SAS Education

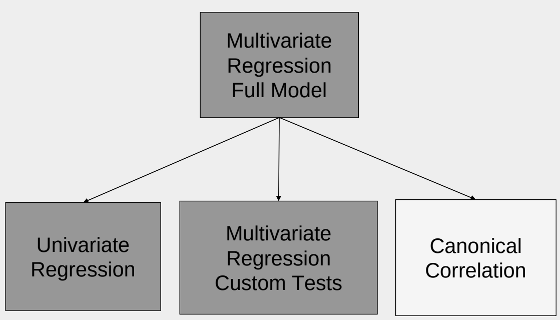
Section 3.3

Canonical Correlation

Objectives

- Define canonical correlation.
- Identify the relationship between canonical correlation analysis and multivariate regression.
- Relate eigenvalues to canonical correlation.
- Perform canonical variate analysis using PROC REG and PROC CANCORR.
- Interpret canonical correlation output, test for significance of canonical variates, and save out canonical scores.

Interpreting Multivariate Regression



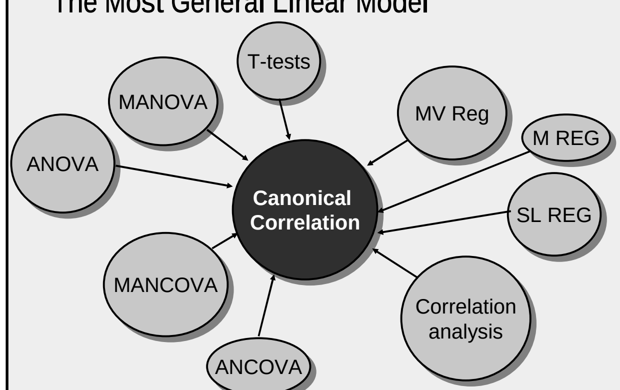
The Connection: R^2 and Eigenvalues

- Just as $\mathbf{HE}^{-1}$ is related to the *F*-ratio for a univariate model, $\mathbf{HT}^{-1}$ is related to R^2.
- Recall that R^2 is the square of the multiple correlation between the set of Xs and Y in a linear model. It is computed as SS_B/SS_T.
- In a multivariate linear model, the eigenvalues of $\mathbf{HT}^{-1}$ are a multivariate generalization of the R^2 in univariate models.
- The eigenvalues of $\mathbf{HT}^{-1}$ are the square of the canonical correlation coefficients.

More on the Eigenvalues of HT^{-1}

- There are *min(p,k)* nonzero eigenvalues in multivariate regression.
- The eigenvalues of $\mathbf{HT}^{-1}$ are equal to the square of the canonical correlations.
- Therefore, the square root of each eigenvalue is a canonical correlation.
- There are *min(p,k)* canonical correlations that can be computed for a data set.
- In the univariate case, *min(p,k)* = 1, therefore, there is always one value for R^2.

The Most General Linear Model



Canonical Correlation in PROC REG

You can use canonical correlation to understand multivariate multiple regression output.

- In PROC REG, you can use the CANPRINT option with the MTEST statement.
- You will see canonical correlations, likelihood ratio significance tests, and information about **HT**$^{-1}$ and **HE**$^{-1}$.

7

Canonical Correlation Using PROC REG

ch3s3d1.sas

This demonstration illustrates how to obtain information about canonical correlation in PROC REG

8

What Have You Learned?

- Of the two possible dimensions of association between the predictors and the responses, only one of them is statistically significant.
- One canonical correlation is significantly different from zero, the other is not.

9

More Detailed Analysis

- The REG procedure provides canonical correlation information from a multivariate multiple regression.
- You may want to investigate the canonical correlations in greater detail.
- The CANCORR procedure provides a great deal of useful information and statistics for canonical correlation analysis.
- Canonical correlation may provide more useful and interesting information than multivariate regression analysis.

10

What Is Really Going On?

From variable sets **X** and **Y**, find the linear combination of each set of variables that is most highly correlated with a linear combination of the other set. These combinations are the first canonical variates:

$$w_1 = a_1 x_1 + a_2 x_2 + ... + a_k x_k$$
$$v_1 = b_1 y_1 + b_2 y_2 + ... + b_p y_p$$
$$r_{w_1 v_1} = \text{ first canonical correlation}$$

Repeat this for a second pair of variates, uncorrelated with the first set.

11

For Example…

Look at the X and Y matrices below:

$$\begin{array}{cc} Y_1\ Y_2\ Y_3\ Y_4 & X_1\ X_2\ X_3 \\ \begin{bmatrix} 2 & 3 & 5 & 8 \\ 3 & 4 & 7 & 7 \\ 2 & 4 & 5 & 7 \\ 1 & 3 & 4 & 6 \\ 2 & 4 & 5 & 9 \\ 4 & 3 & 6 & 6 \\ 4 & 6 & 4 & 7 \\ 2 & 5 & 6 & 7 \end{bmatrix} & \begin{bmatrix} 5 & 2 & 3 \\ 4 & 1 & 4 \\ 6 & 4 & 4 \\ 7 & 6 & 5 \\ 5 & 4 & 3 \\ 5 & 2 & 5 \\ 5 & 5 & 5 \\ 8 & 5 & 4 \end{bmatrix} \end{array}$$

12

For Example…

The best set of coefficients, from canonical correlation analysis:

$$w_1 = .34x_1 - .49x_2 + 1.22x_3$$

$$v_1 = .33y_1 - .11y_2 + .17y_3 - .83y_4$$

$$r_{v_1 w_1} = .99$$

- v_1 and w_1 make up the first pair of canonical variates.
- min(p,k), = 3 pairs of variates can be computed for this data set.

13

Canonical Correlation Using PROC CANCORR

In addition to the canonical correlations, you see the following:

- raw and standardized canonical coefficients for each canonical variate
- simple correlations for all variable and variate combinations
- step-down likelihood-ratio tests for canonical correlations
- multivariate tests for H_0:all canonical correlations = 0
- adjusted canonical correlations and standard errors of the canonical correlations.

14

Likelihood-ratio Tests of Canonical Variates

The likelihood-ratio test of $\mathbf{HE}^{-1}$ is equal to Wilks' Lambda and can be converted to an approximate F test.

- This is a step-down method.
- The null hypothesis is that the canonical correlation $\mathbf{R}_i$ and all those smaller than $\mathbf{R}_i$ = 0.
- This test is performed sequentially for all canonical variate pairs.

15

Interpreting the Canonical Correlation

Canonical coefficients are just like regression coefficients:

- They depend on the other variables in the model.
- They depend on the scale of measurement.
- Standardized coefficients address the scaling issue, but they do not address the problem of dependencies among variables.
- Coefficients are useful for prediction but not for interpretation.

16

Interpreting the Canonical Correlation

A more useful way to interpret the canonical correlation in terms of the input variables is to look at the simple correlation statistics. For each pair of variates, look at

- the correlation between each variable and its canonical variate
- the correlation between each variable and the canonical variate for the other set of variables.

17

Assumptions of Canonical Correlation

Goal: Description

- Minimal measurement error
- Unrestricted variances among variables
- Similar distributional shapes among variables

Goal: Inference

- Assumptions for Description, and at least one set of variables follows a multivariate normal distribution

18

The CANCORR Procedure

General form of the CANCORR procedure:

```
PROC CANCORR <options>;
      <VAR v-variables;>
      WITH w-variables;
RUN;
```

19

Pathological Gambling Example

Two questionnaires:

12 Items:
DSM-IV Criteria

↔

20 Items:
Gamblers Anonymous

Research questions:

- Are they measuring the same dimensions?
- How many dimensions?

20

Canonical Correlation Analysis with PROC CANCORR

ch3s3d2.sas

This demonstration illustrates the CANCORR procedure for canonical correlation analysis

21

What Have You Learned?

- Of the twelve canonical correlations, four are significantly greater than zero.
- Each pair of canonical variates has meaningful interpretation.
- The variables that are most correlated with their own variates are also most correlated with the corresponding variate from the other set of variables.

22

Exercises

This exercise reinforces the concepts discussed previously.

23