}_{\mathbf{z}}^{(1)\prime} + \hat{\mathbf{a}}_{\mathbf{z}}^{(2)}\hat{\mathbf{a}}_{\mathbf{z}}^{(2)\prime} + \cdots + \hat{\mathbf{a}}_{\mathbf{z}}^{(p)}\hat{\mathbf{a}}_{\mathbf{z}}^{(p)\prime}) = p \quad (10\text{-}36a)$$

Total (standardized) sample variance in second set

$$= \operatorname{tr}(\mathbf{R}_{22}) = \operatorname{tr}(\hat{\mathbf{b}}_{\mathbf{z}}^{(1)}\hat{\mathbf{b}}_{\mathbf{z}}^{(1)\prime} + \hat{\mathbf{b}}_{\mathbf{z}}^{(2)}\hat{\mathbf{b}}_{\mathbf{z}}^{(2)\prime} + \cdots + \hat{\mathbf{b}}_{\mathbf{z}}^{(q)}\hat{\mathbf{b}}_{\mathbf{z}}^{(q)\prime}) = q \quad (10\text{-}36b)$$

Proportions of Total Standardized Sample Variance Explained by 1st Canonical Variates

the contributions of the first r canonical variates to the total (standardized) sample variances as

$$\operatorname{tr}\left(\hat{\mathbf{a}}_{\mathbf{z}}^{(1)}\hat{\mathbf{a}}_{\mathbf{z}}^{(1)\prime} + \hat{\mathbf{a}}_{\mathbf{z}}^{(2)}\hat{\mathbf{a}}_{\mathbf{z}}^{(2)\prime} + \cdots + \hat{\mathbf{a}}_{\mathbf{z}}^{(r)}\hat{\mathbf{a}}_{\mathbf{z}}^{(r)\prime}\right) = \sum_{i=1}^{r}\sum_{k=1}^{p} r_{\hat{U}_i, z_k^{(1)}}^2$$

and

$$\operatorname{tr}\left(\hat{\mathbf{b}}_{\mathbf{z}}^{(1)}\hat{\mathbf{b}}_{\mathbf{z}}^{(1)\prime} + \hat{\mathbf{b}}_{\mathbf{z}}^{(2)}\hat{\mathbf{b}}_{\mathbf{z}}^{(2)\prime} + \cdots + \hat{\mathbf{b}}_{\mathbf{z}}^{(r)}\hat{\mathbf{b}}_{\mathbf{z}}^{(r)\prime}\right) = \sum_{i=1}^{r}\sum_{k=1}^{p} r_{\hat{V}_i, z_k^{(2)}}^2$$

The *proportions* of total (standardized) sample variances "explained by" the first canonical variates then become

$$R^2_{\mathbf{z}^{(1)}|\hat{U}_1,\hat{U}_2,\ldots,\hat{U}_r} = \begin{pmatrix}\text{proportion of total standardized} \\ \text{sample variance in first set} \\ \text{explained by } \hat{U}_1, \hat{U}_2, \ldots, \hat{U}_r\end{pmatrix}$$

$$= \frac{\operatorname{tr}\left(\hat{\mathbf{a}}_{\mathbf{z}}^{(1)}\hat{\mathbf{a}}_{\mathbf{z}}^{(1)\prime} + \cdots + \hat{\mathbf{a}}_{\mathbf{z}}^{(r)}\hat{\mathbf{a}}_{\mathbf{z}}^{(r)\prime}\right)}{\operatorname{tr}(\mathbf{R}_{11})}$$

$$= \frac{\sum_{i=1}^{r}\sum_{k=1}^{p} r_{\hat{U}_i, z_k^{(1)}}^2}{p} \tag{10-37}$$

How Well Do Canonical Variates Represent Their Respective Sets: Evidenced by Matrices of Errors

and

$$R^2_{\mathbf{z}^{(2)}|\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_r} = \begin{pmatrix} \text{proportion of total standardized} \\ \text{sample variance in second set} \\ \text{explained by } \hat{V}_1, \hat{V}_2, \ldots, \hat{V}_r \end{pmatrix}$$

$$= \frac{\operatorname{tr}\left(\hat{\mathbf{b}}_{\mathbf{z}}^{(1)}\hat{\mathbf{b}}_{\mathbf{z}}^{(1)\prime} + \cdots + \hat{\mathbf{b}}_{\mathbf{z}}^{(r)}\hat{\mathbf{b}}_{\mathbf{z}}^{(r)\prime}\right)}{\operatorname{tr}\left(\mathbf{R}_{22}\right)}$$

$$= \frac{\sum_{i=1}^{r} \sum_{k=1}^{q} r^2_{\hat{V}_i, z_k^{(2)}}}{q}$$

Descriptive measures (10-37) provide some indication of how well the canonical variates represent their respective sets. They provide single-number descriptions of the matrices of errors. In particular,

$$\frac{1}{p}\operatorname{tr}\left[\mathbf{R}_{11} - \hat{\mathbf{a}}_{\mathbf{z}}^{(1)}\hat{\mathbf{a}}_{\mathbf{z}}^{(1)\prime} - \hat{\mathbf{a}}_{\mathbf{z}}^{(2)}\hat{\mathbf{a}}_{\mathbf{z}}^{(2)\prime} - \cdots - \hat{\mathbf{a}}_{\mathbf{z}}^{(r)}\hat{\mathbf{a}}_{\mathbf{z}}^{(r)\prime}\right] = 1 - R^2_{\mathbf{z}^{(1)}|\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_r}$$

$$\frac{1}{q}\operatorname{tr}\left[\mathbf{R}_{22} - \hat{\mathbf{b}}_{\mathbf{z}}^{(1)}\hat{\mathbf{b}}_{\mathbf{z}}^{(1)\prime} - \hat{\mathbf{b}}_{\mathbf{z}}^{(2)}\hat{\mathbf{b}}_{\mathbf{z}}^{(2)\prime} - \cdots - \hat{\mathbf{b}}_{\mathbf{z}}^{(r)}\hat{\mathbf{b}}_{\mathbf{z}}^{(r)\prime}\right] = 1 - R^2_{\mathbf{z}^{(2)}|\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_r}$$

according to (10-36) and (10-37).

Calculating Proportions of Sample Variance Explained by Canonical Variates

Example 10.7 (Calculating proportions of sample variance explained by canonical variates) Consider the job characteristic–job satisfaction data discussed in Example 10.5. Using the table of sample correlation coefficients presented in that example, we find that

$$R^2_{\mathbf{z}^{(1)}|\hat{U}_1} = \frac{1}{5}\sum_{k=1}^{5} r^2_{\hat{U}_1, z_k^{(1)}} = \frac{1}{5}[(.83)^2 + (.74)^2 + \cdots + (.85)^2] = .58$$

$$R^2_{\mathbf{z}^{(2)}|\hat{V}_1} = \frac{1}{7}\sum_{k=1}^{7} r^2_{\hat{V}_1, z_k^{(2)}} = \frac{1}{7}[(.75)^2 + (.65)^2 + \cdots + (.50)^2] = .37$$

The first sample canonical variate $\hat{U}_1$ of the job characteristics set accounts for 58% of the set's total sample variance. The first sample canonical variate $\hat{V}_1$ of the job satisfaction set explains 37% of the set's total sample variance. We might thus infer that $\hat{U}_1$ is a "better" representative of its set than $\hat{V}_1$ is of its set. The interested reader may wish to see how well $\hat{U}_1$ and $\hat{V}_1$ reproduce the correlation matrices $\mathbf{R}_{11}$ and $\mathbf{R}_{22}$, respectively. [See (10-29).] ■