## Clase 22: Resultados Potenciales en Regresiones y Ejemplos Haciendo Economía I Econ 2205

Ignacio Sarmiento-Barbieri

Universidad de los Andes

October 25, 2023

### **Anuncios**

- Las referencias para esta clase siguen siendo:
  - 1 Mastering Metrics de Angrist y Pischke (cap 1)
  - 2 Mostly Harmless Econometrics de Angrist y Pischke (cap 2)
  - 3 Causal Inference: The Mixtape de Cunningham (cap 4) (disponible online en su pagina web)
- ▶ Quiz sobre Resultados Potenciales → Martes 7 de Noviembre

# Plan para hoy

1 Anuncios

- 2 Modelo de Resultados Potenciales Recap
  - Ejemplo: El Experimento STAR
- 3 Análisis de Regresión en Experimentos
  - Ejemplo: El Experimento STAR (cont.)
  - Ejemplo: Aprendiendo sobre resultados de prueba de HIV

▶ Una variable binaria de tratamiento (e.g., va al hospital, no va):

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

▶ Resultado (Outcome) para el individuo *i* (e.g., salud):

$$Y_i$$
 (2)

- ▶ El modelo tiene varios nombres...
  - Splawa-Neyman marco de resultados potenciales
  - Rubin modelo causal
  - ► Neyman-Rubin (Splawa-Neyman-Rubin)

October 25, 2023

3/42

- Pregunta de investigación:  $D_i$  afecta a  $Y_i$ ?
- ▶ Para cada individuo *i*, hay dos **resultados potenciales** (con D<sub>i</sub> binario)
  - ▶  $Y_{1i}$  si  $D_i$  = 1 Resultado de i si va al hospital
  - Y<sub>0i</sub> si  $D_i = 0$  Resultado de i si no va al hospital
- La diferencia entre los dos resultados nos da el efecto causal del tratamiento de ir al hospital, i.e.,

$$\tau_i = Y_{1i} - Y_{0i} \tag{3}$$

4/42

Sarmiento-Barbieri (Uniandes) Clase 22 October 25, 2023

#### **Problemas**

► Esta ecuación simple:

$$\tau_i = \mathbf{Y}_{1i} - \mathbf{Y}_{0i} \tag{4}$$

- ▶ nos lleva al **problema fundamental de inferencia causal**.
- ▶ Nunca podemos observar simultáneamente  $Y_{1i}$  and  $Y_{0i}$ .



Source: https://tinyurl.com/yv3v5tum

#### Solución Propuesta

- Comparamos
  - resultado para los que van al hospital  $(Y_{1i} | D_i = 1)$
  - resultados para los que **no** van al hospital  $(Y_{0i} | D_i = 0)$

$$E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0]$$
(5)

• que nos da la diferencia observada en los resultados de salud

6 / 42

Solución Propuesta

Entonces

$$E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] = E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]$$
(6)

Haciendo un poco de matemática creativa

$$= E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 1] + E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]$$
(7)

Solución Propuesta

### Que nos dice esto entonces?

$$E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] =$$
(8)

$$E[Y_{1i} | D_i = 1] - E[Y_{0i} | D_i = 1] + E[Y_{0i} | D_i = 1] - E[Y_{0i} | D_i = 0]$$
(9)

Efecto Promedio sobre los tratados (ATT) ©

8 / 42

Solución Propuesta

#### Que nos dice esto entonces?

$$E[Y_i \mid \mathbf{D}_i = 1] - E[Y_i \mid \mathbf{D}_i = 0] = \tag{10}$$

$$=\underbrace{E\left[Y_{1i}\mid D_{i}=1\right]-E\left[Y_{0i}\mid D_{i}=1\right]}_{\text{Efecto Promedio sobre los tratados (ATT) }\odot}+\underbrace{E\left[Y_{0i}\mid D_{i}=1\right]-E\left[Y_{0i}\mid D_{i}=0\right]}_{\text{Sesgo de Selección }\odot}$$
(11)

#### Solución Propuesta

▶ El **primer término** es la *variación buena* que es la respuesta que queremos

$$E[Y_{1i} | D_i = 1] - E[Y_{0i} | D_i = 1]$$
(12)

$$= E[Y_{1i} - Y_{0i} \mid D_i = 1]$$
 (13)

$$= E\left[\tau_i \mid D_i = 1\right] \tag{14}$$

► El efecto causal promedio de la hospitalización para los *individuos hospitalizados* (ATT).

Solución Propuesta

▶ El **segundo término** es la *variación mala* que no nos dejan obtener la respuesta que queremos

$$E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]$$
(15)

- ▶ La diferencia en el promedio de resultados para los no tratados entre el tratamiento y el control.
- ▶ Sesgo de selección cuan malo es el grupo de control como contra factual.

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

#### 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." Three programs did not track any outcome measures at the federal level in fiscal year 2009. "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"	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.

Ejemplo: Programas de entrenamiento

# Downsizing the Federal Government

YOUR GUIDE TO CUTTING FEDERAL SPENDING



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

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

#### Ejemplo: Programas de entrenamiento

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

$$E\left[\text{Salario}_{i} \mid \text{Programa}_{i} = 1\right] - E\left[\text{Salario}_{i} \mid \text{Programa}_{i} = 0\right] = \tag{16}$$

$$E[Salario_{1i} | Programa_i = 1] - E[Salario_{0i} | Programa_i = 1] + (17)$$

Efecto causal promedio del programa de entrenamiento en los salarios de los participantes i.e.,  $\overline{ au}$ 

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

Sesgo de Seleccion



#### Ejemplo: Tamaño de la clase y aprendizaje de estudiantes

- Uno de los insumos más costoso en la función de producción educativa es el tamaño de la clase:
- ► Clases mas pequeñas → mas maestros
- Fundamental entender el beneficio de tamaño de la clase más pequeño
- 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.

Ejemplo: Tamaño de la clase y aprendizaje de estudiantes

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

$$E[Calificacion_i | Clase Pequeña_i = 1] - E[Calificacion_i | Clase Pequeña_i = 0] = (19)$$

#### Experimentos

▶ Como hacen los experimentos para resolver el sesgo de selección?

#### **Experimentos**

- ► Como hacen los experimentos para resolver el sesgo de selección?
- Los experimentos rompen el link entre los resultados potenciales y el tratamiento

#### Experimentos

- ► Como hacen los experimentos para resolver el sesgo de selección?
- ▶ Los experimentos rompen el link entre los resultados potenciales y el tratamiento
- ► En otras palabras, asignando aleatoriamente  $D_i$  hace que  $D_i$  sea independiente de que resultado observamos (es decir  $Y_{1i}$  o  $Y_{0i}$ )

$$\{\mathbf{Y}_{1i}, \mathbf{Y}_{0i}\} \perp D_i \tag{20}$$

- La asignación aleatoria implica que los grupos de tratamiento y control salen de la misma población subyacente.
- ▶ Son iguales en todos los sentidos, incluyendo sus  $E[Y_{0i}]$

Experimentos

- 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

- 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

- ► 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.

Lo primero que tenemos que preguntarnos: Funcionó la aleatorización?

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

Lo primero que tenemos que preguntarnos: Funcionó la aleatorización?

	Variable	Pequeña	Regular	Regular con Asistente	P-value conjunto
Α.	lmorra Cratic	0.47	0.49	0.50	0.09
	lmuerzo Gratis lanco/Asiático	0.47 0.68	$0.48 \\ 0.67$	0.50 0.66	0.09
	Edad en 1985	5.44	5.43	5.42	0.32
Ta	imaño de clase	15.10	22.40	22.80	0.00

Lo primero que tenemos que preguntarnos: Funcionó la aleatorización?

Variable	Pequeña	Regular	Regular con Asistente	P-value conjunto
Almuerzo Gratis	0.47	0.48	0.50	0.09
Blanco/Asiático	0.68	0.40	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

1 Anuncios

- 2 Modelo de Resultados Potenciales Recap
  - Ejemplo: El Experimento STAR

- 3 Análisis de Regresión en Experimentos
  - Ejemplo: El Experimento STAR (cont.)
  - Ejemplo: Aprendiendo sobre resultados de prueba de HIV

- La tabla previa estima/compara el efecto del tratamiento (ATE) haciendo diferencias de medias .
- Podemos hacer lo mismo con una regresión.
- Específicamente, si hacemos la regresión del "outcome" (percentil de la prueba) en una variable dummy para cada grupo de tratamiento.

► Asumimos que el efecto del tratamiento es constante ("homogeneo")

25 / 42

#### El Experimento STAR

Variable

Clase Pequeña

Regular + Asistente

Blanco/Asiatico

Mujer

Almuerzo Gratis

Efecto Fijo Escuela

Fuente: Adaptación tabla 2.2.2 MHE

#### El Experimento STAR

Variable	(1)
Class Passass	4.92
Clase Pequeña	4.82 (2.19)
Regular + Asistente	0.12 (2.23)
Blanco/Asiatico	(2.23)
Mujer	
Almuerzo Gratis	
Efecto Fijo Escuela	No
Euente: Adaptación ta	bla 2 2 2 MHE

Fuente: Adaptación tabla 2.2.2 MHE

#### El Experimento STAR

Variable	(1)	(2)
Clase Pequeña Regular + Asistente	4.82 (2.19) 0.12 (2.23)	5.37 (1.26) 0.29 (1.13)
Blanco/Asiatico	()	(====)
Mujer		
Almuerzo Gratis		
Efecto Fijo Escuela	No	Si

Fuente: Adaptación tabla 2.2.2 MHE

## El Experimento STAR

▶ Si se asigno aleatoriamente y esto elimina el sesgo de selección por qué la regresión incluye controles?

## El Experimento STAR

Si se asigno aleatoriamente y esto elimina el sesgo de selección por qué la regresión incluye controles?

Los controles juegan dos roles en los análisis de regresión de datos experimentales.

Il El diseño experimental STAR utilizó asignación aleatoria condicional: la asignación a clases de diferentes tamaños fue aleatoria dentro de las escuelas, pero no entre escuelas.

Start with raw data. Correlation between X and Y: -0.039

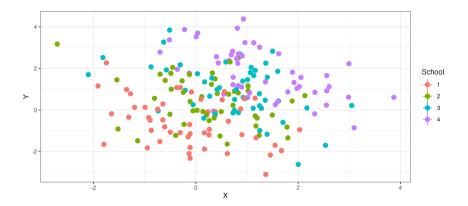
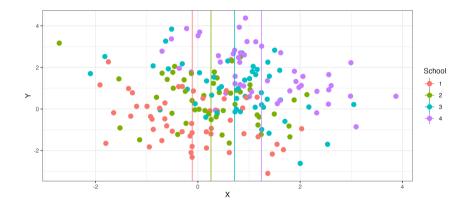


Figure out any between-School differences in X



Remove all between-School differences in X

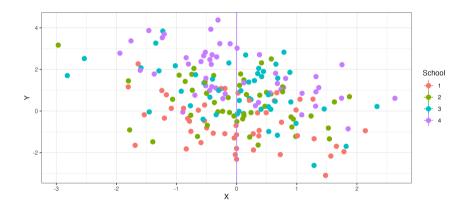
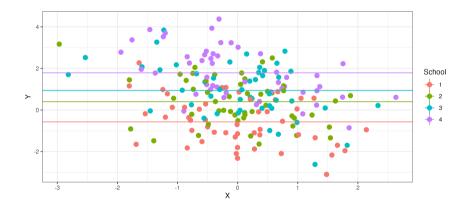
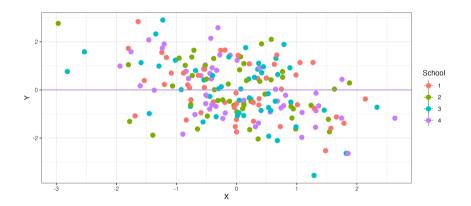


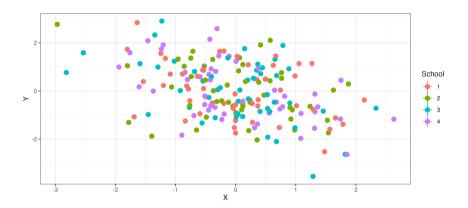
Figure out any between-School differences in Y



Remove all between-School differences in Y



Analyze what's left! Within-School Correlation Between X and Y: -0.429



#### Análisis de Regresión en Experimentos

#### El Experimento STAR

Variable	(1)	(2)	(3)
Clase Pequeña	4.82	5.37	5.36
	(2.19)	(1.26)	(1.21)
Regular + Asistente	0.12	0.29	0.53
	(2.23)	(1.13)	(1.09)
Blanco/Asiatico			8.35
			(1.35)
Mujer			4.48
			(0.63)
Almuerzo Gratis			-13.15
			(0.77)
Efecto Fijo Escuela	No	Si	Si

Fuente: Adaptación tabla 2.2.2 MHE

## El Experimento STAR

Si se asigno aleatoriamente y esto elimina el sesgo de selección por qué la regresión incluye controles?

Los controles juegan dos roles en los análisis de regresión de datos experimentales.

- El diseño experimental STAR utilizó asignación aleatoria condicional: la asignación a clases de diferentes tamaños fue aleatoria dentro de las escuelas, pero no entre escuelas.
- 2 Como vimos en la tabla  $Cov(D_i, X_i) = 0$ , sin embargo la inclusión de control aumenta la precisión.

36 / 42

1 Anuncios

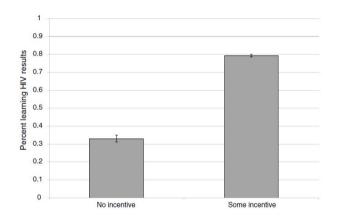
- 2 Modelo de Resultados Potenciales Recap
  - Ejemplo: El Experimento STAR

- 3 Análisis de Regresión en Experimentos
  - Ejemplo: El Experimento STAR (cont.)
  - Ejemplo: Aprendiendo sobre resultados de prueba de HIV

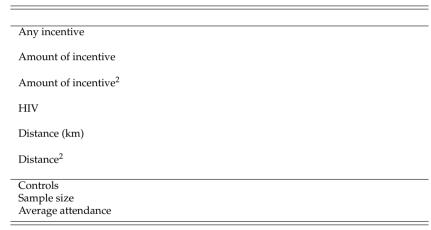
- La idea es que si las personas conocen su estado sobre VIH tomarían precauciones en caso de ser positivo, reduciendo la tasa de infección
- ▶ Si usamos datos observacionales, el problema que surge es que las personas se auto seleccionan a aprender sobre su estado de salud.
- Individuos que se testean también son mas probables que tengan conductas menos riesgosas.
- ▶ Para romper esta dependencia es necesario un experimento

37 / 42

- ▶ Thornton fue hasta Malawi rural e hizo un experimento
- ▶ Fueron puerta a puerta ofreciendo pruebas de HIV gratuitos
- Les dieron aleatoriamente vouchers (o no) entre \$ 1 y \$ 3
- La gente los podía cambiar un vez que visitaban el centro de pruebas mas cercanos



39 / 42



	1
Any incentive	0.431***
•	(0.023)
Amount of incentive	
Amount of incentive <sup>2</sup>	
HIV	-0.055*
	(0.031)
Distance (km)	
Distance <sup>2</sup>	
Controls	Yes
Sample size	2,812
Average attendance	0.69

Nota: \*\*\* Significativamente diferente de cero a un nivel de confianza del 99 por ciento. \*\* Significativamente diferente de cero a un nivel de confianza del 95 por ciento. \* Significativamente diferente de cero a un nivel de confianza del 90 por ciento.

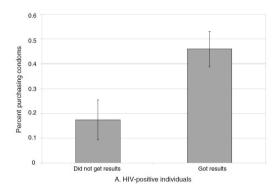
40 / 42

	1	2
Any incentive	0.431***	0.309***
•	(0.023)	(0.026)
Amount of incentive		0.091***
		(0.012)
Amount of incentive <sup>2</sup>		
HIV	-0.055*	-0.052
	(0.031)	(0.032)
Distance (km)	, ,	
Distance <sup>2</sup>		
Controls	Yes	Yes
Sample size	2,812	2,812
Average attendance	0.69	0.69

	1	2	3
Any incentive	0.431***	0.309***	0.219***
	(0.023)	(0.026)	(0.029)
Amount of incentive		0.091***	0.274***
		(0.012)	(0.036)
Amount of incentive <sup>2</sup>			-0.063***
			(0.011)
HIV	-0.055*	-0.052	-0.05
	(0.031)	(0.032)	(0.032)
Distance (km)			
Distance <sup>2</sup>			
G . 1			
Controls	Yes	Yes	Yes
Sample size	2,812	2,812	2,812
Average attendance	0.69	0.69	0.69

	1	2	3	4
Any incentive	0.431***	0.309***	0.219***	0.220***
	(0.023)	(0.026)	(0.029)	(0.029)
Amount of incentive		0.091***	0.274***	0.274***
		(0.012)	(0.036)	(0.035)
Amount of incentive <sup>2</sup>			-0.063***	-0.063***
			(0.011)	(0.011)
HIV	-0.055*	-0.052	-0.05	-0.058*
	(0.031)	(0.032)	(0.032)	(0.031)
Distance (km)				-0.076***
				(0.027)
Distance <sup>2</sup>				0.010**
				(0.005)
Controls	Yes	Yes	Yes	Yes
Sample size	2,812	2,812	2,812	2,812
Average attendance	0.69	0.69	0.69	0.69

	1	2	3	4	5
Any incentive	0.431***	0.309***	0.219***	0.220***	0.219 ***
•	(0.023)	(0.026)	(0.029)	(0.029)	(0.029)
Amount of incentive		0.091***	0.274***	0.274***	0.273***
		(0.012)	(0.036)	(0.035)	(0.036)
Amount of incentive <sup>2</sup>			-0.063***	-0.063***	-0.063***
			(0.011)	(0.011)	(0.011)
HIV	-0.055*	-0.052	-0.05	-0.058*	-0.055*
	(0.031)	(0.032)	(0.032)	(0.031)	(0.031)
Distance (km)				-0.076***	
				(0.027)	
Distance <sup>2</sup>				0.010**	
				(0.005)	
Controls	Yes	Yes	Yes	Yes	Yes
Sample size	2,812	2,812	2,812	2,812	2,812
Average attendance	0.69	0.69	0.69	0.69	0.69



Dependent variables:	Bought condoms	Number of condoms bought		
	1	2		
Got results	-0.022	-0.193		
	(0.025)	(0.148)		
Got results x HIV	0.418***	1.778**		
	(0.143)	(0.564)		
HIV	-0.175**	-0.873		
	(0.085)	(0.275)		
Controls	Yes	Yes		
Sample size	1,008	1,008		
Mean	0.26	0.95		