MAT-269: Sesión 14
Regresión Multivariada II

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

El objetivo de esta sección es estimar $\boldsymbol{B}$ sujeto a restricciones del tipo:

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{C},$$

donde $\boldsymbol{A}$ es matriz $r \times p$ de rango r y $\boldsymbol{C}$ es matriz $t \times k$. Sabemos que $\boldsymbol{A}$ puede ser particionada como:

$$\boldsymbol{A} = (\boldsymbol{A}_r, \boldsymbol{A}_s),$$

donde $\boldsymbol{A}_r$ es no singular. De este modo,

$$\boldsymbol{A}\boldsymbol{B} = (\boldsymbol{A}_r, \boldsymbol{A}_s)\begin{pmatrix}\boldsymbol{B}_r \\ \boldsymbol{B}_s\end{pmatrix} = \boldsymbol{A}_r\boldsymbol{B}_r + \boldsymbol{A}_s\boldsymbol{B}_s = \boldsymbol{C},$$

es decir,

$$\boldsymbol{B}_r = \boldsymbol{A}_r^{-1}(\boldsymbol{C} - \boldsymbol{A}_s\boldsymbol{B}_s).$$

Substituyendo en el modelo, tenemos

$$\begin{aligned}
\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{U} &= (\boldsymbol{X}_r, \boldsymbol{X}_s)\begin{pmatrix}\boldsymbol{B}_r\\ \boldsymbol{B}_s\end{pmatrix} + \boldsymbol{U},\\
&= \boldsymbol{X}_r\boldsymbol{B}_r + \boldsymbol{X}_s\boldsymbol{B}_s + \boldsymbol{U},\\
&= \boldsymbol{X}_r\boldsymbol{A}_r^{-1}(\boldsymbol{C} - \boldsymbol{A}_s\boldsymbol{B}_s) + \boldsymbol{X}_s\boldsymbol{B}_s + \boldsymbol{U},\\
&= \boldsymbol{X}_r\boldsymbol{A}_r^{-1}\boldsymbol{C} + (\boldsymbol{X}_s - \boldsymbol{X}_r\boldsymbol{A}_r^{-1}\boldsymbol{A}_s)\boldsymbol{B}_s + \boldsymbol{U},
\end{aligned}$$

que puede ser escrito como:

$$\boldsymbol{Y}_R = \boldsymbol{X}_R\boldsymbol{B}_s + \boldsymbol{U}, \qquad \boldsymbol{U} \sim \mathsf{N}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}),$$

con

$$\boldsymbol{Y}_R = \boldsymbol{Y} - \boldsymbol{X}_r\boldsymbol{A}_r^{-1}\boldsymbol{C}, \qquad \boldsymbol{X}_R = \boldsymbol{X}_s - \boldsymbol{X}_r\boldsymbol{A}_r^{-1}\boldsymbol{A}_s.$$

De este modo,

$$\widetilde{\boldsymbol{B}}_s = (\boldsymbol{X}_R^\top \boldsymbol{X}_R)^{-1} \boldsymbol{X}_R^\top \boldsymbol{Y},$$
$$\widetilde{\boldsymbol{B}}_r = \boldsymbol{A}_r^{-1}(\boldsymbol{C} - \boldsymbol{A}_s \widetilde{\boldsymbol{B}}_s)$$

Además, como $\boldsymbol{U} \sim \mathsf{N}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ sigue que

$$\boldsymbol{Y}_R \sim \mathsf{N}(\boldsymbol{X}_R \boldsymbol{B}_s, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}),$$

y por tanto,

$$\widetilde{\boldsymbol{B}}_s \sim \mathsf{N}(\boldsymbol{B}_s, (\boldsymbol{X}_R^\top \boldsymbol{X}_R)^{-1} \otimes \boldsymbol{\Sigma}).$$

Como

$$\begin{aligned} \widetilde{\boldsymbol{B}} = \begin{pmatrix} \widetilde{\boldsymbol{B}}_r \\ \widetilde{\boldsymbol{B}}_s \end{pmatrix} &= \begin{pmatrix} \boldsymbol{A}_r^{-1}(\boldsymbol{C} - \boldsymbol{A}_s \widetilde{\boldsymbol{B}}_s) \\ \widetilde{\boldsymbol{B}}_s \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{A}_r^{-1}\boldsymbol{C} \\ \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} -\boldsymbol{A}_r^{-1}\boldsymbol{A}_s \\ \boldsymbol{I} \end{pmatrix} \widetilde{\boldsymbol{B}}_s. \end{aligned}$$

Así, $\widetilde{\boldsymbol{B}}$ sigue una distribución normal con

$$\mathsf{E}(\widetilde{\boldsymbol{B}}) = \begin{pmatrix} \boldsymbol{A}_r^{-1}\boldsymbol{C} \\ \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} -\boldsymbol{A}_r^{-1}\boldsymbol{A}_s \\ \boldsymbol{I} \end{pmatrix} \mathsf{E}(\widetilde{\boldsymbol{B}}_s) = \begin{pmatrix} \boldsymbol{B}_r \\ \boldsymbol{B}_s \end{pmatrix}$$

y

$$\operatorname{vec} \widetilde{\boldsymbol{B}} = \operatorname{vec} \begin{pmatrix} \boldsymbol{A}_r^{-1}\boldsymbol{C} \\ \boldsymbol{0} \end{pmatrix} + \left(\boldsymbol{I} \otimes \begin{pmatrix} -\boldsymbol{A}_r^{-1}\boldsymbol{A}_s \\ \boldsymbol{I} \end{pmatrix} \right) \operatorname{vec} \widetilde{\boldsymbol{B}}_r,$$

de donde sigue que

$$\begin{aligned} \mathsf{Cov}(\operatorname{vec} \widetilde{\boldsymbol{B}}) &= \left(\boldsymbol{I} \otimes \begin{pmatrix} -\boldsymbol{A}_r^{-1}\boldsymbol{A}_s \\ \boldsymbol{I} \end{pmatrix} \right) \mathsf{Cov}(\operatorname{vec} \widetilde{\boldsymbol{B}}_s) \left(\boldsymbol{I} \otimes \begin{pmatrix} -\boldsymbol{A}_r^{-1}\boldsymbol{A}_s \\ \boldsymbol{I} \end{pmatrix} \right)^\top \\ &= (\boldsymbol{X}_R^\top \boldsymbol{X}_R)^{-1} \otimes \begin{pmatrix} -\boldsymbol{A}_r^{-1}\boldsymbol{A}_s \\ \boldsymbol{I} \end{pmatrix} \boldsymbol{\Sigma} \begin{pmatrix} -\boldsymbol{A}_r^{-1}\boldsymbol{A}_s \\ \boldsymbol{I} \end{pmatrix}^\top \end{aligned}$$

Estimación bajo restricciones lineales

Por otro lado,

$$
\begin{aligned}
\boldsymbol{Y}_R - \boldsymbol{X}_R \widetilde{\boldsymbol{B}}_s &= \boldsymbol{Y} - \boldsymbol{X}_r \boldsymbol{A}_r^{-1} \boldsymbol{C} - (\boldsymbol{X}_s - \boldsymbol{X}_r \boldsymbol{A}_r^{-1} \boldsymbol{A}_s) \widetilde{\boldsymbol{B}}_s \\
&= \boldsymbol{Y} - \boldsymbol{X}_r \boldsymbol{A}_r^{-1} (\boldsymbol{C} - \boldsymbol{A}_r^{1} \boldsymbol{A}_s \widetilde{\boldsymbol{B}}_s) - \boldsymbol{X}_s \widetilde{\boldsymbol{B}}_s \\
&= \boldsymbol{Y} - \boldsymbol{X} \widetilde{\boldsymbol{B}},
\end{aligned}
$$

lo que lleva a

$$
\begin{aligned}
\widetilde{\boldsymbol{\Sigma}} &= \frac{1}{n} (\boldsymbol{Y}_R - \boldsymbol{X}_R \widetilde{\boldsymbol{B}}_s)^\top (\boldsymbol{Y}_R - \boldsymbol{X}_R \widetilde{\boldsymbol{B}}_s) \\
&= \frac{1}{n} (\boldsymbol{Y} - \boldsymbol{X} \widetilde{\boldsymbol{B}})^\top (\boldsymbol{Y} - \boldsymbol{X} \widetilde{\boldsymbol{B}}) \\
&= \frac{1}{n} Q(\widetilde{\boldsymbol{B}}).
\end{aligned}
$$

Test de hipótesis lineales

Considere

$$
\begin{aligned}
Q(\widehat{\boldsymbol{B}}) &= (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}})^\top(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}) = \boldsymbol{R} \\
Q(\widetilde{\boldsymbol{B}}) &= (\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{B}})^\top(\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{B}}) = \boldsymbol{S}
\end{aligned}
$$

Cuando $H_0 : \boldsymbol{A}\boldsymbol{B} = \boldsymbol{C}$ es verdadera, $\boldsymbol{R}$ y $\boldsymbol{H} = \boldsymbol{S} - \boldsymbol{R}$ son independientemente distribuídos $\mathsf{W}_k(n-p, \boldsymbol{\Sigma})$ y $\mathsf{W}_k(r, \boldsymbol{\Sigma})$, respectivamente.

Además, podemos escribir

$$
\boldsymbol{H} = (\boldsymbol{A}\widehat{\boldsymbol{B}} - \boldsymbol{C})^\top[\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top]^{-1}(\boldsymbol{A}\widehat{\boldsymbol{B}} - \boldsymbol{C}),
$$

como $n - k \geq p$, ambos $\boldsymbol{R}$ y $\boldsymbol{S}$ son definidas positivas con probabilidad 1.

Test de hipótesis lineales

Sea $L(\boldsymbol{B}, \boldsymbol{\Sigma})$ la función de verosimilitud para las filas de $\boldsymbol{Y}$, el test de razón de verosimilitudes para $H_0 : \boldsymbol{A}\boldsymbol{B} = \boldsymbol{C}$, es

$$\Lambda = \frac{L(\widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\Sigma}})}{L(\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Sigma}})} = \frac{|\widetilde{\boldsymbol{\Sigma}}|^{-n/2}}{|\widehat{\boldsymbol{\Sigma}}|^{-n/2}},$$

de este modo

$$T = \Lambda^{2/n} = \frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widetilde{\boldsymbol{\Sigma}}|} = \frac{|\boldsymbol{R}|}{|\boldsymbol{S}|} = \frac{|\boldsymbol{R}|}{|\boldsymbol{R} + \boldsymbol{H}|} = |\boldsymbol{I} - \boldsymbol{V}|,$$

donde $\boldsymbol{V} = \boldsymbol{S}^{-1/2}\boldsymbol{H}\boldsymbol{S}^{-1/2}$. Cuando H_0 es verdadera $T \sim \Lambda(k, r, n-p)$ y por el principio de razón de verosimilitudes, rechazamos $H_0 : \boldsymbol{A}\boldsymbol{B} = \boldsymbol{C}$ si T es muy pequeño, es decir, si $|\boldsymbol{S}|$ es mucho mayor que $|\boldsymbol{R}|$.

MAT-269: Sesión 13
Ecuaciones Simultáneas

Note que podemos escribir el modelo de regresión multivariado como:

$$\boldsymbol{y}_i = \boldsymbol{B}^\top \boldsymbol{x}_i + \boldsymbol{u}_i, \qquad i = 1, \dots, n, \tag{1}$$

donde $\boldsymbol{u}_i \sim \mathsf{N}_k(\boldsymbol{0}, \boldsymbol{\Sigma})$. A continuación, introducimos una extensión del modelo dado en Ecuación (1) como:

$$\boldsymbol{\Gamma}^\top \boldsymbol{y}_i + \boldsymbol{B}^\top \boldsymbol{x}_i = \boldsymbol{u}_i, \qquad i = 1, \dots, n, \tag{2}$$

donde $\boldsymbol{\Gamma}$ es matriz $k \times k$. Además, asumiremos que $\boldsymbol{\Gamma}$ es no singular.[1]

El modelo de ecuaciones simultaneas en (2) es dado por:

$$\boldsymbol{Y}\boldsymbol{\Gamma} + \boldsymbol{X}\boldsymbol{B} = \boldsymbol{U}, \qquad \boldsymbol{U} \sim \mathsf{N}(\boldsymbol{0}, \boldsymbol{I} \otimes \boldsymbol{\Sigma}).$$

[1] Aunque esto puede ser considerado como una consecuencia de asumir que $\{\boldsymbol{u}_i\}$ siguen una distribución normal.

Haciendo $\boldsymbol{\Theta} = -\boldsymbol{B}\boldsymbol{\Gamma}^{-1}$, lleva a escribir

$$\begin{aligned} \boldsymbol{Y}\boldsymbol{\Gamma} &= -\boldsymbol{X}\boldsymbol{B} + \boldsymbol{U} \\ \boldsymbol{Y} &= \boldsymbol{X}(-\boldsymbol{B}\boldsymbol{\Gamma}^{-1}) + \boldsymbol{U}\boldsymbol{\Gamma}^{-1} \\ &= \boldsymbol{X}\boldsymbol{\Theta} + \boldsymbol{V}, \end{aligned}$$

en cuyo caso las filas de $\boldsymbol{V}$ son

$$\boldsymbol{v}_i \stackrel{\text{IID}}{\sim} \mathsf{N}_k(\boldsymbol{0}, \boldsymbol{\Omega}), \qquad \boldsymbol{\Omega} = \boldsymbol{\Gamma}^{-\top}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1}.$$

La función de log-verosimilitud en términos de $(\boldsymbol{\Theta}, \boldsymbol{\Omega})$ es dada por:

$$\ell_n(\boldsymbol{\Theta}, \boldsymbol{\Omega}) = -\frac{nk}{2}\log 2\pi - \frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\operatorname{tr}\boldsymbol{\Omega}^{-1}\boldsymbol{W},$$

donde

$$\boldsymbol{W} = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Theta})^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Theta})$$

Escribiendo la log-verosimilitud en términos de $(\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$, usando $\boldsymbol{\Theta} = -\boldsymbol{B}\boldsymbol{\Gamma}^{-1}$ y $\boldsymbol{\Omega} = \boldsymbol{\Gamma}^{-\top}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1}$, obtenemos:

$$\ell_n(\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) = -\frac{nk}{2}\log 2\pi + \frac{n}{2}\log|\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}| - \frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_*,$$

donde

$$\boldsymbol{W}_* = (\boldsymbol{Y}\boldsymbol{\Gamma} - \boldsymbol{X}\boldsymbol{B})^\top(\boldsymbol{Y}\boldsymbol{\Gamma} - \boldsymbol{X}\boldsymbol{B})$$

Observación:

Este modelo tiene problemas de identificabilidad. En efecto, considere

$$\boldsymbol{G}\boldsymbol{\Gamma}^\top\boldsymbol{y}_i + \boldsymbol{G}\boldsymbol{B}^\top\boldsymbol{x}_i = \boldsymbol{G}\boldsymbol{u}_i, \qquad i = 1, \dots, n,$$

con $\boldsymbol{G}$ matriz no singular. De este modo,

$$\boldsymbol{G}\boldsymbol{u}_i \sim \mathsf{N}_k(\boldsymbol{0}, \boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^\top).$$

Definiendo

$$\boldsymbol{B}_0 = \boldsymbol{B}\boldsymbol{G}^\top, \qquad \boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}\boldsymbol{G}^\top, \qquad \boldsymbol{\Sigma}_0 = \boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^\top,$$

sigue que

$$\boldsymbol{\Theta}_0 = -\boldsymbol{B}_0\boldsymbol{\Gamma}_0^{-1} = -\boldsymbol{B}\boldsymbol{G}^\top(\boldsymbol{\Gamma}\boldsymbol{G}^\top)^{-1} = -\boldsymbol{B}\boldsymbol{\Gamma} = \boldsymbol{\Theta},$$

mientras que

$$\begin{aligned}\boldsymbol{\Omega}_0 &= \boldsymbol{\Gamma}_0^{-\top}\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0^{-1} = (\boldsymbol{G}^{-\top}\boldsymbol{\Gamma}^{-1})^\top\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^\top\boldsymbol{G}^{-\top}\boldsymbol{\Gamma}^{-1} \\ &= \boldsymbol{\Gamma}^{-\top}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1} = \boldsymbol{\Omega},\end{aligned}$$

de ahí que $(\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\Omega})$ no son identificados.

Observación:

Existen diversos procedimientos de estimación para el modelo de ecuaciones simultáneas basados esencialmente el perfilar la verosimilitud (FIML, LIML, 2SLS).