• Recall: Regression analysis --- concerned with the relationship between a single response and a set of predictors

\[
Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \epsilon
\]

connection between regression model and conditional normal distribution

\[
R^2 = \text{cor}(Y, \hat{Y})^2
\]

• Q: What if we have more than one response?

\[
\hat{Y} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \epsilon
\]

Examples
- variables related to arithmetic power & variables related to reading power
- variables related to governmental policy & variables related to economic goal
- variables related to college performance & variables related to precollege achievement

CCA seeks to identify and quantify the linear associations between two sets of variables

• Population Canonical Variables and Canonical Correlations

Data: two groups of variables

The first group, of \( p \) variables, is represented by the \( (p \times 1) \) random vector \( \mathbf{X}^{(1)} \). The second group, of \( q \) variables, is represented by the \( (q \times 1) \) random vector \( \mathbf{X}^{(2)} \). We assume, in the theoretical development, that \( \mathbf{X}^{(1)} \) represents the smaller set, so that \( p \leq q \).

Some notations

\[
\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1^{(1)} \\ \mathbf{X}_2^{(1)} \\ \vdots \\ \mathbf{X}_p^{(1)} \\ \mathbf{X}_1^{(2)} \\ \mathbf{X}_2^{(2)} \\ \vdots \\ \mathbf{X}_q^{(2)} \end{bmatrix}, \quad \mathbf{\mu} = E(\mathbf{X}) = \begin{bmatrix} E(\mathbf{X}^{(1)}) \\ E(\mathbf{X}^{(2)}) \end{bmatrix} = \begin{bmatrix} \mathbf{\mu}^{(1)} \\ \mathbf{\mu}^{(2)} \end{bmatrix}
\]

\[
\Sigma = E(\mathbf{X} - \mathbf{\mu})(\mathbf{X} - \mathbf{\mu})' = \begin{bmatrix} \Sigma_{11}^{(p \times p)} & \Sigma_{12}^{(p \times q)} \\ \Sigma_{21}^{(q \times p)} & \Sigma_{22}^{(q \times q)} \end{bmatrix}
\]

the \( pq \) elements of \( \Sigma_{12} \) measure the association between the two sets.

- \( p \) and \( q \) are relatively large, hopeless to interpret elements of \( \Sigma_{12} \) collectively
- moreover, it’s often linear combinations of variables that are interesting and useful for predictive or comparative purposes
CCA finds linear functions of one set of variables that maximally correlated with linear function of the other set of variables

- Set \( U = a'X^{(1)} \) and \( V = b'X^{(2)} \) for some pair of coefficient vectors \( a \) and \( b \).

\[
\text{Var}(U) = a' \text{Cov}(X^{(1)}) a = a' \Sigma_{11} a \\
\text{Var}(V) = b' \text{Cov}(X^{(2)}) b = b' \Sigma_{22} b \\
\text{Cov}(U, V) = a' \text{Cov}(X^{(1)}, X^{(2)}) b = a' \Sigma_{12} b
\]

- We shall seek coefficient vectors \( a \) and \( b \) such that

\[
\text{Corr}(U, V) = \frac{a' \Sigma_{12} b}{\sqrt{a' \Sigma_{11} a \sqrt{b' \Sigma_{22} b}}}
\]

is as large as possible.

- The first pair of canonical variables, or first canonical variate pair, is the pair of linear combinations \( U_1, V_1 \) having unit variances, which maximize the correlation (\( \varphi \)).

- The second pair of canonical variables, or second canonical variate pair, is the pair of linear combinations \( U_2, V_2 \) having unit variances, which maximize the correlation (\( \varphi \)) among all choices that are uncorrelated with the first pair of canonical variables.

- The \( k \)th pair of canonical variables, or \( k \)th canonical variate pair, is the pair of linear combinations \( U_k, V_k \) having unit variances, which maximize the correlation (\( \varphi \)) among all choices uncorrelated with the previous \( k - 1 \) canonical variable pairs.

\[\text{Result 10.1.} \quad \text{Suppose} \quad p \leq q \quad \text{and let the random vectors} \quad X^{(1)} \quad \text{and} \quad X^{(2)} \quad \text{have} \quad \text{Cov}(X^{(1)}) = \Sigma_{11}, \quad \text{Cov}(X^{(2)}) = \Sigma_{22} \quad \text{and} \quad \text{Cov}(X^{(1)}, X^{(2)}) = \Sigma_{12}, \quad \text{where} \quad \Sigma \quad \text{has full rank.} \quad \text{For coefficient vectors} \quad a \quad \text{and} \quad b, \quad \text{form the linear combinations} \quad U = a'X^{(1)} \quad \text{and} \quad V = b'X^{(2)}. \quad \text{Then} \]

\[
\max_{a,b} \text{Corr}(U, V) = \rho_1^*
\]

attained by the linear combinations (first canonical variate pair)

\[
U_1 = e_1' \Sigma_{11}^{-1/2} X^{(1)} \quad \text{and} \quad V_1 = f_1' \Sigma_{22}^{-1/2} X^{(2)}
\]

The \( k \)th pair of canonical variates, \( k = 2, 3, \ldots, p \),

\[
U_k = e_k' \Sigma_{11}^{-1/2} X^{(1)} \quad V_k = f_k' \Sigma_{22}^{-1/2} X^{(2)}
\]

maximizes

\[
\text{Corr}(U_k, V_k) = \rho_k^*
\]

among those linear combinations uncorrelated with the preceding \( 1, 2, \ldots, k - 1 \) canonical variables.

Here \( \rho_1^2 \geq \rho_2^2 \geq \cdots \geq \rho_p^2 \) are the eigenvalues of \( \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \Sigma_{11}^{-1/2}, \) and \( e_1, e_2, \ldots, e_p \) are the associated \((p \times 1)\) eigenvectors [The quantities \( \rho_1^2, \rho_2^2, \ldots, \rho_p^2 \) are also the \( p \) largest eigenvalues of the matrix \( \Sigma_{22}^{-1/2} \Sigma_{11} \Sigma_{12} \Sigma_{22}^{-1/2} \) with corresponding \((q \times 1)\) eigenvectors \( f_1, f_2, \ldots, f_p \). Each \( f_i \) is proportional to \( \Sigma_{22}^{-1/2} \Sigma_{11}^{-1/2} e_i \).]
Q: why $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{21}\Sigma_{22}^{-1/2}$ (or $\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}$)?

- geometrical interpretation

Let $A_{(p \times p)} = [a_1, a_2, \ldots, a_p]'$ and $B_{(q \times q)} = [b_1, b_2, \ldots, b_q]'$, so that the vectors of canonical variables are

$$U_{(p \times 1)} = AX_{(1)}^{(1)} \quad V_{(q \times 1)} = BX_{(2)}^{(2)}$$

$$A = E'\Sigma_{11}^{-1/2} = EP_1\Lambda_1^{-1/2}P_1'$$

where $E'$ is an orthogonal matrix with row $e_i'$.

The canonical variates have the properties

$$\text{Var}(U_k) = \text{Var}(V_k) = 1$$

$$\text{Cov}(U_k, U_\ell) = \text{Corr}(U_k, U_\ell) = 0 \quad k \neq \ell$$

$$\text{Cov}(V_k, V_\ell) = \text{Corr}(V_k, V_\ell) = 0 \quad k \neq \ell$$

$$\text{Cov}(U_k, V_\ell) = \text{Corr}(U_k, V_\ell) = 0 \quad k \neq \ell$$

for $k, \ell = 1, 2, \ldots, p$.

$$\rho_{U, X_{(1)}} = \text{Cov}(U, X_{(1)}) = \text{Cov}(U, V_{11}^{-1/2}X_{(1)}) = \text{Cov}(AX_{(1)}, V_{11}^{-1/2}X_{(1)}) = A\Sigma_{11}V_{11}^{-1/2}$$

Similar calculations for the pairs $(U, X_{(2)})$, $(V, X_{(2)})$ and $(V, X_{(1)})$ yield

$$\rho_{U, X_{(1)}} = A\Sigma_{11}V_{11}^{-1/2} \quad \rho_{V, X_{(2)}} = B\Sigma_{22}V_{22}^{-1/2}$$

$$\rho_{U, X_{(2)}} = A\Sigma_{12}V_{22}^{-1/2} \quad \rho_{V, X_{(1)}} = B\Sigma_{21}V_{11}^{-1/2}$$
To ease the computation burden, many people prefer to get the canonical correlations from

\[ |\Sigma_{12}^{-1}\Sigma_{21}^{-1}\Sigma_{21} - \rho^{*2}I| = 0 \]

The coefficient vectors \(a\) and \(b\) follow directly from the eigenvector equations

\[ \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}a = \rho^{*2}a \]
\[ \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{12}^{-1}\Sigma_{12}b = \rho^{*2}b \]

- Interpreting the population canonical variables
- Identify the canonical variables
  - The linear combinations can be interpreted much as in principal components
  - It is helpful to compute the correlation between the canonical variables and the original variables as a way to determine the relative importance of original variables to a particular canonical variable
- Canonical correlation
  - Canonical correlation generalizes the correlation between 2 variables
  - \( |\text{Corr}(X_i^{(1)}, X_k^{(2)})| = |\text{Corr}(a^tX^{(1)}, b^tX^{(2)})| \leq \max_{a,b} \text{Corr}(a^tX^{(1)}, b^tX^{(2)}) = \rho_k^* \)
  - An \( R^2 \) type statistic can be obtained to explain the total variance explained by a given set of canonical variables:

Because of its multiple correlation coefficient interpretation, the \( k \)th squared canonical correlation \( \rho_k^{*2} \) is the proportion of the variance of canonical variate \( U_k \) “explained” by the set \( X^{(2)} \). It is also the proportion of the variance of canonical variate \( V_k \) “explained” by the set \( X^{(1)} \).

First \( r \) canonical variables as a summary of variability
- Remember that the linear combinations are chosen to maximize correlation between the canonical variables
- These linear combinations may be quite different from those obtained from principal component within a particular set of variables
- The canonical may not necessarily explain the total variation within a set of original variables

CCA for standardized data
- If the original variables are standardized with \( Z^{(1)} = [Z_1^{(1)}, Z_2^{(1)}, \ldots, Z_p^{(1)}] \) and \( Z^{(2)} = [Z_1^{(2)}, Z_2^{(2)}, \ldots, Z_q^{(2)}] \), from first principles, the canonical variates are of the form

\[ U_k = a_k^tZ^{(1)} = e_k\rho_k^{-1/2}Z^{(1)} \]
\[ V_k = b_k^tZ^{(2)} = f_k\rho_k^{-1/2}Z^{(2)} \]

Here, \( \text{Cov}(Z^{(1)}) = \rho_{11} \), \( \text{Cov}(Z^{(2)}) = \rho_{22} \), \( \text{Cov}(Z^{(1)}, Z^{(2)}) = \rho_{12} = \rho_{21}^* \), and \( e_k \) and \( f_k \) are the eigenvectors of \( \rho_{11}^{-1/2}\rho_{21}\rho_{21}^{-1/2} \) and \( \rho_{22}^{-1/2}\rho_{21}\rho_{21}^{-1/2} \), respectively. The canonical correlations, \( \rho_k \), satisfy

\[ \text{Corr}(U_k, V_k) = \rho_k^* \quad k = 1, 2, \ldots, p \]

Where \( \rho_1^{*2} \geq \rho_2^{*2} \geq \cdots \geq \rho_p^{*2} \) are the nonzero eigenvalues of the matrix

\[ \rho_1^{1/2}\rho_{21}\rho_{21}^{1/2} \rho_2^{1/2}\rho_{21}\rho_{21}^{1/2} \] (or, equivalently, the largest eigenvalues of \( \rho_{22}^{-1/2}\rho_{21}\rho_{21}^{-1/2} \rho_{21}\rho_{21}^{-1/2} \)).

The canonical correlations are unchanged by the standardization (cf. PCA)

Reading: Reference, 10.1, 10.2, 10.3