

# Clase 22: Resultados Potenciales en Regresiones y Ejemplos

Haciendo Economía I  
Econ 2205

Ignacio Sarmiento-Barbieri

Universidad de los Andes

October 27, 2022

- ▶ Martes **Resultados Potenciales y el Experimento Ideal**
- ▶ Las referencias para esta clase siguen siendo:
  - 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)

## Tarea 8: Resultados Potenciales y el Experimento Ideal

**Instrucciones:** Esta tarea debe realizarse de forma individual y será el insumo para la discusión en grupo en clase.

- 1 Describa su PPR en términos del experimento ideal. Es decir, ¿cuál es el experimento que le permitiría responder causalmente su PPR?
- 2 ¿Porqué utilizar datos observacionales no le permitirían responder su PPR?
- 3 Busque un artículo academico publicado que trate de responder su PPR causalmente.

# 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

$$E[\text{Calificacion}_i \mid \text{Clase Pequeña}_i = 1] - E[\text{Calificacion}_i \mid \text{Clase Pequeña}_i = 0] = \quad (1)$$

# Modelo de Resultados Potenciales

Ejemplo: Tamaño de la clase y aprendizaje de estudiantes

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

$$E[\text{Calificacion}_i \mid \text{Clase Pequeña}_i = 1] - E[\text{Calificacion}_i \mid \text{Clase Pequeña}_i = 0] = \quad (1)$$

$$\underbrace{E[\text{Calificacion}_{1i} \mid \text{Clase Pequeña}_i = 1] - E[\text{Calificacion}_{0i} \mid \text{Clase Pequeña}_i = 1]}_{\text{Efecto causal promedio i.e., } \bar{\tau}} + \quad (2)$$

$$\underbrace{E[\text{Calificacion}_{0i} \mid \text{Clase Pequeña}_i = 1] - E[\text{Calificacion}_{0i} \mid \text{Clase Pequeña}_i = 0]}_{\text{Sesgo de Seleccion}} \quad (3)$$

- ▶ Si las clases pequeña atrae/selecciona individuos que en promedio, tienen calificaciones menores en ausencia de clase pequeñas, entonces tenemos selección negativa
- ▶ Para evaluar es importante romper el sesgo de selección → Experimentos

# 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

- ▶ 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  $Y_{0i}$ )

$$\{Y_{1i}, Y_{0i}\} \perp D_i \quad (4)$$

- ▶ 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  $E[Y_{0i}]$

# 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?

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?

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?

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

- Ejemplo: El Experimento STAR

## 2 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")

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

- ▶ Entonces podemos escribir

$$Y_i = Y_{0i} + D_i (Y_{1i} - Y_{0i}) \quad (6)$$

- ▶ como

$$Y_i = \underbrace{\alpha}_{=E[Y_{0i}]} + D_i \underbrace{\rho}_{Y_{1i} - Y_{0i}} + \underbrace{\eta_i}_{Y_{0i} - E[Y_{0i}]} \quad (7)$$

# Análisis de Regresión en Experimentos

$$Y_i = \alpha + D_i \rho + \eta_i \quad (8)$$

- ▶ Ahora escribamos la esperanza condicional de  $Y_i$  para los dos niveles de  $D_i$ :

$$E[Y_i | D_i = 1] = E[\alpha + \rho + \eta_i | D_i = 1] \quad (9)$$

$$= \alpha + \rho + E[\eta_i | D_i = 1] \quad (10)$$

$$E[Y_i | D_i = 0] = E[\alpha + \eta_i | D_i = 0] \quad (11)$$

$$= \alpha + E[\eta_i | D_i = 0] \quad (12)$$

- ▶ tomando diferencias

$$E[Y_i | D_i = 1] - E[Y_i | D_i = 0] = \rho + \underbrace{E[\eta_i | D_i = 1] - E[\eta_i | D_i = 0]}_{\text{Sesgo de selección}} \quad (13)$$

Sesgo de selección

# Análisis de Regresión en Experimentos

$$E[Y_i | D_i = 1] - E[Y_i | D_i = 0] = \rho + E[\eta_i | D_i = 1] - E[\eta_i | D_i = 0] \quad (14)$$

- ▶ Nuestra estimación del **efecto de tratamiento** ( $\rho$ ) es tan bueno como nuestra habilidad de "shut down" el *sesgo de selección*
- ▶ El sesgo de selección en la regresión es:

$$E[\eta_i | D_i = 1] - E[\eta_i | D_i = 0] \quad (15)$$

- ▶ Este luce mucho como OVB, hay algo en termino de error  $\eta_i$  que esta afectando  $Y_i$  y también esta correlacionado con  $D_i$  ( $D_i$  es endógeno)

# Análisis de Regresión en Experimentos

- ▶ El sesgo de selección en la regresión es:

$$E[\eta_i | D_i = 1] - E[\eta_i | D_i = 0] \quad (16)$$

- ▶ Si asignamos aleatoriamente  $D_i$ , rompemos esta correlación ( $D_i \perp \eta_i$ )

# Análisis de Regresión en Experimentos

- ▶ El sesgo de selección en la regresión es:

$$E[\eta_i | D_i = 1] - E[\eta_i | D_i = 0] \quad (16)$$

- ▶ Si asignamos aleatoriamente  $D_i$ , rompemos esta correlación ( $D_i \perp \eta_i$ )
- ▶ Otra ruta potencial es que la asignación es aleatoria condicional en otras variables  $D_i | X_i \perp \eta_i$ , con la esperanza de que rompa la correlación

# 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

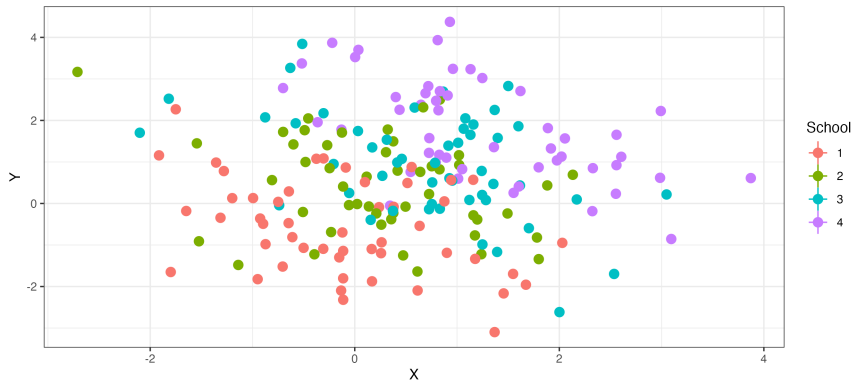
- ▶ Si se asigno aleatoriamente y esto elimina el sesgo de selección por qué la regresión incluye controles?

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.

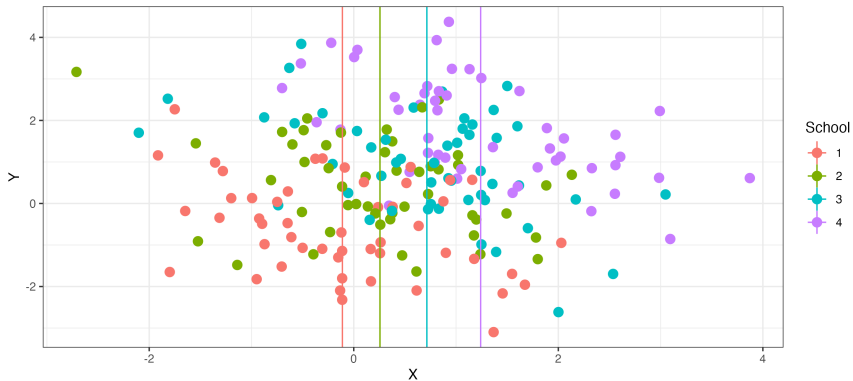
# Detour: Efectos Fijos

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



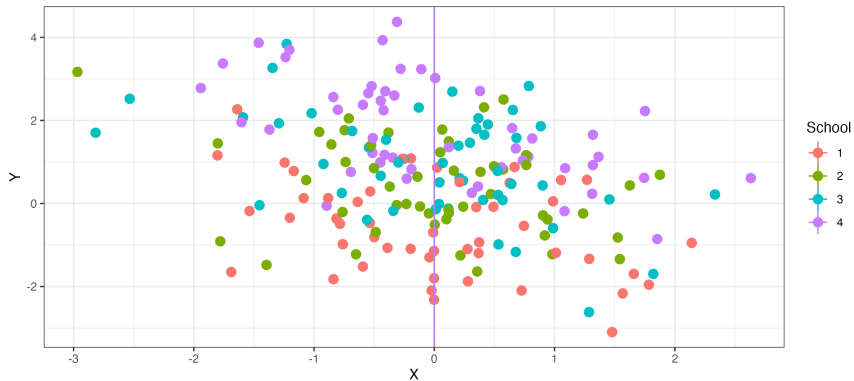
# Detour: Efectos Fijos

Figure out any between-School differences in  $X$



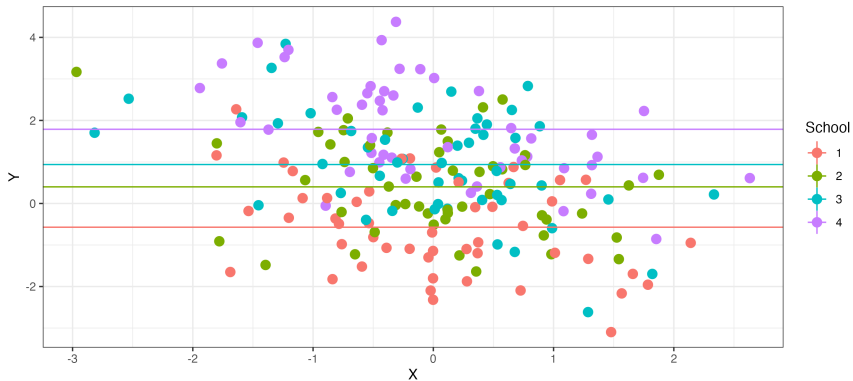
# Detour: Efectos Fijos

Remove all between-School differences in  $X$



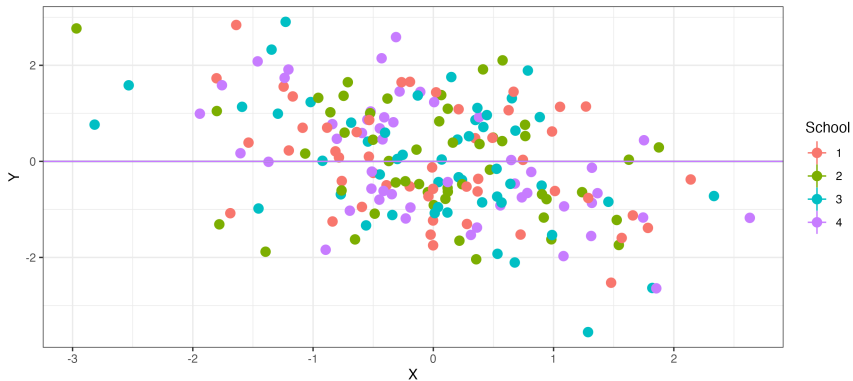
# Detour: Efectos Fijos

Figure out any between-School differences in Y



# Detour: Efectos Fijos

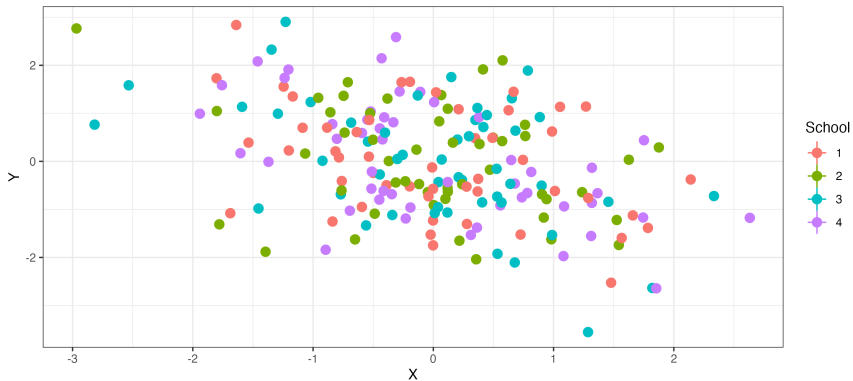
Remove all between-School differences in Y





# 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

- Ejemplo: El Experimento STAR

## 2 Análisis de Regresión en Experimentos

- Ejemplo: El Experimento STAR (cont.)
- Ejemplo: Aprendiendo sobre resultados de prueba de HIV

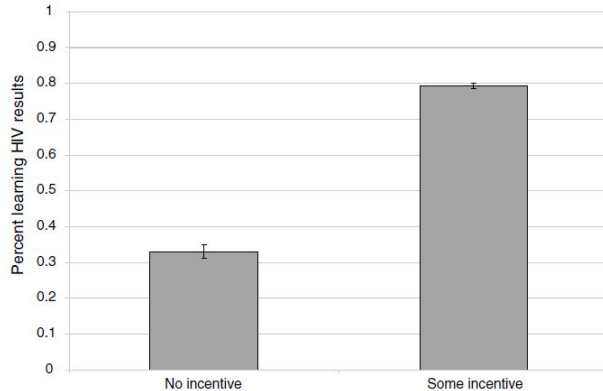
## Aprendiendo sobre resultados de prueba de HIV. Thornton (2008)

- ▶ 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 incentive<sup>2</sup>

HIV

Distance (km)

Distance<sup>2</sup>

---

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.



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

	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.

# 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 <sup>2</sup>		
HIV	-0.055* (0.031)	-0.052 (0.032)
Distance (km)		
Distance <sup>2</sup>		
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 <sup>2</sup>			-0.063*** (0.011)
HIV	-0.055* (0.031)	-0.052 (0.032)	-0.05 (0.032)
Distance (km)			
Distance <sup>2</sup>			
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 <sup>2</sup>			-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 <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

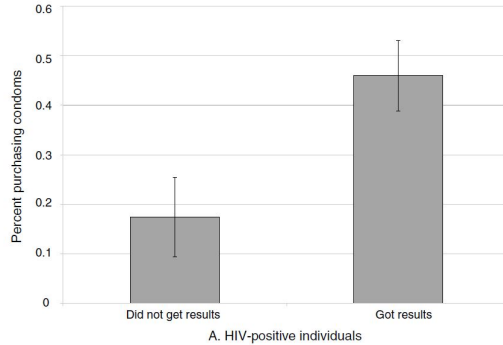
Nota: \*\*\* Significativamente diferente de cero a un nivel de confianza del 99 por ciento. \*\* Significativamente diferente de cero a un nivel de confianza del 95 por ciento. \* Significativamente diferente de cero a un nivel de confianza del 90 por ciento.

# 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 <sup>2</sup>			-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 <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

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