Canonical Correlations

- Like Principal Components Analysis, Canonical Correlation Analysis looks for interesting linear combinations of multivariate observations.

- In Canonical Correlation Analysis, a multivariate response $X$ is partitioned into two groups:

$$X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix}$$

- We look for linear combinations $U = a'X^{(1)}$ and $V = b'X^{(2)}$ that are highly correlated.
• The covariance matrix of $X$ is correspondingly partitioned:

$$\text{Cov}(X) = \Sigma = \begin{bmatrix}
\Sigma_{1,1} & \Sigma_{1,2} \\
\Sigma_{2,1} & \Sigma_{2,2}
\end{bmatrix}$$

• Then

$$\text{Corr}(U, V) = \frac{a'\Sigma_{1,2} b}{\sqrt{a'\Sigma_{1,1} a} \sqrt{b'\Sigma_{2,2} b}}$$

• Without loss of generality, $\text{Var}(U) = a'\Sigma_{1,1} a$ and $\text{Var}(V) = b'\Sigma_{2,2} b$ can both be constrained to equal 1.

• The first pair of canonical variables are the $U_1$ and $V_1$ that maximize the correlation, subject to this constraint.
Differentiation shows that the maximizing $a$ and $b$ satisfy

$$\Sigma_{1,2}b - \alpha \Sigma_{1,1}a = 0$$
$$\Sigma_{2,1}a - \beta \Sigma_{2,2}b = 0$$

for some $\alpha$ and $\beta$.

Then

$$\text{Corr}(U, V) = a'\Sigma_{1,2}b = \alpha a'\Sigma_{1,1}a = \alpha$$
$$= \beta b'\Sigma_{2,2}b = \beta$$

so $\alpha = \beta = \sqrt{\lambda}$, say.
• We assume that $\Sigma_{1,1}$ and $\Sigma_{2,2}$ are invertible, and then

$$\Sigma_{1,1}^{-1} \Sigma_{1,2} b = \sqrt{\lambda} a$$
$$\Sigma_{2,2}^{-1} \Sigma_{2,1} a = \sqrt{\lambda} b$$

• Eliminate $b$:

$$\Sigma_{1,1}^{-1} \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1} a = \lambda a.$$

• That is, $a$ is an eigenvector of $\Sigma_{1,1}^{-1} \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1}$ and $b$ is a multiple of $\Sigma_{2,2}^{-1} \Sigma_{2,1} a$.

• Consequently, $b$ is an eigenvector of $\Sigma_{2,2}^{-1} \Sigma_{2,1} \Sigma_{1,1}^{-1} \Sigma_{1,2}$ with the same eigenvalue as $a$. 
• Since $\text{Corr}(U, V) = \sqrt{\lambda}$, the first canonical pair, $U_1$ and $V_1$, is determined by the largest eigenvalue $\lambda_1 = \text{Corr}(U_1, V_1)^2 = \rho_1^*$.  

• We then look for the pair $U_2$ and $V_2$ with the greatest correlation, subject to $\text{Corr}(U_1, U_2) = \text{Corr}(V_1, V_2) = 0$.  

• As a bonus, $\text{Corr}(U_1, V_2) = \text{Corr}(U_2, V_1) = 0$.  

• Not surprisingly, this second canonical pair, $U_2$ and $V_2$, is determined by the next largest eigenvalue $\lambda_2 = \text{Corr}(U_2, V_2)^2 = \rho_2^*$.  

• Continuing the process, we find $r = \min(p, q)$ canonical pairs.
Example: head and leg measurements of chicken bones

- Head group:
  - skull length;
  - skull breadth.

- Leg group:
  - femur length;
  - tibia length.

- SAS proc cancorr program and output.
The number of canonical pairs

- In sample data, all \( r = \min(p, q) \) canonical correlations will be positive, even if the data are drawn from a population with only \( m < r \) positively correlated pairs.

- `proc cancorr` tests the hypothesis

  \[
  H_0: \rho_m^* = \rho_{m+1}^* = \cdots = \rho_r^* = 0
  \]

  for each value of \( m, m = 1, 2, \ldots, r \).

- The first null hypothesis is that the two groups are entirely uncorrelated. If that is rejected, the second is that the correlation between them is carried by a single linear combination of each group, etc.
Standardization

• The coefficient vectors $a_i$ and $b_i$ may be difficult to interpret, especially when the variables are in different physical units or have very different variances.

• Consequently, the data are often standardized, or equivalently, the correlation matrix is used instead of the covariance matrix.

• Unlike in Principal Components Analysis, the canonical variables are unchanged by any such scaling.

• proc cancorr reports both raw and standardized forms.
Correlations

- Sometimes the canonical variables can be interpreted as representing characteristics of the items being measured.

- The interpretation may be made based on the (raw or standardized) coefficients.

- The correlations between the original variables and the canonical variables can also be informative, but “must be interpreted with caution.”

- `proc cancorr` reports the correlations of each original variable with both sets of canonical variables.
Special Cases

\( p = q = 1 \): the “canonical” correlation between single variables \( X_1 \) and \( X_2 \) is just the absolute value of their simple Pearson correlation \( \rho \).

\( p = 1, q > 1 \): the canonical correlation of a single variable \( X_1 \) and a group of variables \( X^{(2)} \) is the multiple correlation \( R \) in the regression of \( X_1 \) on \( X^{(2)} \), so \( \rho_1^2 = R^2 \).

- Also, in general, \( \rho_k^2 \) is the \( R^2 \) in the regression of \( U_k \) on \( X^{(2)} \), and also in the regression of \( V_k \) on \( X^{(1)} \): canonical correlations are closely related to \( R^2 \).