To ease the computation burden, many people prefer to get the canonical correlations from

\[ \text{Var}(U_k) = \text{Var}(V_k) = 1 \]

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

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

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

for \( k, \ell = 1, 2, \ldots, p. \)

\[ \text{corr}((U_k, V_\ell)) = 0^* \]

\[ \text{Cov}(U, X^{(1)}) = \text{Cov}(AX^{(1)}, X^{(1)}) = \mathbf{A}\Sigma_{11} \]

\[ \mathbf{A} \leftarrow \mathbf{\rho}_{U, X^{(1)}} = \text{Cov}(U, X^{(1)}) = \text{Cov}(U, \sqrt{11}X^{(1)}) = \text{Cov}(AX^{(1)}, V_{11}^{1/2}X^{(1)}) \]

\[ \text{cov}(U) = \mathbf{I} \quad \text{diag}(\Sigma_{11}) = \mathbf{A}\Sigma_{11} V_{11}^{1/2} \]

Similar calculations for the pairs \((U, X^{(2)}), (V, X^{(2)})\) and \((V, X^{(1)})\) yield

\[ \mathbf{\rho}_{U, X^{(2)}} = \mathbf{A}\Sigma_{12} V_{12}^{1/2} \]

\[ \mathbf{\rho}_{V, X^{(1)}} = \mathbf{B}\Sigma_{22} V_{22}^{1/2} \]

Their components may give the canonical variates an interpretation.

To ease the computation burden, many people prefer to get the canonical correlations from

\[ |\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \rho^2^*| = 0 \]

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

\[ e_k \times \mathbf{A} \xi_k \]

\[ \mathbf{a} = \rho^2_a \xi_k \]

\[ \mathbf{b} = \rho^2_b \xi_k \]

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 way to determine the relative important of original variables to a particular canonical variable

- canonical correlation

- canonical correlation generalizes the correlation between 2 variables

\[ |\text{Corr}(X^{(1)}_i, X^{(2)}_k)| = |\text{Corr}(\mathbf{a}X^{(1)}, \mathbf{b}X^{(2)})| \leq \max_a \text{Corr}(\mathbf{a}X^{(1)}, \mathbf{b}X^{(2)}) = \rho^*_k \]

- 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 \(\rho^2_k\) 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 \(X^{(1)}\).
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

\[
\Sigma_{11} \rightarrow E_{11}, \quad \Sigma_{12} \rightarrow E_{12}
\]

\[
\Sigma_{22} \rightarrow E_{22}, \quad \Sigma_{21} \rightarrow E_{21}
\]

\[
U_k = a_k^* Z^{(1)} = e_k \rho_1^{1/2} Z^{(1)}
\]

\[
V_k = b_k^* Z^{(2)} = f_k \rho_2^{1/2} Z^{(2)}
\]

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 \( \rho_{11}^{1/2} \rho_{12} \rho_{21} \rho_{11}^{1/2} \) and \( \rho_{22}^{1/2} \rho_{11} \rho_{12} \rho_{22}^{1/2} \), respectively. The canonical correlations, \( \rho_k \), satisfy

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

where \( \rho_1^2 \geq \rho_2^2 \geq \ldots \geq \rho_p^2 \) are the nonzero eigenvalues of the matrix \( \rho_{11}^{1/2} \rho_{12} \rho_{21} \rho_{11}^{1/2} \) (or, equivalently, the largest eigenvalues of \( \rho_{22}^{1/2} \rho_{11} \rho_{12} \rho_{22}^{1/2} \)).

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

**Reading**: Reference, 10.1, 10.2, 10.3

---

**Sample Canonical Variables and Canonical Correlations**

**Data:**

\[
X = \begin{bmatrix}
X^{(1)} & X^{(2)}
\end{bmatrix}
\]

\[
\begin{bmatrix}
x_1^{(1)} & x_2^{(1)} & \cdots & x_p^{(1)}
x_1^{(2)} & x_2^{(2)} & \cdots & x_p^{(2)}
\vdots & \vdots & \ddots & \vdots
x_1^{(n)} & x_2^{(n)} & \cdots & x_p^{(n)}
\end{bmatrix}
\]

Finding canonical variables and canonical correlations

- replace population distribution by empirical distribution
- replace \( \Sigma \) by \( S \) (or \( S_n \))
- replace \( \rho \) by \( R \)
- then, the rests the same as above

Scatter plots of the first pair may reveal atypical observations \( x_i \), requiring further study. If the canonical correlations \( \rho_2^*, \rho_3^*, \ldots \) are also moderately large, scatter plots of the pairs \( (\hat{U}_2, \hat{V}_2), (\hat{U}_3, \hat{V}_3), \ldots \) may also be helpful in this respect.