Canonical Correlation: Equations

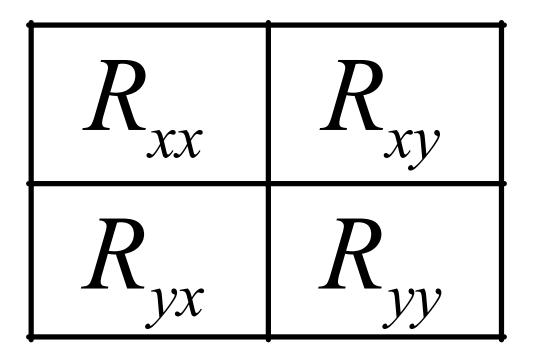
Psy 524 Andrew Ainsworth

### **Data for Canonical Correlations**

- CanCorr actually takes raw data and computes a correlation matrix and uses this as input data.
- You can actually put in the correlation matrix as data (e.g. to check someone else's results)

#### Data

#### • The input correlation set up is:



#### To find the canonical correlations:

 First create a canonical input matrix. To get this the following equation is applied:

# $R = R_{yy}^{-1} R_{yx} R_{xx}^{-1} R_{xy}$

 To get the canonical correlations, you calculate the eigenvalues of R and take the square root

$$r_{ci} = \sqrt{\lambda_i}$$

 In this context the eigenvalues represent percent of overlapping variance accounted for in all of the variables by the two canonical variates

o i.e. it is the squared correlation

#### Testing Canonical Correlations

 Since there will be as many CanCorrs as there are variables in the smaller set, not all will be meaningful (or useful).

 Wilk's Chi Square test – tests whether a CanCorr is significantly different than zero.

$$\chi^2 = -\left[N - 1 - \left(\frac{k_x + k_y + 1}{2}\right)\right] \ln \Lambda_m$$

Where N is number of cases,  $k_{\rm x}\,$  is number of x variables and  $k_{\rm y}\,$  is number of y variables

$$\Lambda_m = \prod_{i=1}^m (1 - \lambda_i)$$

Lamda,  $\Lambda$ , is the product of difference between eigenvalues and 1, generated across m canonical correlations.

 From the text example - For the first canonical correlation:

 $\Lambda_{2} = (1 - .84)(1 - .58) = .07$   $\chi^{2} = -\left[8 - 1 - \left(\frac{2 + 2 + 1}{2}\right)\right] \ln .07$   $\chi^{2} = -(4.5)(-2.7) = 12.15$  $df = (k_{x})(k_{y}) = (2)(2) = 4$ 

 The second CanCorr is tested as  $\Lambda_1 = (1 - .58) = .42$  $\chi^2 = -\left| 8 - 1 - \left( \frac{2 + 2 + 1}{2} \right) \right| \ln .42$  $\chi^2 = -(4.5)(-.87) = 3.92$  $df = (k_x - 1)(k_y - 1) = (2 - 1)(2 - 1) = 1$ 

#### Canonical Coefficients

- Two sets of Canonical Coefficients are required
  - One set to combine the Xs
  - One to combine the Ys
  - Similar to regression coefficients

 $B_{y} = (R_{yy}^{-1/2})'\hat{B}_{y}$ Where  $(R_{yy}^{-1/2})'$  is the transpose of the inverse of the "special" matrix form of square root that keeps all of the eigenvalues positive and  $\hat{B}_{y}$ is a normalized matrix of eigen vectors for yy

 $\mathbf{B}_{\mathbf{x}} = R_{\mathbf{x}\mathbf{x}}^{-1}R_{\mathbf{x}\mathbf{y}}B_{\mathbf{y}}^{*}$ 

Where  $B_y^*$  is  $B_y$  from above dividing each entry by their corresponding canonical correlation.

#### Canonical Variate Scores

- Like factor scores (we'll get there later)
- What a subject would score if you could measure them directly on the canonical variate

$$X = Z_x B_x$$
$$Y = Z_y B_y$$

 Matrices of Correlations between variables and canonical variates; also called loadings or loading matrices

$$A_{x} = R_{xx}B_{x}$$
$$A_{y} = R_{yy}B_{y}$$

		Canonical Variate Pairs	
		First	Second
First Set	TS	74	.68
	TC	.79	.62
Second Set	BS	44	.90
	BC	.88	.48

- Percent of variance in a single variable accounted for by it's own canonical variate
  - This is simply the squared loading for any variable
  - e.g. The percent of variance in Top Shimmies explained by the first canonical variate is -.74<sup>2</sup> ≈ 55%

#### Redundancy

• Within – <u>Average</u> percent of variance in a set of variables explained by their own canonical variate

$$pv_{xc} = \sum_{i=1}^{k_x} \frac{a_{ixc}^2}{k_x}$$

$$pv_{yc} = \sum_{i=1}^{k_y} \frac{a_{iyc}^2}{k_y}$$

$$pv_{xc_1} = \frac{(-.74)^2 + (.79)^2}{2} = .58$$

#### Redundancy

 Across – <u>average</u> percent of variance in the set of Xs explained by the Y canonical variate and vice versa

$$rd = (pv)(r_c^2)$$
$$rd_{x_1 \to y} = \left[\frac{(-.74)^2 + .79^2}{2}\right](.84) = .48$$