

MAT-266: Estimación sujeto a restricciones lineales

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM



Problema:

Abordar la estimación de β y σ^2 sujeto a restricciones lineales del tipo:

$$G\beta = g, \tag{1}$$

donde $G \in \mathbb{R}^{q \times p}$ con $\text{rg}(G) = q$ y $g \in \mathbb{R}^q$.

Objetivo:

Consideraremos dos procedimientos para obtener estimadores restringidos

- ▶ Método de reducción.
- ▶ Método de multiplicadores de Lagrange.

Además, estudiaremos las propiedades estadísticas de los estimadores.



Estimación bajo restricciones lineales

Sea $G = (G_r, G_q)$ donde $G_q \in \mathbb{R}^{q \times q}$ de rango q . De este modo, podemos escribir las restricciones en (1) como:

$$G\beta = (G_r, G_q) \begin{pmatrix} \beta_r \\ \beta_q \end{pmatrix} = G_r\beta_r + G_q\beta_q = g,$$

como G_q es no singular, tenemos

$$\beta_q = G_q^{-1}(g - G_r\beta_r).$$

Particionando X del mismo modo que $\beta = (\beta_r^\top, \beta_q^\top)^\top$, sigue que

$$\begin{aligned} X\beta &= (X_r, X_q) \begin{pmatrix} \beta_r \\ \beta_q \end{pmatrix} = X_r\beta_r + X_q\beta_q \\ &= X_r\beta_r + X_q G_q^{-1}(g - G_r\beta_r) \\ &= (X_r - X_q G_q^{-1} G_r)\beta_r + X_q G_q^{-1} g \end{aligned}$$



Estimación bajo restricciones lineales

Así, podemos escribir el modelo lineal

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

como

$$\mathbf{Y} = (\mathbf{X}_r - \mathbf{X}_q \mathbf{G}_q^{-1} \mathbf{G}_r) \boldsymbol{\beta}_r + \mathbf{X}_q \mathbf{G}_q^{-1} \mathbf{g} + \boldsymbol{\epsilon},$$

es decir, obtenemos el **modelo reducido**, dado por

$$\mathbf{Y}_R = \mathbf{X}_R \boldsymbol{\beta}_r + \boldsymbol{\epsilon},$$

donde

$$\mathbf{Y}_R = \mathbf{Y} - \mathbf{X}_q \mathbf{G}_q^{-1} \mathbf{g}, \quad \mathbf{X}_R = \mathbf{X}_r - \mathbf{X}_q \mathbf{G}_q^{-1} \mathbf{G}_r.$$

En cuyo caso, sabemos que

$$\tilde{\boldsymbol{\beta}}_r = (\mathbf{X}_R^\top \mathbf{X}_R)^{-1} \mathbf{X}_R^\top \mathbf{Y}_R,$$

$$s_r^2 = \frac{1}{n-r} Q_R(\tilde{\boldsymbol{\beta}}_r),$$

con

$$Q_R(\tilde{\boldsymbol{\beta}}_r) = \mathbf{Y}_R^\top (\mathbf{I} - \mathbf{X}_R (\mathbf{X}_R^\top \mathbf{X}_R)^{-1} \mathbf{X}_R^\top) \mathbf{Y}_R$$



Además,

$$\begin{aligned}\tilde{\beta} &= \begin{pmatrix} \tilde{\beta}_r \\ \tilde{\beta}_q \end{pmatrix} = \begin{pmatrix} \tilde{\beta}_r \\ \mathbf{G}_q^{-1}(\mathbf{g} - \mathbf{G}_r \tilde{\beta}_r) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{0} \\ \mathbf{G}_q^{-1} \mathbf{g} \end{pmatrix} + \begin{pmatrix} \mathbf{I} \\ -\mathbf{G}_q^{-1} \mathbf{G}_r \end{pmatrix} \tilde{\beta}_r,\end{aligned}\tag{2}$$

Mientras que

$$\begin{aligned}\mathbf{Y}_R - \mathbf{X}_R \tilde{\beta}_r &= \mathbf{Y} - \mathbf{X}_q \mathbf{G}_q^{-1} \mathbf{g} - (\mathbf{X}_r - \mathbf{X}_q \mathbf{G}_q^{-1} \mathbf{g}) \tilde{\beta}_r \\ &= \mathbf{Y} - (\mathbf{X}_r, \mathbf{X}_q) \begin{pmatrix} \tilde{\beta}_r \\ \tilde{\beta}_q \end{pmatrix} = \mathbf{Y} - \mathbf{X} \tilde{\beta},\end{aligned}$$

de este modo

$$\|\mathbf{Y}_R - \mathbf{X}_R \tilde{\beta}_r\|^2 = \|\mathbf{Y} - \mathbf{X} \tilde{\beta}\|^2.\tag{3}$$



Resultado 1:

Para el modelo lineal $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ sujeto a las restricciones $\mathbf{G}\boldsymbol{\beta} = \mathbf{g}$ con $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. el MLE restringido de $\boldsymbol{\beta}$ es dado por (2) y tenemos que

$$\tilde{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \text{Cov}(\tilde{\boldsymbol{\beta}})),$$

donde

$$\text{Cov}(\tilde{\boldsymbol{\beta}}) = \sigma^2 \begin{pmatrix} \mathbf{I} \\ -\mathbf{G}_q^{-1} \mathbf{G}_r \end{pmatrix} (\mathbf{X}_R^\top \mathbf{X}_R)^{-1} \begin{pmatrix} \mathbf{I} \\ -\mathbf{G}_q^{-1} \mathbf{G}_r \end{pmatrix}^\top.$$

Mientras que

$$s_r^2 = \frac{1}{n-r} Q_R(\tilde{\boldsymbol{\beta}}_r) = \frac{1}{n-r} Q(\tilde{\boldsymbol{\beta}}),$$

donde

$$\frac{(n-r)s_r^2}{\sigma^2} \sim \chi^2(n-r),$$

y $\tilde{\boldsymbol{\beta}}$, $Q(\tilde{\boldsymbol{\beta}})$ son independientes.



Demostración:

Sabemos que

$$\begin{aligned}\tilde{\beta}_r &\sim N_r(\beta_r, \sigma^2(\mathbf{X}_R^\top \mathbf{X}_R)^{-1}), \\ \frac{(n-r)s_r^2}{\sigma^2} &= \frac{Q_R(\tilde{\beta}_r)}{\sigma^2} = \frac{\mathbf{Y}_R^\top (\mathbf{I} - \mathbf{H}_R) \mathbf{Y}_R}{\sigma^2} \sim \chi^2(n-r).\end{aligned}$$

Así, por (2), tenemos

$$\begin{aligned}E(\tilde{\beta}) &= \begin{pmatrix} \mathbf{0} \\ \mathbf{G}_q^{-1} \mathbf{g} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_r \\ -\mathbf{G}_q^{-1} \mathbf{G}_r \end{pmatrix} E(\tilde{\beta}_r) \\ &= \begin{pmatrix} \beta_r \\ \mathbf{G}_q^{-1}(\mathbf{g} - \mathbf{G}_r \beta_r) \end{pmatrix} = \begin{pmatrix} \beta_r \\ \beta_q \end{pmatrix} = \beta\end{aligned}$$



Además,

$$\begin{aligned}\text{Cov}(\tilde{\beta}) &= \begin{pmatrix} \mathbf{I}_r \\ -\mathbf{G}_q^{-1}\mathbf{G}_r \end{pmatrix} \text{Cov}(\tilde{\beta}_r) \begin{pmatrix} \mathbf{I}_r \\ -\mathbf{G}_q^{-1}\mathbf{G}_r \end{pmatrix}^\top \\ &= \sigma^2 \begin{pmatrix} \mathbf{I}_r \\ -\mathbf{G}_q^{-1}\mathbf{G}_r \end{pmatrix} (\mathbf{X}_R^\top \mathbf{X}_R)^{-1} \begin{pmatrix} \mathbf{I}_r \\ -\mathbf{G}_q^{-1}\mathbf{G}_r \end{pmatrix}^\top.\end{aligned}$$

como $\tilde{\beta}$ es una función lineal de $\tilde{\beta}_r$ la normalidad sigue. La independencia entre $\tilde{\beta}$ y $Q(\tilde{\beta})$ sigue por el [Resultado 1](#) en [Slides 7](#).



Estimación bajo restricciones lineales

Considere

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}).$$

La **función Langrangiana** asociada a la restricción lineal $\mathbf{G}\boldsymbol{\beta} = \mathbf{g}$ es dada por:

$$F(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \ell(\boldsymbol{\theta}) + \frac{1}{\sigma^2} \boldsymbol{\lambda}^\top (\mathbf{G}\boldsymbol{\beta} - \mathbf{g}),$$

con $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$. De este modo,

$$\frac{\partial F(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} \mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \frac{1}{\sigma^2} \mathbf{G}^\top \boldsymbol{\lambda}$$

$$\frac{\partial F(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \{Q(\boldsymbol{\beta}) - 2\boldsymbol{\lambda}^\top (\mathbf{G}\boldsymbol{\beta} - \mathbf{g})\}$$

$$\frac{\partial F(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \mathbf{G}\boldsymbol{\beta} - \mathbf{g}$$



Desde la condición de primer orden, obtenemos las ecuaciones de estimación,

$$\begin{aligned}\mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \mathbf{G}^\top \boldsymbol{\lambda} &= \mathbf{0}, \\ n\sigma^2 - \{Q(\boldsymbol{\beta}) - 2\boldsymbol{\lambda}^\top (\mathbf{G}\boldsymbol{\beta} - \mathbf{g})\} &= 0, \\ \mathbf{G}\boldsymbol{\beta} &= \mathbf{g},\end{aligned}$$

es decir,

$$\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = \mathbf{X}^\top \mathbf{Y} + \mathbf{G}^\top \boldsymbol{\lambda}, \quad (4)$$

$$\sigma^2 = \frac{1}{n} Q(\boldsymbol{\beta}) \quad (5)$$

$$\mathbf{G}\boldsymbol{\beta} = \mathbf{g}, \quad (6)$$

Resolviendo la Ecuación (4) con relación a $\boldsymbol{\beta}$ tenemos

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{Y} + \mathbf{G}^\top \boldsymbol{\lambda})$$



Substituyendo este resultado en (6) y resolviendo para λ , sigue que

$$G\tilde{\beta} = G(X^\top X)^{-1}(X^\top Y + G^\top \lambda) = g,$$

es decir,

$$G(X^\top X)^{-1}X^\top Y + G(X^\top X)^{-1}G^\top \lambda = g,$$

por tanto,

$$\tilde{\lambda} = (G(X^\top X)^{-1}G^\top)^{-1}(g - G\hat{\beta}).$$

Reemplazando este resultado en $\tilde{\beta}$ resulta

$$\begin{aligned}\tilde{\beta} &= (X^\top X)^{-1}(X^\top Y + G^\top \tilde{\lambda}) \\ &= (X^\top X)^{-1}X^\top Y + (X^\top X)^{-1}G^\top (G(X^\top X)^{-1}G^\top)^{-1}(g - G\hat{\beta}) \\ &= \hat{\beta} + (X^\top X)^{-1}G^\top (G(X^\top X)^{-1}G^\top)^{-1}(g - G\hat{\beta})\end{aligned}$$



Que puede ser reorganizado como:

$$\tilde{\beta} = A\hat{\beta} + Bg = \hat{\beta} - B(G\hat{\beta} - g),$$

donde $\hat{\beta}$ corresponde al MLE no restringido para β , con

$$B = (X^T X)^{-1} G^T (G(X^T X)^{-1} G^T)^{-1} \quad (7)$$

$$A = I - BG \quad (8)$$

y el estimador insesgado para σ^2 es dado por

$$s_r^2 = \frac{1}{n-r} Q(\tilde{\beta}).$$

Para estudiar las propiedades de este MLE restringido, considere primeramente el siguiente lema.



Lema 1:

La matriz \mathbf{A} definida en (8) tiene las siguientes propiedades:

- (i) \mathbf{A} es idempotente con $\text{rg}(\mathbf{A}) = r$.
- (ii) $\mathbf{X}\mathbf{A}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ es idempotente y simétrica con rango r .
- (iii) $\mathbf{A}(\mathbf{X}^\top\mathbf{X})^{-1} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{A}^\top = \mathbf{A}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{A}^\top$.

Demostración:

Se deja de **tarea** para el lector.



Resultado 2:

Para el modelo lineal

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}).$$

El MLE de $\boldsymbol{\beta}$ bajo las restricciones lineales $\mathbf{G}\boldsymbol{\beta} = \mathbf{g}$, es dado por¹

$$\tilde{\boldsymbol{\beta}} = \mathbf{A}\hat{\boldsymbol{\beta}} + \mathbf{B}\mathbf{g},$$

con distribución

$$\tilde{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top).$$

El MLE restringido de σ^2 es

$$s_r^2 = \frac{1}{n-r} Q(\tilde{\boldsymbol{\beta}}),$$

donde

$$\frac{Q(\tilde{\boldsymbol{\beta}})}{\sigma^2} \sim \chi^2(n-r),$$

y $\tilde{\boldsymbol{\beta}}$ es independiente de $Q(\tilde{\boldsymbol{\beta}})$.

¹También $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top (\mathbf{G}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{G}^\top)^{-1} (\mathbf{g} - \mathbf{G}\hat{\boldsymbol{\beta}})$.

Demostración:

La normalidad sigue desde la linealidad con relación a $\hat{\beta}$. Ahora,

$$E(\tilde{\beta}) = A E(\hat{\beta}) + Bg = (I - BG)\beta + Bg = \beta - B(G\beta - g) = \beta,$$

y

$$\text{Cov}(\tilde{\beta}) = A \text{Cov}(\hat{\beta}) A^T = \sigma^2 A(X^T X)^{-1} A^T.$$

Notando que

$$\begin{aligned}\tilde{\beta} &= A\hat{\beta} + Bg = A(X^T X)^{-1} X^T Y + Bg \\ &= A(X^T X)^{-1} X^T (X\beta + \epsilon) + Bg \\ &= A\beta + A(X^T X)^{-1} X^T \epsilon + Bg \\ &= \beta + A(X^T X)^{-1} X^T \epsilon\end{aligned}$$



De este modo, podemos escribir

$$\begin{aligned} Y - X\tilde{\beta} &= Y - X\beta - XA(X^\top X)^{-1}X^\top \epsilon \\ &= \epsilon - XA(X^\top X)^{-1}X^\top \epsilon \\ &= (I - XA(X^\top X)^{-1}X^\top)\epsilon, \end{aligned}$$

por la parte (ii) del [Lema 1](#), sigue que

$$\frac{Q(\tilde{\beta})}{\sigma^2} = \frac{\epsilon^\top (I - XA(X^\top X)^{-1}X^\top)\epsilon}{\sigma^2} \sim \chi^2(n-r).$$

Para notar la independencia entre $\tilde{\beta}$ y $Q(\tilde{\beta})$ debemos tener²

$$A(X^\top X)^{-1}X^\top (I - XA(X^\top X)^{-1}X^\top) = 0$$

²lo que es consecuencia de escribir $\tilde{\beta}$ y $Q(\tilde{\beta})$ en términos de ϵ .

En efecto,

$$\begin{aligned} A(X^\top X)^{-1}X^\top(I - XA(X^\top X)^{-1}X^\top) \\ = A(X^\top X)^{-1} - A(X^\top X)^{-1}X^\top XA(X^\top X)^{-1}X^\top. \end{aligned}$$

Notando que $A(X^\top X)^{-1} = (X^\top X)^{-1}A^\top$, obtenemos

$$\begin{aligned} A(X^\top X)^{-1}X^\top XA(X^\top X)^{-1}X^\top &= A(X^\top X)^{-1}X^\top X(X^\top X)^{-1}A^\top X^\top \\ &= A(X^\top X)^{-1}A^\top X^\top = A^2(X^\top X)^{-1}X^\top = A(X^\top X)^{-1}X^\top, \end{aligned}$$

lo que concluye la demostración.

