Multivariate Data Analysis

(MGT513, BAT531, TIM711)

Lecture 7

Ch.9 Canonical Correlation

Introduction

- **Goal:** To understand the relationship between two sets of variables $\boldsymbol{X}$ and $\boldsymbol{Y}$
  – Canonical correlation analysis is used to identify and measure the associations among two sets of variables
  – Canonical correlation is appropriate in the same situations where multiple regression would be, but where are there are multiple intercorrelated outcome variables
  – Canonical correlation analysis determines a set of canonical variates, orthogonal linear combinations of the variables within each set that best explain the variability both within and between sets

Source: https://stats.idre.ucla.edu/r/dae/canonical-correlation-analysis/

Examples of canonical correlation analysis

- A researcher has collected data on three psychological variables, four academic variables (standardized test scores) and gender for 600 college freshman. She is interested in how the set of psychological variables relates to the academic variables and gender. In particular, the researcher is interested in how many dimensions (canonical variables) are necessary to understand the association between the two sets of variables.
- A researcher is interested in exploring associations among factors from two multidimensional personality tests, the MMPI and the NEO. She is interested in what dimensions are common between the tests and how much variance is shared between them. She is specifically interested in finding whether the neuroticism dimension from the NEO can account for a substantial amount of shared variance between the two tests.

Source: https://stats.idre.ucla.edu/r/dae/canonical-correlation-analysis/

Examples of canonical correlation analysis

- In psychology, canonical correlation has been used to study the relationship between personality characteristics and vocational interests (Cooley and Lohnes, 1971). Researchers found three pairs of canonical variates (from dozens of different measures of personality and vocational interests) that enabled them to succinctly describe the types of individuals most interested in certain types of jobs. In a similar study, Thorndike, Dawis, and Weiss (1968) used canonical correlation to study the multivariate relationships between vocational interests (based on the Minnesota Vocational Interest Inventory, a set of measures of the interests of nonprofessional men) and vocational needs (as measured by the Minnesota Importance Questionnaire). The authors found evidence for two pairs of canonical variates that held up to a holdout cross-validation study.

Examples of canonical correlation analysis

- In the operations literature, Pisharodi and Langley (1991) examined the relationship between customer service and market response, using data from supplier-customer pairs in the grocery industry. The study allowed the authors to identify which service factors (e.g.. design, delivery, and communication) were most important in explaining the variation in consumer response.

- In a study of the adventure-recreation experience. Ewert and Hollenhorst (1994) examined the relationship between measures of the individual (reflecting experience, skills, involvement, and desire for control) with recreation setting (e.g., naturalness, social orientation, and equipment requirements). In two different samples of subjects (one group of white-waler boaters and another group of rock climbers), they found evidence of significant correlation.

Examples of canonical correlation analysis

- Sometimes, canonical correlation is useful just to establish whether a significant relationship exists between two sets of variables. For example. Champoux (1991) used canonical correlation to see whether a significant multivariate relationship existed among variables describing job characteristics and work motivation before performing more specific tests of theory using multiple regression analysis.

Intuition

- In a sense, a direct generalization of multiple regression
  - Univariate Y and Multivariate $\mathbf{X}$

- Object function of canonical correlation is different from that of OLS multiple regression
  - Instead of minimizing the sum of the squared deviations, we look to find the linear combination of the independent variables *X* that *maximizes* the correlation with the single dependent variable *Y*.

Intuition

- For simple discussion, let's use standardized version of *univariate* Y and *multivariate* $\mathbf{X}$
- Let $\mathbf{Xb}$ denote the linear combination we seek
- The correlation between Y and multivariate $\mathbf{Xb}$:

$$r(Y, \mathbf{Xb}) = \frac{\mathrm{cov}(Y, \mathbf{Xb})}{\sqrt{\mathrm{var}(Y)\mathrm{var}(\mathbf{Xb})}} = \mathrm{cov}(Y, \mathbf{Xb})$$

where $\mathrm{var}(Y) = 1$

- **Claim**: Choose $\mathbf{b}$ to maximize $\mathrm{cov}(Y, \mathbf{Xb}) = \frac{1}{(n-1)} y^{\mathrm{T}}\mathbf{Xb}$

such that $\mathrm{var}(\mathbf{Xb}) = \frac{1}{(n-1)} \mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Xb} = 1$

Intuition

Solution: Lagrangian Method - Let λ be a Lagrange multiplier

$$L = \mathrm{cov}(Y, \mathbf{Xb}) - \lambda(\mathrm{var}(\mathbf{Xb}) - 1)$$
$$= \frac{1}{(n-1)} y^{\mathrm{T}}\mathbf{Xb} - \lambda\left(\frac{1}{(n-1)}\mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Xb} \ - 1\right)$$

- The first order condition with respect to $\mathbf{b}$ is

$$\frac{\vartheta L}{\vartheta \mathbf{b}} = \frac{1}{(n-1)}\mathbf{y}^{\mathrm{T}}\mathbf{X} - \frac{2\lambda}{(n-1)}\mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X} = 0$$

- Rearranging the above gives

$$\mathbf{b} = 2\lambda(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

- In other words,

$$\mathbf{b} \propto (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

where the right term is OLS estimate of multiple linear regression

Intuition

Multivariate $\mathbf{Y}$ instead of univariate Y

– What linear combination of $\mathbf{X}$ and linear combination of $\mathbf{Y}$ produces the highest correlation?

- Let $\mathbf{t} = \mathbf{Ya}$ denote the linear combination of $\mathbf{Y}$
- Let $\mathbf{u} = \mathbf{Xb}$ denote the linear combination of $\mathbf{X}$
- Choose $\mathbf{a}$ and $\mathbf{b}$ to maximize $r(\mathbf{t}, \mathbf{u}) = \frac{1}{(n-1)}\mathbf{a}^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}\mathbf{Xb}$

such that $\mathrm{var}(\mathbf{t}) = \mathrm{var}(\mathbf{Ya}) = \frac{1}{(n-1)}\mathbf{a}^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}\mathbf{Ya} = 1$ and

$$\mathrm{var}(\mathbf{u}) = \mathrm{var}(\mathbf{Xb}) = \frac{1}{(n-1)}\mathbf{b}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Xb} = 1$$

- Canonical variates(variables/scores): the new variables $\mathbf{t}$ and $\mathbf{u}$
- Canonical correlation: the correlation between $\mathbf{t}$ and $\mathbf{u}$ [$r(\mathbf{t}, \mathbf{u})$]

Intuition

Simple example $\mathbf{X} = (X_1, X_2)$, $\mathbf{Y} = (Y_1, Y_2)$

TABLE 9.1 Correlation matrix for four variables: Y_1, Y_2, X_1, and X_2

	Y_1	Y_2	X_1	X_2
Y_1	1.000	−0.307	0.221	0.445
Y_2	−0.307	1.000	0.316	0.168
X_1	0.221	0.316	1.000	−0.176
X_2	0.445	0.168	−0.176	1.000

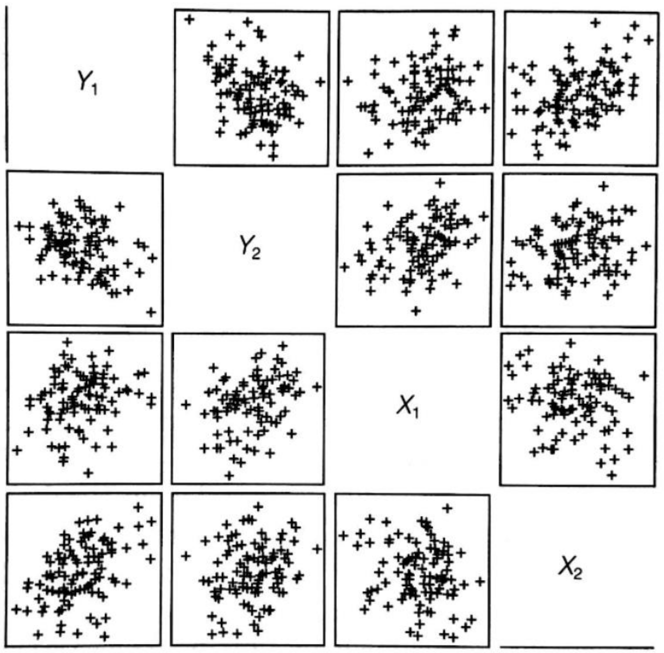


FIGURE 9.1 Pairwise scatter plots for Y_1, Y_2, X_1, and X_2

Intuition

Rough method - Grid search (exhaustive numerical search)

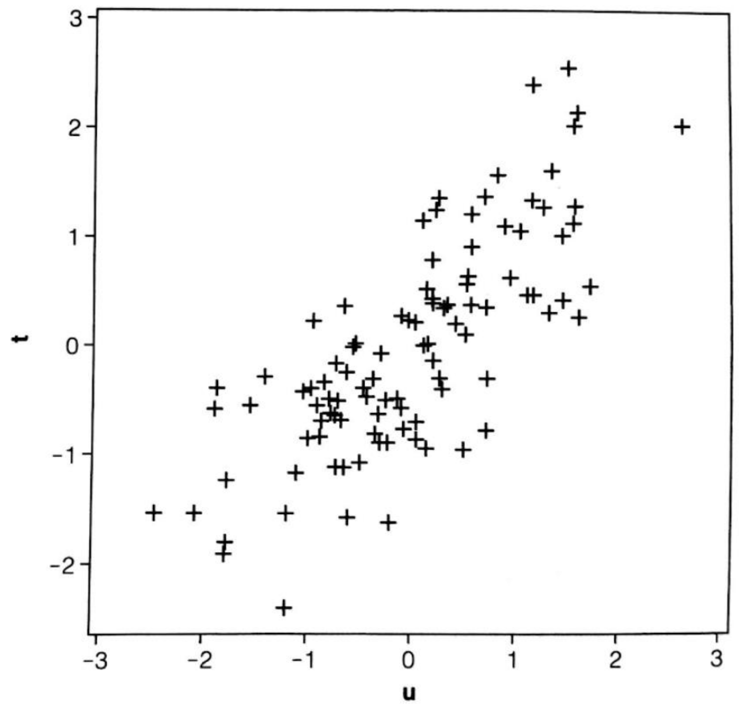
TABLE 9.2 Grid showing correlations between different combinations of *Y* (columns) and *X* (rows)

	100/0	90/10	80/20	70/30	60/40	50/50	40/60	30/70	20/80	10/90	0/100
100/0	0.221	0.343	0.390	0.422	0.444	0.456	0.460	0.455	0.440	0.410	0.316
90/10	0.370	0.520	0.574	0.607	0.625	0.632	0.626	0.608	0.575	0.522	0.374
80/20	0.428	0.584	0.639	0.672	0.688	0.691	0.681	0.658	0.618	0.556	0.386
70/30	0.468	0.628	0.682	0.713	0.728	0.728	0.715	0.687	0.642	0.573	0.390
60/40	0.497	0.657	0.711	0.740	0.753	0.750	0.734	0.702	0.654	0.580	0.386
50/50	0.518	0.677	0.728	0.755	0.766	0.761	0.742	0.707	0.655	0.578	0.377
40/60	0.532	0.686	0.735	0.760	$\boxed{0.767}$	0.760	0.739	0.702	0.648	0.568	0.363
30/70	0.538	0.685	0.731	0.753	0.758	0.748	0.725	0.686	0.630	0.548	0.343
20/80	0.536	0.673	0.714	0.732	0.734	0.722	0.697	0.657	0.600	0.518	0.315
10/90	0.520	0.642	0.677	0.690	0.689	0.675	0.647	0.607	0.550	0.470	0.274
0/100	0.445	0.526	0.544	0.547	0.539	0.520	0.492	0.453	0.402	0.332	0.168

Intuition

Scatter plot of canonical variates – the first pair: $(\mathbf{t}_1, \mathbf{u}_1)$

FIGURE 9.2
Scatter plot of canonical variates from illustration



Intuition

- When each set has more than one variable, we may be able to capture more of the relationship between the two sets of variables by finding additional pairs of canonical variates.

- Canonical variates – the second pair: $(\mathbf{t}_2, \mathbf{u}_2)$
  - Choose $\mathbf{a}_2$ and $\mathbf{b}_2$ to maximize $\frac{1}{(n-1)}\mathbf{a}_2^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}\mathbf{X}\mathbf{b}_2$

    Such that $\frac{1}{(n-1)}\mathbf{a}_2^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\mathbf{a}_2 = 1$ and $\frac{1}{(n-1)}\mathbf{b}_2^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b}_2 = 1$

    $r(\mathbf{t}_1, \mathbf{t}_2) = \mathbf{0}$ and $r(\mathbf{u}_1, \mathbf{u}_2) = \mathbf{0}$

Mechanics

- 1st stage
  - $\mathbf{u}_1 = \mathbf{X}\mathbf{b}_1$: the first linear combination of the first set of variables
  - $\mathbf{t}_1 = \mathbf{Y}\mathbf{a}_1$: the first linear combination of the second set of variables
  - Choose $\mathbf{a}_1$ and $\mathbf{b}_1$ to maximize $r(\mathbf{t}_1, \mathbf{u}_1)$
- 2nd stage
  - $\mathbf{u}_2 = \mathbf{X}\mathbf{b}_2$: the second linear combination of the first set of variables
  - $\mathbf{t}_2 = \mathbf{Y}\mathbf{a}_2$: the second linear combination of the second set of variables
  - Choose $\mathbf{a}_2$ and $\mathbf{b}_2$ to maximize $r(\mathbf{t}_2, \mathbf{u}_2)$

    such that $r(\mathbf{t}_1, \mathbf{t}_2) = \mathbf{0}$ and $r(\mathbf{u}_1, \mathbf{u}_2) = \mathbf{0}$
- The number of stages: $\min(p, q)$
  - p: the number of $\mathbf{X}$ variables
  - q: the number of $\mathbf{Y}$ variables

Mechanics

In matrix form:

- Choose **a** and **b** to maximize $r(\mathbf{t},\mathbf{u}) = \frac{\mathrm{cov}(\mathbf{t},\mathbf{u})}{\sqrt{\mathrm{var}(\mathbf{t})\mathrm{var}(\mathbf{u})}}$

$$\mathrm{cov}(\mathbf{t},\mathbf{u}) = \frac{[\mathbf{t}^{\mathrm{T}}\mathbf{u}]}{(n-1)} = \frac{[\mathbf{a}^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}\mathbf{X}\mathbf{b}]}{(n-1)} = \mathbf{a}^{\mathrm{T}}\mathbf{R}_{\mathbf{YX}}\mathbf{b}$$

$$\mathrm{var}(\mathbf{t}) = 1 \Rightarrow \frac{[\mathbf{t}^{\mathrm{T}}\mathbf{t}]}{(n-1)} = 1$$

$$\Rightarrow \frac{\mathbf{a}^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\mathbf{a}}{(n-1)} = 1$$

$$\Rightarrow \mathbf{a}^{\mathrm{T}}\mathbf{R}_{\mathbf{YY}}\mathbf{a} = 1$$

Similarly, setting $\mathrm{var}(\mathbf{u}) = 1$ is the same as setting $\mathbf{b}^{\mathrm{T}}\mathbf{R}_{\mathbf{XX}}\mathbf{b} = 1$

Mechanics

- Choose $\mathbf{a}$ and $\mathbf{b}$ to maximize $\mathbf{a}^{\mathrm{T}}\mathbf{R}_{\mathbf{YX}}\mathbf{b}$
  subject to $\mathbf{a}^{\mathrm{T}}\mathbf{R}_{\mathbf{YY}}\mathbf{a} = 1$ and $\mathbf{b}^{\mathrm{T}}\mathbf{R}_{\mathbf{XX}}\mathbf{b} = 1$

Lagrangian Method: Using $\alpha/2$ and $\beta/2$ as Lagrange multipliers

$$L = \mathbf{a}^{\mathrm{T}}\mathbf{R}_{\mathbf{YX}}\mathbf{b} - \frac{\alpha}{2}(\mathbf{a}^{\mathrm{T}}\mathbf{R}_{\mathbf{YY}}\mathbf{a} - 1) - \frac{\beta}{2}(\mathbf{b}^{\mathrm{T}}\mathbf{R}_{\mathbf{XX}}\mathbf{b} - 1)$$

$$\frac{\partial L}{\partial \mathbf{a}} = 0 \Rightarrow \mathbf{R}_{\mathbf{YX}}\mathbf{b} - \alpha\mathbf{R}_{\mathbf{YY}}\mathbf{a} = 0$$

$$\frac{\partial L}{\partial \mathbf{b}} = 0 \Rightarrow \mathbf{R}_{\mathbf{XY}}\mathbf{a} - \beta\mathbf{R}_{\mathbf{XX}}\mathbf{b} = 0$$

Premultiplying $\mathbf{a}^{\mathrm{T}}$

$$\mathbf{a}^{\mathrm{T}}\mathbf{R}_{\mathbf{YX}}\mathbf{b} - \alpha(\mathbf{a}^{\mathrm{T}}\mathbf{R}_{\mathbf{YY}}\mathbf{a}) = 0$$

$\alpha = r(\mathbf{t}, \mathbf{u})$ = the canonical correlation since $\mathbf{a}^{\mathrm{T}}\mathbf{R}_{\mathbf{YY}}\mathbf{a} = 1$

Mechanics

Similarly, Premultiplying $\mathbf{b}^{\mathrm{T}}$

$$\mathbf{b}^{\mathrm{T}}\mathbf{R}_{\mathbf{XY}}\mathbf{a} - \beta(\mathbf{b}^{\mathrm{T}}\mathbf{R}_{\mathbf{XX}}\mathbf{b}) = 0$$

$$\beta = r(\mathbf{t}, \mathbf{u}) = \alpha$$

To summarize, since $\mathrm{var}(\mathbf{u}) = 1$ and $\mathrm{var}(\boldsymbol{t}) = 1$

$$r(\mathbf{t}, \mathbf{u}) = \frac{\mathrm{cov}(\mathbf{t}, \mathbf{u})}{\sqrt{\mathrm{var}(\mathbf{t})\mathrm{var}(\mathbf{u})}} = \alpha = \beta$$

Mechanics

Write $\mathbf{a}$ as a function of $\mathbf{b}$

Since $\mathbf{R_{YX}b} - \alpha\mathbf{R_{YY}a} = 0$ (*first order condition*) and $\alpha = r(\mathbf{t},\mathbf{u})$

$$\mathbf{R_{YX}b} = r(\mathbf{t},\mathbf{u})\mathbf{R_{YY}a}$$

$$\mathbf{a} = [\frac{\mathbf{1}}{r(\mathbf{t},\mathbf{u})}]\mathbf{R_{YY}}^{-1}\mathbf{R_{YX}b}$$

Since $\mathbf{R_{XY}a} - \beta\mathbf{R_{XX}b} = 0$ (*first order condition*) and $\beta = r(\mathbf{t},\mathbf{u})$

After rearranging,

$$\mathbf{R_{XY}}\left\{\frac{\mathbf{1}}{r(\mathbf{t},\mathbf{u})}\mathbf{R_{YY}}^{-1}\mathbf{R_{YX}b}\right\} = r(\mathbf{t},\mathbf{u})\mathbf{R_{XX}b}$$

$$\left[\mathbf{R_{XX}}^{-1}\mathbf{R_{XY}}\mathbf{R_{YY}}^{-1}\mathbf{R_{YX}}\right]\mathbf{b} = r^{2}(\mathbf{t},\mathbf{u})\mathbf{b}$$

=> <u>*Eigenvector-Eigenvalue problem!*</u>

Mechanics

Equivalently, write $\mathbf{b}$ as a function of $\mathbf{a}$

$$\mathbf{b} = [\frac{\mathbf{1}}{r(\mathbf{t},\mathbf{u})}]\mathbf{R_{XX}}^{-1}\mathbf{R_{XY}}\mathbf{a}$$

After rearranging

$$\mathbf{R_{YX}}\left\{\frac{\mathbf{1}}{r(\mathbf{t},\mathbf{u})}\mathbf{R_{XX}}^{-1}\mathbf{R_{XY}}\mathbf{a}\right\} = r(\mathbf{t},\mathbf{u})\mathbf{R_{YY}}\mathbf{a}$$

$$\left[\mathbf{R_{YY}}^{-1}\mathbf{R_{YX}}\mathbf{R_{XX}}^{-1}\mathbf{R_{XY}}\right]\mathbf{a} = r^{2}(\mathbf{t},\mathbf{u})\mathbf{a}$$

=> *<u>Eigenvector-Eigenvalue problem again!</u>*

Mechanics

Therefore,

- The canonical weights $\mathbf{b}$ are the eigenvectors of the matrix product:

$$\mathbf{R_{XX}}^{-1}\mathbf{R_{XY}}\mathbf{R_{YY}}^{-1}\mathbf{R_{YX}}$$

- The canonical weights $\mathbf{a}$ are the eigenvectors of the matrix product:

$$\mathbf{R_{YY}}^{-1}\mathbf{R_{YX}}\mathbf{R_{XX}}^{-1}\mathbf{R_{XY}}$$

- The canonical correlations are given by *the square roots* of the associated eigenvalues.

Mechanics

The correlation matrix can be portioned as follows:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{\mathbf{YY}} & \mathbf{R}_{\mathbf{YX}} \\ \mathbf{R}_{\mathbf{XY}} & \mathbf{R}_{\mathbf{XX}} \end{bmatrix} = \begin{bmatrix} 1.000 & -0.307 & 0.221 & 0.445 \\ -0.307 & 1.000 & 0.316 & 0.168 \\ 0.221 & 0.316 & 1.000 & -0.176 \\ 0.445 & 0.168 & -0.176 & 1.000 \end{bmatrix}$$

b: $\mathbf{R}_{\mathbf{XX}}{}^{-1}\mathbf{R}_{\mathbf{XY}}\mathbf{R}_{\mathbf{YY}}{}^{-1}\mathbf{R}_{\mathbf{YX}} = \begin{bmatrix} 0.260 & 0.289 \\ 0.828 & 0.588 \end{bmatrix}$ - *Asymmetric! Eigenvector?*

a: $\mathbf{R}_{\mathbf{YY}}{}^{-1}\mathbf{R}_{\mathbf{YX}}\mathbf{R}_{\mathbf{XX}}{}^{-1}\mathbf{R}_{\mathbf{XY}} = \begin{bmatrix} 0.382 & 0.252 \\ 0.299 & 0.229 \end{bmatrix}$ - *Asymmetric! Eigenvector?*

Trick: Define $\mathbf{Q} = \mathbf{R}_{\mathbf{YY}}{}^{-1/2}\mathbf{R}_{\mathbf{YX}}\mathbf{R}_{\mathbf{XX}}{}^{-1/2}$ where rank of $\mathbf{Q}$ is $k = \min(p, q)$

Mechanics

Singular value decomposition of an *asymmetric* matrix $\mathbf{Q}$

$$\mathbf{Q} = \mathbf{R_{YY}}^{-1/2}\mathbf{R_{YX}}\mathbf{R_{XX}}^{-1/2}$$

$$\mathbf{Q} = \mathbf{MDN^T} = (m_1, m_2, \dots, m_k)\mathbf{D}(n_1, n_2, \dots, n_k)^\mathbf{T}$$

$$\mathbf{D} = \begin{bmatrix} \lambda_1{}^{1/2} & 0 & 0 & 0 \\ 0 & \lambda_2{}^{1/2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_k{}^{1/2} \end{bmatrix}$$

- Canonical coefficients:
  - $\mathbf{a} = \mathbf{R_{YY}}^{-1/2}\mathbf{M}$
  - $\mathbf{b} = \mathbf{R_{XX}}^{-1/2}\mathbf{N}$
- Canonical correlations:
  - $(\lambda_1{}^{1/2}, \lambda_2{}^{1/2}, \dots, \lambda_k{}^{1/2})$

Mechanics

Alternatively, Spectral decomposition (eigenvalue decomposition) of a *symmetric* matrix $\mathbf{QQ^T}$ and $\mathbf{Q^TQ}$

- $\mathbf{QQ^T}$
  - Eigenvalues: $(\lambda_1, \lambda_2, \dots, \lambda_k)$
  - Eigenvectors: $(m_1, m_2, \dots, m_k)$
- $\mathbf{Q^TQ}$
  - Eigenvalues: $(\lambda_1, \lambda_2, \dots, \lambda_k)$
  - Eigenvectors: $(n_1, n_2, \dots, n_k)$
- Canonical coefficients:
  - $\mathbf{a} = \mathbf{R_{YY}}^{-1/2}\mathbf{M}$
  - $\mathbf{b} = \mathbf{R_{XX}}^{-1/2}\mathbf{N}$
- Canonical correlations:
  - $({\lambda_1}^{1/2}, {\lambda_2}^{1/2}, \dots, {\lambda_k}^{1/2})$

Mechanics

The correlation matrix can be portioned as follows:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R_{YY}} & \mathbf{R_{YX}} \\ \mathbf{R_{XY}} & \mathbf{R_{XX}} \end{bmatrix} = \begin{bmatrix} 1.000 & -0.307 & 0.221 & 0.445 \\ -0.307 & 1.000 & 0.316 & 0.168 \\ 0.221 & 0.316 & 1.000 & -0.176 \\ 0.445 & 0.168 & -0.176 & 1.000 \end{bmatrix}$$

- Eigenvalues: $(0.59, 0.02)$
- Eigenvectors:
  - $\mathbf{a}_1' = (0.92, 0.76)$, $\mathbf{a}_2' = (-0.50, 0.72)$
  - $\mathbf{b}_1' = (0.73, 0.83)$, $\mathbf{b}_2' = (0.71, -0.59)$

Canonical Loadings

- Canonical Loadings: the correlations between the original variables and the canonical variates.
- The correlation between $\mathbf{X}$ and $\mathbf{u}$:

$$\mathbf{f} = \frac{1}{(n-1)}\mathbf{X}^{\mathrm{T}}\mathbf{u} = \frac{1}{(n-1)}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{b}) = \mathbf{R}_{\mathbf{XX}}\mathbf{b}$$

- The correlation between $\mathbf{Y}$ and $\mathbf{t}$:

$$\mathbf{g} = \frac{1}{(n-1)}\mathbf{Y}^{\mathrm{T}}\mathbf{t} = \frac{1}{(n-1)}\mathbf{Y}^{\mathrm{T}}(\mathbf{Y}\mathbf{a}) = \mathbf{R}_{\mathbf{YY}}\mathbf{a}$$

TABLE 9.3 Canonical loadings for simple four-variable illustration

	$\mathbf{u}_1$	$\mathbf{u}_2$
X_1	0.5793	0.8151
X_2	0.7004	−0.7137
	$\mathbf{t}_1$	$\mathbf{t}_2$
Y_1	0.6877	−0.7260
Y_2	0.4798	0.8774

Redundancies

- $r^2(\mathbf{t}, \mathbf{u})$ does not tell us how much of the variance in $\boldsymbol{Y}$ is explained by $\boldsymbol{X}$. In fact, it explains how much of the variance in $\boldsymbol{t}$ is explained by $\boldsymbol{u}$.
- To measure how much of the variance in $\boldsymbol{Y}$ is explained by $\boldsymbol{X}$

$$Rd(\boldsymbol{t}|\boldsymbol{u}) = \left[\frac{\text{variance in } \mathbf{t} \text{ explained by } \mathbf{u}}{\text{variance in } \mathbf{t}}\right] \times \left[\frac{\text{variance in } \mathbf{t}}{\text{variance in } \mathbf{Y}}\right]$$

$$Rd(\boldsymbol{t}|\boldsymbol{u}) = \frac{r^2(\boldsymbol{t}, \boldsymbol{u})(\boldsymbol{g}^T\boldsymbol{g})}{q}$$

TABLE 9.3 Canonical loadings for simple four-variable illustration

	$\mathbf{u}_1$	$\mathbf{u}_2$
X_1	0.5793	0.8151
X_2	0.7004	−0.7137
	$\mathbf{t}_1$	$\mathbf{t}_2$
Y_1	0.6877	−0.7260
Y_2	0.4798	0.8774

TABLE 9.4 Measures of redundancy for simple four-variable illustration

	Proportion of Variance Accounted for by **t**	Proportion of Variance Accounted for in Y		
		Canonical R-Squared	$Rd(\mathbf{t}\|\mathbf{u})$	Cumulative Proportion
1	0.3516	0.5900	0.2074	0.2074
2	0.6484	0.0207	0.0134	0.2208

Sample problem: Fader and Lodish (1990)

- Studying the determinants of promotional activity across different supermarket product categories
- Fader and Lodish (1990) wanted to know whether promotional activity (i.e., the types of promotions offered by marketers) differed across categories according to the characteristics of the category.

TABLE 9.5 Correlation matrix for Fader and Lodish data

Structural Variables (X)	
PENET	Percentage of households making at least one category purchase
PCYCLE	Average interpurchase time
PRICE	Average dollars spent in the category per purchase occasion
PVTSH	Combined market share for all private-label and generic products
PUR/HH	Average number of purchase occasions per household during the year
Promotional Variables (Y)	
FEAT	Percent of volume sold on feature (advertised in local newspaper)
DISP	Percent of volume sold on display (e.g., end of aisle)
PCUT	Percent of volume sold at a temporary reduced price
SCOUP	Percent of volume purchased using a retailer's store coupon
MCOUP	Percent of volume purchased using a manufacturer's coupon

Sample problem: Fader and Lodish (1990)

- Structural Variables (***X***): structural characteristics of the category
- Promotional Variables (***Y***): promotional activities

TABLE 9.6 Correlation matrix for Fader and Lodish data

	PENET	*PCYCLE*	*PRICE*	*PVTSH*	*PURHH*	*FEAT*	*DISP*	*PCUT*	*SCOUP*	*MCOUP*
PENET	1.000									
PCYCLE	−0.478	1.000								
PRICE	−0.222	−0.146	1.000							
PVTSH	0.409	−0.127	−0.280	1.000						
PURHH	0.617	−0.719	0.068	0.246	1.000					
FEAT	0.580	−0.379	−0.001	0.270	0.373	1.000				
DISP	0.461	−0.252	−0.111	0.132	0.213	0.535	1.000			
PCUT	0.569	−0.394	−0.108	0.295	0.368	0.918	0.515	1.000		
SCOUP	0.389	−0.178	0.074	0.223	0.261	0.674	0.375	0.588	1.000	
MCOUP	0.053	0.049	0.237	−0.215	−0.026	−0.044	−0.038	−0.040	−0.065	1.000

Sample problem: Fader and Lodish (1990)

- The objective of the canonical correlation analysis is to see the extent to which the structural variables can be used to explain the observed variation in the promotional variables, as well as to better understand the patterns of covariation in the data.

TABLE 9.7 Results from canonical correlation analysis of Fader and Lodish data

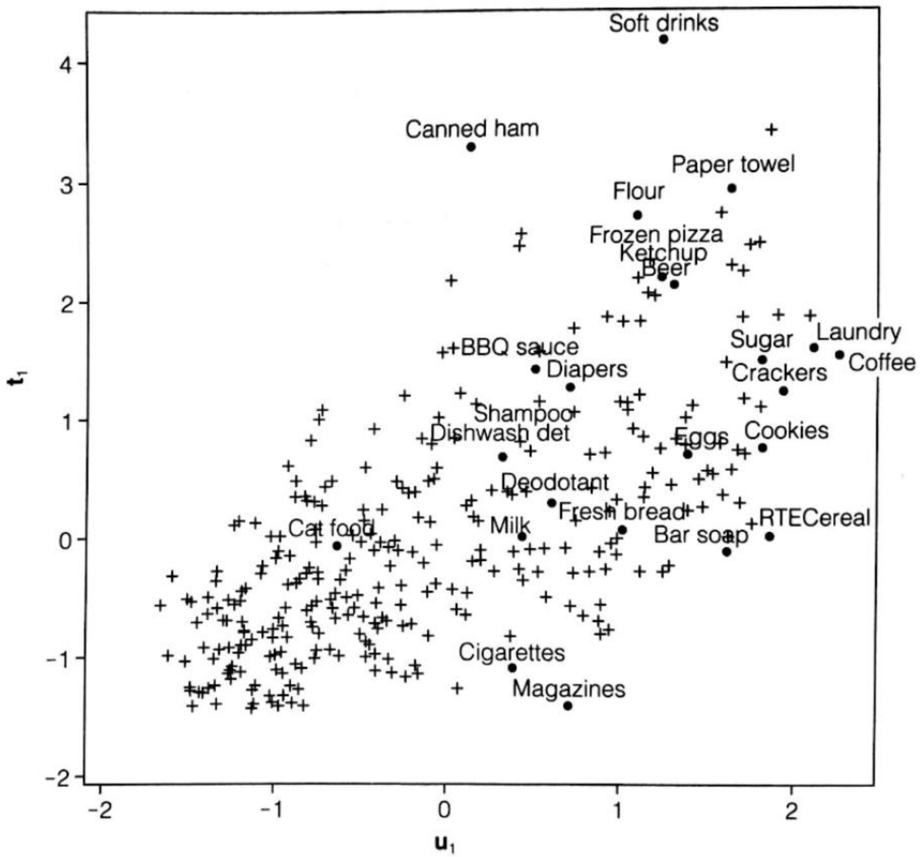
	Canonical Correlations				
	$r(\mathbf{t}_1, \mathbf{u}_1)$	$r(\mathbf{t}_2, \mathbf{u}_2)$	$r(\mathbf{t}_3, \mathbf{u}_3)$	$r(\mathbf{t}_4, \mathbf{u}_4)$	$r(\mathbf{t}_5, \mathbf{u}_5)$
	0.642	0.483	0.265	0.114	0.032

	Canonical Loadings		
	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$
PENET	0.96	−0.11	0.04
PURHH	0.55	−0.15	0.39
PCYCLE	−0.58	0.32	−0.06
PRICE	−0.01	0.77	0.28
PVTSH	0.34	−0.47	0.71
	$\mathbf{t}_1$	$\mathbf{t}_2$	$\mathbf{t}_3$
FEAT	0.94	−0.07	0.29
DISP	0.73	−0.14	−0.38
PCUT	0.90	−0.32	0.18
SCOUP	0.62	0.17	0.61
MCOUP	0.16	0.72	−0.43

Sample problem: Fader and Lodish (1990)

- First pair

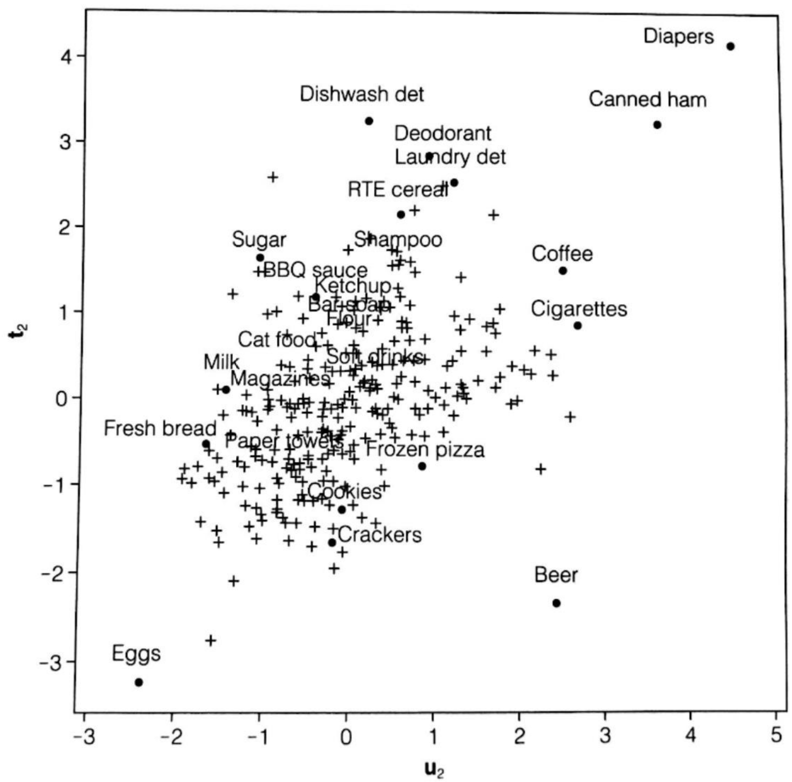
FIGURE 9.3
Scatter plot of first pair of canonical variates from Fader and Lodish data



Sample problem: Fader and Lodish (1990)

- Second pair

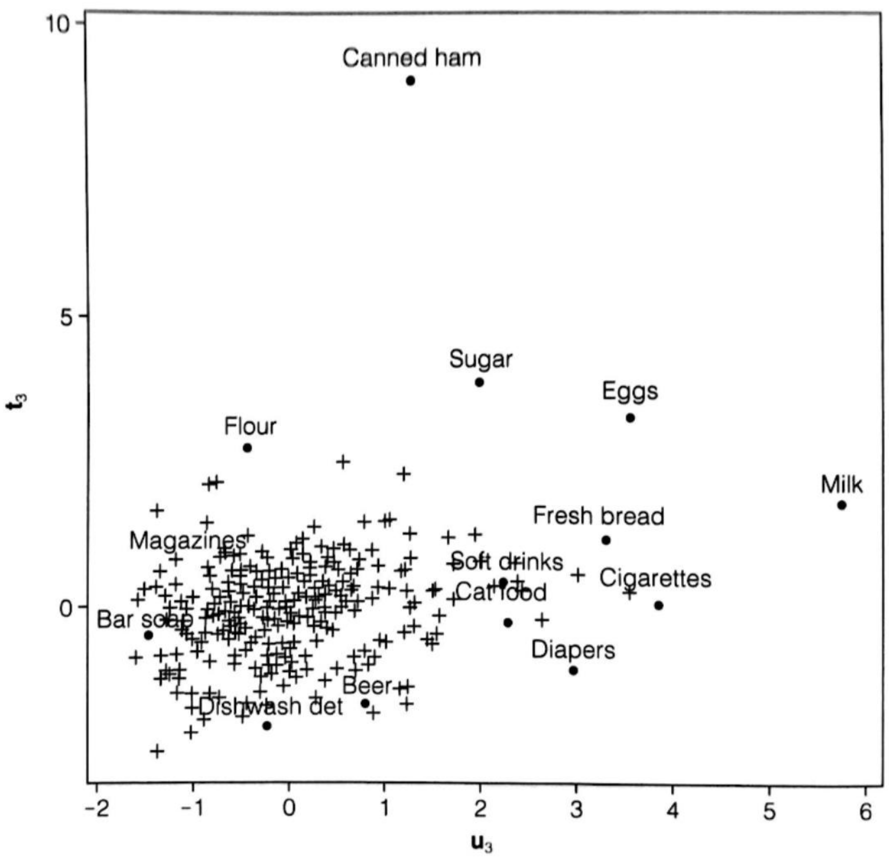
FIGURE 9.4
Plot of second pair of canonical variates from Fader and Lodish data



Sample problem: Fader and Lodish (1990)

- Third pair

FIGURE 9.5
Plot of third pair of canonical variates from Fader and Lodish data



Is the Relationship between the $\boldsymbol{X}$'s and the $\boldsymbol{Y}$'s Significant?

- Regression Analysis: *F* statistics
- Canonical Correlation: Wilks's Λ
  - $\mathbf{S}_{\mathrm{E}}$: the matrix of sum of squared errors (the residual error sum of squares matrix)
  - $\mathbf{S}_{\mathrm{T}}$: the total sum of squares matrix for $\boldsymbol{Y}$
  - $\mathbf{S}_{\mathrm{H}}$: the sum of squares matrix for the variance in $\boldsymbol{Y}$ explained by $\boldsymbol{X}$
  - $\mathbf{S}_{\mathrm{E}} = \mathbf{S}_{\mathrm{T}} - \mathbf{S}_{\mathrm{H}}$

$$\Lambda = \frac{|\mathbf{S}_{\mathbf{E}}|}{|\mathbf{S}_{\mathbf{T}}|} = \frac{|\mathbf{S}_{\mathbf{T}} - \mathbf{S}_{\mathbf{H}}|}{|\mathbf{S}_{\mathbf{T}}|}$$

  - **Idea**: Small Λ implies a strong and significant model

Is the Relationship between the $\boldsymbol{X}$'s and the $\boldsymbol{Y}$'s Significant?

- Some algebra gives
  - $\mathbf{S_T} = \mathbf{Y}^\mathrm{T}\mathbf{Y} \propto \mathbf{R_{YY}}$
  - $\mathbf{S_H} = \mathbf{Y}^\mathrm{T}\mathbf{XB} = \mathbf{Y}^\mathrm{T}\mathbf{X}(\mathbf{X}^\mathrm{T}\mathbf{X})^{-1}\mathbf{X}^\mathrm{T}\mathbf{Y} \propto \mathbf{R_{YY}}\mathbf{R}_{\boldsymbol{XX}}{}^{-1}\mathbf{R_{XY}}$
- Wilks's Λ is given by

$$\Lambda = \frac{|\mathbf{R_{YY}} - \mathbf{R_{YX}}\mathbf{R}_{\boldsymbol{XX}}{}^{-1}\mathbf{R_{XY}}|}{|\mathbf{R_{YY}}|} = |\mathbf{R_{YY}}{}^{-1}||\mathbf{R_{YY}} - \mathbf{R_{YX}}\mathbf{R}_{\boldsymbol{XX}}{}^{-1}\mathbf{R_{XY}}|$$

$$= \left|\mathbf{I} - \mathbf{R_{YY}}{}^{-1}\mathbf{R_{YX}}\mathbf{R}_{\boldsymbol{XX}}{}^{-1}\mathbf{R_{XY}}\right|$$

  - The eigenvalues of $\mathbf{R_{YY}}{}^{-1}\mathbf{R_{YX}}\mathbf{R}_{\boldsymbol{XX}}{}^{-1}\mathbf{R_{XY}}$: $\lambda_1, \lambda_2, \ldots, \lambda_{\min(p,q)}$
  - The eigenvalues of $\mathbf{I} - \mathbf{R_{YY}}{}^{-1}\mathbf{R_{YX}}\mathbf{R}_{\boldsymbol{XX}}{}^{-1}\mathbf{R_{XY}}$: $1 - \lambda_1, 1 - \lambda_2, \ldots, 1 - \lambda_{\min(p,q)}$

$$\Lambda = \prod_{i=1}^{\min(p,q)} (1 - \lambda_i)$$

Is the Relationship between the $\boldsymbol{X}$'s and the $\boldsymbol{Y}$'s Significant?

- Bartlett's Chi-Square Test

$$V = -\left[(n-1) - \frac{(p+q+1)}{2}\right] \ln \Lambda$$

  - where n = number of observations
  - p = number of X variables (rank of **X** matrix)
  - q = number of Y variables (rank of **Y** matrix)
  - V is approximately χ^2 distributed with pq degrees of freedom

Is the Relationship between the $\boldsymbol{X}$'s and the $\boldsymbol{Y}$'s Significant?

- Rao's *F*-Test

$$Ra = \left[\frac{1-\Lambda^{1/s}}{\Lambda^{1/s}}\right] \times \left[\frac{\left(1+ts-\frac{pq}{2}\right)}{pq}\right]$$

  - where $t = (n-1) - \frac{(p+q+1)}{2}$
  - $s = 1$ if $p^2 + q^2 \leq 5$; otherwise $s = \sqrt{\frac{(p^2q^2-4)}{(p^2+q^2-5)}}$
  - n = number of observations
  - p = number of X variables (rank of **X** matrix)
  - q = number of Y variables (rank of **Y** matrix)
  - Ra has an approximate F-distribution with pq degrees of freedom in the numerator and $1 + ts - \frac{pq}{2}$ degrees of freedom in the denominator. If p= 1 or 2 or if q= 1 or 2, then Ra has an exact F-distribution

How Many Pairs of Canonical Variates Are Significant?

- How do we decide how many pairs to keep?
  - Sequential application of Bartlett's chi-square test

TABLE 9.8 Sequential testing of pairs of canonical variates

Test of H_0: The canonical correlations in the current row and all that follow are zero.

	Wilks's Λ	Approx χ^2	*df*	$Pr > \chi^2$
1	0.4126	287.3	25	0.0001
2	0.7026	114.7	16	0.0001
3	0.9168	28.4	9	0.0009
4	0.9860	4.6	4	NS
5	0.9990	0.3	1	NS

  - Drawback: Because the chi-square is quite sensitive to sample size, the null hypothesis of no relationship is relatively easily rejected.

How Many Pairs of Canonical Variates Are Significant?

- How do we decide how many pairs to keep?
  - Scree-type plot of Canonical R^2

FIGURE 9.6
Scree-type plot of canonical R^2s from Fader and Lodish data