

# Introduccion al Business Analytics

## Clase 2: Tareas y Roles de BA

Eduard F. Martinez Gonzalez, Ph.D.

Departamento de Economía, Universidad Icesi

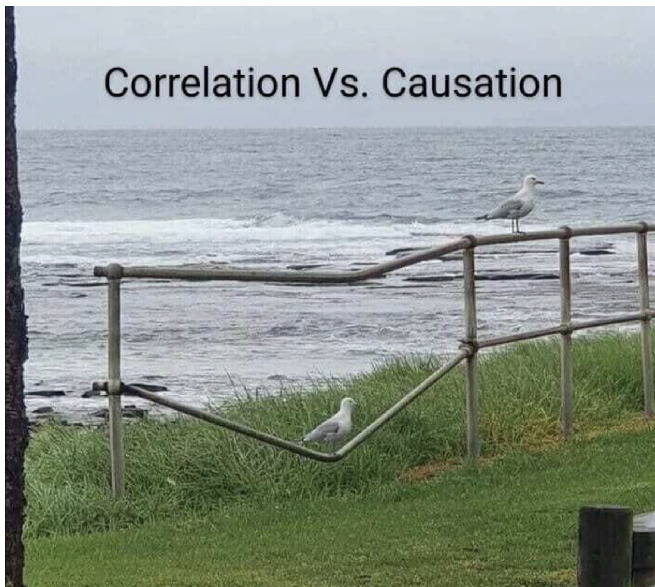
August 27, 2025

- 1 Motivación
- 2 Acerca de este curso
- 3 Marco conceptual
- 4 Estimador de OLS
- 5 El problema de selección
- 6 Definición de parámetros de impacto del tratamiento

# ¿Por qué inferencia causal en políticas públicas?

- ❶ ¿Cuál es el efecto causal de un tratamiento/política sobre un resultado?
  - ▶ ¿Los programas de transferencias condicionadas reducen la pobreza?
  - ▶ ¿Las escuelas de jornada extendida mejoran el rendimiento académico?
  - ▶ ¿Subir el salario mínimo destruye empleos?
- ❷ El riesgo de confundir correlación con causalidad
  - ▶ Podríamos **implementar políticas costosas** que no funcionan.
  - ▶ O **descartar programas efectivos** porque la evidencia estaba mal interpretada.
  - ▶ Terminar con **malas asignaciones de recursos públicos**.
  - ▶ **Ejemplo:** Observar que los niños que asisten a tutorías tienen mejores notas.  $\Rightarrow$  ¿es la tutoría la causa o son familias más motivadas las que inscriben a sus hijos?

# Correlación vs Causalidad



# ¿Por qué inferencia causal en políticas públicas?

- 1 La pregunta central en políticas públicas:

## Contrafactual

**¿Qué hubiera pasado si la política no se hubiera implementado?**

- 2 El reto: no podemos observar al mismo individuo en dos mundos distintos (con política y sin política).
- 3 La inferencia causal nos da **métodos** para aproximar ese mundo contrafactual:
  - ▶ Experimentos aleatorios (RCTs).
  - ▶ Métodos cuasi-experimentales (DiD, RD, IV, Matching).

- 1 Motivación
- 2 Acerca de este curso
- 3 Marco conceptual
- 4 Estimador de OLS
- 5 El problema de selección
- 6 Definición de parámetros de impacto del tratamiento

- 1 Motivación
- 2 Acerca de este curso
- 3 Marco conceptual
- 4 Estimador de OLS
- 5 El problema de selección
- 6 Definición de parámetros de impacto del tratamiento

- 1 Motivación
- 2 Acerca de este curso
- 3 Marco conceptual
- 4 Estimador de OLS**
- 5 El problema de selección
- 6 Definición de parámetros de impacto del tratamiento



# El Estimador de Mínimos Cuadrados Ordinarios (OLS)

**Modelo lineal:**

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

**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  $\beta$  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  $\beta$ :

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

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

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

4. Solución (ecuaciones normales):

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

# Supuestos Fundamentales del Modelo Lineal

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

## Resultado

Bajo estos supuestos, el estimador OLS es **insesgado** y **consistente**.

# Insesgadez del Estimador OLS

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

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

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

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:

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

## Conclusión

El estimador OLS es **insesgado**.

- 1 Motivación
- 2 Acerca de este curso
- 3 Marco conceptual
- 4 Estimador de OLS
- 5 El problema de selección**
- 6 Definición de parámetros de impacto del tratamiento

- 1 Motivación
- 2 Acerca de este curso
- 3 Marco conceptual
- 4 Estimador de OLS
- 5 El problema de selección
- 6 Definición de parámetros de impacto del tratamiento

# Notación y resultados potenciales

- Unidades  $i = 1, \dots, n$ ; tratamiento binario  $D_i \in \{0, 1\}$ .
- Resultados potenciales:  $Y_i(1)$  y  $Y_i(0)$ .
- Resultado observado (consistencia):

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

# Problema fundamental de la inferencia causal

- Para cada  $i$ , observamos solo uno de  $\{Y_i(1), Y_i(0)\}$ .
- Efecto individual:  $\tau_i = Y_i(1) - Y_i(0)$  (inobservable).
- Objetos de interés:

$$\text{ATE} = E[Y(1) - Y(0)]$$

$$\text{ATT} = E[Y(1) - Y(0) \mid D = 1]$$

$$\text{ATC} = E[Y(1) - Y(0) \mid D = 0]$$

$$\text{CATE}(x) = E[Y(1) - Y(0) \mid X = x]$$



# Referencias esenciales

- Rubin, D. B. (1974). *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*. J. Educ. Psychology.
- Imbens, G., & Rubin, D. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. CUP.
- Angrist, J., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics*. Princeton.
- Hernán, M., & Robins, J. (2020). *Causal Inference*. Chapman & Hall/CRC (libro abierto).
- Rosenbaum, P., & Rubin, D. (1983). *The central role of the propensity score*. Biometrika.