# Statistics 135

# Chapter 10

# Canonical Correlations

Chris Drake
*Department of Statistics*
*University of California, Davis*

# Purpose of Canonical Correlations

Principal components (PCA) is used to investigate one set of variables. The original variables are replaced by a set of variables called principal components which are created to account for maximal variation among the original variables. Canonical correlations is a generalization of PCA with the goal to identify and quantify the associations between two sets of variables.

1. Two vectors $\mathbf{X}^{(1)}_{p\times 1}$ and $\mathbf{X}^{(2)}_{q\times 1}$ with $p \leq q$.
2. Focus is on the correlation between a linear combination of one set of variables $\mathbf{X}^{(1)}$ and the linear combination of another set of variables $\mathbf{X}^{(2)}$.
3. Determine the pair of linear combinations with the largest correlation
4. Next, determine a pair of linear combinations having the largest correlation among pairs uncorrelated with the initial pair and so on.

We assume the following model:

$$E\big(\mathbf{X}^{(1)}\big) = \mu^{(1)} \qquad\qquad Cov\big(\mathbf{X}^{(1)}\big) = \mathbf{\Sigma}^{11}$$
$$E\big(\mathbf{X}^{(2)}\big) = \mu^{(2)} \qquad\qquad Cov\big(\mathbf{X}^{(2)}\big) = \mathbf{\Sigma}^{22}$$

$$Cov\big(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}\big) = \mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{21}'$$

Let $U, V$ be linear combinations of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ where

$$U = \mathbf{a}'\mathbf{X}^{(1)} \qquad and \qquad V = \mathbf{b}'\mathbf{X}^{(2)}$$

then

$$Var(U) = \mathbf{a}'\mathbf{\Sigma}_{11}\mathbf{a} \qquad and \qquad Var(V) = \mathbf{b}'\mathbf{\Sigma}_{22}\mathbf{b}$$
$$Cov(U, V) = \mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b}$$

We are interested in finding $\mathbf{a}$ and $\mathbf{b}$ which maximize

$$Corr(U, V) = \frac{\mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{\Sigma}_{11}\mathbf{a}}\sqrt{\mathbf{b}'\mathbf{\Sigma}_{22}\mathbf{b}}} \tag{1}$$

1 The first pair of canonical variables $U_1, V_1$ maximizes (1) subject to $Var(U_1) = Var(V_1) = 1$

2 The second pair of canonical variables $U_2, V_2$ maximizes (1) subject to

$$Var(U_2) = Var(V_2) = 1$$

$$Cov(U_1, U_2) = Cov(V_1, V_2) = Cov(U_1, V_2) = Cov(U_2, V_1) = 0$$

3 The k$^{th}$ canonical pair is given by

$$U_k = \mathbf{e}_k' \mathbf{\Sigma}_{11}^{-1/2} \mathbf{X}^{(1)} \qquad and \qquad V_k = \mathbf{f}_k' \mathbf{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}$$

$$Corr(U_k, V_k) = \rho^*$$

4 $\rho_1^* > \rho_2^* > ... > \rho_p^*$ are the eigenvalues of $\mathbf{\Sigma}_{11}^{-1/2} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1/2}$ and $\mathbf{e}_1, \mathbf{e}_2, .., \mathbf{e}_p$ are the associated eigenvectors;

5 $\rho_1^* > \rho_2^* > ... > \rho_p^*$ are also the p largest eigenvalues (out of q>p) of $\mathbf{\Sigma}_{22}^{-1/2} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1/2}$ with eigenvectors $\mathbf{f}_1, \mathbf{f}_2, .., \mathbf{f}_q$

# Calculation and Interpretation

1 Split the $(p+q) \times (p+q)$ covariance matrix of $(\mathbf{X}_1', \mathbf{X}_2')'$ into $\mathbf{\Sigma}_{11}$, $\mathbf{\Sigma}_{12}$ and $\mathbf{\Sigma}_{22}$

2 Calculate $\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1/2}$

3 Obtain the eigenvalues $\rho_j^*$ for $j = 1, ..., p$ and the associated eigenvectors $\mathbf{e}_j$.

4 For the eigenvalues in (3) obtain the eigenvectors $\mathbf{f}_1, .., \mathbf{f}_q$ (only the first p are needed) of $\mathbf{\Sigma}_{22}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1/2}$, this matrix as its p largest has eigenvalues $\rho_1^*, ..., \rho_p^*$, the other eigenvalues are $\rho_{p+1}^*, ..., \rho_q^*$.

5 Form the linear combinations $U_k = \mathbf{e}_k\mathbf{\Sigma}_{11}^{-1/2}\mathbf{X}^{(1)}$ and $V_k = \mathbf{f}_k\mathbf{\Sigma}_{22}^{-1/2}\mathbf{X}^{(2)}$ for $k = 1, .., p$.

6 Continue until the canonical correlation $\rho_k^*$ is very small.

7 Let $\mathbf{U} = \mathbf{A}\mathbf{X}^{(1)}$ and $\mathbf{V} = \mathbf{B}\mathbf{X}^{(2)}$ where $\mathbf{A} = \mathbf{E}'\mathbf{\Sigma}_{11}^{-1/2}$ and $\mathbf{B} = \mathbf{F}'\mathbf{\Sigma}_{22}^{-1/2}$. Then

$$Cov(\mathbf{U}) = \mathbf{A}\mathbf{\Sigma}_{11}\mathbf{A}' = \mathbf{I} \qquad and \qquad Cov(\mathbf{V}) = \mathbf{B}\mathbf{\Sigma}_{11}\mathbf{B}' = \mathbf{I}$$

8 Since $\mathbf{\Sigma}_{11} = \mathbf{P}_1\mathbf{\Lambda}\mathbf{P}_1'$ we can write

$$\mathbf{U} = \mathbf{A}\mathbf{X}^{(1)} = \mathbf{E}'\mathbf{P}_1\mathbf{\Lambda}^{-1/2}\mathbf{P}_1'\mathbf{X}^{(1)}$$

and we have a similar expression for V. Therefore, $\mathbf{U}$ can be interpreted as

a a transformation of $\mathbf{X}^{(1)}$ to uncorrelated principal components followed by

b a rigid rotation $\mathbf{P}_1$ determined by $\mathbf{\Sigma}_{11}$ followed by

c another rotation $\mathbf{E}'$ determined from the full covariance matrix $\mathbf{\Sigma}$ of $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$.

# Estimation

When we have a sample $(\mathbf{x}_1^{(1)}, \mathbf{x}_1^{(2)}), ..., (\mathbf{x}_1^{(1)}, \mathbf{x}_1^{(2)})$, then we calculate the sample mean vector $\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}$ and the sample covariance matrix

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$$

The calculations outlined in items (1) through (6) in the calculation section are then carried out for the sample covariance or correlation matrix.

A test for a diagonal covariance matrix $H_0: \ \mathbf{\Sigma}_{12} = \mathbf{0}$ is given by

$$-2ln\Lambda = n \ \ ln\Big(\frac{\mid \mathbf{S}_{11} \mid\mid \mathbf{S}_{22} \mid}{\mid \mathbf{S} \mid}\Big) = -n \ \ ln \prod_{i=1}^{p} (1 - \hat{\rho}_i^{*2})$$

which has a $\chi^2_{pq}$ distribution. An improved version is

$$-\Big(n - 1 - \frac{1}{2}(p + q + 1)\Big) \ \ ln \prod_{i=1}^{p} (1 - \hat{\rho}_i^{*2})$$