

# Resultados Potenciales (Cont.)

Haciendo Economía I  
Econ 2205

Ignacio Sarmiento-Barbieri

Universidad de los Andes

October 8, 2024

# Anuncios

- ▶ Las referencias para estas clases van a ser:
  - 1 Mastering Metrics de Angrist y Pischke
  - 2 Mostly Harmless Econometrics de Angrist y Pischke
  - 3 Causal Inference: The Mixtape de Cunningham (disponible online en su pagina web)
- ▶ Quiz sobre Resultados Potenciales → 21 de Octubre

# Plan para hoy

- 1 Anuncios
- 2 Correlación no implica Causalidad Recap
- 3 Modelo de Resultados Potenciales
  - Ejemplo: El Experimento STAR

- 1 Anuncios
- 2 Correlación no implica Causalidad Recap
- 3 Modelo de Resultados Potenciales
  - Ejemplo: El Experimento STAR

# Correlación no implica Causalidad Recap

- ▶ En datos observacionales, es casi seguro que las correlaciones no reflejan una relación causal.
- ▶ Porqué?

# Correlación no implica Causalidad Recap

- ▶ En datos observacionales, es casi seguro que las correlaciones no reflejan una relación causal.
- ▶ Porqué?
  - 1 Simultaneidad
  - 2 Variables Omitidas.

# Correlación no implica Causalidad Recap

- ▶ En datos observacionales, es casi seguro que las correlaciones no reflejan una relación causal.
- ▶ Porqué?
  - 1 Simultaneidad
  - 2 Variables Omitidas.
  - 3 Comportamiento no aleatorio

# Correlación no implica Causalidad Recap

- ▶ En datos observacionales, es casi seguro que las correlaciones no reflejan una relación causal.
- ▶ Porqué?
  - 1 Simultaneidad
  - 2 Variables Omitidas.
  - 3 Comportamiento no aleatorio
    - ▶ Las variables fueron elegidas endógenamente por personas que estaban tomando decisiones que pensaban que eran las mejores.
    - ▶ Al maximizar sujeto a restricciones, eligieron ciertas cosas que crean una correlación falsa con otras cosas.
    - ▶ Esto se ve especialmente en el modelo de resultados potenciales (HOY)



# Modelo de Resultados Potenciales

- Pregunta de investigación: ¿ $D_i$  afecta a  $Y_i$ ?

# Modelo de Resultados Potenciales

- ▶ Pregunta de investigación: ¿ $D_i$  afecta a  $Y_i$ ?
- ▶ Tenemos:
  - ▶ Una variable binaria de tratamiento:

$$D_i = \{0, 1\} \tag{1}$$

- ▶ Resultado (Outcome) para el individuo  $i$ :

$$Y_i \tag{2}$$

- ▶ Note que el resultado se puede escribir como la combinación de 2 resultados potenciales

# Modelo de resultados potenciales

- Note que el resultado se puede escribir como la combinación de 2 resultados potenciales

$$Y_i = Y_{0i} + D_i (Y_{1i} - Y_{0i}) \quad (3)$$

- Para cada individuo  $i$ , hay dos **resultados potenciales** (con  $D_i$  binario)

# Modelo de resultados potenciales

- Note que el resultado se puede escribir como la combinación de 2 resultados potenciales

$$Y_i = Y_{0i} + D_i (Y_{1i} - Y_{0i}) \quad (3)$$

- Para cada individuo  $i$ , hay dos **resultados potenciales** (con  $D_i$  binario)
  - $Y_{1i}$  si  $D_i = 1$  Resultado de  $i$  si tiene seguro de salud

# Modelo de resultados potenciales

- Note que el resultado se puede escribir como la combinación de 2 resultados potenciales

$$Y_i = Y_{0i} + D_i (Y_{1i} - Y_{0i}) \quad (3)$$

- Para cada individuo  $i$ , hay dos **resultados potenciales** (con  $D_i$  binario)
  - $Y_{1i}$  si  $D_i = 1$  Resultado de  $i$  si tiene seguro de salud
  - $Y_{0i}$  si  $D_i = 0$  Resultado de  $i$  si no tiene seguro de salud

# Resultados potenciales

- ▶ Los resultados observables o "reales",  $Y_i$ , son distintos de los resultados potenciales.
- ▶ La forma en que pasamos de los resultados potenciales a los resultados reales es un movimiento filosófico importante. El resultado observable de una unidad es una función de sus resultados potenciales:

$$Y_i = Y_{0i} + D_i (Y_{1i} - Y_{0i}) \quad (4)$$

- ▶ Donde  $D_i = 1$  si la unidad recibió el tratamiento, 0 de lo contrario
- ▶ Notar que:
  - ▶ Cuando  $D_i = 1$ , entonces  $Y_{1i}$
  - ▶ Cuando  $D_i = 0$ ,  $Y_{0i}$ .

# Resultados potenciales

- La diferencia entre los dos resultados nos da el efecto causal del tratamiento  $D_i$ , i.e.,

$$\tau_i = Y_{1i} - Y_{0i} \quad (5)$$

# Modelo de Resultados potenciales

$$\tau_i = Y_{1i} - Y_{0i} \quad (6)$$

Individuo	Con Seguro de Salud $Y_{1i}$	Sin Seguro de Salud $Y_{0i}$	Efecto causal $\tau_i = Y_{1i} - Y_{0i}$
1	5	1	4
2	4	5	-1
3	4	1	3
4	3	4	-1
5	4	2	2



# Modelo de Resultados potenciales

## Problemas

- ▶ Esta ecuación simple:

$$\tau_i = Y_{1i} - Y_{0i} \quad (7)$$

- ▶ nos lleva al **problema fundamental de inferencia causal**.

# Modelo de Resultados Potenciales

## Problemas

- Nunca podemos observar simultáneamente  $Y_{1i}$  and  $Y_{0i}$ .



Source: <https://tinyurl.com/yv3v5tum>

# Modelo de Resultados potenciales

## Solución Propuesta

- ▶ Comparamos
  - ▶ resultado para los tienen seguro de salud ( $D_i = 1$ )
  - ▶ resultados para los que **no** tienen seguro de salud ( $D_i = 0$ )

Personas	$Y_i$	$D_i$
1	5	1
2	5	0
3	4	1
4	4	0
5	4	1

# Modelo de Resultados potenciales

## Solución Propuesta

- ▶ Comparamos
  - ▶ resultado para los tienen seguro de salud ( $D_i = 1$ )
  - ▶ resultados para los que **no** tienen seguro de salud ( $D_i = 0$ )

Personas	$Y_i$	$D_i$
1	5	1
2	5	0
3	4	1
4	4	0
5	4	1

- ▶ Es decir, esta comparación nos da *una* respuesta, pero es esta **la** respuesta que queremos?

# Modelo de Resultados potenciales

## Solución Propuesta

- ▶ Ahora supongamos que el seguro de salud mejora la salud de todos en una cantidad constante  $\tau$

$$Y_{1i} = Y_{0i} + \tau \quad (8)$$

- ▶ O de forma equivalente

$$Y_{1i} - Y_{0i} = \tau \quad (9)$$

- ▶ En otras palabras,  $\tau$ , es tanto el efecto individual como el efecto causal promedio del seguro de salud en la salud.

# Modelo de Resultados potenciales

## Solución Propuesta

- ▶ La pregunta que surge entonces es como se relaciona la comparacion que hicimos con el efecto causal
- ▶ Partiendo de

$$Y_{1i} = Y_{0i} + \tau \quad (10)$$

- ▶ tomando esperanzas condicionadas

$$E[Y_{1i} | D_i = 1] = E[Y_{0i} | D_i = 1] + \tau \quad (11)$$

# Modelo de Resultados potenciales

## Solución Propuesta

► Entonces

$$\underbrace{E[Y_i | D_i = 1] - E[Y_i | D_i = 0]}_{\text{Diferencia en el promedio de los grupos}} = \quad (12)$$

$$\underbrace{\tau}_{\text{Efecto Promedio Causal ☺}} + E[Y_{0i} | D_i = 1] - E[Y_{0i} | D_i = 0] \quad (13)$$

# Modelo de Resultados Potenciales

## Solución Propuesta

► Entonces

$$\underbrace{E[Y_i | D_i = 1] - E[Y_i | D_i = 0]}_{\text{Diferencia en el promedio de los grupos}} = \quad (14)$$

$$\underbrace{\tau}_{\text{Efecto Promedio Causal ☺}} + \underbrace{E[Y_{0i} | D_i = 1] - E[Y_{0i} | D_i = 0]}_{\text{Sesgo de Selección ☹}} \quad (15)$$



# Modelo de Resultados Potenciales

## Ejemplo: Programas de entrenamiento

- ▶ Los gobiernos suelen hacer programas de entrenamiento para ayudar a trabajadores necesitados
- ▶ La pregunta es: Estos programas tienen los efectos deseados (por ej. subir salarios)?
- ▶ Los estudios observacionales que comparan los datos de salarios de participantes y no participantes suelen encontrar que aquellos que terminan estos programas terminan con salarios mas bajos.

# Modelo de Resultados Potenciales

## Ejemplo: Programas de entrenamiento

Nearly All Programs Track Multiple Outcome Measures, but Little is Known about Program Effectiveness

Almost all programs tracked multiple outcome measures related to employment and training, and many programs tracked similar measures. Forty-one of the 47 programs tracked at least three outcome measures in fiscal year 2009, according to officials. The most frequently tracked outcome measure was “entered employment”—the number of program participants who found jobs (see table 2). Many programs also tracked “employment retention” and “wage gain or change.” These are the types of measures developed under the Office of Management and Budget’s (OMB) common measures initiative, which sought to unify definitions for performance across programs with similar goals.<sup>19</sup> Three programs did not track any outcome measures at the federal level in fiscal year 2009.<sup>20</sup> For a detailed list of outcome measures tracked by federal employment and training programs, see appendix V.

Table 2: Outcome Measures Tracked Most Frequently by Programs in Fiscal Year 2009

Outcome measures	Number of programs measuring this outcome
Entered employment	38
Employment retention	29
Wage gain or change	23
Credential attainment	19
Other “positive outcomes” <sup>a</sup>	17
Educational attainment	16
Customer satisfaction	8
Other outcomes <sup>b</sup>	23
No outcome measures	3

Source: Government Accountability Office, “Multiple Employment and Training Programs,” GAO-11-92, January 2011, p. 11.

# Modelo de Resultados Potenciales

Ejemplo: Programas de entrenamiento

## Downsizing the Federal Government

YOUR GUIDE TO CUTTING FEDERAL SPENDING

HOME

ABOUT

BLOG

FEDERAL POLICY BASICS

SPENDING CHARTS

### FEDERAL DEPARTMENTS

Agriculture

Commerce

Defense

## Employment and Training Programs: Ineffective and Unneeded

by Chris Edwards and Daniel J. Murphy

June 1, 2011

Source: <https://www.downsizinggovernment.org/labor/employment-training-programs>

# Modelo de Resultados Potenciales

## Ejemplo: Programas de entrenamiento

- ▶ Los gobiernos suelen hacer programas de entrenamiento para ayudar a trabajadores necesitados
- ▶ La pregunta es: Estos programas tienen los efectos deseados (por ej. subir salarios)?
- ▶ Los estudios observacionales que comparan los datos de salarios de participantes y no participantes suelen encontrar que aquellos que terminan estos programas terminan con salarios mas bajos.
- ▶ El problema es que los participantes se auto seleccionan
- ▶ y que los participantes suelen tener salarios mas bajos

# Modelo de Resultados Potenciales

## Ejemplo: Programas de entrenamiento

- ▶ Como formalizamos esta preocupación en nuestro modelo?
- ▶ Evaluación observacional de programas

$$E[\text{Salario}_i \mid \text{Programa}_i = 1] - E[\text{Salario}_i \mid \text{Programa}_i = 0] = \quad (16)$$

# Modelo de Resultados Potenciales

## Ejemplo: Programas de entrenamiento

- ▶ Como formalizamos esta preocupación en nuestro modelo?
- ▶ Evaluación observacional de programas

$$E[\text{Salario}_i \mid \text{Programa}_i = 1] - E[\text{Salario}_i \mid \text{Programa}_i = 0] = \quad (16)$$

$$\underbrace{E[\text{Salario}_{1i} \mid \text{Programa}_i = 1] - E[\text{Salario}_{0i} \mid \text{Programa}_i = 1]} + \quad (17)$$

Efecto causal promedio del programa de entrenamiento en los salarios de los participantes *i.e.*,  $\tau$

$$\underbrace{E[\text{Salario}_{0i} \mid \text{Programa}_i = 1] - E[\text{Salario}_{0i} \mid \text{Programa}_i = 0]} \quad (18)$$

Sesgo de Selección

- ▶ Si el programa atrae/selecciona individuos que en promedio, tienen salarios menores en ausencia del programa, entonces tenemos selección negativa

# Modelo de Resultados Potenciales

Ejemplo: Programas de entrenamiento

$$\underbrace{E[\text{Salario}_{1i} \mid \text{Programa}_i = 1] - E[\text{Salario}_{0i} \mid \text{Programa}_i = 1]} + \quad (19)$$

Efecto causal promedio del programa de entrenamiento en los salarios de los participantes *i.e.*,  $\tau$

$$\underbrace{E[\text{Salario}_{0i} \mid \text{Programa}_i = 1] - E[\text{Salario}_{0i} \mid \text{Programa}_i = 0]}_{\text{Sesgo de Selección}} \quad (20)$$

- ▶ Aun si el programa tiene, en promedio, un efecto positivo sobre el salario (en el grupo de participantes),  $\tau > 0$ , vamos a detectar efectos menores debido al sesgo de selección
- ▶ Si los efectos son lo suficientemente grandes (relativo al tratamiento), podemos tener el signo equivocado
- ▶ Entonces para evaluar es importante romper el sesgo de selección

# Modelo de Resultados potenciales

## Solución Propuesta

*The goal of most empirical economic research is to overcome selection bias, and therefore to say something about the causal effect of a variable like  $D_i$ .*

*Angrist and Pischke (MHE, p. 15),*

- ▶ Los experimentos nos van a ayudar a resolver el problema de sesgo de selección



# Ejemplo: Tamaño de la clase y aprendizaje de estudiantes

## El Experimento STAR

- ▶ Estudios **observacionales** (no experimentales) sugieren que existe poca o ninguna relación entre el tamaño de la clase y el aprendizaje de los estudiantes.
- ▶ De ser cierto, esto implicaría que se puede ahorrar dinero contratando menos maestros sin la consecuente reducción en el rendimiento.
- ▶ Problema: estudiantes más débiles suelen ser colocados en clases más pequeñas

# Ejemplo: Tamaño de la clase y aprendizaje de estudiantes

## El Experimento STAR

- ▶ Estudios **observacionales** (no experimentales) sugieren que existe poca o ninguna relación entre el tamaño de la clase y el aprendizaje de los estudiantes.
- ▶ De ser cierto, esto implicaría que se puede ahorrar dinero contratando menos maestros sin la consecuente reducción en el rendimiento.
- ▶ Problema: estudiantes más débiles suelen ser colocados en clases más pequeñas
- ▶ Un **experimento** aleatorio puede solucionar esto: aleatorizar estudiantes a clases de diferentes tamaños
- ▶ Esta es la idea del proyecto STAR de Tennessee

# Ejemplo: Tamaño de la clase y aprendizaje de estudiantes

## El Experimento STAR

- ▶ El experimento STAR fue ambicioso e influyente
- ▶ El experimento asignó a los estudiantes a uno de tres tratamientos:
  - 1 Clases **pequeñas** con 13 – 17 niños,
  - 2 Clases **regulares** con 22 – 25 estudiantes y un profesor asistente tiempo parcial (**grupo control**),
  - 3 Clases **regulares** con un profesor asistente de tiempo completo
- ▶ Costó alrededor de \$12 millones y se implementó para una cohorte de niños de jardín de infantes en 1985/86.
- ▶ El estudio duró cuatro años e involucró a unos 11.600 niños.

# Ejemplo: Tamaño de la clase y aprendizaje de estudiantes

## El Experimento STAR

- ▶ Lo *primero* que tenemos que preguntarnos: Funcionó la aleatorización?
- ▶ Idealmente, tendríamos data pretratamiento sobre el resultado.
- ▶ Desafortunadamente, ellos solo tenían info demográfica.

Variable	Pequeña	Regular	Regular con Asistente	P-value conjunto
Almuerzo Gratis	0.47	0.48	0.50	0.09
Blanco/Asiático	0.68	0.67	0.66	0.26
Edad en 1985	5.44	5.43	5.42	0.32

# Ejemplo: Tamaño de la clase y aprendizaje de estudiantes

## El Experimento STAR

- ▶ Lo *primero* que tenemos que preguntarnos: Funcionó la aleatorización?
- ▶ Idealmente, tendríamos data pretratamiento sobre el resultado.
- ▶ Desafortunadamente, ellos solo tenían info demográfica.

Variable	Pequeña	Regular	Regular con Asistente	P-value conjunto
Almuerzo Gratis	0.47	0.48	0.50	0.09
Blanco/Asiático	0.68	0.67	0.66	0.26
Edad en 1985	5.44	5.43	5.42	0.32
Tamaño de clase	15.10	22.40	22.80	0.00

# Ejemplo: Tamaño de la clase y aprendizaje de estudiantes

## El Experimento STAR

- ▶ Lo *primero* que tenemos que preguntarnos: Funcionó la aleatorización?
- ▶ Idealmente, tendríamos data pretratamiento sobre el resultado.
- ▶ Desafortunadamente, ellos solo tenían info demográfica.

Variable	Pequeña	Regular	Regular con Asistente	P-value conjunto
Almuerzo Gratis	0.47	0.48	0.50	0.09
Blanco/ Asiático	0.68	0.67	0.66	0.26
Edad en 1985	5.44	5.43	5.42	0.32
Tamaño de clase	15.10	22.40	22.80	0.00
Percentil Resultado de Prueba	54.70	48.90	50.00	0.00

Fuente: Adaptación tabla 2.2.1 MHE