Interpretación de las variables canónicas poblacionales

- A pesar de que las variables canónicas pueden resultar artificiales (i.e, no tienen un significado físico), frecuentemente pueden ser *interpretadas o identificadas* en términos del significado de las variables originales.
- Una herramienta para facilitar su interpretación, es calcular las correlaciones entre las variables canónicas y las variables originales.
- Sin embargo, estas correlaciones deben ser intepretadas con cuidado, debido a que ellas proporcionan únicamente información univariada, en el sentido de que no indican como contribuyen *conjuntamente* las variables originales al análisis canónico.
- Por esta razón, se prefiere evaluar las contribuciones de las variables originales directamente de los coeficientes estandarizados.

Sean $\underset{(pxp)}{\mathbf{A}} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p]^{'}$ y $\underset{(qxq)}{\mathbf{B}} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_q]^{'}$, entonces los vectores de variables canónicas son

$$\underset{(px1)}{\mathbf{U}} = \mathbf{A}\mathbf{x}^{(1)} \qquad \underset{(qx1)}{\mathbf{V}} = \mathbf{B}\mathbf{x}^{(2)},$$

donde estamos interesados en las primeras p variables canónicas en $\mathbf{V}$. Por tanto

$$Cov(\boldsymbol{U}, \boldsymbol{x}^{(1)}) = Cov(\mathbf{A}\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) = \mathbf{A}\boldsymbol{\Sigma}_{11}$$

Debido a que $Var(U_i) = 1$, $Corr(U_i, x_k^{(1)})$ se obtiene dividiendo $Cov(U_i, x_k^{(1)})$ entre $\sqrt{Var(x_k^{(1)})} = \sigma_{kk}^{1/2}$. Equivalentemente $Corr(U_i, x_k^{(1)}) = Cov(U_i, \sigma_{kk}^{-1/2} x_k^{(1)})$

Interpretación de las variables canónicas poblacionales

Introduciendo la matriz diagonal (pxp) $\mathbf{V}_{11}^{-1/2}$ con el k-ésimo elemento diagonal $\sigma_{kk}^{-1/2}$, tenemos en términos matriciales

$$\underset{(pxp)}{\rho_{(\mathbf{U},\mathbf{x}^{(1)})}} = Corr(\mathbf{U},\mathbf{x}^{(1)}) = Cov(\mathbf{U},\mathbf{V}_{11}^{-1/2}\mathbf{x}^{(1)}) = Cov(\mathbf{A}\mathbf{x}^{(1)},\mathbf{V}_{11}^{-1/2}\mathbf{x}^{(1)}) = \mathbf{A}\mathbf{\Sigma}_{11}\mathbf{V}_{11}^{-1/2}$$

Haciendo cálculos similares obtenemos las correlaciones entre los pares $(\mathbf{U},\mathbf{x}^{(2)})$,$(\mathbf{V},\mathbf{x}^{(2)})$ y $(\mathbf{V},\mathbf{x}^{(1)})$, obteniendo

$$\underset{(pxp)}{\rho_{(\mathbf{U},\mathbf{x}^{(1)})}} = \mathbf{A}\mathbf{\Sigma}_{11}\mathbf{V}_{11}^{-1/2} \qquad \underset{(qxq)}{\rho_{(\mathbf{V},\mathbf{x}^{(2)})}} = \mathbf{B}\mathbf{\Sigma}_{22}\mathbf{V}_{22}^{-1/2}$$

$$\underset{(pxq)}{\rho_{(\mathbf{U},\mathbf{x}^{(2)})}} = \mathbf{A}\mathbf{\Sigma}_{12}\mathbf{V}_{22}^{-1/2} \qquad \underset{(qxp)}{\rho_{(\mathbf{V},\mathbf{x}^{(1)})}} = \mathbf{B}\mathbf{\Sigma}_{21}\mathbf{V}_{11}^{-1/2}$$

donde $\mathbf{V}_{22}^{-1/2}$ es la matrix diagonal (qxq) con i-ésimo elemento diagonal $Var(x_i^{(2)})$

Las variables canónicas derivadas de variables estandarizadas se pueden interpretar también mediante el cálculo de las correlaciones. En este caso,

$$\rho_{\mathbf{U},\mathbf{Z}^{(1)}} = \mathbf{A}_z\rho_{11} \qquad \rho_{\mathbf{V},\mathbf{Z}^{(2)}} = \mathbf{B}_z\rho_{22}$$
$$\rho_{\mathbf{U},\mathbf{Z}^{(2)}} = \mathbf{A}_z\rho_{12} \qquad \rho_{\mathbf{V},\mathbf{Z}^{(1)}} = \mathbf{B}_z\rho_{21}$$

donde $\underset{(pxp)}{\mathbf{A}_z}$ y $\underset{(qxq)}{\mathbf{B}_z}$ son las matrices cuyos renglones contienen los coeficientes canónicos para los conjuntos $\mathbf{Z}^{(1)}$ y $\mathbf{Z}^{(2)}$ respectivamente. Las correlaciones en las matrices mostradas en la ecuación anterior tienen los mismos valores númericos que los de la ecuación de la lámina anterior. Esto es, $\rho_{(\mathbf{U},\mathbf{X}^{(1)})} = \rho_{(\mathbf{U},\mathbf{Z}^{(1)})}$ y así sucesivamente. Esto se debe a que por ejemplo

$$\rho_{(\mathbf{U},\mathbf{X}^{(1)})} = \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2} = \mathbf{A}\mathbf{V}_{11}^{1/2}\mathbf{V}_{11}^{-1/2}\boldsymbol{\Sigma}_{11}\mathbf{V}_{11}^{-1/2} = \mathbf{A}_z\rho_{11} = \rho_{(\mathbf{U},\mathbf{Z}^{(1)})}$$

Es decir las correlaciones no se ven afectadas por la estandarización.

- Las correlaciones $\rho_{(\mathbf{U},\mathbf{x}^{(1)})}$ y $\rho_{(\mathbf{V},\mathbf{x}^{(2)})}$ pueden ayudar en la interpretación de las variables canónicas.
- El espíritu es el mismo que en el análisis de componentes principales cuando las correlaciones entre los componentes principales y sus variables asociadas pueden proporcionar interpretaciones de los componentes, basadas en el área de estudio de las variables.

Ejemplo: Calcular las correlaciones entre las variables canonicas y sus variables originales del ejemplo anterior.

Correlaciones canónicas como generalizaciones de otros coeficientes de correlación

La correlación canónica generaliza la correlación entre dos variables. Cuando $\mathbf{X}^{(1)}$ y $\mathbf{X}^{(2)}$ consisten en una única variable, de modo que $p = q = 1$,

$$\left| Corr\left(X_1^{(1)}, X_2^{(2)}\right)\right| = \left| Corr\left(aX_1^{(1)}, bX_2^{(2)}\right)\right| \quad \forall\, a, b \neq 0$$

Por lo tanto, las variables canónicas $U_1 = X_1^{(1)}$ y $V_1 = X_1^{(2)}$ tienen correlación $\rho_1^* = \left| Corr\left(X_1^{(1)}, X_2^{(2)}\right)\right|$. Cuando $\mathbf{X}^{(1)}$ y $\mathbf{X}^{(2)}$ tienen más componentes, fijando $\mathbf{a}^{'} = [0, \ldots, 0, 1, 0, \ldots, 0]$ con 1 en la i-ésima posición y $\mathbf{b}^{'} = [0, \ldots, 0, 1, 0, \ldots, 0]$ con 1 en la k-ésima posición tenemos

$$\left| Corr\left(X_i^{(1)}, X_k^{(2)}\right)\right| = \left| Corr\left(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)}\right)\right| \leq \max_{\mathbf{a},\mathbf{b}} Corr\left(\mathbf{a}'\mathbf{X}^{(1)}, \mathbf{b}'\mathbf{X}^{(2)}\right) = \rho_1^*$$

Es decir, la primera correlación canónica es más grande que el valor absoluto de cualquier valor en $\rho_{12} = \mathbf{V}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{V}_{22}^{-1/2}$

Correlaciones canónicas como generalizaciones de otros Coeficientes de Correlación

- Además, el coeficiente de correlación múltiple, que se define como $R = +\sqrt{R^2} = \rho_{1(X^{(2)})}$, es un caso especial de una correlación canónica cuando $\mathbf{X}^{(1)}$ tiene el único elemento $X_1^{(1)}(p = 1)$. Recordando que

$$\rho_{1(X^{(2)})} = \max_{\mathbf{b}} Corr(X_1^{(1)}, \mathbf{b}'\mathbf{X}^{(2)}) = \rho_1^* \qquad p = 1$$

  Por tanto, el coeficiente de correlacion multiple es la correlación canonica para el caso en que $\mathbf{X}^{(1)}$ tiene un solo elemento.

- Cuando $p > 1$, ρ_1^* es mayor que cada una de las correlaciones múltiples de $X_i^{(1)}$ con $\mathbf{X}^{(2)}$ o las correlaciones múltiples de $X_i^{(2)}$ con $\mathbf{X}^{(1)}$.

Correlaciones canónicas como generalizaciones de otros Coeficientes de Correlación

Finalmente, observamos que

$$\rho_{U_k(X^{(2)})} = \max_{\mathbf{b}} Corr(U_k, \mathbf{b}\,'\mathbf{X}^{(2)}) = Corr(U_k, V_k) = \rho_k^* \qquad k = 1, 2, \ldots, p$$

De manera similar

$$\rho_{V_k(X^{(1)})} = \max_{\mathbf{a}} Corr(\mathbf{a}\,'\mathbf{X}^{(1)}, V_k) = Corr(U_k, V_k) = \rho_k^* \qquad k = 1, 2, \ldots, p$$

Es decir, las correlaciones canónicas son también los coeficientes de correlación múltiple de U_k con $\mathbf{X}^{(2)}$ o los coeficientes de correlación múltiple de V_k con $\mathbf{X}^{(1)}$.

Interpretación de las correlaciones canónicas poblacionales

- Debido a su interpretación como coeficientes de correlación múltiple, la k-ésima correlación canónica al cuadrado ρ_k^{*2} se interpreta como el coeficiente de determinación (R^2), es decir, es la proporción de la varianza de la variable canónica U_k "explicada" por el conjunto $\mathbf{X}^{(2)}$.
- Es también la proporción de la varianza de la variable canónica V_k "explicada" por el conjunto $\mathbf{X}^{(1)}$.
- Por lo tanto, ρ_k^{*2} se denomina frecuentemente como la varianza compartida entre los dos conjuntos $\mathbf{X}^{(1)}$ y $\mathbf{X}^{(2)}$.
- El valor más grande, ρ_1^{*2}, a veces se considera como una medida de la *superposición de los conjuntos.*

Variables canonica muestrales y correlaciones canónicas muestrales

Supongamos que se elige una muestra aleatoria de n observaciones sobre cada una de las $p+q$ variables $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$ y estas se organizan en la matriz de datos $n \times (p+q)$

$$\mathbf{X} = \left[\mathbf{X}^{(1)} | \mathbf{X}^{(2)} \right]$$

$$= \left[\begin{array}{cccc|cccc} x_{11}^{(1)} & x_{12}^{(1)} & \cdots & x_{1p}^{(1)} & x_{11}^{(2)} & x_{12}^{(2)} & \cdots & x_{1q}^{(2)} \\ x_{21}^{(1)} & x_{22}^{(1)} & \cdots & x_{2p}^{(1)} & x_{21}^{(2)} & x_{22}^{(2)} & \cdots & x_{2q}^{(2)} \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{n1}^{(1)} & x_{n2}^{(1)} & \cdots & x_{np}^{(1)} & x_{n1}^{(2)} & x_{n2}^{(2)} & \cdots & x_{nq}^{(2)} \end{array} \right] = \left[\begin{array}{c|c} \mathbf{x}_1^{(1)\prime} & \mathbf{x}_1^{(2)\prime} \\ \vdots & \vdots \\ \mathbf{x}_n^{(1)\prime} & \mathbf{x}_n^{(2)\prime} \end{array} \right]$$

El vector de medias muestrales puede organizarse como

$$\underset{(p+q)\times 1}{\overline{\mathbf{x}}} = \left[\begin{array}{c} \overline{\mathbf{x}}^{(1)} \\ \hline \overline{\mathbf{x}}^{(2)} \end{array} \right] \quad \text{donde} \quad \begin{array}{l} \overline{\mathbf{x}}^{(1)} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j^{(1)} \\ \overline{\mathbf{x}}^{(2)} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j^{(2)} \end{array}$$

Variables canonicas muestrales y correlaciones canónicas muestrales

Del mismo modo, la matriz de covarianza muestral puede ser organizada de manera análoga a la representación de $\mathbf{\Sigma}$. Por tanto

$$\underset{(p+q)\times(p+q)}{\mathbf{S}} = \left[\begin{array}{c|c} \underset{(\mathbf{p}\times\mathbf{p})}{\mathbf{S}_{11}} & \underset{(\mathbf{p}\times\mathbf{q})}{\mathbf{S}_{12}} \\ \hline \underset{(\mathbf{q}\times\mathbf{p})}{\mathbf{S}_{21}} & \underset{(\mathbf{q}\times\mathbf{q})}{\mathbf{S}_{22}} \end{array}\right]$$

donde

$$\mathbf{S}_{kl} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{x}_j^{(k)} - \overline{\mathbf{x}}^{(k)})(\mathbf{x}_j^{(l)} - \overline{\mathbf{x}}^{(l)})', \quad k, l = 1, 2.$$

Las combinaciones lineales

$$\hat{U} = \hat{\mathbf{a}}'\mathbf{x}^{(1)}, \quad \hat{V} = \hat{\mathbf{b}}'\mathbf{x}^{(2)}$$

tienen correlación muestral

$$\mathbf{r}_{\hat{U},\hat{V}} = \frac{\hat{\mathbf{a}}'\mathbf{S}_{12}\hat{\mathbf{b}}}{\sqrt{\hat{\mathbf{a}}'\mathbf{S}_{11}\hat{\mathbf{a}}}\sqrt{\hat{\mathbf{b}}'\mathbf{S}_{22}\hat{\mathbf{b}}}}$$

Variables canonicas muestrales y correlaciones canónicas muestrales

- El primer par de variables canónicas muestrales es el par de combinaciones lineales $\hat{U}_1, \hat{V}_1$ que tienen varianzas unitarias muestrales que maximizan la correlación muestral $\mathbf{r}_{\hat{U},\hat{V}}$.
- En general, el k-ésimo par de variables canónicas muestrales es el par de combinaciones lineales $\hat{U}_k, \hat{V}_k$ que tienen varianzas unitarias muestrales que maximizan la proporción $\mathbf{r}_{\hat{U},\hat{V}}$. entre aquellas combinaciones lineales no correlacionadas con las anterior $k-1$ variables canónicas muestrales.
- La correlación muestral entre $\hat{U}_k$ y $\hat{V}_k$ se denomina la k-ésima correlación canónica muestral.
- Las variables canónicas muestrales y las correlaciones canónicas muestrales se obtienen a partir de las matrices de covarianza muestrales $\mathbf{S}_{11}, \mathbf{S}_{12} = \mathbf{S_{21}}'$ y $\mathbf{S}_{22}$ de una manera consistente con el caso de poblacional descrito antes.

Variables canonicas muestrales y correlaciones canónicas muestrales

Resultado: Sean $\widehat{\rho_1^*}^2 \geq \widehat{\rho_2^*}^2 \geq \ldots \geq \widehat{\rho_p^*}^2$ los p valores propios ordenados de $\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$ con los correspondientes vectores propios $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \ldots . \hat{\mathbf{e}}_p$, donde $\boldsymbol{S_{kl}}$ está definida como antes y $p \leq q$. Sea $\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \ldots, \hat{\mathbf{f}}_p$ los vectores propios de $\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}$. Cada $\hat{\mathbf{f}}_i$ es proporcional a $\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\hat{\mathbf{e}}_i$, así los primeros p vectores propios $\hat{\mathbf{f}}'s$ se pueden obtener como $\hat{\mathbf{f}}_k = (1/\widehat{\rho_k^*})\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\hat{\mathbf{e}}_k$, $k = 1, 2, \ldots, p$. Entonces, el k-ésimo par de variables canónicas muestrales es

$$\hat{U}_k = \underbrace{\hat{\mathbf{e}}_k' \mathbf{S}_{11}^{-1/2}}_{\hat{\mathbf{a}}'_k} \mathbf{x}^{(1)} \qquad \hat{V}_k = \underbrace{\hat{\mathbf{f}}_k' \mathbf{S}_{22}^{-1/2}}_{\hat{\mathbf{b}}'_k} \mathbf{x}^{(2)}$$

donde $\mathbf{x}^{(1)}$ y $\mathbf{x}^{(2)}$ son los valores de las variables $\mathbf{X}^{(1)}$ y $\mathbf{X}^{(2)}$ para una unidad experimental particular.

Variables canonicas muestrales y correlaciones canónicas muestrales

El primer par de variables canónicas muestrales tiene la correlación máxima

$$r_{\hat{U}_1,\hat{V}_1} = \widehat{\rho_1^*}$$

y para el k-ésimo par de variables canónicas

$$r_{\hat{U}_k,\hat{V}_k} = \widehat{\rho_k^*}$$

es la correlación más grande posible entre combinaciones lineales no correlacionadas con las $k-1$ variables canónicas anteriores.

$\widehat{\rho_1^*}, \widehat{\rho_2^*}, \ldots, \widehat{\rho_p^*}$ son las correlaciones canónicas muestrales.

Las variables canónicas muestrales tienen varianzas muestrales unitarias

$$s_{\hat{U}_k,\hat{U}_k} = s_{\hat{V}_k,\hat{V}_k} = 1$$

y sus correlaciones muestrales son

$$r_{\hat{U}_k,\hat{U}_l} = r_{\hat{V}_k,\hat{V}_l} = 0, \; k \neq l$$
$$r_{\hat{U}_k,\hat{V}_l} = 0, \; k \neq l$$

Variables canonicas muestrales y correlaciones canónicas muestrales

La interpretación de $\hat{U}_k, \hat{V}_k$ es más sencilla mediante el cálculo de las correlaciones muestrales entre las variables canónicas y las variables en los conjuntos $\mathbf{X}^{(1)}$ y $\mathbf{X}^{(2)}$. Definimos las matrices

$$\underset{(pxp)}{\hat{\mathbf{A}}} = [\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \ldots, \hat{\mathbf{a}}_p]' \qquad \underset{(qxq)}{\hat{\mathbf{B}}} = \left[\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \ldots, \hat{\mathbf{b}}_q\right]'$$

cuyos renglones son los vectores de coeficientes para las variables canónicas muestrales. De manera análoga al caso de las variables canónicas poblacionales, tenemos

$$\underset{(px1)}{\hat{\mathbf{U}}} = \hat{\mathbf{A}}\mathbf{x}^{(1)} \quad \underset{(qx1)}{\hat{\mathbf{V}}} = \hat{\mathbf{B}}\mathbf{x}^{(2)}$$

Variables canonicas muestrales y correlaciones canónicas muestrales

Podemos definir

$$
\begin{aligned}
\mathbf{R}_{\hat{\mathbf{U}},\mathbf{x}^{(1)}} &= \text{matriz de correlaciones muestrales de } \hat{\mathbf{U}} \quad \text{con} \quad \mathbf{x}^{(1)} \\
\mathbf{R}_{\hat{\mathbf{V}},\mathbf{x}^{(2)}} &= \text{matriz de correlaciones muestrales de } \hat{\mathbf{V}} \quad \text{con} \quad \mathbf{x}^{(2)} \\
\mathbf{R}_{\hat{\mathbf{U}},\mathbf{x}^{(2)}} &= \text{matriz de correlaciones muestrales de } \hat{\mathbf{U}} \quad \text{con} \quad \mathbf{x}^{(2)} \\
\mathbf{R}_{\hat{\mathbf{V}},\mathbf{x}^{(1)}} &= \text{matriz de correlaciones muestrales de } \hat{\mathbf{V}} \quad \text{con} \quad \mathbf{x}^{(1)}
\end{aligned}
$$

De acuerdo a la definición para el caso poblacional, obtenemos que

$$
\begin{aligned}
\mathbf{R}_{\hat{\mathbf{U}},\mathbf{x}^{(1)}} &= \hat{\mathbf{A}}\mathbf{S}_{11}\mathbf{D}_{11}^{-1/2} \\
\mathbf{R}_{\hat{\mathbf{V}},\mathbf{x}^{(2)}} &= \hat{\mathbf{B}}\mathbf{S}_{22}\mathbf{D}_{22}^{-1/2} \\
\mathbf{R}_{\hat{\mathbf{U}},\mathbf{x}^{(2)}} &= \hat{\mathbf{A}}\mathbf{S}_{12}\mathbf{D}_{22}^{-1/2} \\
\mathbf{R}_{\hat{\mathbf{V}},\mathbf{x}^{(1)}} &= \hat{\mathbf{B}}\mathbf{S}_{21}\mathbf{D}_{11}^{-1/2}
\end{aligned}
$$

donde $\mathbf{D}_{11}^{-1/2}$ es la matriz diagonal (pxp) con i-ésimo elemento diagonal (la varianza muestral, $var(x_i^{(1)})^{-1/2}$) y $\mathbf{D}_{22}^{-1/2}$ es la matriz diagonal (qxq) con i-ésimo elemento diagonal (la varianza muestral, $var(x_i^{(2)})^{-1/2}$)

Variables canonicas muestrales y correlaciones canónicas muestrales estandarizadas

Si las observaciones son estandarizadas, la matriz de datos se convierte en

$$\mathbf{Z} = \left[\mathbf{Z}^{(1)} | \mathbf{Z}^{(2)}\right] = \left[\begin{array}{c|c} \mathbf{z}_1^{(1)\,\prime} & \mathbf{z}_1^{(2)\,\prime} \\ \vdots & \vdots \\ \mathbf{z}_n^{(1)\,\prime} & \mathbf{z}_n^{(2)\,\prime} \end{array}\right]$$

y las variables canónicas muestrales se convierten en

$$\underset{(px1)}{\hat{\mathbf{U}}} = \hat{\mathbf{A}}_z \mathbf{z}^{(1)} \qquad \underset{(qx1)}{\hat{\mathbf{V}}} = \hat{\mathbf{B}}_z \mathbf{z}^{(2)}$$

donde $\hat{\mathbf{A}}_z = \hat{\mathbf{A}}\mathbf{D}_{11}^{1/2}$ y $\hat{\mathbf{B}}_z = \hat{\mathbf{B}}\mathbf{D}_{22}^{1/2}$. Las correlaciones canónicas muestrales no son afectadas por la estandarización. Las correlaciones mostradas en la lámina anterior permanecen igual y pueden ser calculadas mediante observaciones estandarizadas, por medio de la substitución de $\hat{\mathbf{A}}_z$ por $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}_z$ por $\hat{\mathbf{B}}$, y $\mathbf{R}$ por $\mathbf{S}$. Note que $\mathbf{D}_{11}^{-1/2} = \underset{(pxp)}{\mathbf{I}}$ y $\mathbf{D}_{22}^{-1/2} = \underset{(qxq)}{\mathbf{I}}$ para observaciones $\mathbf{Z}$

- Si las variables canónicas representan un buen resumen de sus respectivos conjuntos de variables, entonces las asociaciones entre variable se pueden describir en términos de las variables canónicas y sus correlaciones.
- Es útil tener medidas que resuman el grado en que las variables canónicas explican la variación en sus respectivos conjuntos.
- También es útil, en ocasiones, calcular la proporción de varianza en un conjunto de variables explicada por las variables canónicas del otro conjunto.

Matrices de errores de aproximación

Dadas las matrices $\hat{\mathbf{A}}$ y $\hat{\mathbf{B}}$ definidas anteriormente, sean $\hat{\mathbf{a}}^{(i)}$ y $\hat{\mathbf{b}}^{(i)}$ la i-ésima columna de $\hat{\mathbf{A}}^{-1}$ y $\hat{\mathbf{B}}^{-1}$, respectivamente. Puesto que $\hat{\mathbf{U}} = \hat{\mathbf{A}}\mathbf{x}^{(1)}$ y $\hat{\mathbf{V}} = \hat{\mathbf{B}}\mathbf{x}^{(2)}$ podemos escribir

$$\underset{(px1)}{\mathbf{x}^{(1)}} = \underset{(pxp)}{\hat{\mathbf{A}}}^{-1} \underset{(px1)}{\hat{\mathbf{U}}} \qquad \underset{(qx1)}{\mathbf{x}^{(2)}} = \underset{(qxq)}{\hat{\mathbf{B}}}^{-1} \underset{(qx1)}{\hat{\mathbf{V}}}$$

Debido a que la covarianza muestral $Cov(\hat{\mathbf{U}}, \hat{\mathbf{V}}) = \hat{\mathbf{A}}\mathbf{S}_{12}\hat{\mathbf{B}}'$, y

1. $Cov(\hat{\mathbf{U}}) = \hat{\mathbf{A}}\mathbf{S}_{11}\hat{\mathbf{A}}' = \underset{(pxp)}{\mathbf{I}}$
2. $Cov(\hat{\mathbf{V}}) = \hat{\mathbf{B}}\mathbf{S}_{22}\hat{\mathbf{B}}' = \underset{(qxq)}{\mathbf{I}}$

Se deduce que

$$\mathbf{S}_{12} = \hat{\mathbf{A}}^{-1}\left[\begin{array}{cccc|c} \widehat{\rho_1^*} & 0 & \cdots & 0 & \\ 0 & \widehat{\rho_2^*} & \cdots & 0 & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \widehat{\rho_p^*} & \end{array}\right](\hat{\mathbf{B}}^{-1})' = \widehat{\rho_1^*}\hat{\mathbf{a}}^{(1)}\hat{\mathbf{b}}^{(1)\,\prime} + \widehat{\rho_2^*}\hat{\mathbf{a}}^{(2)}\hat{\mathbf{b}}^{(2)\,\prime} + \ldots + \widehat{\rho_p^*}\hat{\mathbf{a}}^{(p)}\hat{\mathbf{b}}^{(p)\,\prime}$$

$$\begin{aligned} \mathbf{S}_{11} &= (\hat{\mathbf{A}}^{-1})(\hat{\mathbf{A}}^{-1})' = \hat{\mathbf{a}}^{(1)}\hat{\mathbf{a}}^{(1)\,\prime} + \hat{\mathbf{a}}^{(2)}\hat{\mathbf{a}}^{(2)\,\prime} + \cdots + \hat{\mathbf{a}}^{(p)}\hat{\mathbf{a}}^{(p)\,\prime} \\ \mathbf{S}_{22} &= (\hat{\mathbf{B}}^{-1})(\hat{\mathbf{B}}^{-1})' = \hat{\mathbf{b}}^{(1)}\hat{\mathbf{b}}^{(1)\,\prime} + \hat{\mathbf{b}}^{(2)}\hat{\mathbf{b}}^{(2)\,\prime} + \cdots + \hat{\mathbf{b}}^{(q)}\hat{\mathbf{b}}^{(q)\,\prime} \end{aligned}$$

Como $\mathbf{x}^{(1)} = \hat{\mathbf{A}}^{-1}\hat{\mathbf{U}}$ y $\hat{\mathbf{U}}$ tiene covarianza muestral $\mathbf{I}$, las primeras r columnas de $\hat{\mathbf{A}}^{-1}$ contienen las covarianzas muestrales de las primeras r variables canónicas $\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_r$ con sus variables componentes $X_1^{(1)}, X_2^{(1)}, \ldots, X_p^{(1)}$. Similarmente las primeras r columnas de $\hat{\mathbf{B}}^{-1}$ contienen las covarianzas muestrales de $\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_r$ con sus variables componentes $X_1^{(2)}, X_2^{(2)}, \ldots, X_p^{(2)}$.

Si solo utilizamos los primeros r pares canónicos, entonces definimos,

$$\widetilde{\mathbf{x}}^{(1)} = \left[\begin{array}{c|c|c|c} \hat{\mathbf{a}}^{(1)} & \hat{\mathbf{a}}^{(2)} & \cdots & \hat{\mathbf{a}}^{(r)} \end{array}\right] \begin{bmatrix} \hat{U}_1 \\ \hat{U}_2 \\ \vdots \\ \hat{U}_r \end{bmatrix}$$

y

$$\widetilde{\mathbf{x}}^{(2)} = \left[\begin{array}{c|c|c|c} \hat{\mathbf{b}}^{(1)} & \hat{\mathbf{b}}^{(2)} & \cdots & \hat{\mathbf{b}}^{(r)} \end{array}\right] \begin{bmatrix} \hat{V}_1 \\ \hat{V}_2 \\ \vdots \\ \hat{V}_r \end{bmatrix}$$

entonces $\mathbf{S}_{12}$ se aproxima por
$Cov(\widetilde{\mathbf{x}}^{(1)}, \widetilde{\mathbf{x}}^{(2)}) = \widehat{\rho_1^*}\hat{\mathbf{a}}^{(1)}\hat{\mathbf{b}}^{(1)\,\prime} + \widehat{\rho_2^*}\hat{\mathbf{a}}^{(2)}\hat{\mathbf{b}}^{(2)\,\prime} + \ldots + \widehat{\rho_r^*}\hat{\mathbf{a}}^{(r)}\hat{\mathbf{b}}^{(r)\,\prime}$

Entonces las matrices de errores de aproximación son

$$\mathbf{S}_{11}-(\hat{\mathbf{a}}^{(1)}\hat{\mathbf{a}}^{(1)\,\prime}+\hat{\mathbf{a}}^{(2)}\hat{\mathbf{a}}^{(2)\,\prime}+\cdots+\hat{\mathbf{a}}^{(r)}\hat{\mathbf{a}}^{(r)\,\prime})=\hat{\mathbf{a}}^{(r+1)}\hat{\mathbf{a}}^{(r+1)\,\prime}+\cdots+\hat{\mathbf{a}}^{(p)}\hat{\mathbf{a}}^{(p)\,\prime}$$
$$\mathbf{S}_{22}-(\hat{\mathbf{b}}^{(1)}\hat{\mathbf{b}}^{(1)\,\prime}+\hat{\mathbf{b}}^{(2)}\hat{\mathbf{b}}^{(2)\,\prime}+\cdots+\hat{\mathbf{b}}^{(r)}\hat{\mathbf{b}}^{(r)\,\prime})=\hat{\mathbf{b}}^{(r+1)}\hat{\mathbf{b}}^{(r+1)\,\prime}+\cdots+\hat{\mathbf{b}}^{(q)}\hat{\mathbf{b}}^{(q)\,\prime}$$

$$\mathbf{S}_{12}-(\widehat{\rho_1^*}\hat{\mathbf{a}}^{(1)}\hat{\mathbf{b}}^{(1)\,\prime}+\widehat{\rho_2^*}\hat{\mathbf{a}}^{(2)}\hat{\mathbf{b}}^{(2)\,\prime}+\ldots+\widehat{\rho_r^*}\hat{\mathbf{a}}^{(r)}\hat{\mathbf{b}}^{(r)\,\prime})=\widehat{\rho^*}_{r+1}\hat{\mathbf{a}}^{(r+1)}\hat{\mathbf{b}}^{(r+1)\,\prime}+\cdots+\widehat{\rho^*}_{p}\hat{\mathbf{a}}^{(p)}\hat{\mathbf{b}}^{(p)\,\prime}$$

- Las matrices de error de aproximación se pueden interpretar como resúmenes descriptivos de qué tan bien las primeras r variables canónicas muestrales reproducen a las matrices de covarianzas de muestrales.
- Valores grandes en las entradas de las matrices de error de aproximación indican un *ajuste pobre* a las variables correspondientes.

Cuando las observaciones son estandarizadas, las matriz de correlación muestral $\mathbf{R}_{kl}$ sustituye a $\mathbf{S}_{kl}$ y $\hat{\mathbf{a}_z}^{(k)}$, $\hat{\mathbf{b}_z}^{(k)}$ sustituyen a $\hat{\mathbf{a}}^{(k)}$, $\hat{\mathbf{b}}^{(k)}$ en la relación de arriba

Ejemplo: Matrices de errores de aproximación

Los datos representan mediciones de cráneos ($\boldsymbol{X}^{(1)}$) y de huesos de piernas ($\boldsymbol{X}^{(2)}$) de aves de corral. Tomamos una muestra de aves y se les mide

$$Craneo(\boldsymbol{X}^{(1)}) = \begin{cases} X_1^{(1)} = \text{longitud del craneo} \\ X_2^{(1)} = \text{Anchura del craneo} \end{cases}$$

$$Pierna(\boldsymbol{X}^{(2)}) = \begin{cases} X_1^{(2)} = \text{longitud del femur} \\ X_2^{(2)} = \text{longitud de la tibia} \end{cases}$$

tienen la matriz de correlación muestral

$$\boldsymbol{R} = \left[\begin{array}{c|c} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \hline \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{array}\right] = \left[\begin{array}{cc|cc} 1.0 & 0.505 & 0.569 & 0.602 \\ 0.505 & 1.0 & .422 & 0.467 \\ \hline 0.569 & 0.422 & 1.0 & 0.926 \\ 0.602 & 0.467 & 0.926 & 1.0 \end{array}\right]$$

Obtener las matrices de errores de aproximación

- Cuando las observaciones son estandarizadas, las matrices de covarianza muestrales $\mathbf{S}_{kl}$ son las matrices de correlación $\mathbf{R}_{kl}$.
- Los vectores de coeficientes canónicos son los renglones de las matrices $\hat{\mathbf{A}}_z$ y $\hat{\mathbf{B}}_z$ y las columnas $\hat{\mathbf{A}}_z^{-1}$ y $\hat{\mathbf{B}}_z^{-1}$ son las correlaciones muestrales entres las variables canónicas y sus variables componentes estandarizadas. Específicamente

$$\text{Cov muestral}(z^{(1)}, \hat{\mathbf{U}}) = \text{Cov muestral}(\hat{\mathbf{A}}_z^{-1}\hat{\mathbf{U}}, \hat{\mathbf{U}}) = \hat{\mathbf{A}}_z^{-1}$$

y

$$\text{Cov muestral}(z^{(2)}, \hat{\mathbf{V}}) = \text{Cov muestral}(\hat{\mathbf{B}}_z^{-1}\hat{\mathbf{V}}, \hat{\mathbf{V}}) = \hat{\mathbf{B}}_z^{-1}$$

Proporciones de varianza muestral explicada

Entonces

$$\hat{\mathbf{A}}_z^{-1} = \left[\hat{\mathbf{a}}_z^{(1)}, \hat{\mathbf{a}}_z^{(2)}, \ldots, \hat{\mathbf{a}}_z^{(p)}\right] = \begin{bmatrix} r_{\hat{U}_1, z_1^{(1)}} & r_{\hat{U}_2, z_1^{(1)}} & \cdots & r_{\hat{U}_p, z_1^{(1)}} \\ r_{\hat{U}_1, z_2^{(1)}} & r_{\hat{U}_2, z_2^{(1)}} & \cdots & r_{\hat{U}_p, z_2^{(1)}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\hat{U}_1, z_p^{(1)}} & r_{\hat{U}_2, z_p^{(1)}} & \cdots & r_{\hat{U}_p, z_p^{(1)}} \end{bmatrix}$$

$$\hat{\mathbf{B}}_z^{-1} = \left[\hat{\mathbf{b}}_z^{(1)}, \hat{\mathbf{b}}_z^{(2)}, \ldots, \hat{\mathbf{b}}_z^{(q)}\right] = \begin{bmatrix} r_{\hat{V}_1, z_1^{(2)}} & r_{\hat{V}_2, z_1^{(2)}} & \cdots & r_{\hat{V}_q, z_1^{(2)}} \\ r_{\hat{V}_1, z_2^{(2)}} & r_{\hat{V}_2, z_2^{(2)}} & \cdots & r_{\hat{V}_q, z_2^{(2)}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{\hat{V}_1, z_q^{(2)}} & r_{\hat{V}_2, z_q^{(2)}} & \cdots & r_{\hat{V}_q, z_q^{(2)}} \end{bmatrix}$$

donde $r_{\hat{U}_i, z_k^{(1)}}$ son los coeficientes de correlaciones muestrales entre U_i y la k-esima variable estandarizada del grupo 1, $z_k^{(1)}$. $r_{\hat{V}_i, z_k^{(2)}}$ son los coeficientes de correlaciones muestrales entre la V_i y la k-esima variable estandarizada del grupo 2, $z_k^{(2)}$.

Para las observaciones estandarizadas también se cumple que

$$\mathbf{R_{11}} = (\hat{\mathbf{A}}_z^{-1})(\hat{\mathbf{A}}_z^{-1})' = \hat{\mathbf{a}}_z^{(1)}\hat{\mathbf{a}}_z^{(1)\,\prime} + \hat{\mathbf{a}}_z^{(2)}\hat{\mathbf{a}}_z^{(2)\,\prime} + \cdots + \hat{\mathbf{a}}_z^{(p)}\hat{\mathbf{a}}_z^{(p)\,\prime}$$

$$\mathbf{R_{22}} = (\hat{\mathbf{B}}_z^{-1})(\hat{\mathbf{B}}_z^{-1})' = \hat{\mathbf{b}}_z^{(1)}\hat{\mathbf{b}}_z^{(1)\,\prime} + \hat{\mathbf{b}}_z^{(2)}\hat{\mathbf{b}}_z^{(2)\,\prime} + \cdots + \hat{\mathbf{b}}_z^{(q)}\hat{\mathbf{b}}_z^{(q)\,\prime}$$

de lo cual se obtiene:

Varianza total muestral (estandarizada) en el primer conjunto

$$tr(\mathbf{R}_{11}) = tr(\hat{\mathbf{a}}_z^{(1)}\hat{\mathbf{a}}_z^{(1)\,\prime} + \hat{\mathbf{a}}_z^{(2)}\hat{\mathbf{a}}_z^{(2)\,\prime} + \cdots + \hat{\mathbf{a}}_z^{(p)}\hat{\mathbf{a}}_z^{(p)\,\prime}) = p$$

Varianza total muestral (estandarizada) en el segundo conjunto

$$tr(\mathbf{R}_{22}) = tr(\hat{\mathbf{b}}_z^{(1)}\hat{\mathbf{b}}_z^{(1)\,\prime} + \hat{\mathbf{b}}_z^{(2)}\hat{\mathbf{b}}_z^{(2)\,\prime} + \cdots + \hat{\mathbf{b}}_z^{(q)}\hat{\mathbf{b}}_z^{(q)\,\prime}) = q$$

Proporciones de varianza muestral explicada

Debido a que las correlaciones en las primeras $r < p$ columnas de $\hat{\mathbf{A}}_z^{-1}$ y $\hat{\mathbf{B}}_z^{-1}$ solo incluyen las variables canónicas muestrales $\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_r$ y $\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_r$ respectivamente.
Las contribuciones de las primeras r variables canonicas a las varianzas totales muestrales (estandarizadas) del grupo 1 se definen como

$$tr(\hat{\mathbf{a}}_z^{(1)}\hat{\mathbf{a}}_z^{(1)\,\prime} + \hat{\mathbf{a}}_z^{(2)}\hat{\mathbf{a}}_z^{(2)\,\prime} + \cdots + \hat{\mathbf{a}}_z^{(r)}\hat{\mathbf{a}}_z^{(r)\,\prime}) = \sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\hat{U}_i, z_k^{(1)}}$$

Las contribuciones de las primeras r variables canonicas a las varianzas totales muestrales (estandarizadas) del grupo 2 se definen como

$$tr(\hat{\mathbf{b}}_z^{(1)}\hat{\mathbf{b}}_z^{(1)\,\prime} + \hat{\mathbf{b}}_z^{(2)}\hat{\mathbf{b}}_z^{(2)\,\prime} + \cdots + \hat{\mathbf{b}}_z^{(r)}\hat{\mathbf{b}}_z^{(r)\,\prime}) = \sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\hat{V}_i, z_k^{(2)}}$$

Proporciones de varianza muestral explicada

Las proporciones de varianzas muestrales totales (estandarizadas) *explicadas* por las primeras variables canónicas están dadas por

$$R^2_{z^{(1)}|\hat{U}_1,\hat{U}_2,\ldots,\hat{U}_r} = \left(\begin{array}{c} \text{Proporción de varianza total} \\ \text{muestral estandarizada en el primer} \\ \text{conjunto explicada por } \hat{U}_1, \hat{U}_2, \ldots, \hat{U}_r \end{array} \right)$$

$$= \frac{tr(\hat{\mathbf{a}}_z^{(1)}\hat{\mathbf{a}}_z^{(1)\,\prime} + \cdots + \hat{\mathbf{a}}_z^{(r)}\hat{\mathbf{a}}_z^{(r)\,\prime})}{tr(\mathbf{R}_{11})}$$

$$= \frac{\sum_{i=1}^{r}\sum_{k=1}^{p} r^2_{\hat{U}_i, z_k^{(1)}}}{p}$$

y

$$R^2_{z^{(2)}|\hat{V}_1,\hat{V}_2,\ldots,\hat{V}_r} = \left(\begin{array}{c} \text{Proporción de varianza total} \\ \text{muestral estandarizada en el segundo} \\ \text{conjunto explicada por } \hat{V}_1, \hat{V}_2, \ldots, \hat{V}_r \end{array} \right)$$

$$= \frac{tr(\hat{\mathbf{b}}_z^{(1)}\hat{\mathbf{b}}_z^{(1)\,\prime} + \hat{\mathbf{b}}_z^{(2)}\hat{\mathbf{b}}_z^{(2)\,\prime} + \cdots + \hat{\mathbf{b}}_z^{(r)}\hat{\mathbf{b}}_z^{(r)\,\prime})}{tr(\mathbf{R}_{22})}$$

$$= \frac{\sum_{i=1}^{r}\sum_{k=1}^{q} r^2_{\hat{V}_i, z_k^{(2)}}}{q}$$

Cuando $\mathbf{\Sigma}_{12} = \mathbf{0}$,$\mathbf{a}'\mathbf{X}^{(1)}$ y $\mathbf{b}'\mathbf{X}^{(2)}$ tienen covarianza $\mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b} = 0$ para todos los vectores $\mathbf{a}$ y $\mathbf{b}$. En consecuencia, todas las correlaciones canónicas deben ser cero, y no tiene sentido seguir un análisis de correlación canónica. El siguiente resultado proporciona una forma de probar $\mathbf{\Sigma}_{12} = 0$ para muestras grandes.

Resultado: Sea

$$\mathbf{X}_j = \left[\frac{\mathbf{X}_j^{(1)}}{\mathbf{X}_j^{(2)}} \right], \quad j = 1, 2, \ldots, n$$

una muestra aleatoria de una población $N_{p+q}(\mu, \mathbf{\Sigma})$ con

$$\mathbf{\Sigma} = \begin{bmatrix} \underset{(pxp)}{\mathbf{\Sigma}_{11}} & \underset{(pxq)}{\mathbf{\Sigma}_{12}} \\ \underset{(qxp)}{\mathbf{\Sigma}_{21}} & \underset{(qxq)}{\mathbf{\Sigma}_{22}} \end{bmatrix}$$

Entonces la prueba de la razón de verosimilitud de $H_0 : \mathbf{\Sigma}_{12} = \underset{(pxq)}{\mathbf{0}}$ vs $H_1 : \mathbf{\Sigma}_{12} \neq \underset{(pxq)}{\mathbf{0}}$ rechaza H_0 para valores grandes de

$$-2\ln\Lambda = n\ln\left(\frac{|\mathbf{S}_{11}||\mathbf{S}_{22}|}{|\mathbf{S}|}\right) = -n\ln\prod_{i=1}^{p}(1-\widehat{\rho_i^*}^2)$$

donde

$$\mathbf{S} = \left[\begin{array}{c|c} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \hline \mathbf{S}_{21} & \mathbf{S}_{22} \end{array}\right]$$

es el estimador insesgado de $\mathbf{\Sigma}$. Para n grande, la distribución del estadístico de prueba se aproxima a una variable chi-cuadrada con pq grados de libertad.

Cuando n y $n-(p+q)$ son ambos grandes, el estadístico modificado

$$-\left[n-1-\frac{1}{2}(p+q+1)\right]\ln\prod_{i=1}^{p}(1-\widehat{\rho_i^{*}}^{2})\sim\chi^2_{pq}$$

Por tanto para n y $n-(p+q)$ grandes, se rechaza $H_0:\mathbf{\Sigma}_{12}=\underset{(pxq)}{\mathbf{0}}\,(\rho_1^*=\rho_2^*=\cdots=\rho_p^*)$ en un nivel de significancia α, si

$$-\left[n-1-\frac{1}{2}(p+q+1)\right]\ln\prod_{i=1}^{p}(1-\widehat{\rho_i^{*}}^{2})>\chi^2_{pq}(\alpha)$$

donde $\chi^2_{pq}(\alpha)$ es el percentil superior del α% de una Chi-cuadrada con pq grados de libertad