Canonical Correlation Analysis

Nathaniel E. Helwig

Associate Professor of Psychology and Statistics
University of Minnesota

April 7, 2024

Table of Contents

Content adapted from:
Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis (6th ed).

Table of Contents

Purpose of Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) connects *two sets of variables* by finding linear combinations of variables that maximally correlate.

There are two typical purposes of CCA:

1. Data reduction: explain covariation between two sets of variables using small number of linear combinations
2. Data interpretation: find features (i.e., canonical variates) that are important for explaining covariation between sets of variables

Linear Combinations of Two Sets of Variables

Let $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_q)^\top$ denote random vectors with mean vectors $\boldsymbol{\mu}_X$ and $\boldsymbol{\mu}_Y$ and covariance matrices $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$.

Let $\boldsymbol{Z}^\top = (\boldsymbol{X}^\top, \boldsymbol{Y}^\top)$ and note $\boldsymbol{Z} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}^\top = (\boldsymbol{\mu}_X^\top, \boldsymbol{\mu}_Y^\top)$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_X & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_Y \end{pmatrix}$$

where $\boldsymbol{\Sigma}_{XY} = E[(\boldsymbol{X} - \boldsymbol{\mu}_X)(\boldsymbol{Y} - \boldsymbol{\mu}_Y)^\top]$ is the covariance of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Define new variables U and V via linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$

$$U = \mathbf{a}^\top \boldsymbol{X}$$
$$V = \mathbf{b}^\top \boldsymbol{Y}$$

Defining Canonical Variates (and Correlations)

Note that $U = \mathbf{a}^\top \boldsymbol{X}$ and $V = \mathbf{b}^\top \boldsymbol{Y}$ have properties

$$\begin{aligned} \text{Var}(U) &= \mathbf{a}^\top \boldsymbol{\Sigma}_X \mathbf{a} \\ \text{Var}(V) &= \mathbf{b}^\top \boldsymbol{\Sigma}_Y \mathbf{b} \\ \text{Cov}(U, V) &= \mathbf{a}^\top \boldsymbol{\Sigma}_{XY} \mathbf{b} \end{aligned}$$

The first pair of canonical variates (U_1, V_1) is defined via the pair of linear combination vectors $\{\mathbf{a}_1, \mathbf{b}_1\}$ that maximize

$$\text{Cor}(U, V) = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U)}\sqrt{\text{Var}(V)}} = \frac{\mathbf{a}^\top \boldsymbol{\Sigma}_{XY} \mathbf{b}}{\sqrt{\mathbf{a}^\top \boldsymbol{\Sigma}_X \mathbf{a}}\sqrt{\mathbf{b}^\top \boldsymbol{\Sigma}_Y \mathbf{b}}}$$

subject to U_1 and V_1 having unit variance.

Remaining canonical variates (U_ℓ, V_ℓ) maximize the above subject to having unit variance and being uncorrelated with (U_k, V_k) for all $k < \ell$.

Computing Canonical Variates (and Correlations)

The k-th pair of canonical variates is given by

$$U_k = \underbrace{\mathbf{u}_k^\top \boldsymbol{\Sigma}_X^{-1/2}}_{\mathbf{a}_k^\top} \boldsymbol{X} \quad \text{and} \quad V_k = \underbrace{\mathbf{v}_k^\top \boldsymbol{\Sigma}_Y^{-1/2}}_{\mathbf{b}_k^\top} \boldsymbol{Y}$$

where

- $\mathbf{u}_k$ is the k-th eigenvector of $\boldsymbol{\Sigma}_X^{-1/2}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_X^{-1/2}$
- $\mathbf{v}_k$ is the k-th eigenvector of $\boldsymbol{\Sigma}_Y^{-1/2}\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_X^{-1}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1/2}$

The k-th canonical correlation is given by

$$\text{Cor}(U_k, V_k) = \rho_k$$

where ρ_k^2 is the k-th eigenvalue of $\boldsymbol{\Sigma}_X^{-1/2}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_X^{-1/2}$
[ρ_k^2 is also the k-th eigenvalue of $\boldsymbol{\Sigma}_Y^{-1/2}\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_X^{-1}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1/2}$]

Covariance of Original and Canonical Variables

$\boldsymbol{U} = \mathbf{A}^\top \boldsymbol{X}$ and $\boldsymbol{V} = \mathbf{B}^\top \boldsymbol{Y}$ with $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_p]$ and $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_q]$.

- $\boldsymbol{U} = (U_1, \ldots, U_p)^\top$ contains the p canonical variates from $\boldsymbol{X}$
- $\boldsymbol{V} = (V_1, \ldots, V_q)^\top$ contains the q canonical variates from $\boldsymbol{Y}$
- If $p \leq q$, we are interested in first p canonical variates from $\boldsymbol{Y}$

The canonical variates and original variables have covariance matrices

$$\begin{aligned}
\text{Cov}(\boldsymbol{U}, \boldsymbol{X}) &= \text{Cov}(\mathbf{A}^\top \boldsymbol{X}, \boldsymbol{X}) = \mathbf{A}^\top \boldsymbol{\Sigma}_X \\
\text{Cov}(\boldsymbol{U}, \boldsymbol{Y}) &= \text{Cov}(\mathbf{A}^\top \boldsymbol{X}, \boldsymbol{Y}) = \mathbf{A}^\top \boldsymbol{\Sigma}_{XY} \\
\text{Cov}(\boldsymbol{V}, \boldsymbol{X}) &= \text{Cov}(\mathbf{B}^\top \boldsymbol{Y}, \boldsymbol{X}) = \mathbf{B}^\top \boldsymbol{\Sigma}_{YX} \\
\text{Cov}(\boldsymbol{V}, \boldsymbol{Y}) &= \text{Cov}(\mathbf{B}^\top \boldsymbol{Y}, \boldsymbol{Y}) = \mathbf{B}^\top \boldsymbol{\Sigma}_Y
\end{aligned}$$

Correlation of Original and Canonical Variables

The canonical variates and original variables have correlation matrices

$$\text{Cor}(\boldsymbol{U}, \boldsymbol{X}) = \text{Cov}(\mathbf{A}^\top \boldsymbol{X}, \tilde{\boldsymbol{\Sigma}}_X^{-1/2} \boldsymbol{X}) = \mathbf{A}^\top \boldsymbol{\Sigma}_X \tilde{\boldsymbol{\Sigma}}_X^{-1/2}$$
$$\text{Cor}(\boldsymbol{U}, \boldsymbol{Y}) = \text{Cov}(\mathbf{A}^\top \boldsymbol{X}, \tilde{\boldsymbol{\Sigma}}_Y^{-1/2} \boldsymbol{Y}) = \mathbf{A}^\top \boldsymbol{\Sigma}_{XY} \tilde{\boldsymbol{\Sigma}}_Y^{-1/2}$$
$$\text{Cor}(\boldsymbol{V}, \boldsymbol{X}) = \text{Cov}(\mathbf{B}^\top \boldsymbol{Y}, \tilde{\boldsymbol{\Sigma}}_X^{-1/2} \boldsymbol{X}) = \mathbf{B}^\top \boldsymbol{\Sigma}_{YX} \tilde{\boldsymbol{\Sigma}}_X^{-1/2}$$
$$\text{Cor}(\boldsymbol{V}, \boldsymbol{Y}) = \text{Cov}(\mathbf{B}^\top \boldsymbol{Y}, \tilde{\boldsymbol{\Sigma}}_Y^{-1/2} \boldsymbol{Y}) = \mathbf{B}^\top \boldsymbol{\Sigma}_Y \tilde{\boldsymbol{\Sigma}}_Y^{-1/2}$$

given that $\text{Var}(U_k) = \text{Var}(V_\ell) = 1$ for all k, ℓ.

- $\tilde{\boldsymbol{\Sigma}}_X = \text{diag}(\boldsymbol{\Sigma}_X)$ is a diagonal matrix containing $\boldsymbol{X}$ variances
- $\tilde{\boldsymbol{\Sigma}}_Y = \text{diag}(\boldsymbol{\Sigma}_Y)$ is a diagonal matrix containing $\boldsymbol{Y}$ variances

Canonical Variates and Summarizing Variability

The linear transformations $\boldsymbol{U} = \mathbf{A}^\top \boldsymbol{X}$ and $\boldsymbol{V} = \mathbf{B}^\top \boldsymbol{Y}$ are defined to maximize the correlations between the canonical variables.

NOT the same as maximizing the explained variance in $\boldsymbol{\Sigma}_X$ or $\boldsymbol{\Sigma}_Y$.

If the first few pairs of canonical variables do not well explain the variability in $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$, then the interpretation becomes less clear.

Moving to the Sample Situation

Assume that $\mathbf{z}_i = (\mathbf{x}_i^\top, \mathbf{y}_i^\top)^\top \overset{\text{iid}}{\sim} \text{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_X & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_Y \end{pmatrix}$$

and let the sample mean vector and covariance matrix be denoted by

$$\bar{\mathbf{z}} = \begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} \mathbf{S}_X & \mathbf{S}_{XY} \\ \mathbf{S}_{YX} & \mathbf{S}_Y \end{pmatrix}$$

where

- $\bar{\mathbf{x}} = (1/n)\sum_{i=1}^n \mathbf{x}_i \quad \text{and} \quad \bar{\mathbf{y}} = (1/n)\sum_{i=1}^n \mathbf{y}_i$
- $\mathbf{S}_X = \frac{1}{n-1}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$
- $\mathbf{S}_Y = \frac{1}{n-1}\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top$
- $\mathbf{S}_{XY} = \mathbf{S}_{YX}^\top = \frac{1}{n-1}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top$

Defining Canonical Variates (and Correlations)

Note that $U = \mathbf{a}^\top \boldsymbol{X}$ and $V = \mathbf{b}^\top \boldsymbol{Y}$ have sample properties

$$\widehat{\text{Var}}(U) = \mathbf{a}^\top \mathbf{S}_X \mathbf{a}$$
$$\widehat{\text{Var}}(V) = \mathbf{b}^\top \mathbf{S}_Y \mathbf{b}$$
$$\widehat{\text{Cov}}(U, V) = \mathbf{a}^\top \mathbf{S}_{XY} \mathbf{b}$$

The first pair of sample canonical variates (U_1, V_1) is defined via the pair of linear combination vectors $\{\mathbf{a}_1, \mathbf{b}_1\}$ that maximize

$$\widehat{\text{Cor}}(U, V) = \frac{\widehat{\text{Cov}}(U, V)}{\sqrt{\widehat{\text{Var}}(U)}\sqrt{\widehat{\text{Var}}(V)}} = \frac{\mathbf{a}^\top \mathbf{S}_{XY} \mathbf{b}}{\sqrt{\mathbf{a}^\top \mathbf{S}_X \mathbf{a}}\sqrt{\mathbf{b}^\top \mathbf{S}_Y \mathbf{b}}}$$

subject to U_1 and V_1 having unit variance.

Remaining canonical variates (U_ℓ, V_ℓ) maximize the above subject to having unit variance and being uncorrelated with (U_k, V_k) for all $k < \ell$.

Calculating Canonical Variates (and Correlations)

The sample estimate of the k-th pair of canonical variates is given by

$$\hat{U}_k = \underbrace{\hat{\mathbf{u}}_k^\top \mathbf{S}_X^{-1/2}}_{\hat{\mathbf{a}}_k^\top} \boldsymbol{X} \quad \text{and} \quad \hat{V}_k = \underbrace{\hat{\mathbf{v}}_k \mathbf{S}_Y^{-1/2}}_{\hat{\mathbf{b}}_k^\top} \boldsymbol{Y}$$

where

- $\hat{\mathbf{u}}_k$ is the k-th eigenvector of $\mathbf{S}_X^{-1/2}\mathbf{S}_{XY}\mathbf{S}_Y^{-1}\mathbf{S}_{YX}\mathbf{S}_X^{-1/2}$
- $\hat{\mathbf{v}}_k$ is the k-th eigenvector of $\mathbf{S}_Y^{-1/2}\mathbf{S}_{YX}\mathbf{S}_X^{-1}\mathbf{S}_{XY}\mathbf{S}_Y^{-1/2}$

The sample estimate of the k-th canonical correlation is given by

$$\widehat{\text{Cor}}(U_k, V_k) = \hat{\rho}_k$$

where $\hat{\rho}_k^2$ is the k-th eigenvalue of $\mathbf{S}_X^{-1/2}\mathbf{S}_{XY}\mathbf{S}_Y^{-1}\mathbf{S}_{YX}\mathbf{S}_X^{-1/2}$
[$\hat{\rho}_k^2$ is also the k-th eigenvalue of $\mathbf{S}_Y^{-1/2}\mathbf{S}_{YX}\mathbf{S}_X^{-1}\mathbf{S}_{XY}\mathbf{S}_Y^{-1/2}$]

Covariance of Original and Canonical Variables

$\hat{\boldsymbol{U}} = \hat{\mathbf{A}}^\top \boldsymbol{X}$ and $\hat{\boldsymbol{V}} = \hat{\mathbf{B}}^\top \boldsymbol{Y}$ with $\hat{\mathbf{A}} = [\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_p]$ and $\hat{\mathbf{B}} = [\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_q]$.

- $\hat{\boldsymbol{U}} = (\hat{U}_1, \ldots, \hat{U}_p)^\top$ contains the p canonical variates from $\boldsymbol{X}$
- $\hat{\boldsymbol{V}} = (\hat{V}_1, \ldots, \hat{V}_q)^\top$ contains the q canonical variates from $\boldsymbol{Y}$
- If $p \leq q$, we are interested in first p canonical variates from $\boldsymbol{Y}$

The sample canonical variates and original variables have covariances

$$\widehat{\text{Cov}}(\hat{\boldsymbol{U}}, \boldsymbol{X}) = \widehat{\text{Cov}}(\hat{\mathbf{A}}^\top \boldsymbol{X}, \boldsymbol{X}) = \hat{\mathbf{A}}^\top \mathbf{S}_X$$
$$\widehat{\text{Cov}}(\hat{\boldsymbol{U}}, \boldsymbol{Y}) = \widehat{\text{Cov}}(\hat{\mathbf{A}}^\top \boldsymbol{X}, \boldsymbol{Y}) = \hat{\mathbf{A}}^\top \mathbf{S}_{XY}$$
$$\widehat{\text{Cov}}(\hat{\boldsymbol{V}}, \boldsymbol{X}) = \widehat{\text{Cov}}(\hat{\mathbf{B}}^\top \boldsymbol{Y}, \boldsymbol{X}) = \hat{\mathbf{B}}^\top \mathbf{S}_{YX}$$
$$\widehat{\text{Cov}}(\hat{\boldsymbol{V}}, \boldsymbol{Y}) = \widehat{\text{Cov}}(\hat{\mathbf{B}}^\top \boldsymbol{Y}, \boldsymbol{Y}) = \hat{\mathbf{B}}^\top \mathbf{S}_Y$$

Correlation of Original and Canonical Variables

The sample canonical variates and original variables have correlations

$$\widehat{\text{Cor}}(\hat{\boldsymbol{U}}, \boldsymbol{X}) = \widehat{\text{Cov}}(\hat{\mathbf{A}}^\top \boldsymbol{X}, \tilde{\mathbf{S}}_X^{-1/2} \boldsymbol{X}) = \hat{\mathbf{A}}^\top \mathbf{S}_X \tilde{\mathbf{S}}_X^{-1/2}$$
$$\widehat{\text{Cor}}(\hat{\boldsymbol{U}}, \boldsymbol{Y}) = \widehat{\text{Cov}}(\hat{\mathbf{A}}^\top \boldsymbol{X}, \tilde{\mathbf{S}}_Y^{-1/2} \boldsymbol{Y}) = \hat{\mathbf{A}}^\top \mathbf{S}_{XY} \tilde{\mathbf{S}}_Y^{-1/2}$$
$$\widehat{\text{Cor}}(\hat{\boldsymbol{V}}, \boldsymbol{X}) = \widehat{\text{Cov}}(\hat{\mathbf{B}}^\top \boldsymbol{Y}, \tilde{\mathbf{S}}_X^{-1/2} \boldsymbol{X}) = \hat{\mathbf{B}}^\top \mathbf{S}_{YX} \tilde{\mathbf{S}}_X^{-1/2}$$
$$\widehat{\text{Cor}}(\hat{\boldsymbol{V}}, \boldsymbol{Y}) = \widehat{\text{Cov}}(\hat{\mathbf{B}}^\top \boldsymbol{Y}, \tilde{\mathbf{S}}_Y^{-1/2} \boldsymbol{Y}) = \hat{\mathbf{B}}^\top \mathbf{S}_Y \tilde{\mathbf{S}}_Y^{-1/2}$$

given that $\text{Var}(\hat{U}_k) = \text{Var}(\hat{V}_\ell) = 1$ for all k, ℓ.

- $\tilde{\mathbf{S}}_X = \text{diag}(\mathbf{S}_X)$ is a diagonal matrix containing $\boldsymbol{X}$ variances
- $\tilde{\mathbf{S}}_Y = \text{diag}(\mathbf{S}_Y)$ is a diagonal matrix containing $\boldsymbol{Y}$ variances

Covariance Matrix Implied by CCA for $\boldsymbol{X}$

Note that we have the following properties

$$\widehat{\text{Cov}}(\hat{\mathbf{U}}) = \hat{\mathbf{A}}^\top \mathbf{S}_X \hat{\mathbf{A}} = \mathbf{I}_p$$

This implies that we can write

$$\begin{aligned}
\hat{\mathbf{A}}^\top \mathbf{S}_X \hat{\mathbf{A}} &= \mathbf{I}_p \\
(\hat{\mathbf{A}}^\top)^{-1} \hat{\mathbf{A}}^\top \mathbf{S}_X \hat{\mathbf{A}} (\hat{\mathbf{A}}^{-1}) &= (\hat{\mathbf{A}}^\top)^{-1} (\hat{\mathbf{A}}^{-1}) \\
\mathbf{S}_X &= (\hat{\mathbf{A}}^{-1})^\top (\hat{\mathbf{A}}^{-1}) \\
&= \sum_{j=1}^{p} (\hat{\mathbf{a}}^{(j)})(\hat{\mathbf{a}}^{(j)})^\top
\end{aligned}$$

where $\mathbf{a}^{(j)}$ denotes the j-th column of $(\hat{\mathbf{A}}^{-1})^\top$.

Covariance Matrix Implied by CCA for $\boldsymbol{Y}$

Note that we have the following properties

$$\widehat{\text{Cov}}(\hat{\mathbf{V}}) = \hat{\mathbf{B}}^\top \mathbf{S}_Y \hat{\mathbf{B}} = \mathbf{I}_q$$

This implies that we can write

$$\begin{aligned}
\hat{\mathbf{B}}^\top \mathbf{S}_Y \hat{\mathbf{B}} &= \mathbf{I}_q \\
(\hat{\mathbf{B}}^\top)^{-1} \hat{\mathbf{B}}^\top \mathbf{S}_Y \hat{\mathbf{B}} (\hat{\mathbf{B}}^{-1}) &= (\hat{\mathbf{B}}^\top)^{-1} (\hat{\mathbf{B}}^{-1}) \\
\mathbf{S}_Y &= (\hat{\mathbf{B}}^{-1})^\top (\hat{\mathbf{B}}^{-1}) \\
&= \sum_{j=1}^{q} (\hat{\mathbf{b}}^{(j)})(\hat{\mathbf{b}}^{(j)})^\top
\end{aligned}$$

where $\mathbf{b}^{(j)}$ denotes the j-th column of $(\hat{\mathbf{B}}^{-1})^\top$.

Covariance Matrix Implied by CCA for $(\boldsymbol{X}, \boldsymbol{Y})$

Note that we have the following properties (assuming $p < q$)

$$\widehat{\text{Cov}}(\hat{\mathbf{U}}, \hat{\mathbf{V}}) = \hat{\mathbf{A}}^\top \mathbf{S}_{XY} \hat{\mathbf{B}} = \begin{pmatrix} \hat{\rho}_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \hat{\rho}_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \hat{\rho}_p & 0 & \cdots & 0 \end{pmatrix} = \hat{\boldsymbol{\rho}}$$

This implies that we can write

$$\begin{aligned} \hat{\mathbf{A}}^\top \mathbf{S}_{XY} \hat{\mathbf{B}} &= \hat{\boldsymbol{\rho}} \\ (\hat{\mathbf{A}}^\top)^{-1} \hat{\mathbf{A}}^\top \mathbf{S}_{XY} \hat{\mathbf{B}} (\hat{\mathbf{B}}^{-1}) &= (\hat{\mathbf{A}}^\top)^{-1} \hat{\boldsymbol{\rho}} (\hat{\mathbf{B}}^{-1}) \\ \mathbf{S}_{XY} &= (\hat{\mathbf{A}}^{-1})^\top \hat{\boldsymbol{\rho}} (\hat{\mathbf{B}}^{-1}) \\ &= \sum_{j=1}^{p} \hat{\rho}_j (\hat{\mathbf{a}}^{(j)}) (\hat{\mathbf{b}}^{(j)})^\top \end{aligned}$$

CCA Error of Approximation Matrices

Using $r < p$ canonical variates, the approximation error matrices are

$$\mathbf{E}_X = \mathbf{S}_X - \sum_{j=1}^{r} (\hat{\mathbf{a}}^{(j)})(\hat{\mathbf{a}}^{(j)})^\top = \sum_{k=r+1}^{p} (\hat{\mathbf{a}}^{(k)})(\hat{\mathbf{a}}^{(k)})^\top$$

$$\mathbf{E}_Y = \mathbf{S}_Y - \sum_{j=1}^{r} (\hat{\mathbf{b}}^{(j)})(\hat{\mathbf{b}}^{(j)})^\top = \sum_{k=r+1}^{q} (\hat{\mathbf{b}}^{(k)})(\hat{\mathbf{b}}^{(k)})^\top$$

$$\mathbf{E}_{XY} = \mathbf{S}_{XY} - \sum_{j=1}^{r} \hat{\rho}_j (\hat{\mathbf{a}}^{(j)})(\hat{\mathbf{b}}^{(j)})^\top = \sum_{k=r+1}^{q} \hat{\rho}_k (\hat{\mathbf{a}}^{(k)})(\hat{\mathbf{b}}^{(k)})^\top$$

The error matrices provide a descriptive measure of how well the first r pairs of canonical variates explain the covariation in the data.

Likelihood Ratio Test: Is CCA Worthwhile?

If $\mathbf{\Sigma}_{XY} = \mathbf{0}_{p\times q}$, then $\text{Cov}(U, V) = \mathbf{a}^\top \mathbf{\Sigma}_{12}\mathbf{b} = 0$ for any $(\mathbf{a}, \mathbf{b})$.

- Implies that all canonical correlations must be zero
- Then there is no point in pursuing CCA

For large n, we reject $H_0 : \mathbf{\Sigma}_{XY} = \mathbf{0}_{p\times q}$ in favor of $H_1 : \mathbf{\Sigma}_{XY} \neq \mathbf{0}_{p\times q}$ if

$$-2\ln(\Lambda) = n\ln\left(\frac{|\mathbf{S}_X||\mathbf{S}_Y|}{|\mathbf{S}|}\right) = -n\sum_{j=1}^{p}\ln(1-\hat{\rho}_j^2)$$

is larger than $\chi^2_{pq}(\alpha)$.

For an improvement to the χ^2 approximation, Bartlett suggested replacing the scaling factor of n by $n - 1 - (1/2)(p + q + 1)$

$$-2\ln(\Lambda) \approx -[n - 1 - (1/2)(p + q + 1)]\sum_{j=1}^{p}\ln(1-\hat{\rho}_j^2)$$

Table of Contents

Men's Olympic Decathlon Data from 1988

Data from men's 1988 Olympic decathlon

- Total of $n = 34$ athletes
- Have $p = 10$ variables giving score for each decathlon event
- Have overall decathlon score also (`score`)

```
> decathlon[1:9,]
         run100 long.jump  shot high.jump run400 hurdle discus pole.vault javelin run1500 score
Schenk    11.25      7.43 15.48      2.27  48.90  15.13  49.28        4.7   61.32  268.95  8488
Voss      10.87      7.45 14.97      1.97  47.71  14.46  44.36        5.1   61.76  273.02  8399
Steen     11.18      7.44 14.20      1.97  48.29  14.81  43.66        5.2   64.16  263.20  8328
Thompson  10.62      7.38 15.02      2.03  49.06  14.72  44.80        4.9   64.04  285.11  8306
Blondel   11.02      7.43 12.92      1.97  47.44  14.40  41.20        5.2   57.46  256.64  8286
Plaziat   10.83      7.72 13.58      2.12  48.34  14.18  43.06        4.9   52.18  274.07  8272
Bright    11.18      7.05 14.12      2.06  49.34  14.39  41.68        5.7   61.60  291.20  8216
De.Wit    11.05      6.95 15.34      2.00  48.21  14.36  41.32        4.8   63.00  265.86  8189
Johnson   11.15      7.12 14.52      2.03  49.15  14.66  42.36        4.9   66.46  269.62  8180
```

Resigning Running Events

For the running events (`run100`, `run400`, `run1500`, and `hurdle`), lower scores correspond to better performance, whereas higher scores represent better performance for other events.

To make interpretation simpler, we will resign the running events:

```
> decathlon[,c(1,5,6,10)] <- (-1)*decathlon[,c(1,5,6,10)]
> decathlon[1:9,]
         run100 long.jump  shot high.jump run400 hurdle discus pole.vault javelin run1500 score
Schenk   -11.25      7.43 15.48      2.27 -48.90 -15.13  49.28        4.7   61.32 -268.95  8488
Voss     -10.87      7.45 14.97      1.97 -47.71 -14.46  44.36        5.1   61.76 -273.02  8399
Steen    -11.18      7.44 14.20      1.97 -48.29 -14.81  43.66        5.2   64.16 -263.20  8328
Thompson -10.62      7.38 15.02      2.03 -49.06 -14.72  44.80        4.9   64.04 -285.11  8306
Blondel  -11.02      7.43 12.92      1.97 -47.44 -14.40  41.20        5.2   57.46 -256.64  8286
Plaziat  -10.83      7.72 13.58      2.12 -48.34 -14.18  43.06        4.9   52.18 -274.07  8272
Bright   -11.18      7.05 14.12      2.06 -49.34 -14.39  41.68        5.7   61.60 -291.20  8216
De.Wit   -11.05      6.95 15.34      2.00 -48.21 -14.36  41.32        4.8   63.00 -265.86  8189
Johnson  -11.15      7.12 14.52      2.03 -49.15 -14.66  42.36        4.9   66.46 -269.62  8180
```

Split Events into Two Sets: Arms versus Legs

We will split the decathlon events into two different sets:

- $\boldsymbol{X}$: shot, discus, javelin, pole.vault
- $\boldsymbol{Y}$: run100, run400, run1500, hurdle, long.jump, high.jump

Note that $\boldsymbol{X}$ are "arm events" (throwing/vaulting), whereas $\boldsymbol{Y}$ are "leg events" (running/jumping).

R code to split the `decathlon` data into two sets of events

```
> X <- as.matrix(decathlon[,c("shot", "discus", "javelin", "pole.vault")])
> Y <- as.matrix(decathlon[,c("run100", "run400", "run1500", "hurdle", "long.jump", "high.jump")])
> n <- nrow(X)       # n = 34
> p <- ncol(X)       # p = 4
> q <- ncol(Y)       # q = 6
```

CCA of Decathlon Data in R

R code to conduct canonical correlation analysis

```
# canonical correlations of covariance (unstandardized data)
> cca <- cancor(X, Y)

# cca (the normal way)
> Sx <- cov(X)
> Sy <- cov(Y)
> Sxy <- cov(X,Y)
> Sxeig <- eigen(Sx, symmetric=TRUE)
> Sxisqrt <- Sxeig$vectors %*% diag(1/sqrt(Sxeig$values)) %*% t(Sxeig$vectors)
> Syeig <- eigen(Sy, symmetric=TRUE)
> Syisqrt <- Syeig$vectors %*% diag(1/sqrt(Syeig$values)) %*% t(Syeig$vectors)
> Xmat <- Sxisqrt %*% Sxy %*% solve(Sy) %*% t(Sxy) %*% Sxisqrt
> Ymat <- Syisqrt %*% t(Sxy) %*% solve(Sx) %*% Sxy %*% Syisqrt
> Xeig <- eigen(Xmat, symmetric=TRUE)
> Yeig <- eigen(Ymat, symmetric=TRUE)

# compare correlations (same)
> cca$cor
[1] 0.7702006 0.5033532 0.4184145 0.3052556
> rho <- sqrt(Xeig$values)
> rho
[1] 0.7702006 0.5033532 0.4184145 0.3052556
> sqrt(Yeig$values[1:p])
[1] 0.7702006 0.5033532 0.4184145 0.3052556
```

CCA of Decathlon Data in R (continued)

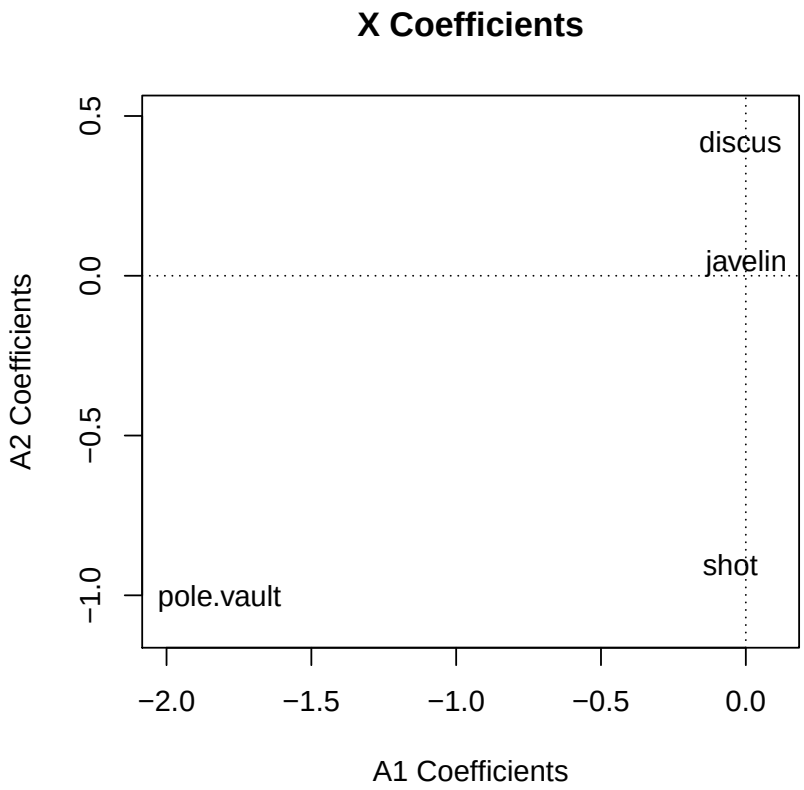
R code to compare the CCA coefficients:

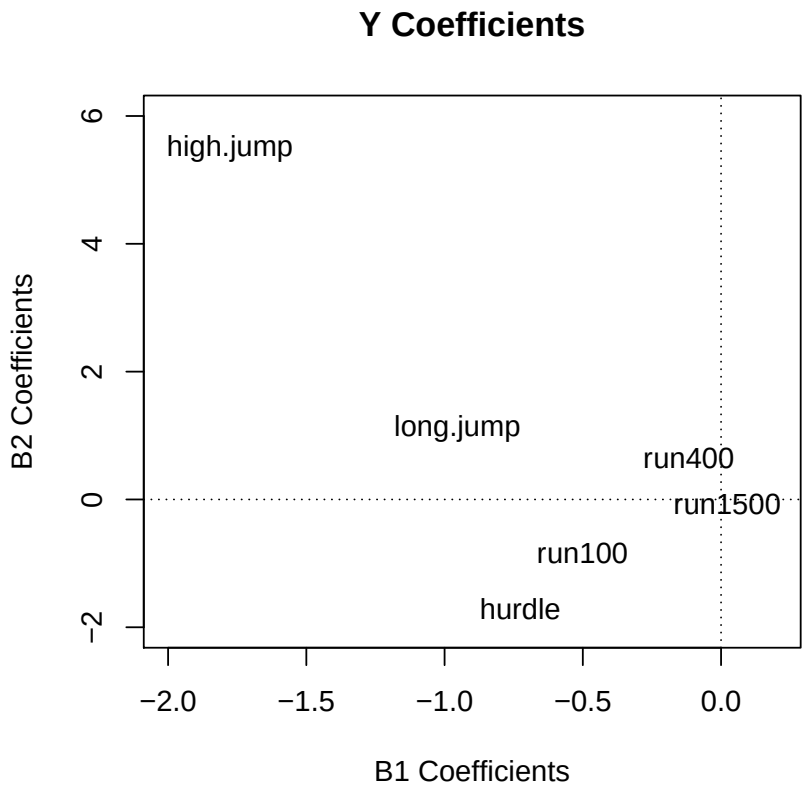
```
# compare linear combinations (different!)
> Ahat <- Sxisqrt %*% Xeig$vectors
> Bhat <- Syisqrt %*% Yeig$vectors
> sum((cca$xcoef - Ahat)^2)
[1] 6.710414
> sum((cca$ycoef[,1:p] - Bhat[,1:p])^2)
[1] 42.98483

# NOTE: you need to multiply R's xcoef and ycoef by sqrt(n-1)
#       to obtain the results we are expecting...

# compare linear combinations (same!)
> Ahat <- Sxisqrt %*% Xeig$vectors
> Bhat <- Syisqrt %*% Yeig$vectors
> sum((cca$xcoef * sqrt(n-1) - Ahat)^2)
[1] 3.031301e-28
> sum((cca$ycoef[,1:p] * sqrt(n-1) - Bhat[,1:p])^2)
[1] 2.414499e-25
```

Plot the CCA Coefficients





R code for left plot:

```
plot(Ahat[,1:2], xlab="A1 Coefficients", ylab="A2 Coefficients",
     type="n", main="X Coefficients", xlim=c(-2, 0.1), ylim=c(-1.1, 0.5))
text(Ahat[,1:2], labels=colnames(X))
```

Define the Canonical Variables

If $\mathbf{X} = \{x_{ij}\}_{n\times p}$ and $\mathbf{Y} = \{y_{ij}\}_{n\times q}$, then

- $\hat{\mathbf{U}} = \mathbf{X}\hat{\mathbf{A}} = \{\hat{u}_{ij}\}_{n\times p}$ where columns of $\hat{\mathbf{U}}$ contain the canonical variables for the X set
- $\hat{\mathbf{V}} = \mathbf{Y}\hat{\mathbf{B}} = \{\hat{v}_{ij}\}_{n\times q}$ where columns of $\hat{\mathbf{V}}$ contain the canonical variables for the Y set

R code to define canonical variables:

```
> U <- X %*% Ahat
> V <- Y %*% Bhat
```

Covariance Matrices of Canonical Variables

R code to check covariance matrices of the canonical variables:

```
# canonical variable covariances
> round(cov(U),4)
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    1    0    0
[3,]    0    0    1    0
[4,]    0    0    0    1

> round(cov(V),4)
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0    0    0    0    0
[2,]    0    1    0    0    0    0
[3,]    0    0    1    0    0    0
[4,]    0    0    0    1    0    0
[5,]    0    0    0    0    1    0
[6,]    0    0    0    0    0    1

> round(cov(U,V),4)
       [,1]   [,2]   [,3]   [,4] [,5] [,6]
[1,] 0.7702 0.0000 0.0000 0.0000    0    0
[2,] 0.0000 0.5034 0.0000 0.0000    0    0
[3,] 0.0000 0.0000 0.4184 0.0000    0    0
[4,] 0.0000 0.0000 0.0000 0.3053    0    0

> rho
[1] 0.7702006 0.5033532 0.4184145 0.3052556
```

Covariances of Canonical and Observed Variables

R code to check covariance matrices of the canonical variables:

```
# covariance of original and canonical variables (U and X)
> Ainv <- solve(Ahat)
> sum( ( cov(U, X) - crossprod(Ahat, Sx) )^2 )
[1] 3.396329e-30
> sum( ( Sx - crossprod(Ainv) )^2 )
[1] 4.364327e-27

# covariance of original and canonical variables (V and Y)
> Binv <- solve(Bhat)
> sum( ( cov(V, Y) - crossprod(Bhat, Sy) )^2 )
[1] 1.696269e-28
> sum( ( Sy - crossprod(Binv) )^2 )
[1] 3.027024e-26

# covariance of original and canonical variables (U and Y)
> sum( (cov(U, Y) - crossprod(Ahat, Sxy))^2 )
[1] 2.071712e-29

# covariance of original and canonical variables (V and X)
> sum( (cov(V, X) - crossprod(Bhat, t(Sxy)))^2 )
[1] 2.943246e-28

# covariance of canonical variables (U and V)
> rhomat <- cbind(diag(rho), matrix(0, p, q-p))
> sum( (cov(U, V) - rhomat)^2 )
[1] 1.241068e-27
> sum( (Sxy - crossprod(Ainv, rhomat) %*% Binv)^2 )
[1] 1.355523e-25
```

Error of Approximation Matrices ($r = 2$)

R code to calculate error of approximation matrices with $r = 2$:

```
# error of approximation matrices (with r=2)
> Ainv <- solve(Ahat)
> Binv <- solve(Bhat)
> r <- 2
> Ex <- Sx - crossprod(Ainv[1:r,])
> Ey <- Sy - crossprod(Binv[1:r,])
> Exy <- Sxy - crossprod(diag(rho[1:r]) %*% Ainv[1:r,], Binv[1:r,])

# get norms of error matrices
> sqrt(mean(Ex^2))
[1] 6.561393
> sqrt(mean(Ey^2))
[1] 18.37339
> sqrt(mean(Exy^2))
[1] 1.725392
```

CCA of Standardized Decathlon Data in R

R code to conduct standardized canonical correlation analysis

```
# standardize data
> Xs <- scale(X)
> Ys <- scale(Y)

# canonical correlations of correlation (standardized data)
> ccas <- cancor(Xs, Ys)

# cca (the normal way)
> Sx <- cov(Xs)
> Sy <- cov(Ys)
> Sxy <- cov(Xs,Ys)
> Sxeig <- eigen(Sx, symmetric=TRUE)
> Sxisqrt <- Sxeig$vectors %*% diag(1/sqrt(Sxeig$values)) %*% t(Sxeig$vectors)
> Syeig <- eigen(Sy, symmetric=TRUE)
> Syisqrt <- Syeig$vectors %*% diag(1/sqrt(Syeig$values)) %*% t(Syeig$vectors)
> Xmat <- Sxisqrt %*% Sxy %*% solve(Sy) %*% t(Sxy) %*% Sxisqrt
> Ymat <- Syisqrt %*% t(Sxy) %*% solve(Sx) %*% Sxy %*% Syisqrt
> Xeig <- eigen(Xmat, symmetric=TRUE)
> Yeig <- eigen(Ymat, symmetric=TRUE)

# compare correlations (same)
> cca$cor
[1] 0.7702006 0.5033532 0.4184145 0.3052556
> sqrt(Xeig$values)
[1] 0.7702006 0.5033532 0.4184145 0.3052556
> sqrt(Yeig$values[1:p])
[1] 0.7702006 0.5033532 0.4184145 0.3052556
```

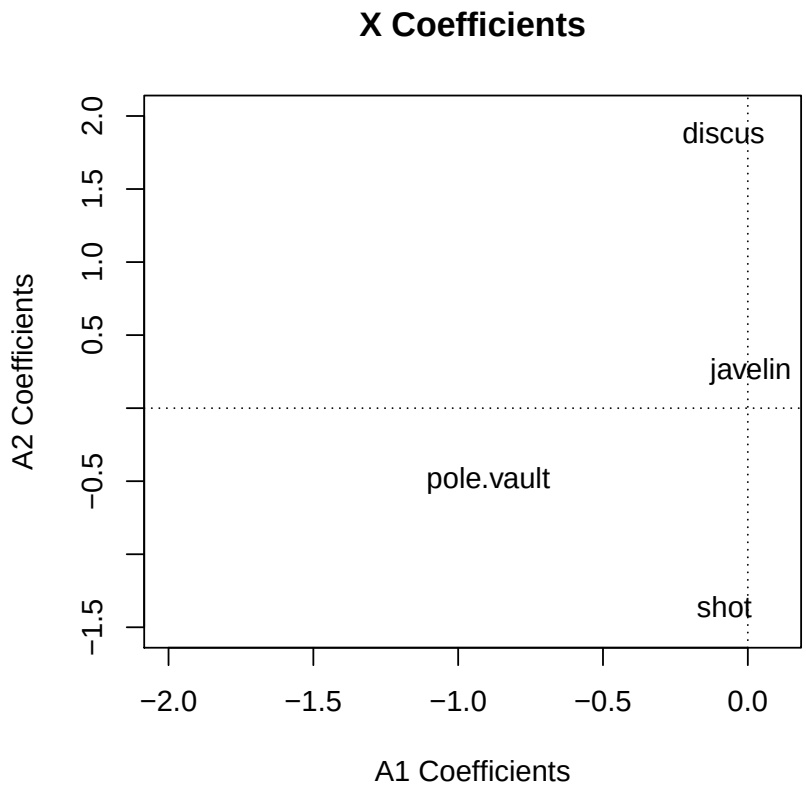
CCA of Standardized Decathlon Data in R (continued)

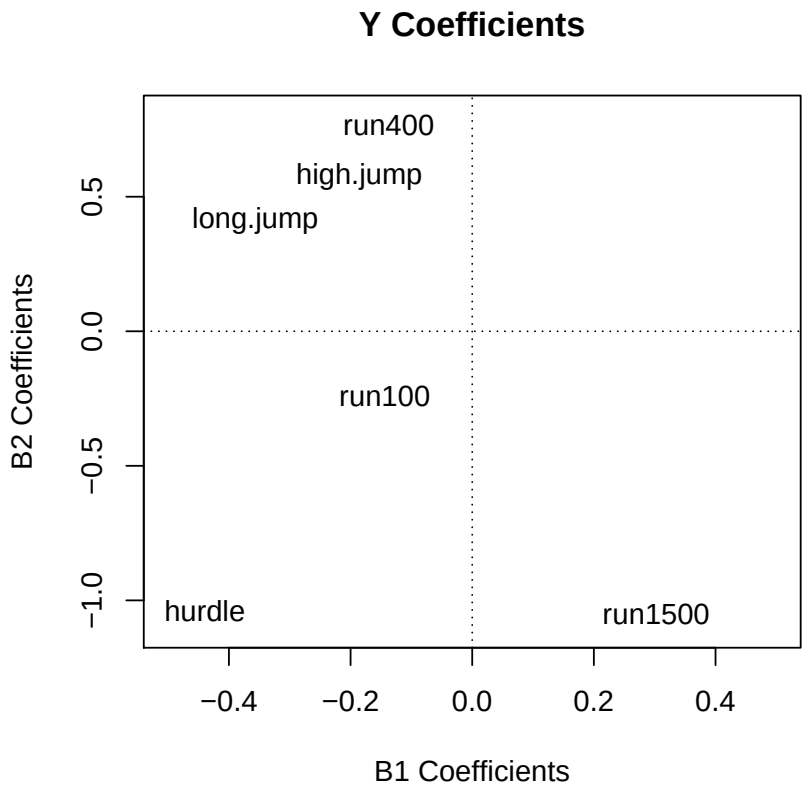
R code to compare the CCA coefficients:

```
# compare linear combinations (different?)
> Ahat <- Sxisqrt %*% Xeig$vectors
> Bhat <- Syisqrt %*% Yeig$vectors
> sum((ccas$xcoef * sqrt(n-1) - Ahat)^2)
[1] 3.332536e-29
> sum((ccas$ycoef[,1:p] * sqrt(n-1) - Bhat[,1:p])^2)
[1] 11.59453

# note that the signing is arbitary!!
> ccas$ycoef[,1:p] * sqrt(n-1)
                [,1]       [,2]       [,3]        [,4]
run100    -0.1439138 -0.2404940  0.5274876 -0.13754449
run400    -0.1373435  0.7655659 -1.2826821  0.96359176
run1500    0.3023537 -1.0519285 -0.1514027 -0.52923644
hurdle    -0.4396044 -1.0374417  0.6303782  0.49905604
long.jump -0.3564702  0.4110878 -0.0253127 -1.09325282
high.jump -0.1855627  0.5731149 -0.2615838 -0.09007821
> Bhat[,1:p]
           [,1]       [,2]       [,3]        [,4]
[1,]  0.1439138 -0.2404940 -0.5274876 -0.13754449
[2,]  0.1373435  0.7655659  1.2826821  0.96359176
[3,] -0.3023537 -1.0519285  0.1514027 -0.52923644
[4,]  0.4396044 -1.0374417 -0.6303782  0.49905604
[5,]  0.3564702  0.4110878  0.0253127 -1.09325282
[6,]  0.1855627  0.5731149  0.2615838 -0.09007821
> Bhat[,1:p] <- Bhat[,1:p] %*% diag(c(-1,1,-1,1))
> sum((ccas$ycoef[,1:p] * sqrt(n-1) - Bhat[,1:p])^2)
[1] 1.132493e-28
```

Plot the Standardized CCA Coefficients





R code for left plot:

```
plot(Ahat[,1:2], xlab="A1 Coefficients", ylab="A2 Coefficients",
     type="n", main="X Coefficients", xlim=c(-2, 0.1), ylim=c(-1.1, 0.5))
text(Ahat[,1:2], labels=colnames(X))
```

Define the Canonical Variables

If $\mathbf{X}_s = \{(x_{ij} - \bar{x}_j)/s_{x_j}\}_{n\times p}$ and $\mathbf{Y}_s = \{(y_{ij} - \bar{y}_j)/s_{y_j}\}_{n\times q}$, then

- $\hat{\mathbf{U}} = \mathbf{X}_s\hat{\mathbf{A}} = \{\hat{u}_{ij}\}_{n\times p}$ where columns of $\hat{\mathbf{U}}$ contain the canonical variables for the X_s set
- $\hat{\mathbf{V}} = \mathbf{Y}_s\hat{\mathbf{B}} = \{\hat{v}_{ij}\}_{n\times q}$ where columns of $\hat{\mathbf{V}}$ contain the canonical variables for the Y_s set

R code to define canonical variables:

```
> U <- Xs %*% Ahat
> V <- Ys %*% Bhat
```

Covariance Matrices of Canonical Variables

R code to check covariance matrices of the canonical variables:

```
# canonical variable covariances
> round(cov(U),4)
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    1    0    0
[3,]    0    0    1    0
[4,]    0    0    0    1

> round(cov(V),4)
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0    0    0    0    0
[2,]    0    1    0    0    0    0
[3,]    0    0    1    0    0    0
[4,]    0    0    0    1    0    0
[5,]    0    0    0    0    1    0
[6,]    0    0    0    0    0    1

> round(cov(U,V),4)
       [,1]   [,2]   [,3]   [,4] [,5] [,6]
[1,] 0.7702 0.0000 0.0000 0.0000    0    0
[2,] 0.0000 0.5034 0.0000 0.0000    0    0
[3,] 0.0000 0.0000 0.4184 0.0000    0    0
[4,] 0.0000 0.0000 0.0000 0.3053    0    0

> rho
[1] 0.7702006 0.5033532 0.4184145 0.3052556
```

Covariances of Canonical and Observed Variables

R code to check covariance matrices of the canonical variables:

```
# covariance of original and canonical variables (U and Xs)
> Ainv <- solve(Ahat)
> sum( ( cov(U, Xs) - crossprod(Ahat, Sx) )^2 )
[1] 2.759323e-31
> sum( ( Sx - crossprod(Ainv) )^2 )
[1] 6.569732e-30

# covariance of original and canonical variables (V and Ys)
> Binv <- solve(Bhat)
> sum( ( cov(V, Ys) - crossprod(Bhat, Sy) )^2 )
[1] 2.406961e-31
> sum( ( Sy - crossprod(Binv) )^2 )
[1] 3.136492e-29

# covariance of original and canonical variables (U and Ys)
> sum( (cov(U, Ys) - crossprod(Ahat, Sxy))^2 )
[1] 5.477785e-32

# covariance of original and canonical variables (V and Xs)
> sum( (cov(V, Xs) - crossprod(Bhat, t(Sxy)))^2 )
[1] 1.336149e-31

# covariance of canonical variables (U and V)
> rhomat <- cbind(diag(rho), matrix(0, p, q-p))
> sum( (cov(U, V) - rhomat)^2 )
[1] 1.272906e-29
> sum( (Sxy - crossprod(Ainv, rhomat) %*% Binv)^2 )
[1] 7.505349e-30
```

Error of Approximation Matrices ($r = 2$)

R code to calculate error of approximation matrices with $r = 2$:

```
# error of approximation matrices (with r=2)
> Ainv <- solve(Ahat)
> Binv <- solve(Bhat)
> r <- 2
> Ex <- Sx - crossprod(Ainv[1:r,])
> Ey <- Sy - crossprod(Binv[1:r,])
> Exy <- Sxy - crossprod(diag(rho[1:r]) %*% Ainv[1:r,], Binv[1:r,])

# get norms of error matrices
> sqrt(mean(Ex^2))
[1] 0.2432351
> sqrt(mean(Ey^2))
[1] 0.2296716
> sqrt(mean(Exy^2))
[1] 0.07458264
```