

Inferencia, Causalidad y Políticas Públicas

ECO-60116

Week 01: Fundamentos de la Inferencia Causal

Eduard F. Martinez Gonzalez, Ph.D.

Departamento de Economía, Universidad Icesi

August 29, 2025

Acerca de este curso

① Al finalizar el curso podrás:

- ▶ Comprender y aplicar el marco conceptual de la inferencia causal.
- ▶ Usar metodologías: OLS, IV, RCT, DiD, RD, Matching.
- ▶ Diseñar e implementar evaluaciones de impacto con datos reales.

② Módulos principales:

- ▶ Fundamentos de la Inferencia Causal
- ▶ Diseños experimentales y cuasi-experimentales
- ▶ Métodos avanzados y robustez
- ▶ Integración y aplicación

③ Metodología:

- ▶ Exposición teórica del profesor.
- ▶ Ejemplos prácticos con R y artículos aplicados.
- ▶ Presentación estudiantil de artículos.

④ Evaluación:

- ▶ Presentación de artículo: 30%.
- ▶ Taller aplicado en parejas: 20%.
- ▶ Proyecto final con defensa: 50%.

- 1 Motivación
- 2 Marco conceptual: Rubin Causal Model
- 3 RECAP: ¿Qué dice y qué no dice una regresión?
- 4 Endogenidad y estimador de OLS
- 5 En nuestra próxima clase...

- 1 Motivación
- 2 Marco conceptual: Rubin Causal Model
- 3 RECAP: ¿Qué dice y qué no dice una regresión?
- 4 Endogenidad y estimador de OLS
- 5 En nuestra próxima clase...

Pregunta de investigación

Se implementa un programa de becas de sostenimiento dirigido a estudiantes universitarios de bajos recursos en universidades públicas.

Pregunta de investigación

Se implementa un programa de becas de sostenimiento dirigido a estudiantes universitarios de bajos recursos en universidades públicas.

Los beneficiarios son seleccionados por orden de inscripción: solo los primeros N estudiantes inscritos reciben el apoyo financiero.

Pregunta de investigación

Se implementa un programa de becas de sostenimiento dirigido a estudiantes universitarios de bajos recursos en universidades públicas.

Los beneficiarios son seleccionados por orden de inscripción: solo los primeros N estudiantes inscritos reciben el apoyo financiero.

Nuestro objetivo es estimar el efecto de este programa sobre los logros educativos, medidos a través del puntaje de las pruebas Saber Pro de los estudiantes (Y).

Notación básica

Unidades y tratamiento:

$$i = 1, \dots, n, \quad D_i \in \{0, 1\}$$

donde $D_i = 1$ si el estudiante recibe la beca y $D_i = 0$ en caso contrario.

Resultados potenciales:

$$Y_i(1), Y_i(0) \quad (\text{bien definidos})$$

$Y_i(1)$ y $Y_i(0)$, puntaje Saber Pro con y sin beca, respectivamente.

Resultado observado:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

Efecto individual (inobservable):

$$\tau_i = Y_i(1) - Y_i(0)$$

Problema fundamental

Para cada i observamos *solo uno* de $\{Y_i(1), Y_i(0)\}$.

¿Qué mide realmente la diferencia en medias?

Estimador naive (diferencia de medias): diferencia promedio del puntaje Saber Pro entre becados y no becados:

$$\hat{\beta}_{\text{naive}} = \mathbb{E}[Y \mid D = 1] - \mathbb{E}[Y \mid D = 0]$$

¿Qué mide realmente la diferencia en medias?

Estimador naive (diferencia de medias): diferencia promedio del puntaje Saber Pro entre becados y no becados:

$$\hat{\beta}_{\text{naive}} = \mathbb{E}[Y \mid D = 1] - \mathbb{E}[Y \mid D = 0]$$

Descomposición:

$$\underbrace{\mathbb{E}[Y \mid D = 1] - \mathbb{E}[Y \mid D = 0]}_{\text{Diferencia observada}} = \underbrace{\mathbb{E}[Y(1) - Y(0)]}_{\text{Efecto causal (ATE)}} + \underbrace{\left(\mathbb{E}[Y(0) \mid D = 1] - \mathbb{E}[Y(0) \mid D = 0] \right)}_{\text{Sesgo de selección}}$$

Interpretación

- **Efecto causal (ATE):** cambio promedio en el puntaje que produciría otorgar la beca.
- **Sesgo de selección:** diferencias *ex ante* entre becados y no becados (p. ej., mayor motivación, mejor preparación o mejor acceso a información entre quienes se inscriben primero).

Implicaciones: Sesgo de selección positivo

Ejemplos de mecanismos: Quienes se inscriben primero tienen mayor motivación/autoeficacia, o pueden tener mejor preparación previa (tutorías, capital académico). Es decir:

$$\mathbb{E}[Y(0) \mid D = 1] > \mathbb{E}[Y(0) \mid D = 0]$$

Entonces el término de selección es positivo:

$$\delta \equiv (\mathbb{E}[Y(0) \mid D = 1] - \mathbb{E}[Y(0) \mid D = 0]) > 0$$

Por tanto el $\hat{\beta}_{naive}$ atribuye al programa diferencias *ex ante* (motivación, preparación, acceso a información) que ya favorecerían a los tratados aun sin beca.

$$\hat{\beta}_{naive} = ATE + \delta$$

Implicaciones: Sesgo de selección negativo

Ejemplos de mecanismos: se inscriben primero estudiantes con mayor necesidad financiera (madres/padres jóvenes o cuidadores) y menor capital académico.

Implica que, *ex ante*, los tratados tenían peores resultados académicos que los no tratados:

$$\mathbb{E}[Y(0) \mid D = 1] < \mathbb{E}[Y(0) \mid D = 0]$$

Entonces el término de selección es negativo:

$$\delta \equiv (\mathbb{E}[Y(0) \mid D = 1] - \mathbb{E}[Y(0) \mid D = 0]) < 0$$

Por tanto el $\hat{\beta}_{naive}$ subestima el efecto causal

$$\hat{\beta}_{naive} = ATE + \delta$$

Si $|\delta| > |ATE|$, la diferencia observada puede incluso *invertir de signo* y sugerir (erróneamente) un efecto negativo.

- 1 Motivación
- 2 Marco conceptual: Rubin Causal Model
- 3 RECAP: ¿Qué dice y qué no dice una regresión?
- 4 Endogenidad y estimador de OLS
- 5 En nuestra próxima clase...

¿Qué es una regresión?

- Hasta ahora no hemos especificado *cómo* estimar la diferencia de medias del Saber Pro entre becados y no becados.

¿Qué es una regresión?

- Hasta ahora no hemos especificado *cómo* estimar la diferencia de medias del Saber Pro entre becados y no becados.
- La forma más simple es una regresión lineal con un indicador de beca.

¿Qué es una regresión?

- Hasta ahora no hemos especificado *cómo* estimar la diferencia de medias del Saber Pro entre becados y no becados.
- La forma más simple es una regresión lineal con un indicador de beca.
- Una **regresión** consiste en explicar una variable en función de otra(s) variables a través de una función (típicamente lineal).

¿Qué es una regresión?

- Hasta ahora no hemos especificado *cómo* estimar la diferencia de medias del Saber Pro entre becados y no becados.
- La forma más simple es una regresión lineal con un indicador de beca.
- Una **regresión** consiste en explicar una variable en función de otra(s) variables a través de una función (típicamente lineal).

$$y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

- Y_{1i} : variable de resultado (puntaje Saber Pro).
- X_{1i} : variable de interés (beca de sostenimiento).
- X_{2i} : variable de control (GPA previo, edad, carrera, etc.).
- ε_i : término de error (toda la variación de y_i no explicada por X_{1i} y X_{2i}).

Interpretación de β_1

$$\beta_1 = \frac{\partial \mathbb{E}[Y \mid X_1, X_2]}{\partial X_1}$$

- Si X_{1i} es recibir la beca (binaria), β_1 es una *diferencia condicional de medias* entre estudiantes comparables en X_{2i} (GPA previo, carrera, cohorte, etc.).

Interpretación de β_1

$$\beta_1 = \frac{\partial \mathbb{E}[Y \mid X_1, X_2]}{\partial X_1}$$

- Si X_{1i} es recibir la beca (binaria), β_1 es una *diferencia condicional de medias* entre estudiantes comparables en X_{2i} (GPA previo, carrera, cohorte, etc.).
- **En introducción a la econometría, se suele leer:**

Interpretación de β_1

$$\beta_1 = \frac{\partial \mathbb{E}[Y \mid X_1, X_2]}{\partial X_1}$$

- Si X_{1i} es recibir la beca (binaria), β_1 es una *diferencia condicional de medias* entre estudiantes comparables en X_{2i} (GPA previo, carrera, cohorte, etc.).
- **En introducción a la econometría, se suele leer:**
 - ▶ “ β_1 es el **efecto** de un cambio de X_{1i} sobre y_i ”

Interpretación de β_1

$$\beta_1 = \frac{\partial \mathbb{E}[Y \mid X_1, X_2]}{\partial X_1}$$

- Si X_{1i} es recibir la beca (binaria), β_1 es una *diferencia condicional de medias* entre estudiantes comparables en X_{2i} (GPA previo, carrera, cohorte, etc.).
- **En introducción a la econometría, se suele leer:**
 - ▶ “ β_1 es el **efecto** de un cambio de X_{1i} sobre y_i ”
 - ▶ “Un aumento de una unidad en X_1 está **asociado** a un aumento en y_i de β_1 unidades”

Interpretación de β_1

$$\beta_1 = \frac{\partial \mathbb{E}[Y \mid X_1, X_2]}{\partial X_1}$$

- Si X_{1i} es recibir la beca (binaria), β_1 es una *diferencia condicional de medias* entre estudiantes comparables en X_{2i} (GPA previo, carrera, cohorte, etc.).
- **En introducción a la econometría, se suele leer:**
 - ▶ “ β_1 es el **efecto** de un cambio de X_{1i} sobre y_i ”
 - ▶ “Un aumento de una unidad en X_1 está **asociado** a un aumento en y_i de β_1 unidades”
- **¿Ven alguna diferencia entre estas dos interpretaciones?**

Interpretación de β_1

$$\beta_1 = \frac{\partial \mathbb{E}[Y \mid X_1, X_2]}{\partial X_1}$$

- Si X_{1i} es recibir la beca (binaria), β_1 es una *diferencia condicional de medias* entre estudiantes comparables en X_{2i} (GPA previo, carrera, cohorte, etc.).
- **En introducción a la econometría, se suele leer:**
 - ▶ “ β_1 es el **efecto** de un cambio de X_{1i} sobre y_i ”
 - ▶ “Un aumento de una unidad en X_1 está **asociado** a un aumento en y_i de β_1 unidades”
- **¿Ven alguna diferencia entre estas dos interpretaciones?**
 - ▶ “Efecto” sugiere **causalidad**.
 - ▶ “Asociado” describe **correlación** (relación estadística) sin afirmar causalidad.

Interpretación de β_1

$$\beta_1 = \frac{\partial \mathbb{E}[Y \mid X_1, X_2]}{\partial X_1}$$

- Si X_{1i} es recibir la beca (binaria), β_1 es una *diferencia condicional de medias* entre estudiantes comparables en X_{2i} (GPA previo, carrera, cohorte, etc.).
- **En introducción a la econometría, se suele leer:**
 - ▶ “ β_1 es el **efecto** de un cambio de X_{1i} sobre y_i ”
 - ▶ “Un aumento de una unidad en X_1 está **asociado** a un aumento en y_i de β_1 unidades”
- **¿Ven alguna diferencia entre estas dos interpretaciones?**
 - ▶ “Efecto” sugiere **causalidad**.
 - ▶ “Asociado” describe **correlación** (relación estadística) sin afirmar causalidad.
- **Clave:** misma ecuación, *dos lecturas distintas*. La causalidad depende del **diseño** y los **supuestos**, no solo de estimar una recta.

Interpretación *predictiva* de una regresión

Idea central: la lectura predictiva es **descriptiva**: no afirma causalidad; resume **correlaciones (condicionales)** presentes en los datos para pronosticar Y .

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots$$

- Optimiza **error de predicción** (p. ej., MSE/RMSE), no un parámetro causal.

Interpretación *predictiva* de una regresión

Idea central: la lectura predictiva es **descriptiva**: no afirma causalidad; resume **correlaciones (condicionales)** presentes en los datos para pronosticar Y .

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots$$

- Optimiza **error de predicción** (p. ej., MSE/RMSE), no un parámetro causal.
- Buena predicción \Rightarrow capta patrones útiles aunque provengan de variables omitidas, simultaneidad o pura asociación.

Interpretación *predictiva* de una regresión

Idea central: la lectura predictiva es **descriptiva**: no afirma causalidad; resume **correlaciones (condicionales)** presentes en los datos para pronosticar Y .

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots$$

- Optimiza **error de predicción** (p. ej., MSE/RMSE), no un parámetro causal.
- Buena predicción \Rightarrow capta patrones útiles aunque provengan de variables omitidas, simultaneidad o pura asociación.

¿Cuándo es útil predecir?

- Clasificar hogares por probabilidad de **pobreza** usando variables simples (techo, material de la vivienda, etc.).
- Anticipar **lluvia** a partir de nubosidad, temperatura, humedad, etc.
- *Ejemplo descriptivo (Colombia)*: correlación ≈ 0.72 entre **educación de padres** y **educación de hijos**; ≈ 0.83 en **hogares pobres**.

Mensaje: la regresión predictiva “no toma partido” por mecanismos; su meta es acertar pronósticos, no identificar efectos.

Interpretación *causal* de una regresión

Más ambiciosa (y arriesgada): intenta decir *cómo funciona el mundo*, no sólo describir patrones.

Ejemplo motivador:

- “Los beneficiarios de la beca obtienen un puntaje Saber Pro **20% mayor** que los no beneficiarios.” (*descriptivo/asociativo*)
- “*Otorgar* la beca **aumenta** el puntaje Saber Pro del estudiante en **20%.**” (*causal*)

Interpretación *causal* de una regresión

Más ambiciosa (y arriesgada): intenta decir *cómo funciona el mundo*, no sólo describir patrones.

Ejemplo motivador:

- “Los beneficiarios de la beca obtienen un puntaje Saber Pro **20% mayor** que los no beneficiarios.” (*descriptivo/asociativo*)
- “*Otorgar* la beca **aumenta** el puntaje Saber Pro del estudiante en **20%.**” (*causal*)

¿Qué necesitamos para leer β_1 como efecto causal?

- Una fuente de variación exógena y supuestos que hagan creíble la interpretación causal.

Interpretación *causal* de una regresión

Más ambiciosa (y arriesgada): intenta decir *cómo funciona el mundo*, no sólo describir patrones.

Ejemplo motivador:

- “Los beneficiarios de la beca obtienen un puntaje Saber Pro **20% mayor** que los no beneficiarios.” (*descriptivo/asociativo*)
- “Otorgar la beca **aumenta** el puntaje Saber Pro del estudiante en **20%.**” (*causal*)

¿Qué necesitamos para leer β_1 como efecto causal?

- Una fuente de variación exógena y supuestos que hagan creíble la interpretación causal.

¿Qué es una variable exógena?

- Un regresor cuya variación es **independiente** de ε_i (*la asignación de la beca no está correlacionada con inobservables que afectan simultáneamente a Y_i*). Es decir, que $(\mathbb{E}[Y(0) \mid D = 1] - \mathbb{E}[Y(0) \mid D = 0]) = 0$
- Por ejemplo, si los cupos se asignan por sorteo entre las personas inscritas, en lugar de por orden de llegada.

- 1 Motivación
- 2 Marco conceptual: Rubin Causal Model
- 3 RECAP: ¿Qué dice y qué no dice una regresión?
- 4 Endogenidad y estimador de OLS
- 5 En nuestra próxima clase...

El Estimador de Mínimos Cuadrados Ordinarios (OLS)

Hasta ahora no hemos dicho *cómo* estimar los parámetros β de la regresión.

El Estimador de Mínimos Cuadrados Ordinarios (OLS)

Hasta ahora no hemos dicho *cómo* estimar los parámetros β de la regresión.

Modelo lineal:

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

Dimensiones:

$$\mathbf{X} : n \times k, \quad \mathbf{y} : n \times 1, \quad \hat{\beta} : k \times 1$$

El Estimador de Mínimos Cuadrados Ordinarios (OLS)

Hasta ahora no hemos dicho *cómo* estimar los parámetros β de la regresión.

Modelo lineal:

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

Dimensiones:

$$\mathbf{X} : n \times k, \quad \mathbf{y} : n \times 1, \quad \hat{\beta} : k \times 1$$

Objetivo: Minimizar la suma de los residuos al cuadrado:

$$\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

El Estimador de Mínimos Cuadrados Ordinarios (OLS)

Hasta ahora no hemos dicho *cómo* estimar los parámetros β de la regresión.

Modelo lineal:

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

Dimensiones:

$$\mathbf{X} : n \times k, \quad \mathbf{y} : n \times 1, \quad \hat{\beta} : k \times 1$$

Objetivo: Minimizar la suma de los residuos al cuadrado:

$$\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

Solución analítica:

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

El Estimador de Mínimos Cuadrados Ordinarios (OLS)

Hasta ahora no hemos dicho *cómo* estimar los parámetros β de la regresión.

Modelo lineal:

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

Dimensiones:

$$\mathbf{X} : n \times k, \quad \mathbf{y} : n \times 1, \quad \hat{\beta} : k \times 1$$

Objetivo: Minimizar la suma de los residuos al cuadrado:

$$\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

Solución analítica:

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

Interpretación

$\hat{\beta}$ es el vector de coeficientes que mejor ajusta \mathbf{y} en el sentido de mínimos cuadrados, siempre que $\mathbf{X}'\mathbf{X}$ sea invertible.

Derivación del Estimador OLS

Problema de optimización: encontrar los coeficientes β que minimizan la suma de los residuos al cuadrado:

$$S(\beta) = \sum_{i=1}^n (y_i - x_i' \beta)^2 = \varepsilon' \varepsilon = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

Derivación del Estimador OLS

Problema de optimización: encontrar los coeficientes β que minimizan la suma de los residuos al cuadrado:

$$S(\beta) = \sum_{i=1}^n (y_i - x_i' \beta)^2 = \varepsilon' \varepsilon = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

1. Expandimos el producto cuadrático:

$$S(\beta) = \mathbf{y}'\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{y} + \beta'(\mathbf{X}'\mathbf{X})\beta$$

Derivación del Estimador OLS

Problema de optimización: encontrar los coeficientes β que minimizan la suma de los residuos al cuadrado:

$$S(\beta) = \sum_{i=1}^n (y_i - x_i' \beta)^2 = \varepsilon' \varepsilon = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

1. Expandimos el producto cuadrático:

$$S(\beta) = \mathbf{y}'\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{y} + \beta'(\mathbf{X}'\mathbf{X})\beta$$

2. Derivamos con respecto a β :

$$\frac{\partial S(\beta)}{\partial \beta} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\beta$$

Derivación del Estimador OLS

Problema de optimización: encontrar los coeficientes β que minimizan la suma de los residuos al cuadrado:

$$S(\beta) = \sum_{i=1}^n (y_i - x_i' \beta)^2 = \varepsilon' \varepsilon = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

1. Expandimos el producto cuadrático:

$$S(\beta) = \mathbf{y}'\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{y} + \beta'(\mathbf{X}'\mathbf{X})\beta$$

2. Derivamos con respecto a β :

$$\frac{\partial S(\beta)}{\partial \beta} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\beta$$

3. Igualamos a cero (condición de primer orden):

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{y}$$

Derivación del Estimador OLS

Problema de optimización: encontrar los coeficientes β que minimizan la suma de los residuos al cuadrado:

$$S(\beta) = \sum_{i=1}^n (y_i - x_i' \beta)^2 = \varepsilon' \varepsilon = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

1. Expandimos el producto cuadrático:

$$S(\beta) = \mathbf{y}'\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{y} + \beta'(\mathbf{X}'\mathbf{X})\beta$$

2. Derivamos con respecto a β :

$$\frac{\partial S(\beta)}{\partial \beta} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\beta$$

3. Igualamos a cero (condición de primer orden):

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{y}$$

4. Solución (ecuaciones normales):

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

Insesgadez del Estimador OLS

¿Bajo que supuestos podemos asumir que $\hat{\beta}$ es un estimador insesgado de β ?

Insesgadez del Estimador OLS

¿Bajo que supuestos podemos asumir que $\hat{\beta}$ es un estimador insesgado de β ?

- Linealidad: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- Exogeneidad: $\mathbb{E}[\varepsilon \mid \mathbf{X}] = \mathbf{0}$
- No multicolinealidad perfecta: $\text{rank}(\mathbf{X}) = k \Rightarrow \mathbf{X}'\mathbf{X}$ invertible

Insesgadez del Estimador OLS

¿Bajo que supuestos podemos asumir que $\hat{\beta}$ es un estimador insesgado de β ?

- Linealidad: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- Exogeneidad: $\mathbb{E}[\varepsilon \mid \mathbf{X}] = \mathbf{0}$
- No multicolinealidad perfecta: $\text{rank}(\mathbf{X}) = k \Rightarrow \mathbf{X}'\mathbf{X}$ invertible

Sabemos que:

- $\mathbf{y} = \mathbf{X}\beta + \varepsilon \quad (1)$

- $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2)$

Insesgadez del Estimador OLS

¿Bajo que supuestos podemos asumir que $\hat{\beta}$ es un estimador insesgado de β ?

- Linealidad: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- Exogeneidad: $\mathbb{E}[\varepsilon \mid \mathbf{X}] = \mathbf{0}$
- No multicolinealidad perfecta: $\text{rank}(\mathbf{X}) = k \Rightarrow \mathbf{X}'\mathbf{X}$ invertible

Sabemos que:

- $\mathbf{y} = \mathbf{X}\beta + \varepsilon \quad (1)$

- $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2)$

Sustituyendo (1) en (2) tenemos:

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$$

Insesgadez del Estimador OLS

¿Bajo que supuestos podemos asumir que $\hat{\beta}$ es un estimador insesgado de β ?

- Linealidad: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- Exogeneidad: $\mathbb{E}[\varepsilon \mid \mathbf{X}] = \mathbf{0}$
- No multicolinealidad perfecta: $\text{rank}(\mathbf{X}) = k \Rightarrow \mathbf{X}'\mathbf{X}$ invertible

Sabemos que:

- $\mathbf{y} = \mathbf{X}\beta + \varepsilon \quad (1)$

- $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2)$

Sustituyendo (1) en (2) tenemos:

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$$

Aplicando esperanza condicional:

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\varepsilon \mid \mathbf{X}]$$

Insesgadez del Estimador OLS

¿Bajo que supuestos podemos asumir que $\hat{\beta}$ es un estimador insesgado de β ?

- Linealidad: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- Exogeneidad: $\mathbb{E}[\varepsilon \mid \mathbf{X}] = \mathbf{0}$
- No multicolinealidad perfecta: $\text{rank}(\mathbf{X}) = k \Rightarrow \mathbf{X}'\mathbf{X}$ invertible

Sabemos que:

- $\mathbf{y} = \mathbf{X}\beta + \varepsilon \quad (1)$

- $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2)$

Sustituyendo (1) en (2) tenemos:

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$$

Aplicando esperanza condicional:

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\varepsilon \mid \mathbf{X}]$$

Por exogeneidad:

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}] = \beta$$

Sesgo por variable omitida

Modelo verdadero (con omitidas):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\phi + \mathbf{u}, \quad \mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}] = \mathbf{0}$$

Sesgo por variable omitida

Modelo verdadero (con omitidas):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\phi + \mathbf{u}, \quad \mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}] = \mathbf{0}$$

Modelo estimado (omite Z):

$$\mathbf{y} = \mathbf{X}\tilde{\beta} + \tilde{\mathbf{u}}$$

Sesgo por variable omitida

Modelo verdadero (con omitidas):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\phi + \mathbf{u}, \quad \mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}] = \mathbf{0}$$

Modelo estimado (omite Z):

$$\mathbf{y} = \mathbf{X}\tilde{\beta} + \tilde{\mathbf{u}}$$

Estimador OLS con omitidas:

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

Sesgo por variable omitida

Modelo verdadero (con omitidas):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\phi + \mathbf{u}, \quad \mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}] = \mathbf{0}$$

Modelo estimado (omite Z):

$$\mathbf{y} = \mathbf{X}\tilde{\beta} + \tilde{\mathbf{u}}$$

Estimador OLS con omitidas:

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

Sustituyendo y del modelo verdadero:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{Z}\phi + \mathbf{u})$$

Sesgo por variable omitida

Modelo verdadero (con omitidas):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\phi + \mathbf{u}, \quad \mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}] = \mathbf{0}$$

Modelo estimado (omite Z):

$$\mathbf{y} = \mathbf{X}\tilde{\beta} + \tilde{\mathbf{u}}$$

Estimador OLS con omitidas:

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

Sustituyendo y del modelo verdadero:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{Z}\phi + \mathbf{u})$$

Resolviendo se obtiene:

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\phi + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

Esperanza del estimador con variable omitida

Aplicando esperanza condicional:

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}, \mathbf{Z}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\phi + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}]$$

Esperanza del estimador con variable omitida

Aplicando esperanza condicional:

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}, \mathbf{Z}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\phi + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}]$$

Por exogenidad: $\mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}] = \mathbf{0}$

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}, \mathbf{Z}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\phi$$

Esperanza del estimador con variable omitida

Aplicando esperanza condicional:

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}, \mathbf{Z}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\phi + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}]$$

Por exogenidad: $\mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}] = \mathbf{0}$

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}, \mathbf{Z}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\phi$$

Por expectativas iteradas (Fijando \mathbf{Z} y condicionando en \mathbf{X}):

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}] = \beta + \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}}_{\text{Cor}(\mathbf{X}, \mathbf{Z})} \phi$$

Esperanza del estimador con variable omitida

Aplicando esperanza condicional:

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}, \mathbf{Z}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\phi + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}]$$

Por exogenidad: $\mathbb{E}[\mathbf{u} \mid \mathbf{X}, \mathbf{Z}] = \mathbf{0}$

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}, \mathbf{Z}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\phi$$

Por expectativas iteradas (Fijando \mathbf{Z} y condicionando en \mathbf{X}):

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}] = \beta + \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}}_{\text{Cor}(\mathbf{X}, \mathbf{Z})} \phi$$

En el caso univariado: $\delta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}] = \beta + \delta \cdot \phi$$

Dirección del sesgo por variable omitida

El sesgo depende de dos factores:

- La correlación entre la variable de interés X y la variable omitida Z (signo de δ).
- El efecto causal de la variable omitida Z sobre Y (signo de ϕ).

Dirección del sesgo por variable omitida

El sesgo depende de dos factores:

- La correlación entre la variable de interés X y la variable omitida Z (signo de δ).
- El efecto causal de la variable omitida Z sobre Y (signo de ϕ).

$$\text{Sesgo} = \delta \cdot \phi$$

Correlación $X-Z$ (δ)	Efecto $Z \rightarrow Y$ (ϕ)	Dirección del sesgo en $\hat{\beta}$
Positiva	Positiva	+ (sobreestima β)
Negativa	Positiva	- (subestima β)
Positiva	Negativa	- (subestima β)
Negativa	Negativa	+ (sobreestima β)

- 1 Motivación
- 2 Marco conceptual: Rubin Causal Model
- 3 RECAP: ¿Qué dice y qué no dice una regresión?
- 4 Endogenidad y estimador de OLS
- 5 En nuestra próxima clase...

Referencias esenciales

- Angrist, J., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics*. Princeton.
- Bernal, R., & Peña, X. (2011). *Guía práctica para la evaluación de impacto*. Universidad de los Andes.
- Gertler, P., Martinez, S., Premand, P., Rawlings, L., & Vermeersch, C. (2017). *La evaluación de impacto en la práctica* (2ª ed.). Banco Mundial.
- Hernán, M., & Robins, J. (2020). *Causal Inference*. Chapman & Hall/CRC (libro abierto).
- Imbens, G., & Rubin, D. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. CUP.
- Rubin, D. B. (1974). *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*. J. Educ. Psychology.