Canonical Correlation Analysis

Edps/Soc 584, Psych 594

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

© Board of Trustees, University of Illinois

Spring 2017

Overview

Canonical Correlation Analysis and
Tests on Correlation & Covariance Matrices

- Introduction
- Testing for Relationship
- General Problem (i.e., multiple linear combinations)
- Matrix Computation
- Describing the relationship between sets (i.e., specific questions asked and answer in canonical analysis)
- SAS
- Some ideas on dealing with More than two sets.
- Summary.

Reading: J&W Chapter 10

Additional Reference: Morrison (2005)

Introduction

We ended MANOVA talking about checking hypotheses and also made assumptions about equality of covariance matrices in discriminant analysis. There are other tests on covariance matrices that are interesting.

For example

- Single sample: $H_o : \mathbf{\Sigma} = \mathbf{\Sigma}_o$ (where $\mathbf{\Sigma}_o$ is some specified matrix) versus $H_a : \mathbf{\Sigma} \neq \mathbf{\Sigma}_o$.
- Tests for special structures, e.g.,

$$H_o : \mathbf{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

When might you want to test this one?

More Tests

- Population Correlation Matrix

$$H_o : \mathcal{R} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

- $H_o : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_K$ versus not not H_o.
- Simultaneously test equality of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ from K samples.
- Testing the independence of sets of variables— what Canonical Correlation analysis deals with:

$$\mathbf{\Sigma} = \left(\begin{array}{c|c} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \hline \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{array}\right)$$

and test $H_o : \mathbf{\Sigma}_{12} = \mathbf{0}$.

Sets of Variables

In Canonical correlation analysis, we're concerned with whether two sets of variables are related or not. For example:

Teacher ratings: X_1 relaxed, X_2 (motivated), X_3 (organized)	&	Student Achievement: Y_1 (reading), Y_2 (language), Y_3 (math)
Psychological health	&	Performance or Behavioral measures
Job performance	&	Job satisfaction
WAIS sub-tests (e.g., digit-span, vocabulary)	&	Various measures of experience (e.g., age, education, etc)

Goal of Canonical Correlation Analysis

(Due to Hotelling about 1935)

Suppose you have (p+q) variables in a vector and partition it into two parts

$$\mathbf{X}_{(p+q)} = \left(\frac{\mathbf{X}_{1,(p\times 1)}}{\mathbf{X}_{2,(q\times 1)}}\right)$$

with covariance matrix $\mathbf{\Sigma}$, which has also been partitioned

$$\mathbf{\Sigma} = \underbrace{\left(\begin{array}{c|c} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \hline \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{array}\right)}_{p \qquad\quad q} \begin{array}{l} \}\ p \\ \}\ q \end{array}$$

Note that $\mathbf{\Sigma}_{12} = \mathbf{\Sigma}'_{21}$.

Goal: Determine the relationship between the two sets of variables $\mathbf{X}_1$ and $\mathbf{X}_2$.

Goal continued

What linear combination of $\mathbf{X}_1$, i.e.,

$$\mathbf{a}'\mathbf{X}_1 = a_1 X_{11} + a_2 X_{12} + \cdots + a_p X_{1p},$$

is most closely related to a linear combination of $\mathbf{X}_2$,

$$\mathbf{b}'\mathbf{X}_2 = b_1 X_{21} + b_2 X_{22} + \cdots + b_p X_{2p}.$$

We want to choose $\mathbf{a}_{p\times 1}$ and $\mathbf{b}_{q\times 1}$ to maximize the correlation

$$\rho(\mathbf{a}'\mathbf{X}_1, \mathbf{b}'\mathbf{X}_2).$$

These linear combinations are called "canonical variates".

Plan:

1. Determine whether $\mathbf{X}_1$ and $\mathbf{X}_2$ are related.
2. If related, find linear combinations that maximize the canonical correlation.

Testing for Relationship

Two methods to test whether the two sets of variables are related.
We'll start with Wilk's likelihood ratio test for the independence of several sets of variables.

This test pertains to

- K set of variables measures on n individuals.
- The i^{th} set of variables consists of p_i variables.
- There are $\binom{2}{K} = K!/(2!(K-2)!)$ "inter-variable" covariance matrices.
- $\mathbf{\Sigma}_{ik}$ which is $(p_i \times p_k)$ covariance matrix whose elements are equal to the covariances between variables in the i^{th} set and the k^{th} set.
- $H_o : \mathbf{\Sigma}_{ik} = \mathbf{0}$ for all $i \neq k$.

This test is more general than what we need for canonical correlation analysis, but might be useful in other contexts.

Assumptions of Wilk's Test

The requirements are

1. The within set covariance matrices are positive definite; that is, $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{22}$, ..., $\boldsymbol{\Sigma}_{KK}$ are all positive definite.
2. A sample of n observations has been drawn from a (single) population and measures taken for $\text{p} = \sum_{i=1}^{K} p_i$ variables.

For each set of variables, compute $\mathbf{S}_{\text{p}\times\text{p}}$, so we have all the within set covariance matrices and between set matrices:

$$\mathbf{S}_{\text{p}\times\text{p}} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \cdots & \mathbf{S}_{1K} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \cdots & \mathbf{S}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{K1} & \mathbf{S}_{K2} & \cdots & \mathbf{S}_{KK} \end{pmatrix}$$

Wilk's Test Statistic

Wilk's test statistic equals

$$V = \frac{\det(\mathbf{S})}{\det(\mathbf{S}_{11})\det(\mathbf{S}_{22})\cdots\det(\mathbf{S}_{KK})} = \frac{\det(\mathbf{R})}{\det(\mathbf{R}_{11})\det(\mathbf{R}_{22})\cdots\det(\mathbf{R}_{KK})}$$

where

- $\mathbf{R}$ is the correlation matrix
- $\mathbf{R}_{ii}$ are the within set correlation matrices.
- The scale of the variables is not important, so we can use either $\mathbf{S}$ or $\mathbf{R}$.

The distribution of V is very complicates; however, Box (1949) gave a good approximation of V's sampling distribution. . . .
When H_o is true and n is large, then

$$-\frac{(n-1)}{c}\ln(V) \approx \chi^2_f$$

Wilk's Test Statistic

When H_o is true and n is large, then

$$-\frac{(n-1)}{c}\ln(V) \approx \chi^2_f$$

- $\frac{1}{c} = 1 - \frac{1}{12f(n-1)}(2\tau_3 + 3\tau_2)$
- $f = (1/2)\tau_2$
- $\tau_2 = (\sum_{i=1}^K p_i)^2 - \sum_{i=1}^K (p_i)^2$
- $\tau_3 = (\sum_{i=1}^K p_i)^3 - \sum_{i=1}^K (p_i)^3$.

Distribution of Wilk's Test continued

- Reject for large values of V.
- For canonical correlation analysis where K_2, the test statistic simplifies to

$$
\begin{aligned}
V &= \frac{\det(\mathbf{S})}{\det(\mathbf{S}_{11})\det(\mathbf{S}_{22})} \\
&= \frac{\det(\mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})}{\det(\mathbf{S}_{11})} = \frac{\det(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})}{\det(\mathbf{S}_{22})} \\
&= \det(\mathbf{I} - \mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}) = \det(\mathbf{I} - \mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})
\end{aligned}
$$

- Time for an example.

Example: WAIS and Age

The data (from Morrison (1990), pp 307–308) are from an investigation of the relationship between the Wechsler Adult Intelligence Scale (WAIS) and age.

Participants: $n = 933$ white men and women aged 25-64

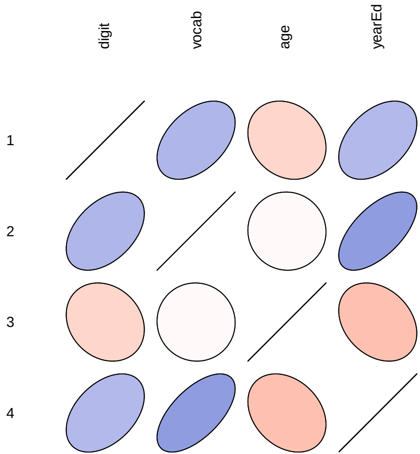
Two sets of variables:

Set 1:	$p = 2$	$X_1 =$	digit span sub-test of WAIS
		$X_2 =$	vocabulary sub-test of WAIS
Set 2:	$q = 2$	$X_3 =$	chronological age
		$X_4 =$	years of formal education

Sample correlation matrix:

$$\mathbf{R} = \left(\begin{array}{c|c} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \hline \mathbf{R}_{21} & \mathbf{R}_{22} \end{array}\right) = \left(\begin{array}{rr|rr} 1.00 & .45 & -.19 & .43 \\ .45 & 1.00 & -.02 & .62 \\ \hline -.19 & -.02 & 1.00 & -.29 \\ .43 & .62 & -.29 & 1.00 \end{array}\right)$$

Picture of the Correlation Matrix



Example of Wilk's Test for Relationship

Testing $H_o : \mathbf{\Sigma}_{12} = \mathbf{0}$: Method I

$$V = \frac{|\mathbf{R}|}{|\mathbf{R}_{11}||\mathbf{R}_{22}|} = \frac{.4015}{(.7975)(.9159)} = .5497$$

When n is large and the null hypothesis is true $-\frac{(n-1)}{c} \ln V$ is approximately distributed as χ^2_f random variable where

$$\begin{aligned}
\tau_2 &= (p+q)^2 - (p^2+q^2) = (2+2)^2 - (2^2+2^2) = 8 \\
\tau_3 &= (p+q)^3 - (p^3+q^3) = (2+2)^3 - (2^3+2^3) = 48 \\
f &= \frac{1}{2}\tau_2 = (.5)(8) = 4 \\
1/c &= 1 - \frac{1}{12f(n-1)}(2\tau_3 + 3\tau_2) = .9973
\end{aligned}$$

Example of Wilk's Test for Relationship

Since

$$-\frac{(n-1)}{c}\ln V = -(.9973)(933-1)\ln(.5497) = 556.20$$

is much larger than a $\chi^2_4(.05)$, we reject the hypothesis; the data support the conclusion that the two sets of variables are related. (p–value$<<$ .00001).

Method II: Canonical Correlation

Another approach to testing $H_o : \mathbf{\Sigma}_{12} = \mathbf{0}$.

Find the vectors **a** and **b** that maximize the correlation

$$\rho(\mathbf{a}'\mathbf{X}_1, \mathbf{b}'\mathbf{X}_2)$$

Define **C** which is $(p+q) \times 2$ matrix

$$\mathbf{C} = \left(\begin{array}{c|c} \mathbf{a} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{b} \end{array}\right) \begin{array}{l} \}\quad p \text{ rows} \\ \}\quad q \text{ rows} \end{array}$$

Consider the linear combination $\mathbf{C}'\mathbf{X}$,

$$\mathbf{C}'\mathbf{X} = \left(\begin{array}{c|c} \mathbf{a}' & \mathbf{0}' \\ \hline \mathbf{0}' & \mathbf{b}' \end{array}\right) \left(\begin{array}{c} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{array}\right) = \left(\begin{array}{c} \mathbf{a}'\mathbf{X}_1 \\ \hline \mathbf{b}'\mathbf{X}_2 \end{array}\right)$$

The next piece that we need is $\text{cov}(\mathbf{C}'\mathbf{X})$...

Covariance matrix for $C'X$

$$\text{cov}(\mathbf{C}'\mathbf{X}) = \mathbf{C}'\mathbf{\Sigma}\mathbf{C} = \left(\begin{array}{c|c} \mathbf{a}' & \mathbf{0}' \\ \hline \mathbf{0}' & \mathbf{b}' \end{array}\right)\left(\begin{array}{c|c} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \hline \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{array}\right)\left(\begin{array}{c|c} \mathbf{a} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{b} \end{array}\right)$$

$$\left(\begin{array}{c|c} \mathbf{a}'\mathbf{\Sigma}_{11}\mathbf{a} & \mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b} \\ \hline \mathbf{b}'\mathbf{\Sigma}_{21}\mathbf{a} & \mathbf{b}'\mathbf{\Sigma}_{22}\mathbf{b} \end{array}\right)$$

and the correlation between $\mathbf{a}'\mathbf{X}_1$ and $\mathbf{b}'\mathbf{X}_2$ is

$$\rho(\mathbf{a}'\mathbf{X}_1, \mathbf{b}'\mathbf{X}_2) = \frac{\mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{\Sigma}_{11}\mathbf{a}}\sqrt{\mathbf{b}'\mathbf{\Sigma}_{22}\mathbf{b}}}$$

If $\mathbf{\Sigma}_{12} = \mathbf{0}$, then $\mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b} = 0$ for all possible choices of $\mathbf{a}$ and $\mathbf{b}$.

The correlation is estimated by

$$\frac{\mathbf{a}'\mathbf{S}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{S}_{11}\mathbf{a}}\sqrt{\mathbf{b}'\mathbf{S}_{22}\mathbf{b}}}$$

The Idea

The idea underlying this method for testing whether there is a relationship between two sets of variables is that if the correlation (in the population) is 0, then let's find the maximum possible value of the correlation $\rho(\mathbf{a}'\mathbf{X}_1, \mathbf{b}'\mathbf{X}_2)$ and use it as a test statistic for the null hypothesis $H_o : \mathbf{\Sigma}_{12} = \mathbf{0}$.

To simplify the problem, we'll constrain **a** and **b** such that

$$\widehat{\text{var}}(\mathbf{a}'\mathbf{X}_1) = \mathbf{a}'\mathbf{S}_{11}\mathbf{a} = 1 \qquad \text{and} \qquad \widehat{\text{var}}(\mathbf{b}'\mathbf{X}_i) = \mathbf{b}'\mathbf{S}_{22}\mathbf{b} = 1$$

Our maximization problem is now to the find the **a** and **b**

$$\max_{\mathbf{a},\mathbf{b}} \left(\mathbf{a}'\mathbf{S}_{12}\mathbf{b}\right)$$

This is the canonical correlation (subject to the constraint that the variances of the linear combinations equal 1).

Solution that Maximizes the Correlation

The largest sample correlation is the square root of the the largest eigenvalue ("root") of

$$\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}$$

or equivalently

$$\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$$

- The matrix products $\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}$ and $\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ have the same characteristic roots (eigenvalues), which we'll call $c_1, c_2, \ldots, c_r$.
- Assume that we've ordered the roots: $c_1 \geq \cdots \geq c_r$.
- The eigenvector of $\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}$ associated with c_1 gives use $\mathbf{a}_1$.
- The eigenvector of $\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ corresponding to c_1 corresponds to $\mathbf{b}_1$.

Solution that Maximizes the Correlation

- The matrix products $\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}$ and $\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ have the same characteristic roots (eigenvalues), which we'll call $c_1, c_2, \ldots, c_r$.
- Assume that we've ordered the roots: $c_1 \geq \cdots \geq c_r$.
- The eigenvector of $\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}$ associated with c_1 gives use $\mathbf{a}_1$.
- The eigenvector of $\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ corresponding to c_1 corresponds to $\mathbf{b}_1$.
- Setting $\mathbf{a} = \mathbf{a}_1$ and $\mathbf{b} = \mathbf{b}_1$ yields the maximum:

$$\sqrt{c_1} = \max_{\mathbf{a},\mathbf{b}}(\mathbf{a}'\mathbf{S}_{12}\mathbf{b}).$$

- The sample correlation between $U_1 = \mathbf{a}_1\mathbf{X}_1$ and $V_1 = \mathbf{b}_1\mathbf{X}_2$ equals $\pm\sqrt{c_1}$ (you have to determine the sign).

The Test

Formal Test of $H_o : \mathbf{\Sigma}_{12} = 0$: Consider c_1 = the largest root (eigenvalue). Reject $H_o : \mathbf{\Sigma}_{12} = 0$ if $c_1 > \theta_{\alpha;s,m.n^*}$ where θ is the $(1-\alpha)\times 100\%$ percentile point of the greatest root distribution with parameters

$$s = \min(p,q), \qquad m = \frac{1}{2}(|p-q|-1), \qquad n^* = \frac{1}{2}(n-p-q-2)$$

There are charts and tables of upper percentile points of the largest root distribution in various multivariate statistics texts (e.g., Morrison).

The Test

There are charts and tables of upper percentile points of the largest root distribution in various multivariate statistics texts (e.g., Morrison).

For online via jstor.org: (Assuming connection to uiuc via netid)

D. L. Heck (1960) Charts of Some Upper Percentage Points of the Distribution of the Largest Characteristic Root. *The Annals of Mathematical Statistics*, *Vol. 31*, No. 3, pp. 625-642.

K. C. Sreedharan Pillai, Celia G. Bantegui (1959). On the Distribution of the Largest of Six Roots a Matrix in Multivariate AnalysisOn the Distribution of the Largest of Six Roots a Matrix in Multivariate Analysis. *Biometrika, vol 46*, pp. 237-24.

Easier Way to Get Percentiles

A method to approximate percentiles was proposed by Johnstone (2002). Approximate null distribution of the largest root in multivariate analysis. *Ann Appl Stat*, *3*(4), 1616-1633.

- R package `RMT`.
- MATLAB is in development.
- SAS IML code that I wrote and put on the web-site. This only computes the 90th, 95th and 99th percentiles.
- The paper provides way to compute approximate p-values which are much more accurate that SAS and R package `car`.
- p-values provided by SAS yields a lower bound (tests are liberal).
- A stand alone program by Lutz (2000).

Example: Method II

Testing $H_o : \boldsymbol{\Sigma}_{12} = \mathbf{0}$: Method II – the largest root distribution

We first find the roots of $\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}$ (which are equal to the roots of $\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}$). So we need

$$\mathbf{R}_{11}^{-1} = \begin{pmatrix} 1.254 & -.564 \\ -.564 & 1.254 \end{pmatrix} \quad \mathbf{R}_{22}^{-1} = \begin{pmatrix} 1.092 & .317 \\ .317 & 1.092 \end{pmatrix}$$

and multiplying the matrices gives us

$$\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21} = \begin{pmatrix} .0937 & .0873 \\ .2130 & .3730 \end{pmatrix}$$

The roots of this matrix product are the solution of the equation

$$\begin{vmatrix} (.0937 - c) & .0873 \\ .2130 & (.3730 - c) \end{vmatrix} = 0$$

$$(.0937 - c)(.3730 - c) - (.2130)(.0873) = 0$$

Example continued

$$\begin{aligned} (.0937 - c)(.3730 - c) - (.2130)(.0873) &= 0 \\ c^2 - .4667c + .0164 &= 0 \\ (c - .4285)(c - .0381) &= 0 \end{aligned}$$

So $c_1 = .4285$ and $c_2 = .0381$.
The largest correlation between a linear combination of variables in set 1 ($U_1 = \mathbf{a}'\mathbf{X}_1$) and a linear combination of variables in set 2 ($V_1 = \mathbf{b}'\mathbf{X}_2$) equals

$$\sqrt{c_1} = \sqrt{.4285} = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{b} = .654$$

To test whether $H_o : \boldsymbol{\Sigma}_{12} = \mathbf{0}$, we have

$$\begin{aligned} s &= \min(p, q) = \min(2, 2) = 2 \\ m &= (1/2)(|p - q| - 1) = .5(|2 - 2| - 1) = -.5 \\ n^* &= (1/2)(n - p - q - 2) = .5(933 - 2 - 2 - 2) = 463.5 \end{aligned}$$

where p = number of variables in set 1, and q = number of variables in set 2.

Finishing the Test & Finding a_1 and b_1

Using the chart (page 628 of Heck (1960)) of the greatest root distribution, we find $\theta_{2,-.5,463.5}(.01) = .02$. Since $c_1 = .4285 > .02$, we reject H_o; there is a dependency between the sets of variables.

The linear combination that Maximizes the correlation.

To compute $\mathbf{a}_1$, we use $\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}$

$$\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{a}_1 = c_1\mathbf{a}_1$$

$$\begin{pmatrix} .0937 & .0873 \\ .2130 & .3738 \end{pmatrix}\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} = .4285\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}$$

Two equations, two unknowns: . . .

Finishing the Test & Finding a_1 and b_1

$$
\begin{aligned}
\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{a}_1 &= c_1\mathbf{a}_1 \\
\begin{pmatrix} .0937 & .0873 \\ .2130 & .3738 \end{pmatrix}\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} &= .4285\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} \\
\text{Two equations, two unknowns:} & \quad \ldots \\
-.3348a_{11} + .0873a_{12} &= 0 \\
.2130a_{11} - .0547a_{12} &= 0
\end{aligned}
$$

For convenience, we'll set $a_{12} = 1$, and solve for $a_{11} = (.0547/.2130)a_{12} = .26$.

Finding a_1 and b_1

Any vector that is proportional to (i.e., is a multiple of) $\mathbf{a} = (.26, 1)'$ is a solution and gives us the correct linear combination.

To compute $\mathbf{b}_1$, we use $\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}$

$$\begin{aligned}
\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{b}_1 &= c_1\mathbf{b}_1 \\
\begin{pmatrix} .0305 & .0798 \\ -.0378 & .4361 \end{pmatrix}\begin{pmatrix} b_{11} \\ b_{12} \end{pmatrix} &= .4285\begin{pmatrix} b_{11} \\ b_{12} \end{pmatrix} \\
\text{Two Equations, Two Unknowns} & \quad \ldots \\
-.3980 b_{11} + .0798 b_{12} &= 0 \\
-.0378 b_{11} + .0076 b_{12} &= 0
\end{aligned}$$

We'll set $b_{12} = 1$ and solve for $b_{11} = (.0076/.0378)b_{12} = .20$.
Any vector proportional to (a multiple of) $\mathbf{b} = (.20, 1)'$ is a solution and gives us the correct linear combination.

Finding a_1 and b_1

$$\begin{aligned} \text{Two Equations, Two Unknowns} & \quad \ldots \\ -.3980 b_{11} + .0798 b_{12} &= 0 \\ -.0378 b_{11} + .0076 b_{12} &= 0 \end{aligned}$$

We'll set $b_{12} = 1$ and solve for $b_{11} = (.0076/.0378) b_{12} = .20$. Any vector proportional to (a multiple of) $\mathbf{b} = (.20, 1)'$ is a solution and gives us the correct linear combination.

Summary and Conclusion (So Far)

Since the correlation matrix, $\mathbf{R}$, was used, to solve for the vectors $\mathbf{a}_1$ and $\mathbf{b}_1$, we use the standardized scores (i.e., z-scores) rather than the original (raw) variables.

$$
\begin{aligned}
U_1 &= .26z_{\text{(digit span)}} + 1.00z_{\text{(vocabulary)}} \\
V_1 &= .20z_{\text{(age)}} + 1.00z_{\text{(years of formal education)}}
\end{aligned}
$$

Interpretation/Summary:

- The correlation between equals U_1 and V_1, which equals $\sqrt{c_1} = \sqrt{.4285} = .654 = (\mathbf{a}_1'\mathbf{R}_{12}\mathbf{b}_1)/(\sqrt{\mathbf{a}_1'\mathbf{R}_{11}\mathbf{a}_1}\sqrt{\mathbf{b}_1'\mathbf{R}_{22}\mathbf{b}_1})$, is the largest possible one for any linear combination of the variables in sets 1 and 2.
- U_1: places four times more weight on vocabulary than on digit span. . . long term versus short term memory.
- V_1: places five times more weight on years of formal education than on chronological age.
- The major link between the two sets of variables is due to education and vocabulary.

The More Usual Scaling of U_1 and V_1

(What I did wasn't the typical way to scale $\mathbf{a}_1$ and $\mathbf{b}_1$).

- The standard or typical way that $\mathbf{a}_1$ and $\mathbf{b}_1$ are scaled is such that the variances of $U_1 = \mathbf{a}_1\mathbf{X}_1$ and $V_1 = \mathbf{b}_1\mathbf{X}_2$ equal 1.
- Since any multiple of $\mathbf{a}_1$ and/or $\mathbf{b}_1$ is a solution, we just need to multiply the vectors by an appropriate constant.
- For example,

$$\mathbf{a}_1^* = \frac{\mathbf{a}_1}{\sqrt{\mathbf{a}_1'\mathbf{R}_{11}\mathbf{a}}}$$

and now

$$\begin{aligned} \text{var}(U_1) = \text{var}(\mathbf{a}_1^*\mathbf{X}_1) &= \mathbf{a}_1^{*'}\mathbf{R}_{11}\mathbf{a}_1^* \\ &= \left(\frac{\mathbf{a}_1'}{\sqrt{\mathbf{a}_1'\mathbf{R}_{11}\mathbf{a}}}\right)\mathbf{R}_{11}\left(\frac{\mathbf{a}_1}{\sqrt{\mathbf{a}_1'\mathbf{R}_{11}\mathbf{a}}}\right) \\ &= 1 \end{aligned}$$

Our Example

$$\mathbf{a}_1'\mathbf{R}_{11}\mathbf{a}_1 = (0.26, 1.00)\begin{pmatrix} 1.00 & .45 \\ .45 & 1.00 \end{pmatrix}\begin{pmatrix} 0.26 \\ 1.00 \end{pmatrix} = 1.3016$$

$$\text{So} \qquad \mathbf{a}_1^{*\prime} = \frac{1}{\sqrt{1.3016}}(0.26, 1.00) = (.2279, .8765)$$

As a check:

$$\text{var}(U_1) = (.2279, .8765)\begin{pmatrix} 1.00 & .45 \\ .45 & 1.00 \end{pmatrix}\begin{pmatrix} .2279 \\ .8765 \end{pmatrix} = 1.00$$

Doing the same thing for **b**:

$$\mathbf{b}_1'\mathbf{R}_{22}\mathbf{b}_1 = (0.20, 1.00)\begin{pmatrix} 1.00 & -.29 \\ -.29 & 1.00 \end{pmatrix}\begin{pmatrix} 0.20 \\ 1.00 \end{pmatrix} = 1.0403$$

$$\text{So} \qquad \mathbf{b}_1^{*\prime} = \left(\frac{1}{\sqrt{1.0403}}\right)\mathbf{b}_1' = (.2081, 1.0403)$$

General Problem

So far we've only looked at the largest correlation possible between linear combinations of the variables from two sets; however, there are more c_i's.

There are more linear combinations:

$$\begin{aligned} U_1 &= \mathbf{a}_1'\mathbf{X}_1 & V_1 &= \mathbf{b}_1'\mathbf{X}_2 \\ U_2 &= \mathbf{a}_2'\mathbf{X}_1 & V_2 &= \mathbf{b}_2'\mathbf{X}_2 \\ &\vdots & &\vdots \\ U_r &= \mathbf{a}_r'\mathbf{X}_1 & V_r &= \mathbf{b}_r'\mathbf{X}_2 \end{aligned}$$

With the property that the sample correlation between U_1 and V_1 is the largest, the sample correlation between U_2 and V_2 is the largest among all linear combinations uncorrelated with U_1 and V_1, etc.

That is, for all $i \neq k$,

$$\begin{aligned} \text{cov}(U_i, U_k) &= \mathbf{a}_i'\mathbf{S}_{11}\mathbf{a}_k = 0 & \text{cov}(V_i, V_k) &= \mathbf{b}_i'\mathbf{S}_{22}\mathbf{b}_k = 0 \\ \text{cov}(U_i, V_k) &= \mathbf{a}_i'\mathbf{S}_{12}\mathbf{b}_k = 0 & \text{cov}(U_k, V_i) &= \mathbf{a}_k'\mathbf{S}_{12}\mathbf{b}_i = 0 \end{aligned}$$

Assumptions and Solution

Assume

1. The elements of $\mathbf{\Sigma}_{(p+q)\times(p+q)}$ are finite.
2. $\mathbf{\Sigma}$ is full rank; that is, rank $= p+q$.
3. The first $r \le \min(p,q)$ characteristic roots of $\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$ are distinct.

Then $\mathbf{a}_i$ and $\mathbf{b}_i$ are estimated from the data by solving

$$(\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} - c_i\mathbf{S}_{11})\mathbf{a}_i = 0$$

and

$$(\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} - c_i\mathbf{S}_{22})\mathbf{b}_i = 0$$

where c_i is the i^{th} root of the equation.

$\det(\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}-c_i\mathbf{S}_{11}) = 0$ or equivalently $\det(\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}-c_i\mathbf{S}_{22}) = 0$

$$c_i = \text{squared sample correlation between } U_i \text{ and } V_i$$

Back to our Example

The linear combination that gives the next largest correlation and is orthogonal to the first one.

Using the second root of the matrix product, $c_2 = .0381$, which is the second largest squared correlation, repeat the process we did previously to get the vectors $\mathbf{a}_2$ and $\mathbf{b}_2$:

$$\mathbf{a}_2' = (-1.00, 0.64) \qquad \text{and} \qquad \mathbf{b}_2' = (1.00, 0.10)$$

or using the more typically scaling, we get

$$\mathbf{a}_2^{*'} = (-1.0953, .7001) \qquad \text{and} \qquad \mathbf{b}_2{}^{*\prime} = (1.0249, .1025)$$

Statistical Hypothesis Test: $H_o : \rho(U_2, V_2) = 0$ vs $H_a : \rho(U_2, V_2) \neq 0$:

$$-\left(n - 1 - \frac{1}{2}(p + q + 1)\right) \ln(1-c_2) = -(932-.5(5)) \ln(1-.0381) = 36.11$$

If the null hypothesis is true (and n large), then this statistic is approximately distributed at χ^2_{pq}. Since $\chi^2_4(.05) = 9.488$, we reject the null and conclude that the second canonical correlation is not zero.

The Canonical Variates

To find $\mathbf{a}_2$ and $\mathbf{b}_2$, we use the same process that we used to find $\mathbf{a}_1$ and $\mathbf{b}_1$, which gives us

$$U_2 = -1.00 Z_{(\text{digit span})} + 0.64 Z_{(\text{vocabulary})}$$

$$V_2 = 1.00 Z_{(\text{age})} + 0.10 Z_{(\text{years of formal education})}$$

- U_2 is a weighted contrast between digit span (performance) and vocabulary (verbal) sub-tests.
- V_2 is nearly all age.
- There's a widening in the gap between performance with advancing age. As people get older, there's a larger difference between accumulated knowledge (vocabulary) and performance skills.

Correlational Structure of Canonical Variates

Note: The following covariances (correlations) are all zero:

$$\text{cov}(U_1, U_2) = \mathbf{a}_1\mathbf{R}_{11}\mathbf{a}_2 = 0 \qquad \text{cov}(V_1, V_2) = \mathbf{b}_1\mathbf{R}_{22}\mathbf{b}_2 = 0$$

$$\text{cov}(U_1, V_2) = \mathbf{a}_1\mathbf{R}_{12}\mathbf{b}_2 = 0 \qquad \text{cov}(V_1, U_2) = \mathbf{b}_1\mathbf{R}_{21}\mathbf{a}_2 = 0$$

In other words,the sample correlation matrix for the canonical variates is

	U_1	U_2	V_1	V_2
U_1	1.000	.000	.654	.000
U_2	.000	1.000	.000	.195
V_1	.654	.000	1.000	.000
V_2	.000	.195	.000	1.000

which is much simpler than the sample correlation matrix back on pages 13–14

Computation

Or the how to do this in IML, R, and/or MATLAB: We symmetrize the matrix so that our solution are the eigenvalues and vectors of

$$(\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2})\mathbf{e}_i = c_i\mathbf{e}_i$$

where $\mathbf{S}_{11}^{-1/2}$ is the inverse of the square root matrix. Then the combination vectors that you want equal

$$\mathbf{a}_i = \mathbf{S}_{11}^{-1/2}\mathbf{e}_i$$

$$\mathbf{b}_i = \frac{1}{\sqrt{c_i}}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{a}_i$$

OR You can find the eigenvalues and eigenvectors of

$$(\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2})\mathbf{f}_i = c_i\mathbf{f}_i$$

Then the combination vectors that you want equal

$$\mathbf{b}_i^* = \mathbf{S}_{22}^{-1/2}\mathbf{f}_i$$

Showing Why this Works

$$\left(\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\right)\mathbf{e} = c^*\mathbf{e}$$

$$\mathbf{S}_{11}^{-1/2}\left(\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\right)\mathbf{e} = c^*\mathbf{S}_{11}^{-1/2}\mathbf{e}$$

$$\left(\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\right)\underbrace{\mathbf{S}_{11}^{-1/2}\mathbf{e}}_{\mathbf{a}} = c^*\underbrace{\mathbf{S}_{11}^{-1/2}\mathbf{e}}_{\mathbf{a}}$$

$$\left(\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\right)\mathbf{a} = c^*\mathbf{a}$$

This is what we did for discriminant analysis to find eigenvalues and vectors of a non-symmetric matrix.

Questions Answered by CCA

1. To what extent can one set of two (or more) variables be predicted by or "explained" by another set of two or more variables?
2. What contributions does a single variable make to the explanatory power of the set of variable to which the variable belongs?
3. To what extent does a single variable contribute to predicting or "explaining" the composite of the variables in the variable set to which the variable does not belong?

We'll talk about how to answer each of these.

Summary of What CCA Does

Started with sample **R** (or **S**), which in our example is

		X_1	X_2	X_3	X_4	
set 1	X_1	1.00				digit span
	X_2	.45	1.00			vocabulary
set 2	X_3	−.19	−.02	1.00		age
	X_4	.43	.62	−.29	1.00	years of education

We found linear transformation, "canonical variates", of the original variables within sets to maximize the between set correlations:

$$U_i = \mathbf{a}_i'\mathbf{X}_1 \quad \text{and} \quad V_i = \mathbf{b}_i'\mathbf{X}_2$$

where $(\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21})\mathbf{a}_i = c_i\mathbf{a}_i$ and $(\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12})\mathbf{b}_i = c_i\mathbf{b}_i$

And scaled them so that the variance of U_i and V_i equal 1

$$\mathbf{a}_i'\mathbf{R}_{11}\mathbf{a}_i = \mathbf{a}_i'\mathbf{R}_{22}\mathbf{b}_i = 1$$

Example: The Simplification of R

	X_1	X_2	X_3	X_4	U_1	U_2	V_1	V_2
X_1	1.00	.45	−.19	.43				
X_2	.45	1.00	−.02	.62				
X_3	−.19	−.02	1.00	−.29				
X_4	.43	.62	−.29	1.00				
U_1					1.00	.00	.654	.00
U_2					0.00	1.00	.00	.195
V_1					.654	.00	1.00	.00
V_2					.00	.195	0.00	1.00

Question 1

	X_1	X_2	X_3	X_4	U_1	U_2	V_1	V_2
X_1	$\mathbf{R}_{11}$		$\mathbf{R}_{12}$					
X_2								
X_3	$\mathbf{R}_{21}$		$\mathbf{R}_{22}$					
X_4								
U_1	Structure		Index		1.00	0.00	$\sqrt{c_1}$	0.00
U_2	coefficients		coefficients		0.00	1.00	0.00	$\sqrt{c_2}$
V_1	Index		Structure		$\sqrt{c_1}$	0.00	1.00	0.00
V_2	coefficients		coefficients		0.00	$\sqrt{c_2}$	0.00	1.00

Question 1: To what extent can one set of two or more variables be explained by another set of two or more variables?

Answer: The (first) canonical correlation $\sqrt{c_1}$.

Question 2

Question 2: What is the contribution between a variable and the canonical (composite) variable for it's set?

Answer: The "structure coefficients".

Variables Within Sets | Canonical Variates

$$\mathbf{X}_1 = \begin{pmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1p} \end{pmatrix} \quad \mathbf{X}_2 = \begin{pmatrix} X_{21} \\ X_{22} \\ \vdots \\ X_{2q} \end{pmatrix}$$

$$U_i = \mathbf{a}_i'\mathbf{X}_1$$

$$V_i = \mathbf{b}_i'\mathbf{X}_2$$

Structure coefficients equal,

$$\text{Set 1: } \underbrace{\text{corr}}_{Vector}(\mathbf{X}_1, U_i) = \text{corr}(\mathbf{X}_1, \mathbf{a}_i'\mathbf{X}_1)$$

$$\text{Set 2: } \underbrace{\text{corr}}_{Vector}(\mathbf{X}_2, V_i) = \text{corr}(\mathbf{X}_2, \mathbf{b}_i'\mathbf{X}_2)$$

The Structure Coefficients

$$\begin{aligned}
\text{Set 1:}\text{corr}_{(p)}(\mathbf{X}_1, U_i) &= \text{corr}(\mathbf{X}_1, \mathbf{a}_i'\mathbf{X}_1) \\
&= \begin{pmatrix} 1/\sqrt{s_{11}} & 0 & \cdots & 0 \\ 0 & 1/\sqrt{s_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/sqrt s_{pp} \end{pmatrix} \mathbf{S}_{11}\mathbf{a}_i \\
&= \text{diag}(1/\sqrt{s_{ii}})\mathbf{S}_{11}\mathbf{a}_i
\end{aligned}$$

When we use $\mathbf{R}$ rather than $\mathbf{S}$ to find U_i and V_i, then

$$\text{set 1:} \qquad \text{corr}(\mathbf{Z}_1, U_i)_{(p\times 1)} = \mathbf{R}_{11}\mathbf{a}_i$$

and

$$\text{set 2:} \qquad \text{corr}(\mathbf{Z}_2, V_i)_{(q\times 1)} = \mathbf{R}_{22}\mathbf{b}_i$$

Question 3

Question 3 To what extent does a single variable contribute to explaining the canonical (composite) variable in the set of variables to which it does not belong?

Answer: The correlation between it and the canonical variate of the other set of variables: "index coefficients".

$$\text{corr}(\mathbf{X}_1, V_i)_{(p\times 1)} = \text{corr}(\mathbf{X}_1, \underline{i}'\mathbf{X}_2) = \text{diag}(1/\sqrt{s_{11(ii)}})\boldsymbol{\Sigma}_{12}\mathbf{b}_i$$

and

$$\text{corr}(\mathbf{X}_2, U_i)_{(q\times 1)} = \text{corr}(\mathbf{X}_2, i'\mathbf{X}_1) = \text{diag}(1/\sqrt{s_{22(ii)}})\boldsymbol{\Sigma}_{21}\mathbf{a}_i$$

Relationship between Index and Structure

$\mathbf{a}_i$ can be written as a linear combination of $\mathbf{b}_i$ (and visa versa).

$$\mathbf{a}_i = \frac{1}{c_i}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{b}_i \quad \text{and} \quad \mathbf{b}_i = \frac{1}{c_i}\boldsymbol{\Sigma}_{22}^{-1}{}_{21}\mathbf{a}_i$$

Re-arrange terms a bit

$$\underbrace{\sqrt{c_i}}\,\boldsymbol{\Sigma}_{11}\mathbf{a}_i = \boldsymbol{\Sigma}_{12}\mathbf{b}_i \qquad \underbrace{\sqrt{c_i}}\,\boldsymbol{\Sigma}_{22}\mathbf{b}_i = \boldsymbol{\Sigma}_{21}\mathbf{a}_i$$

So the Index Coefficients

$$\begin{aligned}
\text{corr}(\mathbf{X}_1, V_i)_{(p\times 1)} &= \text{diag}(1/\sqrt{\sigma_{11(ii)}})\boldsymbol{\Sigma}_{12}\mathbf{b}_i \\
&= \mathbf{D}_1^{-1/2}(\sqrt{c_i}\boldsymbol{\Sigma}_{11}\mathbf{a}_i) \\
&= \underbrace{\sqrt{c_i}}_{\text{cov}(U_i,V_i)}\ \underbrace{\mathbf{D}_1^{-1/2}\boldsymbol{\Sigma}_{11}\mathbf{a}_i}_{\text{corr}(\mathbf{X}_1,U_i)} \\
\text{Index coefficient} &= \underbrace{\text{cov}(U_i, V_i)}_{canonical\ \ correlation} \ \text{(Structure Coefficient)} \\
\text{and} \quad \text{corr}(\mathbf{X}_2, U_i) &= \text{cov}(U_i, v_i)\text{corr}(\mathbf{X}_2, V_i)
\end{aligned}$$

SAS

```
data corrmat (type=corr);
input  TYPE  $  NAME  $ x1 x2 x3 x4;
list;
datalines;
N - 933 933 933 933
CORR x1 1.00 .45 -.19 .43
CORR x2 .45 1.00 -.02 .62
CORR x3 -.19 -.02 1.00 -.29
CORR x4 .43 .62 -.29 1.00
* If you do not input N, default is to assume that N=10,000;
proc cancorr data=corrmat simple corr;
var x1 x2;
with x3 x4;
title 'Canonical Correlation Analysis of WAIS';
```

Edited Output

Correlations Among the VAR Variables

	x1	x2	
x1	1.0000	0.4500	$\leftarrow \mathbf{R}_{11}$
x2	0.4500	1.0000	

Correlations Among the WITH Variables

	x3	x4	
x3	1.0000	-0.2900	$\leftarrow \mathbf{R}_{22}$
x4	-0.2900	1.0000	

Correlations Between the VAR Variables and the WITH Variables

	x3	x4	
x1	-0.1900	0.4300	$\leftarrow \mathbf{R}_{12}$
x2	-0.0200	0.6200	

Canonical Correlations, etc

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation
rho1=	0.654638	0.653850	0.018718	0.428551
rho2=	0.195178	.	0.031508	0.038095

Test of H0: The canonical correlations in the current row and all that follow are zero

	Likelihood Ratio	Approx F Value	Num DF	Den DF	Pr > F	
1	0.54967944	162.01	4	1858	<.0001	$H_o : \rho_1 = \rho_2 = 0$
2	0.96190539	36.83	1	930	<.0001	$H_o : \rho_2 = 0$

Canonical Coefficients

Raw Canonical Coefficients for the VAR Variables

	V1	V2	
x1	0.2285863899	-1.096205618	← columns $= a_i$
x2	0.8760790439	0.6974267017	

Raw Canonical Coefficients for the WITH Variables

	W1	W2	
x3	0.2085960958	1.02387007	← columns $= b_i$
x4	1.0403638062	0.0972902985	

Notes:

- Since we input a correlation matrix,"raw" are same as standardized coefficients.
- SAS "V" is same as lecture notes "U".
- SAS "W" is same as lecture notes "V".

Structure Coefficients

```
Correlations Between the VAR Variables and
Their Canonical Variables

              V1                               V2

x1    0.6228                              -0.7824
x2    0.9789                               0.2041

Correlations Between the WITH Variables and
Their Canonical Variables

              W1                               W2

x3   -0.0931                               0.9957
x4    0.9799                              -0.1996
```

Index Coefficients

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

	W1	W2
x1	0.4077	-0.1527
x2	0.6409	0.0398

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

	V1	V2
x3	-0.0610	0.1943
x4	0.6415	-0.0390

More than Two Sets

For $M > 2$, let p$= \sum_{m=1}^{M} p_m$,

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \\ \hline \vdots \\ \hline \mathbf{X}_m \end{pmatrix} \begin{matrix} \} p_1 \\ \} p_2 \\ \vdots \\ \} p_M \end{matrix} \quad \text{and} \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \cdots & \mathbf{\Sigma}_{1M} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \cdots & \mathbf{\Sigma}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{M1} & \mathbf{\Sigma}_{M2} & \cdots & \mathbf{\Sigma}_{MM} \end{pmatrix}$$

Consider the linear combinations

$$\begin{aligned} Z_{11} &= \mathbf{a}'_{11}\mathbf{X}_1 \\ Z_{12} &= \mathbf{a}'_{12}\mathbf{X}_2 \\ Z_{1M} &= \mathbf{a}'_{1M}\mathbf{X}_M \end{aligned}$$

where $\mathbf{a}_{1m}$ is the $(p_m \times 1)$ vector for the m^{th} canonical variable.

The (estimated) covariance matrix of $(Z_{11}, Z_{12}, \ldots, Z_{1M})$ is $\Phi(\hat{1})_{(M\times M)}$, which equals. . . .

Set up for More than Two Sets

$$\Phi(\hat{1})_{(M\times M)} = \begin{pmatrix} 1 & \hat{\phi}_{12}(1) & \hat{\phi}_{13}(1) & \cdots & \hat{\phi}_{1M}(1) \\ \hat{\phi}_{21}(1) & 1 & \hat{\phi}_{23}(1) & \cdots & \hat{\phi}_{2M}(1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\phi}_{M1}(1) & \hat{\phi}_{M2}(1) & \hat{\phi}_{M3}(1) & \cdots & 1 \end{pmatrix}$$

where $\hat{\phi}_{ii(1)} = \mathbf{a}'_{1i}\mathbf{\Sigma}_{ii}\mathbf{a}_{1i} = 1$ and $\hat{\phi}_{ik(1)} = \mathbf{a}'_{1i}\mathbf{\Sigma}_{ik}\mathbf{a}_{1k}$.

- In the two set case, we only had one off diagonal element that could maximize, i.e., $\hat{\phi}_{12}(1)$.
- There are (at least) five ways to generalized canonical correlation to a multiple set problems.
- For $M = 2$ they are all equivalent to what we've talked about.

Horst's Suggestions

SUMCOR: Horst (1965) "Factor analysis of data matrices"
Horst suggested maximizing

$$\max_{\mathbf{a}_1,\ldots,\mathbf{a}_M} \sum_{i<k} \hat{\phi}_{ik}(1)$$

That is, sum of all off-diagonal elements of the correlation matrix between the M canonical variates $Z_1, \ldots, Z_M$, i.e., $\hat{\Phi}(1)$.

MAXVAR: Horst also suggest maximizing the variance of a linear combination of $\mathbf{Z}' = (Z_{11}, Z_{12}, \ldots, Z_{1M})$, which is the first principal component of $\hat{\Phi}(1)$.
The variance of this maximal linear combination is the largest eigenvalue of $\hat{\Phi}(1)$; that is, λ_1.

Kettenring's Suggestions

Kettenring (1971), *Biometrika*

SSQCOR: Find the linear combinations that maximize

$$\max_{\mathbf{a}_1,\ldots,\mathbf{a}_M} \left(\sum_{i<k} \hat{\phi}_{ik}^2(1)\right) = \sum_{i=1}^{M} \lambda_{ik}^2 - M$$

MINVAR: Kettenring also suggested minimizing the smallest eigenvalue λ_{1M}; that is, the variance of the minimal linear combination.

GENVAR

Steel (1951) in the *Annals of Mathematical Statistics* suggested minimizing the generalized sample variance

$$\det(\hat{\Phi}(1)) = \prod_{m=1}^{M} \lambda_{ij}$$

Once the first canonical variate has been found, all five methods call all be extended to find $\mathbf{Z}_2, \mathbf{Z}_3, \ldots, \mathbf{Z}_M$ such that they are orthogonal to ones previously found.

Summary/Discusion

- Discriminant analysis is a special case of canonical correlation analysis where one set of variables is a dummy coded variable that defines populations. i.e., For individual j in group i,

$$\mathbf{X}_{1j}'{}_{(1\times p)} = (0, \ldots, \underbrace{1}_{i^{th}}, \ldots 0).$$

  The other set of variables are q continuous/numerical ones.
- Discriminant analysis and MANOVA use the same matrix $\mathbf{W}^{-1}\mathbf{B}$ or $\mathbf{E}^{-1}\mathbf{H}$.
- (Binary or Multinomial) Logistic regression is the "flip" side of discriminant analysis and MANOVA (i.e., interchange "response" and "predictor" roles).
  "Conditional Gaussian distribution" or the "location model".
- See handout on similarities and differences between PCA, MANOVA, DA, and canonical correlation analysis.