Estimación por Mínimos Cuadrados Ordinarios (OLS)

Modelo

Consideramos el modelo de regresión lineal clásico:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

donde:

- $\mathbf{y}$ es un vector columna $n \times 1$ de observaciones de la variable dependiente.
- $\mathbf{X}$ es una matriz $n \times k$ de variables explicativas (de rango completo).
- $\boldsymbol{\beta}$ es un vector $k \times 1$ de parámetros desconocidos.
- $\boldsymbol{\varepsilon}$ es un vector de errores aleatorios $n \times 1$.

El estimador de mínimos cuadrados ordinarios (OLS) se obtiene minimizando la suma de cuadrados de los residuos:

$$\min_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Expandimos la expresión cuadrática:

$$S(\boldsymbol{\beta}) = \mathbf{y}^\top \mathbf{y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}$$

Tomamos derivada con respecto a $\boldsymbol{\beta}$, igualamos a cero:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

Resolviendo la ecuación anterior:

$$\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = \mathbf{X}^\top \mathbf{y} \quad \Rightarrow \quad \boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}}$$

La matriz $\mathbf{X}^\top \mathbf{X}$ es invertible si y solo si las columnas de $\mathbf{X}$ son linealmente independientes, es decir, $\operatorname{rank}(\mathbf{X}) = k$.

Inversa generalizada de Moore–Penrose

Cuando $\mathbf{X}^\top \mathbf{X}$ no es invertible (por ejemplo, si $\mathbf{X}$ no tiene rango completo), se usa la inversa de Moore–Penrose:

$$\hat{\boldsymbol{\beta}} = \mathbf{X}^+ \mathbf{y}$$

donde $\mathbf{X}^+ \in \mathbb{R}^{k \times n}$ es la inversa de Moore–Penrose de $\mathbf{X}$, y cumple las siguientes condiciones:

$$\begin{aligned}
&\text{(i)} && \mathbf{X}\mathbf{X}^+\mathbf{X} = \mathbf{X} \\
&\text{(ii)} && \mathbf{X}^+\mathbf{X}\mathbf{X}^+ = \mathbf{X}^+ \\
&\text{(iii)} && (\mathbf{X}\mathbf{X}^+)^\top = \mathbf{X}\mathbf{X}^+ \\
&\text{(iv)} && (\mathbf{X}^+\mathbf{X})^\top = \mathbf{X}^+\mathbf{X}
\end{aligned}$$

Esta solución es la de norma mínima:

$$\hat{\boldsymbol{\beta}} = \operatorname{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^k} \{ \|\boldsymbol{\beta}\| \ : \ \mathbf{X}\boldsymbol{\beta} = \mathbf{y} \}$$

Ejercicios

1. Sea $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$ y $\mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$. Calcule $\hat{\boldsymbol{\beta}}$ usando la inversa de Moore–Penrose.
2. Sea $\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$ y $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Calcule la estimación $\hat{\boldsymbol{\beta}}$ por mínimos cuadrados ordinarios.
3. Verifique que la matriz $\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ no tiene inversa de $\mathbf{X}^\top\mathbf{X}$ y proponga cómo estimar $\boldsymbol{\beta}$.
4. Demuestre que si $\mathbf{X}$ tiene rango completo, entonces la solución obtenida por $\mathbf{X}^+$ coincide con la fórmula usual de OLS.
5. Sea $\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ y $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. ¿Cuál es la estimación de $\boldsymbol{\beta}$ por Moore–Penrose?

Teorema de Gauss-Markov

Sea el modelo de regresión lineal clásico:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

donde $\mathbf{y}$ es un vector $n \times 1$ de observaciones, $\mathbf{X}$ es una matriz de variables explicativas de dimensión $n \times k$ con $\operatorname{rank}(\mathbf{X}) = k$, $\boldsymbol{\beta}$ es un vector de parámetros $k \times 1$ y $\boldsymbol{\varepsilon}$ es un vector de errores aleatorios.

Supóngase que se cumplen los siguientes supuestos:

1. **Linealidad en los parámetros:** el modelo está correctamente especificado como $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
2. **Exogeneidad:** $\mathbb{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$.
3. **Homocedasticidad:** $\operatorname{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\mathbf{I}_n$.
4. **No multicolinealidad perfecta:** las columnas de $\mathbf{X}$ son linealmente independientes.

Entonces, el estimador de Mínimos Cuadrados Ordinarios (OLS):

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$$

es el **mejor estimador lineal insesgado** (BLUE), es decir, tiene la menor varianza entre todos los estimadores lineales e insesgados de $\boldsymbol{\beta}$.

Demostración

Sea $\tilde{\boldsymbol{\beta}}$ otro estimador lineal e insesgado de $\boldsymbol{\beta}$ tal que:

$$\tilde{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y}, \quad \text{con } \mathbf{C} \in \mathbb{R}^{k\times n}$$

y tal que $\mathbb{E}[\tilde{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ para todo $\boldsymbol{\beta}$.

Como $\tilde{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y} = \mathbf{C}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{C}\mathbf{X}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\varepsilon}$, entonces:

$$\mathbb{E}[\tilde{\boldsymbol{\beta}}] = \mathbf{C}\mathbf{X}\boldsymbol{\beta} \quad \Rightarrow \quad \text{Para que sea insesgado: } \mathbf{C}\mathbf{X} = \mathbf{I}_k$$

La varianza condicional de $\tilde{\boldsymbol{\beta}}$ es:

$$\operatorname{Var}[\tilde{\boldsymbol{\beta}}|\mathbf{X}] = \operatorname{Var}[\mathbf{C}\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{C}\operatorname{Var}[\boldsymbol{\varepsilon}|\mathbf{X}]\mathbf{C}^\top = \sigma^2\mathbf{C}\mathbf{C}^\top$$

Recordemos que el estimador OLS es:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$$

Y su varianza condicional es:

$$\text{Var}[\hat{\boldsymbol{\beta}}|\mathbf{X}] = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$$

Sea $\mathbf{d} = \tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}$, entonces:

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \mathbf{d}$$

Como ambos estimadores son insesgados:

$$\text{Var}[\tilde{\boldsymbol{\beta}}] = \text{Var}[\hat{\boldsymbol{\beta}} + \mathbf{d}] = \text{Var}[\hat{\boldsymbol{\beta}}] + \text{Var}[\mathbf{d}] + 2 \cdot \text{Cov}[\hat{\boldsymbol{\beta}}, \mathbf{d}]$$

Pero $\text{Cov}[\hat{\boldsymbol{\beta}}, \mathbf{d}] = 0$ porque $\hat{\boldsymbol{\beta}}$ es el proyector ortogonal de $\mathbf{y}$ sobre el espacio columna de $\mathbf{X}$, y $\mathbf{d}$ es ortogonal a ese espacio.

Por lo tanto:

$$\text{Var}[\tilde{\boldsymbol{\beta}}] = \text{Var}[\hat{\boldsymbol{\beta}}] + \text{Var}[\mathbf{d}] \quad \Rightarrow \quad \text{Var}[\tilde{\boldsymbol{\beta}}] \geq \text{Var}[\hat{\boldsymbol{\beta}}]$$

La igualdad se cumple solo si $\text{Var}[\mathbf{d}] = \mathbf{0}$, es decir, si $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$.

Supuestos clásicos del modelo lineal (CLM)

Para que la estimación por mínimos cuadrados ordinarios (OLS) posea propiedades deseables como insesgadez, eficiencia y consistencia, se requieren los siguientes supuestos fundamentales:

- (A1) **Generación del modelo**: el modelo de datos es correctamente especificado como $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
- (A2) **Condición de exogeneidad**: $\mathbb{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$.
- (A3) **Varianza constante (homocedasticidad)**: $\text{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\mathbf{I}_T$.
- (A4) **Rango completo**: la matriz $\mathbf{X}$ tiene rango completo, es decir, $\text{rank}(\mathbf{X}) = k$ con $T \geq k$.

¿Qué ocurre si el supuesto (A1) no se cumple?

En esta sección se analiza el impacto de errores de especificación estructural, particularmente cuando se omiten o se incluyen variables de forma incorrecta dentro del modelo lineal. Nos enfocamos en la situación de omisión de regresores relevantes.

Error de especificación: variables omitidas

Supongamos que el modelo verdadero está dado por:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

donde $\mathbf{X}_1$ y $\mathbf{X}_2$ representan dos conjuntos distintos de regresores. Sin embargo, se estima un modelo restringido que omite $\mathbf{X}_2$:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$$

a lo que se le denomina la *regresión corta*. Esta especificación da lugar a varias consecuencias importantes:

(1) Insesgadez

El estimador OLS para $\boldsymbol{\beta}_1$ en la regresión corta es:

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1^\top\mathbf{X}_1)^{-1}\mathbf{X}_1^\top\mathbf{y}$$

Tomando esperanza condicional a $\mathbf{X}$, se tiene:

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_1|\mathbf{X}] = \boldsymbol{\beta}_1 + (\mathbf{X}_1^\top\mathbf{X}_1)^{-1}\mathbf{X}_1^\top\mathbf{X}_2\boldsymbol{\beta}_2$$

Este término adicional implica que, salvo que $\mathbf{X}_1^\top\mathbf{X}_2 = 0$, el estimador es sesgado. Este sesgo puede ser considerable, e incluso alterar el signo estimado de un parámetro, como ocurre frecuentemente en ecuaciones de demanda donde el precio se subestima.

(2) Varianza

Curiosamente, la varianza de $\hat{\boldsymbol{\beta}}_1$ bajo la regresión corta es menor que la varianza del estimador cuando se incluye $\mathbf{X}_2$. Es decir:

$$\text{Var}[\hat{\boldsymbol{\beta}}_1|\mathbf{X}] \leq \text{Var}[\hat{\boldsymbol{\beta}}_{1.2}|\mathbf{X}]$$

donde $\hat{\boldsymbol{\beta}}_{1.2}$ denota el estimador de $\boldsymbol{\beta}_1$ en la regresión completa. Esta desigualdad se justifica mediante el uso del proyector de residuos $\mathbf{M}$ y refleja que al omitir $\mathbf{X}_2$, implícitamente estamos imponiendo la restricción $\boldsymbol{\beta}_2 = 0$, lo cual equivale a incorporar información adicional, aunque ésta sea errónea.

(3) Precisión y MSE

A pesar del sesgo, es posible que el estimador en la regresión corta sea más *preciso* si se considera el error cuadrático medio (MSE):

$$\text{MSE}(\hat{\boldsymbol{\beta}}_1) = \text{Var}[\hat{\boldsymbol{\beta}}_1] + \Big(\text{Bias}[\hat{\boldsymbol{\beta}}_1]\Big)^2$$

El sesgo introduce error sistemático, pero la menor varianza puede compensarlo. Si el sesgo es pequeño, se puede preferir el modelo más simple.

Ortogonalidad

Si se cumple que $\mathbf{X}_1^\top \mathbf{X}_2 = 0$, entonces el sesgo desaparece:

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_1] = \boldsymbol{\beta}_1$$

Esto ocurre cuando $\mathbf{X}_2$ no aporta información adicional sobre $\mathbf{y}$ una vez considerada $\mathbf{X}_1$, y por tanto, su omisión no genera sesgo. En este caso, los estimadores $\hat{\boldsymbol{\beta}}_1$ y $\hat{\boldsymbol{\beta}}_{1.2}$ coinciden.

Ejemplo: demanda de gasolina y variables omitidas

Considérese el siguiente modelo de regresión lineal que describe la demanda de gasolina (G) como función del precio de la gasolina (PG) y el ingreso (Y):

$$G = \beta_1 PG + \beta_2 Y + \varepsilon$$

Este modelo es típicamente estimado con datos de series de tiempo. En este contexto, se espera que:

- $\beta_1 < 0$: un mayor precio disminuye la demanda.
- $\beta_2 > 0$: mayores ingresos aumentan la demanda.
- Existe correlación positiva entre PG e Y: $\text{Cov}(PG, Y) > 0$.

¿Qué sucede si se omite el ingreso (Y)?

Si estimamos un modelo que excluye la variable relevante Y, es decir:

$$G = \beta_1 PG + \varepsilon$$

entonces el estimador de β_1 resultará sesgado. En particular, se tiene:

$$\mathbb{E}[\hat{\beta}_1|PG] = \beta_1 + \beta_2 \frac{\text{Cov}(PG, Y)}{\text{Var}(PG)}$$

Dado que $\beta_2 > 0$ y $\text{Cov}(PG, Y) > 0$, el sesgo será positivo. Esto implica que el estimador de β_1 subestimará en magnitud su valor negativo verdadero, pudiendo incluso cambiarle el signo, lo que contradice el principio económico de pendiente negativa en la ecuación de demanda.

Evidencia empírica: regresión múltiple correcta

Se reporta a continuación una estimación correcta de la ecuación de demanda con las variables relevantes:

- Observaciones: 36
- $\bar{G} = 226.09$, desviación estándar = 50.59
- $R^2 = 0.9836$, error estándar de los residuos = 6.68

Resultados:

Variable	Coeficiente	Error estándar	Valor t
Constante	-79.75	8.67	-9.20
Y	0.0369	0.0013	28.02
PG	-15.12	1.88	-8.04

Estos resultados empíricos apoyan la teoría: al incluir el ingreso, se obtiene un coeficiente de precio negativo significativo, lo que valida el modelo económico subyacente.

Variables irrelevantes: el caso opuesto

Supóngase ahora que el modelo verdadero es:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \varepsilon$$

pero se estima una regresión más extensa:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \varepsilon$$

donde $\mathbf{X}_2$ contiene variables irrelevantes tales que $\boldsymbol{\beta}_2 = 0$.
En este caso:

- **No hay sesgo:** $\mathbb{E}[\hat{\boldsymbol{\beta}}_1] = \boldsymbol{\beta}_1$.
- **Menor eficiencia:** se incrementa la varianza de $\hat{\boldsymbol{\beta}}_1$ en relación con la regresión corta.

Esto ocurre porque se está desperdiciando grados de libertad al estimar parámetros que en realidad son nulos. Sin embargo, si $\mathbf{X}_2$ es ortogonal a $\mathbf{X}_1$, entonces no se pierde eficiencia.

Modelos lineales con términos no lineales

En el contexto de la estimación por mínimos cuadrados ordinarios (OLS), es posible introducir formas funcionales no lineales sin abandonar la linealidad en los parámetros. Entre estas formas se encuentran los términos cuadráticos, cúbicos e interacciones, todos los cuales pueden estimarse mediante OLS siempre que las variables correspondientes se incluyan explícitamente en la matriz de diseño.

Por ejemplo, el siguiente modelo incluye tanto un término cuadrático como una interacción:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_2^2 + \beta_4 x_2 x_3 + \varepsilon$$

Efectos parciales

En este tipo de modelos, los efectos marginales ya no son constantes. El efecto parcial de x_2 sobre y se obtiene derivando la expresión respecto de x_2:

$$\frac{\partial y}{\partial x_2} = \beta_2 + 2\beta_3 x_2 + \beta_4 x_3$$

Esto contrasta con el modelo lineal simple:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

donde el efecto parcial de x_2 es simplemente constante e igual a β_2.

Estimación de varianza del efecto parcial

La estimación del efecto parcial en modelos con términos no lineales depende del punto en el que se evalúe. Por lo tanto, su varianza también es función de los datos. Sea:

$$\hat{\delta} = \hat{\beta}_2 + 2\hat{\beta}_3 x_2 + \hat{\beta}_4 x_3$$

La varianza de $\hat{\delta}$ se obtiene aplicando la fórmula de propagación del error:

$$\text{Var}(\hat{\delta}) = \text{Var}(\hat{\beta}_2) + 4x_2^2 \,\text{Var}(\hat{\beta}_3) + x_3^2 \,\text{Var}(\hat{\beta}_4) + 4x_2 \,\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) + 2x_3 \,\text{Cov}(\hat{\beta}_2, \hat{\beta}_4) + 4x_2 x_3 \,\text{Cov}(\hat{\beta}_3, \hat{\beta}_4)$$

Este resultado implica que tanto el efecto marginal como su error estándar dependen del valor específico de las covariables x_2 y x_3. Por ello, en la práctica, es habitual reportar los efectos evaluados en la media de los datos.

Considere una regresión lineal con logaritmo del ingreso como variable dependiente, e inclusión de términos cuadráticos:

$$\log(Y) = \beta_0 + \beta_1 \text{AGE} + \beta_2 \text{AGE}^2 + \ldots + \varepsilon$$

Los resultados estimados son los siguientes:

- Observaciones: 27,322
- $R^2 = 0.1724$
- Error estándar de los residuos: 0.447

Estimaciones:

Variable	Coef.	Error Est.	t-ratio	Media
AGE	0.06225	0.00213	29.19	43.53
AGE^2	−0.00074	0.0000242	−30.58	2022.99

El efecto marginal del aumento de edad sobre el log-ingreso se calcula como:

$$\frac{\partial \log(Y)}{\partial \text{AGE}} = \hat{\beta}_1 + 2\hat{\beta}_2 \cdot \text{AGE} = 0.06225 - 2(0.00074)(43.53) \approx 0.00018$$

La varianza estimada de este efecto parcial incluye los términos:

$$\text{Var}(\hat{\delta}) = \text{Var}(\hat{\beta}_1) + 4(\text{AGE})^2 \,\text{Var}(\hat{\beta}_2) + 4(\text{AGE}) \,\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$$

Usando los valores estimados:

$$\text{Var}(\hat{\delta}) = 4.54799 \times 10^{-6} + 4(43.5272)^2 (5.87973 \times 10^{-10}) + 4(43.5272)(-5.1285 \times 10^{-8}) \approx 7.476 \times 10^{-8}$$

El error estándar es entonces:

$$\sqrt{7.476 \times 10^{-8}} \approx 0.00027$$

Una forma común de modelar relaciones no aditivas entre variables explicativas es mediante términos de interacción. Considere el siguiente modelo estimado por mínimos cuadrados ordinarios (OLS), donde la variable dependiente es el logaritmo del ingreso:

$$\log(Y) = \beta_0 + \beta_1 \cdot \text{AGE} + \beta_2 \cdot \text{FEMALE} + \beta_3 \cdot (\text{AGE} \times \text{FEMALE}) + \varepsilon$$

Los resultados estimados son los siguientes:

- Observaciones: 27,322
- Error estándar residual: 0.4893
- $R^2 = 0.009$

Coeficientes:

Variable	Coef.	Error Est.	t-ratio	Media
Constante	-1.2259	0.0161	-76.38	–
AGE	0.00227	0.00036	6.24	43.53
FEMALE	0.21239	0.02363	8.99	0.48
AGE_FEM	-0.00620	0.00052	-11.82	21.30

La interpretación directa del coeficiente de la variable `FEMALE` (0.2124) podría sugerir que las mujeres ganan más que los hombres. Sin embargo, el término de interacción AGE_FEM indica que la diferencia por género depende de la edad.

El efecto marginal del género (ser mujer) es:

$$\frac{\partial \log(Y)}{\partial \text{FEMALE}} = \beta_2 + \beta_3 \cdot \text{AGE}$$

Evaluado en la edad promedio ($AGE = 43.53$), se obtiene:

$$\hat{\delta} = 0.2124 - 0.00620 \cdot 43.53 \approx -0.0575$$

Por tanto, al considerar la interacción, el efecto estimado de ser mujer sobre el ingreso logarítmico resulta negativo en promedio.

Estimación por mínimos cuadrados restringidos

En muchos contextos teóricos, se imponen restricciones sobre los coeficientes del modelo. Estas restricciones pueden ser de varios tipos:

1. **Eliminación de variables:** imponer que cierto coeficiente sea igual a cero.
2. **Restricciones de suma:** exigir que ciertos coeficientes sumen un valor fijo (por ejemplo, $\beta_1 + \beta_2 + \beta_3 = 1$).
3. **Restricciones de igualdad:** igualar coeficientes entre sí (por ejemplo, $\beta_2 = \beta_3$).

Estas restricciones pueden formularse de forma general como:

$$\text{Minimizar } S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad \text{sujeto a } \mathbf{R}\boldsymbol{\beta} = \mathbf{q}$$

donde $\mathbf{R}$ es una matriz de dimensión $J \times k$ que representa las restricciones lineales, y $\mathbf{q}$ es un vector de dimensión $J \times 1$.

Enfoque de Lagrange

Para resolver el problema anterior, se utiliza el método de Lagrange con multiplicadores $\boldsymbol{\lambda}$:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + 2\boldsymbol{\lambda}^\top (\mathbf{R}\boldsymbol{\beta} - \mathbf{q})$$

Derivando y resolviendo el sistema de primeras condiciones de orden:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + 2\mathbf{R}^\top \boldsymbol{\lambda} = 0$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}} = 2(\mathbf{R}\boldsymbol{\beta} - \mathbf{q}) = 0$$

Despejando, se obtiene la solución restringida:

$$\boldsymbol{\beta}^* = \boldsymbol{\beta} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \left[\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top\right]^{-1} (\mathbf{R}\boldsymbol{\beta} - \mathbf{q})$$

donde $\boldsymbol{\beta}$ es el estimador OLS no restringido.

Ejemplo

Suponga que se quiere imponer la restricción $\beta_2 = \beta_3$. Entonces, podemos redefinir el modelo como:

$$y = \beta_1 x_1 + \beta_2 (x_2 + x_3) + \varepsilon$$

y proceder a estimar este nuevo modelo reformulado, lo cual es equivalente a estimar el modelo original sujeto a la restricción lineal deseada.