

Resultados Potenciales en Regresiones y Ejemplos

Haciendo Economía I
Econ 2205

Ignacio Sarmiento-Barbieri

Universidad de los Andes

October 12, 2024

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 → 21 de Octubre

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

Modelo de Resultados Potenciales

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

$$D_i = \{0, 1\} \quad (1)$$

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

$$Y_i \quad (2)$$

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

Modelo de Resultados Potenciales

- ▶ Pregunta de investigación: D_i afecta a Y_i ?
- ▶ Para cada individuo i , hay dos **resultados potenciales** (con D_i binario)
 - ▶ Y_{1i} si $D_i = 1$ Resultado de i si va al hospital
 - ▶ Y_{0i} 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.,

$$\tau_i = Y_{1i} - Y_{0i} \quad (3)$$

Modelo de Resultados potenciales

Problemas

- ▶ Esta ecuación simple:

$$\tau_i = Y_{1i} - Y_{0i} \quad (4)$$

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



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

Modelo de Resultados potenciales

Solución Propuesta

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

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

- ▶ que nos da la *diferencia observada en los resultados de salud*

Modelo de Resultados potenciales

Solución Propuesta

Modelo de Resultados Potenciales

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.

Modelo de Resultados Potenciales

Ejemplo: Tamaño de la clase y aprendizaje de estudiantes

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

Modelo de Resultados Potenciales

Experimentos

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

Modelo de Resultados Potenciales

Experimentos

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

Modelo de Resultados Potenciales

Experimentos

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

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

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

Análisis de Regresión en Experimentos

- ▶ 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.

Análisis de Regresión en Experimentos

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

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

Análisis de Regresión en Experimentos

El Experimento STAR

Variable

Clase Pequeña

Regular + Asistente

Blanco/ Asiatico

Mujer

Almuerzo Gratis

Efecto Fijo Escuela

Fuente: Adaptación tabla 2.2.2 MHE

Análisis de Regresión en Experimentos

El Experimento STAR

Variable	(1)
Clase Pequeña	4.82 (2.19)
Regular + Asistente	0.12 (2.23)
Blanco/ Asiatico	
Mujer	
Almuerzo Gratis	
Efecto Fijo Escuela	No

Fuente: Adaptación tabla 2.2.2 MHE

Análisis de Regresión en Experimentos

El Experimento STAR

Variable	(1)	(2)
Clase Pequeña	4.82 (2.19)	5.37 (1.26)
Regular + Asistente	0.12 (2.23)	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

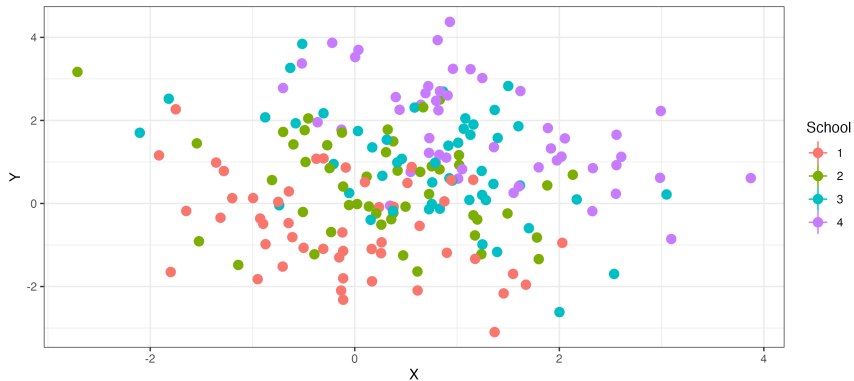
- ▶ Si se asigno aleatoriamente y esto elimina el sesgo de selección por qué la regresión incluye controles?

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

- 1 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.

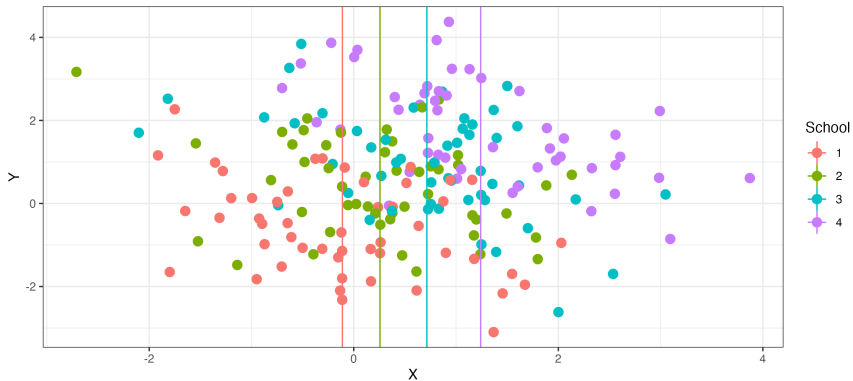
Detour: Efectos Fijos

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



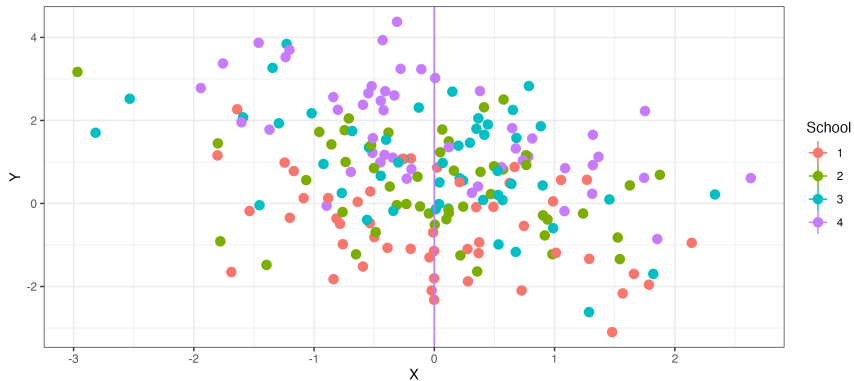
Detour: Efectos Fijos

Figure out any between-School differences in X



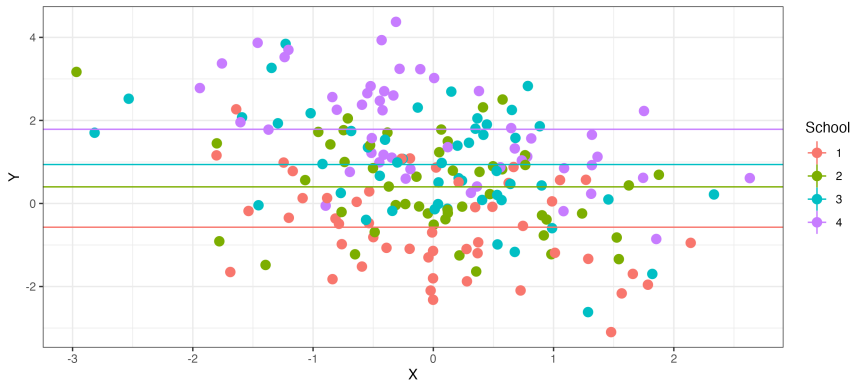
Detour: Efectos Fijos

Remove all between-School differences in X



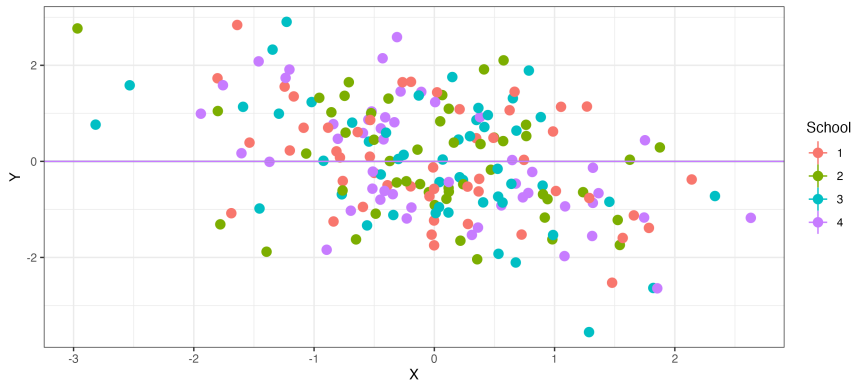
Detour: Efectos Fijos

Figure out any between-School differences in Y



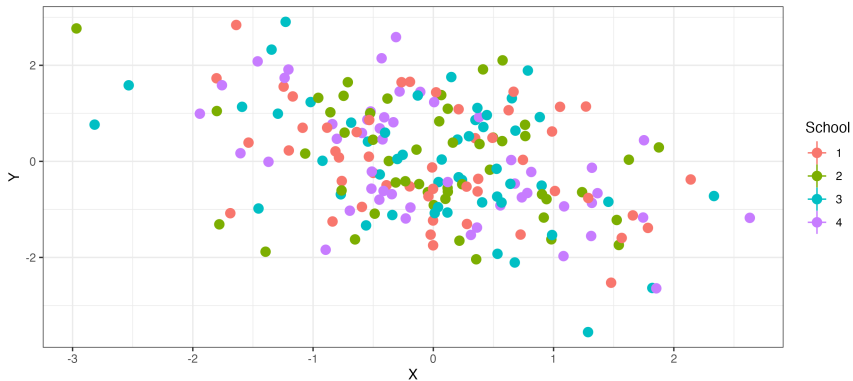
Detour: Efectos Fijos

Remove all between-School differences in Y



Detour: Efectos Fijos

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 (2.19)	5.37 (1.26)	5.36 (1.21)
Regular + Asistente	0.12 (2.23)	0.29 (1.13)	0.53 (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.

- 1 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.

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

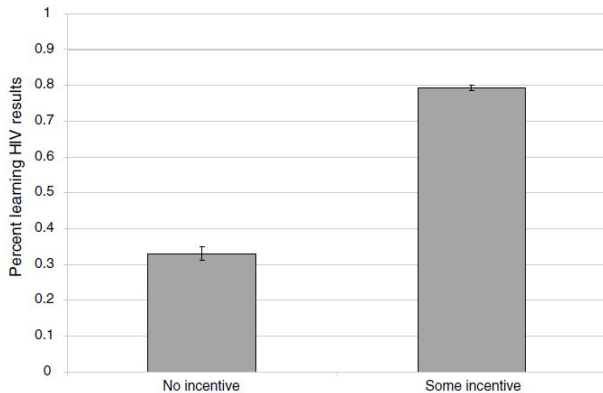
Aprendiendo sobre resultados de prueba de HIV. Thornton (2008)

- ▶ 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

Aprendiendo sobre resultados de prueba de HIV. Thornton (2008)

- ▶ 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

Aprendiendo sobre resultados de prueba de HIV. Thornton (2008)



Aprendiendo sobre resultados de prueba de HIV. Thornton (2008)

Any incentive

Amount of incentive

Amount of incentive²

HIV

Distance (km)

Distance²

Controls

Sample size

Average attendance

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.

Apreniendo sobre resultados de prueba de HIV. Thornton (2008)

	1
Any incentive	0.431*** (0.023)
Amount of incentive	
Amount of incentive ²	
HIV	-0.055* (0.031)
Distance (km)	
Distance ²	
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.

Aprendiendo sobre resultados de prueba de HIV. Thornton (2008)

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

Aprendiendo sobre resultados de prueba de HIV. Thornton (2008)

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

Aprendiendo sobre resultados de prueba de HIV. Thornton (2008)

	1	2	3	4
Any incentive	0.431*** (0.023)	0.309*** (0.026)	0.219*** (0.029)	0.220*** (0.029)
Amount of incentive		0.091*** (0.012)	0.274*** (0.036)	0.274*** (0.035)
Amount of incentive ²			-0.063*** (0.011)	-0.063*** (0.011)
HIV	-0.055* (0.031)	-0.052 (0.032)	-0.05 (0.032)	-0.058* (0.031)
Distance (km)				-0.076*** (0.027)
Distance ²				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

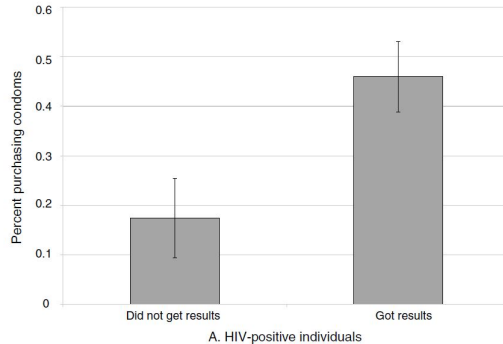
Nota: *** Significativamente diferente de cero a un nivel de confianza del 99 por ciento. ** Significativamente diferente de cero a un nivel de confianza del 95 por ciento. * Significativamente diferente de cero a un nivel de confianza del 90 por ciento.

Aprendiendo sobre resultados de prueba de HIV. Thornton (2008)

	1	2	3	4	5
Any incentive	0.431*** (0.023)	0.309*** (0.026)	0.219*** (0.029)	0.220*** (0.029)	0.219*** (0.029)
Amount of incentive		0.091*** (0.012)	0.274*** (0.036)	0.274*** (0.035)	0.273*** (0.036)
Amount of incentive ²			-0.063*** (0.011)	-0.063*** (0.011)	-0.063*** (0.011)
HIV	-0.055* (0.031)	-0.052 (0.032)	-0.05 (0.032)	-0.058* (0.031)	-0.055* (0.031)
Distance (km)				-0.076*** (0.027)	
Distance ²				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

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.

Apreniendo sobre resultados de prueba de HIV. Thornton (2008)



Aprendiendo sobre resultados de prueba de HIV. Thornton (2008)

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