

Inferencia, Causalidad y Políticas Públicas

ECO-60116

Week 02: Experimentos Aleatorios Controlados (RCTs)

Eduard F. Martinez Gonzalez, Ph.D.

Departamento de Economía, Universidad Icesi

September 1, 2025

RECAP

- **Efecto individual:** $\tau_i = Y_i(1) - Y_i(0)$.
- **Problema fundamental de la inferencia causal:** para cada unidad i sólo se observa uno de los dos resultados potenciales.

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

RECAP

- **Efecto individual:** $\tau_i = Y_i(1) - Y_i(0)$.
- **Problema fundamental de la inferencia causal:** para cada unidad i sólo se observa uno de los dos resultados potenciales.

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

- **Diferencia de medias observada:**

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

RECAP

- **Efecto individual:** $\tau_i = Y_i(1) - Y_i(0)$.
- **Problema fundamental de la inferencia causal:** para cada unidad i sólo se observa uno de los dos resultados potenciales.

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

- **Diferencia de medias observada:**

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

- **Condición clave para interpretación causal:** se requiere *variación exógena* en D_i (diseño/identificación), no basta con estimar una regresión.

$$\mathbb{E}[\varepsilon \mid \mathbf{X}] = 0$$

RECAP

- **Efecto individual:** $\tau_i = Y_i(1) - Y_i(0)$.
- **Problema fundamental de la inferencia causal:** para cada unidad i sólo se observa uno de los dos resultados potenciales.

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

- **Diferencia de medias observada:**

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

- **Condición clave para interpretación causal:** se requiere *variación exógena* en D_i (diseño/identificación), no basta con estimar una regresión.

$$\mathbb{E}[\varepsilon \mid \mathbf{X}] = 0$$

- **Bajo exogeneidad:** el estimador de Mínimos Cuadrados Ordinarios (OLS) es insesgado.

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

1 Experimentos Aleatorios Controlados (RCTs)

RCT Natural: Lotería del Draft de Vietnam

Contexto

- Durante la Guerra de Vietnam, EE. UU. implementó **loterías de reclutamiento** (1969–1972).

RCT Natural: Lotería del Draft de Vietnam

Contexto

- Durante la Guerra de Vietnam, EE. UU. implementó **loterías de reclutamiento** (1969–1972).
- Cada fecha de nacimiento recibió un **Random Sequence Number (RSN)** entre 1 y 365.

RCT Natural: Lotería del Draft de Vietnam

Contexto

- Durante la Guerra de Vietnam, EE. UU. implementó **loterías de reclutamiento** (1969–1972).
- Cada fecha de nacimiento recibió un **Random Sequence Number (RSN)** entre 1 y 365.
- Los hombres con RSN bajo quedaban **elegibles para ser reclutados**.

RCT Natural: Lotería del Draft de Vietnam

Contexto

- Durante la Guerra de Vietnam, EE. UU. implementó **loterías de reclutamiento** (1969–1972).
- Cada fecha de nacimiento recibió un **Random Sequence Number (RSN)** entre 1 y 365.
- Los hombres con RSN bajo quedaban **elegibles para ser reclutados**.
- La lotería generó **variación exógena** en la probabilidad de servicio militar.

Datos: Social Security Earnings

Fuente de datos

- **Social Security Administration** – Continuous Work History Sample (CWHS).

Datos: Social Security Earnings

Fuente de datos

- **Social Security Administration** – Continuous Work History Sample (CWHS).
- Muestra aleatoria del 1% de los números de seguro social.

Datos: Social Security Earnings

Fuente de datos

- **Social Security Administration** – Continuous Work History Sample (CWHS).
- Muestra aleatoria del 1% de los números de seguro social.
- Contiene ingresos anuales (1964–84), cubriendo empleo asalariado y autoempleo.

Datos: Social Security Earnings

Fuente de datos

- **Social Security Administration** – Continuous Work History Sample (CWHS).
- Muestra aleatoria del 1% de los números de seguro social.
- Contiene ingresos anuales (1964–84), cubriendo empleo asalariado y autoempleo.
- Emparejados con RSN de la lotería según fecha de nacimiento.

Datos: Social Security Earnings

Fuente de datos

- **Social Security Administration** – Continuous Work History Sample (CWHS).
- Muestra aleatoria del 1% de los números de seguro social.
- Contiene ingresos anuales (1964–84), cubriendo empleo asalariado y autoempleo.
- Emparejados con RSN de la lotería según fecha de nacimiento.

Datos: Social Security Earnings

Fuente de datos

- **Social Security Administration** – Continuous Work History Sample (CWHS).
- Muestra aleatoria del 1% de los números de seguro social.
- Contiene ingresos anuales (1964–84), cubriendo empleo asalariado y autoempleo.
- Emparejados con RSN de la lotería según fecha de nacimiento.

Estrategia

- Comparar perfiles de ingreso entre elegibles y no elegibles por la lotería.

Datos: Social Security Earnings

Fuente de datos

- **Social Security Administration** – Continuous Work History Sample (CWHS).
- Muestra aleatoria del 1% de los números de seguro social.
- Contiene ingresos anuales (1964–84), cubriendo empleo asalariado y autoempleo.
- Emparejados con RSN de la lotería según fecha de nacimiento.

Estrategia

- Comparar perfiles de ingreso entre elegibles y no elegibles por la lotería.
- Diferencias atribuibles a servicio militar \Rightarrow efecto causal.

Metodología: Draft Eligibility como IV

Problema

- La condición de veterano (D_i) está correlacionada con características no observadas.

Metodología: Draft Eligibility como IV

Problema

- La condición de veterano (D_i) está correlacionada con características no observadas.
- Ejemplo: quienes se enlistan voluntariamente pueden tener distinto potencial laboral.

Metodología: Draft Eligibility como IV

Problema

- La condición de veterano (D_i) está correlacionada con características no observadas.
- Ejemplo: quienes se enlistan voluntariamente pueden tener distinto potencial laboral.

Metodología: Draft Eligibility como IV

Problema

- La condición de veterano (D_i) está correlacionada con características no observadas.
- Ejemplo: quienes se enlistan voluntariamente pueden tener distinto potencial laboral.

Solución

- Usar **elegibilidad por lotería** (Z_i) como **instrumento** para D_i .

Metodología: Draft Eligibility como IV

Problema

- La condición de veterano (D_i) está correlacionada con características no observadas.
- Ejemplo: quienes se enlistan voluntariamente pueden tener distinto potencial laboral.

Solución

- Usar **elegibilidad por lotería** (Z_i) como **instrumento** para D_i .
- Bajo aleatorización: $Z_i \perp\!\!\!\perp \{Y_i(0), Y_i(1)\}$.

Metodología: Draft Eligibility como IV

Problema

- La condición de veterano (D_i) está correlacionada con características no observadas.
- Ejemplo: quienes se enlistan voluntariamente pueden tener distinto potencial laboral.

Solución

- Usar **elegibilidad por lotería** (Z_i) como **instrumento** para D_i .
- Bajo aleatorización: $Z_i \perp\!\!\!\perp \{Y_i(0), Y_i(1)\}$.
- Estimador tipo Wald:

$$\hat{\alpha} = \frac{\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]}{\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0]}$$

Metodología: Draft Eligibility como IV

Problema

- La condición de veterano (D_i) está correlacionada con características no observadas.
- Ejemplo: quienes se enlistan voluntariamente pueden tener distinto potencial laboral.

Solución

- Usar **elegibilidad por lotería** (Z_i) como **instrumento** para D_i .
- Bajo aleatorización: $Z_i \perp\!\!\!\perp \{Y_i(0), Y_i(1)\}$.
- Estimador tipo Wald:

$$\hat{\alpha} = \frac{\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]}{\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0]}$$

- Interpretable como efecto local en los **cumplidores** (LATE).

Resultados principales

- Para hombres blancos nacidos 1950–52:

Resultados principales

- Para hombres blancos nacidos 1950–52:
 - ▶ Veteranos ganaban $\sim 15\%$ menos que no veteranos en los años 1980s.

Resultados principales

- Para hombres blancos nacidos 1950–52:
 - ▶ Veteranos ganaban $\sim 15\%$ menos que no veteranos en los años 1980s.
 - ▶ Pérdida anual $\approx 2,000$ dólares (constantes de 1978).

Resultados principales

- Para hombres blancos nacidos 1950–52:
 - ▶ Veteranos ganaban $\sim 15\%$ menos que no veteranos en los años 1980s.
 - ▶ Pérdida anual $\approx 2,000$ dólares (constantes de 1978).
 - ▶ Magnitud similar a **perder 2 años de experiencia laboral civil**.

Resultados principales

- Para hombres blancos nacidos 1950–52:
 - ▶ Veteranos ganaban $\sim 15\%$ menos que no veteranos en los años 1980s.
 - ▶ Pérdida anual $\approx 2,000$ dólares (constantes de 1978).
 - ▶ Magnitud similar a **perder 2 años de experiencia laboral civil**.
- Para no blancos: resultados imprecisos, sin efectos robustos.

Resultados principales

- Para hombres blancos nacidos 1950–52:
 - ▶ Veteranos ganaban $\sim 15\%$ menos que no veteranos en los años 1980s.
 - ▶ Pérdida anual $\approx 2,000$ dólares (constantes de 1978).
 - ▶ Magnitud similar a **perder 2 años de experiencia laboral civil**.
- Para no blancos: resultados imprecisos, sin efectos robustos.
- Conclusión: el servicio militar en Vietnam redujo el ingreso de largo plazo de los veteranos blancos.

Intuición del RCT

Idea central: Asignar tratamiento por **azar** hace independientes D_i y los inobservables que afectan Y_i .

$$D_i \perp\!\!\!\perp \{Y_i(1), Y_i(0)\}$$

Intuición del RCT

Idea central: Asignar tratamiento por **azar** hace independientes D_i y los inobservables que afectan Y_i .

$$D_i \perp\!\!\!\perp \{Y_i(1), Y_i(0)\}$$

Consecuencia: Los grupos Tratamiento/Control son *en promedio* comparables en todo (observado y no observado).

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

Intuición del RCT

Idea central: Asignar tratamiento por **azar** hace independientes D_i y los inobservables que afectan Y_i .

$$D_i \perp\!\!\!\perp \{Y_i(1), Y_i(0)\}$$

Consecuencia: Los grupos Tratamiento/Control son *en promedio* comparables en todo (observado y no observado).

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

Estimación simple (muestra grande):

$$\widehat{ATE} = \bar{Y}_{\text{trat}} - \bar{Y}_{\text{ctrl}}$$

- Asignación aleatoria: $Z_i \in \{0, 1\}$ (invitación/oferta).
- Tratamiento efectivo: $D_i \in \{0, 1\}$ (lo que la unidad *recibe*).
- Resultado: Y_i .

RCT ideal (cumplimiento perfecto): $D_i = Z_i$

$$\Rightarrow \text{ITT} = \text{ATE.} \widehat{ATE} = \mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]$$

Incumplimiento: ITT, TOT/LATE (IV)

Problema: $D_i \neq Z_i$ por *no-take-up*, *crossovers*, etc.

- **ITT (Intention-to-Treat):** efecto de *ofrecer* el programa:
 $\mathbb{E}[Y \mid Z=1] - \mathbb{E}[Y \mid Z=0]$.
- **TOT (Treatment-on-the-Treated)** bajo supuestos (monotonicidad, exclusión):

$$\text{TOT} = \frac{\mathbb{E}[Y \mid Z=1] - \mathbb{E}[Y \mid Z=0]}{\mathbb{E}[D \mid Z=1] - \mathbb{E}[D \mid Z=0]} = \frac{\text{ITT}}{\text{Compliance Rate}}$$

Lectura: TOT/LATE identifica el efecto *local* en *cumplidores* (*compliers*).

Diseño: unidad de aleatorización y estratificación

Unidad de aleatorización

- Individual (*households, estudiantes*) vs. por grupos/cluster (*escuelas, municipios*).
- Si hay efectos de pares o contagio, considerar **cluster RCT**.

Estratificación/Blocking

- Aleatorizar *dentro* de celdas homogéneas (p. ej., sexo, cuartil de ingreso, escuela).
- Mejora precisión; asegurar tamaños mínimos por estrato.

Regla práctica: si estratificas, **ajusta** por estratos en la regresión/análisis.

Poder estadístico y tamaño muestral (intuición)

Reglas prácticas (para diferencia de medias):

$$n \approx \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{\Delta^2}$$

- σ^2 : varianza del resultado, Δ : MDE (efecto mínimo detectable).
- Balance 50/50 suele maximizar poder.
- En **cluster RCT**: inflación por ICC ρ y tamaño de cluster \bar{m} :

$$\text{DEFF} \approx 1 + (\bar{m} - 1)\rho$$

Amenazas y cómo mitigarlas

- ➊ **Attrition (pérdida de muestra):** verificar balance en attriters; bounds (Lee), seguimiento intensivo.
- ➋ **Spillovers/Contagio:** aleatorizar por cluster o buffers; medir exposición indirecta.
- ➌ **Noncompliance:** reportar ITT; TOT vía IV con Z como instrumento.
- ➍ **Heterogeneidad ex ante:** estratificar; reportar efectos por subgrupos pre-especificados.
- ➎ **Medición/Outcomes múltiples:** pre-registro, índices (p. ej., z-scores), correcciones por multiplicidad.

Chequeos de balance

Bajo aleatorización válida:

$$X \perp Z \Rightarrow \mathbb{E}[X \mid Z=1] = \mathbb{E}[X \mid Z=0]$$

- Tabla de balance: medias por grupo, diferencia, error estándar y p-valor.
- No “sobre-testear” pero reportar variables clave y estratos.
- Si estratificaste, controla por **dummies de estrato** en el análisis.

Estimación básica (sin y con controles)

Diferencia de medias / ITT:

$$\widehat{\text{ITT}} = \bar{Y}_{Z=1} - \bar{Y}_{Z=0}$$

Regresión equivalente:

$$Y_i = \alpha + \beta Z_i + \varepsilon_i \quad (\hat{\beta} = \widehat{\text{ITT}})$$

Con controles (precisión):

$$Y_i = \alpha + \beta Z_i + \gamma' X_i + \varepsilon_i$$

TOT/LATE vía IV:

$$Y_i = \alpha + \theta D_i + u_i, \quad D_i = \pi_0 + \pi_1 Z_i + v_i, \quad \hat{\theta} = \frac{\widehat{\text{ITT}}}{\widehat{\text{First Stage}}}$$

Código R mínimo para el aula

```
# Simular RCT simple
set.seed(123)
n <- 400
Z <- rbinom(n,1,0.5)
Y0 <- rnorm(n, 50, 10)
tau <- 5
Y1 <- Y0 + tau
Y <- ifelse(Z==1, Y1, Y0)
t.test(Y[Z==1], Y[Z==0])$estimate # ITT

# Incumplimiento
D <- ifelse(Z==1, rbinom(n,1,0.8), rbinom(n,1,0.1))
ITT <- mean(Y[Z==1]) - mean(Y[Z==0])
FS <- mean(D[Z==1]) - mean(D[Z==0])
TOT <- ITT/FS
```

Plantilla: Artículo del día (RCT)

- ➊ **Pregunta y contexto** (1 slide)
- ➋ **Diseño experimental**: unidad, asignación, estratos, poder (1–2)
- ➌ **Cumplimiento y amenazas**: attrition, spillovers, placebo (1)
- ➍ **Resultados**: ITT principal; TOT/LATE si aplica (2)
- ➎ **Robustez y heterogeneidad** (1)
- ➏ **Política pública**: magnitudes, costo-efectividad (1)

Ejercicio corto (10 min)

Escenario: Beca universitaria con cupos limitados. 1) Proponer unidad de aleatorización; 2) Una variable para estratificar y por qué; 3) Definir outcome primario y dos secundarios; 4) Escribir la *estimación ITT* que reportarías; 5) Una amenaza y cómo la mitigas.