IECD-325: Modelos lineales

Felipe Osorio

felipe.osorio@uv.cl

Definición 1 (modelo lineal):

Considere la variable aleatoria Y, decimos que sigue un modelo lineal si

$$\mathsf{E}(Y) = \sum_{i=1}^{p} x_j \beta_j = \boldsymbol{x}^\top \boldsymbol{\beta},$$

donde $\boldsymbol{x} = (x_1, \dots, x_p)^\top$ representa un vector de p variables regresoras, mientras que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ denota un vector de parámetros desconocidos.

Definición 1 (modelo lineal):

Considere la variable aleatoria Y, decimos que sigue un modelo lineal si

$$\mathsf{E}(Y) = \sum_{i=1}^{p} x_j \beta_j = \boldsymbol{x}^\top \boldsymbol{\beta},$$

donde $\boldsymbol{x} = (x_1, \dots, x_p)^\top$ representa un vector de p variables regresoras, mientras que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ denota un vector de parámetros desconocidos.

Observación:

Cuando

$$\mathrm{Cov}(\boldsymbol{Y}) = \sum_{r=1}^{K} \phi_r \boldsymbol{\Sigma}_r = \boldsymbol{\Sigma}(\boldsymbol{\phi}),$$

donde $\boldsymbol{\phi} = (\phi_1, \dots, \phi_K)^\top$ son parámetros desconocidos y $\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K$ son matrices (simétricas) conocidas, decimos que $\boldsymbol{Y}$ sigue un modelo lineal general.

Típicamente, asumiremos que $Y_1, \dots, Y_n$ son independientes con varianza constante, en cuyo caso

$$\mathrm{Cov}(\boldsymbol{Y}) = \sigma^2 \boldsymbol{I},$$

y decimos que $\boldsymbol{Y}$ sigue un modelo lineal simple.

Definición 2:

El vector $\boldsymbol{Y}$ se dice seguir un modelo lineal si

$$\mathsf{E}(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}, \qquad \mathsf{Cov}(\boldsymbol{Y}) = \sum_{r=1}^{K} \phi_r \boldsymbol{\Sigma}_r,$$

o equivalentemente

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

con

$$\mathsf{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}, \qquad \mathsf{Cov}(\boldsymbol{\epsilon}) = \sum_{r=1}^{K} \phi_r \boldsymbol{\Sigma}_r.$$

Observación:

Estos supuestos de momentos, suelen ser llamados condiciones de Gauss-Markov. Es usual que sean expresados en términos del modelo lineal simple, como:

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

con

$$\mathsf{E}(\boldsymbol{\epsilon}) = \mathbf{0}, \qquad \mathsf{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n.$$

Definición 3:

Se dice que el vector $\boldsymbol{Y}$ sigue un modelo lineal normal si $\boldsymbol{Y} \sim \mathsf{N}_n(\boldsymbol{X\beta}, \boldsymbol{\Sigma}(\boldsymbol{\phi}))$, o bien

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathsf{N}_n(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{\phi})).$$

Mientras que sigue un modelo normal simple, si

$$\boldsymbol{Y} \sim \mathsf{N}_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}). \tag{1}$$

Observación:

Se debe destacar que el modelo en (1) puede ser escrito como

$$Y_i \overset{\text{ind}}{\sim} \mathsf{N}_1(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2), \qquad i = 1, \dots, n. \tag{2}$$

En efecto, la inferencia estadística para los modelos definidos en Ecuaciones (1) y (2) es equivalente.[1]

[1] Aunque esto NO es verdad en general.

Supuestos (Modelo lineal normal simple):

El modelo lineal descrito por la ecuación:

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

tiene los siguientes supuestos:

A1: $\boldsymbol{X}$ es una matriz no aleatoria $n \times p$ con $n > p$.

A2: La matriz $\boldsymbol{X}$ tiene rango p, es decir, $\boldsymbol{X}$ es rango columna completo.

A3: El vector aleatorio n-dimensional $\boldsymbol{Y}$ tiene elementos que son observables.

A4: El vector aleatorio no observable $\boldsymbol{\epsilon}$ satisface

$$\mathsf{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}, \qquad \mathsf{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}, \quad \sigma^2 > 0.$$

A4*: El vector aleatorio $\boldsymbol{\epsilon}$ satisface $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, o equivalentemente

$$\boldsymbol{Y} \sim \mathsf{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n).$$

Considere la función de verosimilitud proveniente del modelo $\boldsymbol{Y} \sim \mathsf{N}_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}_n)$, dada por

$$L(\boldsymbol{\theta}) = (2\pi\sigma^2)^{-n/2} \exp\Big(-\frac{1}{2\sigma^2} \|\boldsymbol{Y} - \boldsymbol{X\beta}\|^2 \Big),$$

con $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$. De este modo, podemos llevar a cabo la estimación ML, considerando

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \|\boldsymbol{Y} - \boldsymbol{X\beta}\|^2 \\ &= -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} Q(\boldsymbol{\beta}), \end{aligned}$$

donde

$$Q(\boldsymbol{\beta}) = \|\boldsymbol{Y} - \boldsymbol{X\beta}\|^2 = \sum_{i=1}^n (Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2,$$

se denomina suma de cuadrados del error.

Diferenciando con relación a $\boldsymbol{\beta}$ y σ^2, obtenemos

$$\mathsf{d}_\beta\,\ell(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2}\,\mathsf{d}_\beta\,\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\|^2 = \frac{1}{\sigma^2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{X}\,\mathsf{d}\boldsymbol{\beta},$$

y

$$\mathsf{d}_{\sigma^2}\,\ell(\boldsymbol{\theta}) = -\frac{n}{2\sigma^2}\,\mathsf{d}\sigma^2 + \frac{1}{2\sigma^4}\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\|^2\,\mathsf{d}\sigma^2.$$

Es decir,[2]

$$\begin{aligned}\frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} &= \frac{1}{\sigma^2}\boldsymbol{X}^\top(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}),\\ \frac{\partial\ell(\boldsymbol{\theta})}{\partial\sigma^2} &= -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\|^2.\end{aligned}$$

Desde la condición de primer orden, tenemos las ecuaciones de verosimilitud:

$$\begin{aligned}\boldsymbol{X}^\top(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}) &= \boldsymbol{0}\\ n\widehat{\sigma}^2 - \|\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 &= 0.\end{aligned}$$

[2] usando el 1er Teorema de identificación dado en Magnus y Neudecker, 2007, pag. 98

Resolviendo con relación a $\boldsymbol{\beta}$ obtenemos las ecuaciones normales

$$\boldsymbol{X}^\top \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{Y},$$

dado que $\operatorname{rg}(\boldsymbol{X}) = \operatorname{rg}(\boldsymbol{X}^\top \boldsymbol{X}) = p$, entonces el sistema anterior admite solución única dada por:[3]

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y},$$
$$\widehat{\sigma}^2 = \frac{1}{n} Q(\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2.$$

Se define el vector de valores predichos como

$$\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y} = \boldsymbol{H}\boldsymbol{Y},$$

donde $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$.

[3] Note que $\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^+\boldsymbol{Y}$ con $\boldsymbol{X}^+ = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ la inversa Moore-Penrose de $\boldsymbol{X}$.

Inferencia en el modelo lineal

Es fácil notar que $\boldsymbol{H}$ es simétrica e idempotente, en cuyo caso

$$\operatorname{rg}(\boldsymbol{H}) = \operatorname{tr}(\boldsymbol{H}) = \operatorname{tr}(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top) = \operatorname{tr}((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}) = p.$$

Además, el vector de diferencias entre $\boldsymbol{Y}$ y $\widehat{\boldsymbol{Y}}$ se denomina el vector de residuos, es decir

$$\boldsymbol{e} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}.$$

Además, tenemos que $\boldsymbol{I} - \boldsymbol{H}$ también es simétrica e idempotente. Con esta notación podemos escribir

$$\begin{aligned} Q(\widehat{\boldsymbol{\beta}}) &= \|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 = \boldsymbol{e}^\top\boldsymbol{e} = \boldsymbol{Y}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} \\ &= \sum_{i=1}^n (Y_i - \boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}})^2, \end{aligned}$$

que es conocido como suma de cuadrados residual.

Resultado 1:

Considere el modelo:

$$\boldsymbol{Y} \sim \mathsf{N}_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}_n),$$

con los supuestos A1 a A4*. Entonces, tenemos que:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y} \sim \mathsf{N}_p(\boldsymbol{\beta}, \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}), \qquad (3)$$

de donde sigue que $\widehat{\boldsymbol{\beta}}$ es insesgado y $\mathsf{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$. Además, la variable aleatoria

$$\frac{Q(\widehat{\boldsymbol{\beta}})}{\sigma^2} = \frac{\boldsymbol{Y}^\top (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{Y}}{\sigma^2} \sim \chi^2(n-p).$$

Finalmente $\widehat{\boldsymbol{\beta}}$ y $s^2 = Q(\widehat{\boldsymbol{\beta}})/(n-p)$ son independientes.

Demostración:

Notando que $\widehat{\boldsymbol{\beta}}$ es una transformación lineal de un vector aleatorio normal y

$$\mathsf{E}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathsf{E}(\boldsymbol{Y}) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta},$$
$$\mathsf{Cov}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\,\mathsf{Cov}(\boldsymbol{Y})\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1},$$

y el resultado en (3) sigue. Por otro lado,

$$\frac{Q(\widehat{\boldsymbol{\beta}})}{\sigma^2} = \frac{(n-p)s^2}{\sigma^2} = \frac{\boldsymbol{Y}^\top(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}}{\sigma^2},$$

sigue una distribución chi-cuadrado pues $\boldsymbol{I}-\boldsymbol{H}$ es matriz idempotente con

$$\operatorname{rg}(\boldsymbol{I}-\boldsymbol{H}) = \operatorname{tr}(\boldsymbol{I}-\boldsymbol{H}) = \operatorname{tr}(\boldsymbol{I}) - \operatorname{tr}(\boldsymbol{H}) = n-p.$$

Además, el parámetro de no centralidad es dado por

$$\lambda = \frac{1}{2\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{X}\boldsymbol{\beta} = 0.$$

En efecto,

$$\boldsymbol{H}\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X} = \boldsymbol{X} \quad \Rightarrow \quad (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}.$$

La independencia entre $\widehat{\boldsymbol{\beta}}$ y s^2 sigue desde $(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$

Observación:

Es fácil notar que

$$\begin{aligned} \mathsf{E}\{Q(\widehat{\boldsymbol{\beta}})\} &= \mathsf{E}\{\boldsymbol{Y}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}\} = \sigma^2 \operatorname{tr}(\boldsymbol{I} - \boldsymbol{H}) + \boldsymbol{\beta}^\top\boldsymbol{X}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X}\boldsymbol{\beta} \\ &= \sigma^2(n-p), \end{aligned}$$

es decir,

$$\mathsf{E}(\widehat{\sigma}^2) = \Big(\frac{n-p}{n}\Big)\sigma^2,$$

lo que permite sugerir el estimador insesgado:

$$s^2 = \frac{1}{n-p}\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2.$$

Resultado 2:

El vector de valores predichos y el vector de residuos son independientemente distribuídos como

$$\widehat{\boldsymbol{Y}} \sim \mathsf{N}_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{H}), \qquad \boldsymbol{e} \sim \mathsf{N}_n(\boldsymbol{0}, \sigma^2(\boldsymbol{I} - \boldsymbol{H})).$$

Demostración:

Considere

$$\begin{pmatrix} \widehat{\boldsymbol{Y}} \\ \boldsymbol{e} \end{pmatrix} = \begin{pmatrix} \boldsymbol{H} \\ \boldsymbol{I} - \boldsymbol{H} \end{pmatrix} \boldsymbol{Y},$$

luego,

$$\mathsf{E}\begin{pmatrix} \widehat{\boldsymbol{Y}} \\ \boldsymbol{e} \end{pmatrix} = \begin{pmatrix} \boldsymbol{H} \\ \boldsymbol{I} - \boldsymbol{H} \end{pmatrix} \boldsymbol{X\beta} = \begin{pmatrix} \boldsymbol{HX\beta} \\ \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X\beta} \\ \boldsymbol{0} \end{pmatrix}.$$

Además,

$$\begin{aligned}\operatorname{Cov}\begin{pmatrix}\widehat{\boldsymbol{Y}}\\ \boldsymbol{e}\end{pmatrix} &= \sigma^2\begin{pmatrix}\boldsymbol{H}\\ \boldsymbol{I}-\boldsymbol{H}\end{pmatrix}\left(\boldsymbol{H}^\top,(\boldsymbol{I}-\boldsymbol{H})^\top\right)\\ &= \sigma^2\begin{pmatrix}\boldsymbol{H}\boldsymbol{H}^\top & \boldsymbol{H}(\boldsymbol{I}-\boldsymbol{H})^\top\\ (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{H}^\top & (\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{I}-\boldsymbol{H})^\top\end{pmatrix}\\ &= \sigma^2\begin{pmatrix}\boldsymbol{H}^2 & \boldsymbol{0}\\ \boldsymbol{0} & (\boldsymbol{I}-\boldsymbol{H})^2\end{pmatrix}\\ &= \sigma^2\begin{pmatrix}\boldsymbol{H} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}-\boldsymbol{H}\end{pmatrix},\end{aligned}$$

lo que permite establecer el resultado.

Observación:

Debemos resaltar que $\boldsymbol{H}$ y $\boldsymbol{I}-\boldsymbol{H}$ son matrices de rango incompleto y por tanto $\widehat{\boldsymbol{Y}}$ y $\boldsymbol{e}$ siguen distribuciones normales singulares.

Estimación mínimos cuadrados

Considere el siguiente problema

$$\min_{\beta} Q_n(\boldsymbol{\beta}) = \min_{\beta} \|\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta}\|^2,$$

cuya solución, $\widehat{\boldsymbol{\beta}}_n$ es llamado estimador mínimos cuadrados ordinarios (OLS), dado por:

$$\widehat{\boldsymbol{\beta}}_n = (\boldsymbol{X}_n^\top \boldsymbol{X}_n)^{-1}\boldsymbol{X}_n^\top \boldsymbol{Y}_n,$$

Es habitual asumir el modelo

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

con $\mathsf{E}(\boldsymbol{\epsilon}_n) = \mathbf{0}$ y $\mathsf{Cov}(\boldsymbol{\epsilon}_n) = \sigma^2\boldsymbol{I}_n$. Esto lleva a

$$\begin{aligned}
\mathsf{E}(\widehat{\boldsymbol{\beta}}_n) &= (\boldsymbol{X}_n^\top \boldsymbol{X}_n)^{-1}\boldsymbol{X}_n^\top \mathsf{E}(\boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{\epsilon}_n) = \boldsymbol{\beta} \\
\mathsf{Cov}(\widehat{\boldsymbol{\beta}}_n) &= \mathsf{Cov}((\boldsymbol{X}_n^\top \boldsymbol{X}_n)^{-1}\boldsymbol{X}_n^\top(\boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{\epsilon}_n)) = \sigma^2(\boldsymbol{X}_n^\top \boldsymbol{X}_n)^{-1}.
\end{aligned}$$

Supuestos (Estimación mínimos cuadrados):

Considere el modelo,

$$\boldsymbol{Y}_n = \boldsymbol{X}_n \boldsymbol{\beta} + \boldsymbol{\epsilon}_n, \tag{4}$$

y suponga las condiciones:

B1: $\mathrm{E}(\boldsymbol{\epsilon}_n) = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{\epsilon}_n) = \sigma^2 \boldsymbol{I}$.

B2: Sea $h_{kk} = \boldsymbol{x}_{k,n}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_{k,n}$ el k-ésimo elemento de la diagonal de $\boldsymbol{H}_n$ y considere

$$\max_{1 \leq k \leq n} h_{kk} \to 0, \quad \text{conforme } n \to \infty.$$

B3: $\lim_{n \to \infty} \frac{1}{n} \boldsymbol{X}_n^\top \boldsymbol{X}_n = \boldsymbol{K}$ es una matriz no singular (no estocástica).

Usando el supuesto B3, tenemos

$$\lim_{n\to\infty} \operatorname{Cov}(\widehat{\boldsymbol{\beta}}_n) = \sigma^2 \lim_{n\to\infty} \frac{1}{n}\Big(\frac{\boldsymbol{X}_n^\top \boldsymbol{X}_n}{n}\Big)^{-1} = \sigma^2 \lim_{n\to\infty} \frac{1}{n}\boldsymbol{K}^{-1} = \mathbf{0}.$$

Esto implica que $\widehat{\boldsymbol{\beta}}_n \xrightarrow{\text{2nd}} \boldsymbol{\beta}$. Es decir, $\widehat{\boldsymbol{\beta}}_n$ es un estimador consistente de $\boldsymbol{\beta}$.

Resultado 3 (Distribución asintótica del estimador LS):

Considere el modelo (4) bajo los supuestos B1 a B3. Entonces

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\text{D}} \mathsf{N}_p(\mathbf{0}, \sigma^2 \boldsymbol{K}^{-1}).$$

Demostración:

Ver Sen y Singer (1993),[4] pág. 279.

[4] Large Sample Methods in Statistics. Chapman & Hall, London.

Resultado 4 (Teorema de Gauss-Markov):

Suponga $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ y sea $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}$ el estimador mínimos cuadrados. Asuma el supuesto B1. El estimador $\widehat{\gamma} = \boldsymbol{h}^\top \widehat{\boldsymbol{\beta}}$ de $\gamma = \boldsymbol{h}^\top \boldsymbol{\beta}$ es el mejor estimador lineal e insesgado (BLUE).

Demostración:

Sea $\boldsymbol{c}^\top \boldsymbol{Y}$ cualquier otro estimador lineal e insesgado de $\gamma = \boldsymbol{h}^\top \boldsymbol{\beta}$. Dado que $\boldsymbol{c}^\top \boldsymbol{Y}$ es insesgado, sigue que

$$\boldsymbol{h}^\top \boldsymbol{\beta} = \mathsf{E}(\boldsymbol{c}^\top \boldsymbol{Y}) = \boldsymbol{c}^\top \mathsf{E}(\boldsymbol{Y}) = \boldsymbol{c}^\top \boldsymbol{X}\boldsymbol{\beta}, \qquad \forall\, \boldsymbol{\beta},$$

luego, tenemos

$$\boldsymbol{c}^\top \boldsymbol{X} = \boldsymbol{h}^\top.$$

Ahora,

$$\operatorname{var}(\boldsymbol{c}^\top \boldsymbol{Y}) = \boldsymbol{c}^\top \operatorname{Cov}(\boldsymbol{Y})\boldsymbol{c} = \sigma^2 \boldsymbol{c}^\top \boldsymbol{c}, \tag{5}$$

mientras que

$$\operatorname{var}(\boldsymbol{h}^\top \widehat{\boldsymbol{\beta}}) = \boldsymbol{h}^\top \operatorname{Cov}(\widehat{\boldsymbol{\beta}})\boldsymbol{h} = \sigma^2 \boldsymbol{h}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{h} = \sigma^2 \boldsymbol{c}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{c}, \tag{6}$$

desde (5) y (6), tenemos

$$\begin{aligned} \operatorname{var}(\boldsymbol{c}^\top \boldsymbol{Y}) - \operatorname{var}(\boldsymbol{h}^\top \widehat{\boldsymbol{\beta}}) &= \sigma^2(\boldsymbol{c}^\top \boldsymbol{c} - \boldsymbol{c}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{c}) \\ & \sigma^2 \boldsymbol{c}^\top (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X})\boldsymbol{c} \geq 0, \end{aligned}$$

el resultado sigue pues $\boldsymbol{I} - \boldsymbol{H}$ es semidefinida positiva.

Estimación de parámetros en regresión

Ejemplo:

Considere el modelo:

$$Y_{ij} = \mu_i + \epsilon_{ij}, \qquad i = 1, \dots, p; j = 1, \dots, n.$$

El modelo puede ser escrito en la forma

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

donde

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_p \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1} \end{pmatrix},$$

con $\boldsymbol{\beta} = (\mu_1, \dots, \mu_p)^\top$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^\top, \dots, \boldsymbol{\epsilon}_p^\top)^\top$ y $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in})^\top$.

Note que

$$\boldsymbol{X}^\top \boldsymbol{X} = \begin{pmatrix} \mathbf{1}^\top & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}^\top & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}^\top \end{pmatrix} \begin{pmatrix} \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{1}^\top \mathbf{1} & 0 & \dots & 0 \\ 0 & \mathbf{1}^\top \mathbf{1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{1}^\top \mathbf{1} \end{pmatrix} = n\boldsymbol{I}_p$$

Además,

$$\boldsymbol{X}^\top \boldsymbol{Y} = \begin{pmatrix} \mathbf{1}^\top & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}^\top & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}^\top \end{pmatrix} \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_p \end{pmatrix} = \begin{pmatrix} \mathbf{1}^\top \boldsymbol{Y}_1 \\ \mathbf{1}^\top \boldsymbol{Y}_2 \\ \vdots \\ \mathbf{1}^\top \boldsymbol{Y}_p \end{pmatrix} = \begin{pmatrix} n\overline{Y}_1 \\ n\overline{Y}_2 \\ \vdots \\ n\overline{Y}_p \end{pmatrix},$$

con $\overline{Y}_i = \frac{1}{n} \sum_{j=1}^{n} Y_{ij}$, $i = 1, \dots, p$.

De ahí que $\widehat{\boldsymbol{\beta}} = (\overline{Y}_1, \dots, \overline{Y}_p)^\top$. Note que

$$Q(\widehat{\boldsymbol{\beta}}) = \boldsymbol{Y}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} = \boldsymbol{Y}^\top\boldsymbol{Y} - \boldsymbol{Y}^\top\boldsymbol{X}\widehat{\boldsymbol{\beta}}.$$

En nuestro caso,

$$\boldsymbol{Y}^\top\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \dots, \boldsymbol{Y}_p^\top)\begin{pmatrix}\boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_p\end{pmatrix} = \boldsymbol{Y}_1^\top\boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_p^\top\boldsymbol{Y}_p$$

y

$$\boldsymbol{Y}^\top\boldsymbol{X}\widehat{\boldsymbol{\beta}} = (n\overline{Y}_1, \dots, n\overline{Y}_p)\begin{pmatrix}\overline{Y}_1 \\ \vdots \\ \overline{Y}_p\end{pmatrix} = n\overline{Y}_1^2 + \cdots + n\overline{Y}_p^2.$$

Es decir,

$$\begin{aligned}
Q(\widehat{\boldsymbol{\beta}}) &= \boldsymbol{Y}^\top\boldsymbol{Y} - \boldsymbol{Y}^\top\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{Y}_1^\top\boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_p^\top\boldsymbol{Y}_p - n\overline{Y}_1^2 - \cdots - n\overline{Y}_p^2 \\
&= (\boldsymbol{Y}_1^\top\boldsymbol{Y}_1 - n\overline{Y}_1^2) + \cdots + (\boldsymbol{Y}_p^\top\boldsymbol{Y}_p - n\overline{Y}_p^2) \\
&= \Big(\sum_{j=1}^n Y_{1j}^2 - n\overline{Y}_1^2\Big) + \cdots + \Big(\sum_{j=1}^n Y_{pj}^2 - n\overline{Y}_p^2\Big) \\
&= \sum_{j=1}^n (Y_{1j} - \overline{Y}_1)^2 + \cdots + \sum_{j=1}^n (Y_{pj} - \overline{Y}_p)^2 \\
&= \sum_{i=1}^p \sum_{j=1}^n (Y_{ij} - \overline{Y}_i)^2.
\end{aligned}$$

Finalmente, el estimador insesgado de σ^2 adopta la forma:

$$s^2 = \frac{1}{np - p} \sum_{i=1}^p \sum_{j=1}^n (Y_{ij} - \overline{Y}_i)^2.$$