



Semana 3

Introducción a Inferencia Causal

DCDPP - Datos para la evaluación de Políticas Públicas | PUC | 07 de octubre, 2022

 **Pablo A. Celhay** |  pacelhay@uc.cl

Outline

1. ¿Qué es la evaluación de impacto?
2. El modelo causal de Rubin
3. El problema de selección
4. Regresión MCO
5. Experimentos
6. Variables Instrumentales

1. ¿Qué es la evaluación de impacto?

¿Qué es la evaluación de impacto?

- La evaluación de programas es un concepto que abarca distintos tipos de evaluación:
 - La evaluación de la necesidad del programa
 - La evaluación de la lógica o el diseño de un programa - teoría de cambio
 - La evaluación de la implementación del programa - monitoreo
 - La evaluación de impacto del programa - evaluación de impacto
 - La evaluación de los costos y beneficios - análisis de costo beneficio
- Este curso hablará de todo esto pero se concentra en evaluación de impacto

¿Por qué hacemos evaluaciones de impacto?

- Enormes recursos destinados a mejorar resultados: Gasto en Salud, Educación, Vivienda, Policía, etc.

Pero, **hacen alguna diferencia?**

- Los objetivos de una evaluación de programa son:
 - Determinar si los programas son efectivos
 - Proveer "rendición de cuentas" (accountability) a organizaciones que crean estos programas
 - Mejorar programas a futuro en base a lecciones aprendidas.

Evaluación de programa: Ejemplo

Expansión de redes de agua y saneamiento en Bolivia (BID 2015)



● Solución de saneamiento y pileta de agua en construcción.



● Actividades de DESCOM desarrolladas en el marco del proyecto.



● Obras de agua y saneamiento en construcción.

Evaluación de programa: Ejemplo

Expansión de redes de agua y saneamiento en Bolivia (BID 2015)

- Problema: ".... Según datos del Censo 2012, el acceso a agua por cañería de red es del 68,3% de los hogares a nivel nacional, pero sólo del 40,3% en zonas rurales. El censo también reporta que mientras el 30,1% de los hogares de todo el país carece de servicio sanitario de baño o letrina, en zonas rurales la cifra aumenta al 62,4%. Por otro lado, la defecación al aire libre es del 46% en zonas rurales versus un 4% en zonas urbanas" (BID 2016)
- Programa: ".... programa de agua y saneamiento (AyS) focalizado en comunidades rurales cuya población sea inferior a 500 habitantes con el objeto de: i) aumentar la cobertura de AyS en estas comunidades, ii) reducir el riesgo de enfermedades de origen hídrico en niños menores 5 años, iii) disminuir las prácticas de defecación al aire libre, y iv) mejorar la productividad de los adultos en hogares que dedican un promedio de 3 horas diarias a tareas de recolección y acarreo de agua" (BID 2016)

Evaluación de programa: Ejemplo

Expansión de redes de agua y saneamiento en Bolivia (BID 2015)

- El Banco InterAmericano del Desarrollo (BID) realizó (y está realizando) una evaluación de impacto del AyS en Bolivia.
 - Implementó un experimento aleatorio en donde un grupo de comunidades fueron intervenidas y otras no, en el año 2016
 - Han analizado la implementación en detalle
 - Midiendo variables de consumo de agua, cambio de hábitos (higiene), y salud, con datos antes y después de la implementación.
- Es uno de los primeros estudios a gran escala de AyS en Latinoamérica:
 - Pero, ¿porqué realizaron un experimento aleatorio?
 - ¿Cómo deberíamos interpretar sus resultados?
 - ¿Qué métodos se podrían haber utilizado de no haber aleatorización?
 - ¿Qué supuestos están detrás de estos métodos?
- Estas preguntas resumen nuestro semestre a venir.

Resultados vs. Impacto

- Comencemos por una terminología básica que utilizamos en evaluación de impacto. Es importante distinguir entre variables de resultado e impactos
- Variables de resultados: cosas que podemos (potencialmente) medir:
 - Resultados de tests estandarizados (simce)
 - Salarios
 - Edad en el primer nacimiento
 - Muertes por accidentes
- Impactos: cambios en variables de resultado atribuibles o causados por un programa o política pública:
 - Cambio en Simce luego de un programa de capacitación docente
 - Cambio en salarios luego de ir a la Universidad
 - Cambio en la edad en el primer nacimiento debido a una política de anticonceptivos
 - Cambio en muertes por accidentes debido a una mayor regulación de maneja bajo alcohol

2. El modelo causal de Rubin

El modelo causal de Rubin (1974)

- Suponga que:
 - Hay N personas, $i = 1, \dots, N$
 - Son seleccionadas de manera aleatoria de una población
 - Queremos estimar el efecto de una variable binaria de tratamiento, D_i , sobre Y_i **¿cómo lo hacemos?**
- Digamos que hay dos resultados potenciales para cada observación:
 - Y_{0i} = resultado para i bajo control
 - Y_{1i} = resultado para i bajo tratamiento
 - ¿Qué observamos en nuestros datos?

No podemos observar Y_{0i} e Y_{1i} para el mismo i

- Pero teóricamente, definimos el efecto causal de D_i sobre Y_i como:

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

El modelo causal de Rubin (1974)

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

- Notar:
 1. El efecto del tratamiento se define siempre en términos relativos **¿A qué?**
 2. El efecto del tratamiento no necesariamente es el mismo para todos los i **¿Por qué?**
 3. Nunca observamos Y_{1i}, Y_{0i} para el mismo i **¿Por qué?**
- Esto se conoce como el problema fundamental de la inferencia causal (Holland 1986)
- Imposible observar el valor Y_{1i}, Y_{0i} en la misma unidad i
- ¿Podemos estimar entonces τ_i ?
- ¿Entonces para qué hacemos este curso?

El efecto tratamiento promedio (ATE)

- Se puede estimar el efecto de tratamiento promedio (ATE)
- Definamos ATE:

$$\begin{aligned}\tau^{ATE} &= E[Y_{1i} - Y_{0i}] \\ &= E[Y_{1i}] - E[Y_{0i}]\end{aligned}$$

- En evaluación de impacto queremos estimar este parámetro de datos que incluyan:
 - Personas que recibieron el programa
 - Personas que no recibieron el programa - ¿Podemos estimar ATE de estos datos?

El problema de la evaluación

Ejemplo: Cursos de manejo llevan a mayores choques

En Agosto del año 2010, en un país se legisla para que los conductores de auto tengan mayor precaución y se exigen cursos de manejo.



El problema de la evaluación

Ejemplo: Cursos de manejo llevan a mayores choques

En Agosto del año 2010, en un país se legisla para que los conductores de auto tengan mayor precaución y se exigen cursos de manejo.

- El argumento de los(as) legisladores era el siguiente: "un estudio reciente muestra que, en la población menor a 18 años, personas que tomaron el curso de manejo tienen una probabilidad de chocar cuatro veces mayor al compararlas con aquellas que no tomaron el curso"
- Esto efectivamente fue así. Los datos mostraban que alrededor del 5% de los(as) adolescentes que tomaron el curso de manejo reportaron tener algún accidente comparado con solo 1% de aquellos(as) que no lo hicieron.
- **Discusión grupal:** Preguntas acerca de la interpretación de estos datos:
 - Esta conclusión ¿es correcta necesariamente?
 - ¿Cuáles son los supuestos necesarios para llegar a esta discusión?
 - ¿Porque no son creíbles estos supuestos?

El estimador ingenuo

- La diferencia simple de medias observadas para personas tratadas y no tratadas se llama el estimador ingenuo (Naive estimator) del efecto del programa:

$$\tau^N = E[Y_{1i}|T] - E[Y_{0i}|C]$$

- ¿Qué significan estas esperanzas condicionales?
- $E[Y_{1i}|T]$: Promedio del resultado bajo tratamiento para la persona i en el grupo de tratamiento
- $E[Y_{0i}|C]$: Promedio del resultado bajo control para la persona i en el grupo de control
- ¿Algunas preguntas?
 - ¿Podemos observar ambos términos $E[Y_{1i}|T]$ y $E[Y_{0i}|C]$?
 - Recuerden a ATE: $E[Y_{1i}] - E[Y_{0i}]$
 - El estimador ingenuo, ¿nos entrega ATE?

El estimador ingenuo

$$\tau^N = E[Y_{1i}|T] - E[Y_{0i}|C]$$

- Recordemos la definición de ATE:

$$\tau^{ATE} = E[Y_{1i}] - E[Y_{0i}]$$

- Usar el estimador ingenuo para obtener ATE asume entonces:
 - $E[Y_{1i}] = E[Y_{1i}|T]$
 - $E[Y_{0i}] = E[Y_{0i}|C]$
- ¿Cómo interpretamos estos supuestos?
- Lo llamaremos ingenuo porque ignora el **problema de selección**

3. El problema de selección

El problema de selección

- Este es el problema más grande que tenemos en evaluación de impacto
- El problema de selección se genera principalmente porque sólo podemos observar algunas personas bajo tratamiento
 - Las personas que participan en un programa son diferentes a aquellas que no participan
 - Las personas que participan en un programa son también diferentes a si mismas previo al inicio del programa

El problema de selección

- La pregunta clave detrás del problema de selección es: Porqué estas personas fueron tratadas y otras no?
- Bajo el problema de selección, los supuesto detrás del estimador ingenuo son cuestionables: Los resultados para los individuos no tratados son probablemente un mal estimador del estado contrafactual de estos mismos individuos
- Pregunta: ¿Por qué los conductores que no recibieron el curso son una mala comparación para aquellos que si lo recibieron?

Dos tipos de selección

- **Selección en Observables:** Participantes son diferentes a los no participantes en características observables, i.e. edad o nivel de educación. Estas características son medibles y por eso las llamamos observables.
- **Selección en No-Observables:** Participantes son diferentes a los no participantes en características no observables, i.e. aversión al riesgo, habilidad inherente, motivación. En general no tenemos o no podemos medir estas variables y tenerlas en nuestros datos.
- Las características pueden ser observables y no observables dependiendo del contexto y la base de datos que tengamos.

Este curso estudia métodos para estos problemas de selección

- El estimador ingenuo tiene supuestos fuertes que no se sostienen y son comunes
 - Comparar un mismo grupo antes y después del programa
 - Comparar en un mismo periodo personas en distintos grupos
- ¿Qué métodos utilizar para poder corregir este problema?
- Piensen en el caso del curso de manejo y sus efectos en la tasa de accidentes de tránsito?
 - Ya discutimos que el estimador ingenuo no nos sirve
 - ¿Qué métodos habría que proponerles a los legisladores para que saquen mejores conclusiones?

Panorama

- Selección en observables
 - MCO
 - Matching y Propensity Score Matching
- Selección en No-observables
 - Experimentos
 - Datos de panel y diferencias en diferencias
 - Variables instrumentales
 - Regresión discontinua
- Quizás ya los han escuchado pero entender bien estos métodos requiere práctica (talleres).

4. Regresión de MCO

Volvamos a nuestra querida Regresión Lineal

- La regresión lineal que permite estimar efecto de un tratamiento sobre una variable de resultado es:

$$y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$$

- y_i : es una variable de resultado (variable dependiente)
- D_i : es una variable binaria (*dummy*) que es igual a 1 si la unidad de análisis (individuo, hogar, municipio, país) participa o recibe un programa y 0 en caso contrario.
- ε_i : es un error que se distribuye i.i.d. (supuesto)
- ¿Cómo se estima β_1 ? ¿Cuáles son los supuestos de MELI?

Una forma de ver el problema de selección en nuestra regresión lineal

- ¿Cómo se estima β_1 ? ¿Cuáles son los supuestos de MELI?

$$\begin{aligned}\beta_1^{OLS} &= E[y_i | D = 1] - E[y_i | D = 0] \\ &= E[\beta_0 + \beta_1 + \varepsilon_i | D = 1] - E[\beta_0 + \varepsilon_i | D = 0] \\ &= \beta_1 + \underbrace{E[\varepsilon_i | D = 1] - E[\varepsilon_i | D = 0]}_{\text{selección}}\end{aligned}$$

- ¿Qué nos dice $E[\varepsilon_i | D = 1] - E[\varepsilon_i | D = 0]$?

Recordando la última clase

- Sea $D_i \in \{0, 1\}$ un indicador de si la persona i está bajo tratamiento o no. Si está bajo tratamiento, $D_i = 1$, de lo contrario, $D_i = 0$. Sea Y_{1i} la realización de la variable de resultado bajo tratamiento y Y_{0i} la realización de la variable de resultado sin tratamiento. Entonces el impacto para la persona i se define como:

$$\Delta_i = Y_{1i} - Y_{0i}$$

- ¿Cuál es el problema fundamental aquí?

→ El problema fundamental de la evaluación es que no observamos el contrafactual Y_{0i} para aquellas personas que están bajo tratamiento, y tampoco observamos Y_{1i} para aquellas personas que están sin tratamiento

¿Qué son los parámetros de tratamiento?

- La mayoría de los enfoques a la evaluación en ciencias sociales parten del hecho que no podemos calcular Δ_i a nivel individual
- Por lo que intentamos estimar alguna función de Δ_i (hint: el promedio es una función)
- Pero no es claro qué parámetro vamos a estimar
- Tampoco para qué población vamos a estimar este parámetro
 - Depende mucho de la pregunta que queramos responder
- Todo esto tendrá más sentido en las próximas clases pero es bueno tenerlo en mente

¿Qué queremos saber de un programa?

Heckman, Lalonde, Smnith (1999)

- La proporción de la población que se beneficiaría del programa:

→ Si tuviésemos Δ_i , ¿cómo se construye?

$$Pr(\Delta_i > 0)$$

- La proporción de los tratados que se beneficia del programa:

$$Pr(\Delta_i > 0 | D = 1)$$

- La complejidad de estos parámetros es que requiere que sepamos la distribución de los efectos del programa, $F(\Delta_i)$ y la distribución de los efectos del programa sobre las personas que toman el programa $F(\Delta_i | D = 1)$.

¿Qué podemos estimar comúnmente?

- La literatura y la práctica se enfoca en estimar efectos promedio
- Efecto promedio del programa (ATE)

$$\Delta^{ATE} = E(\Delta_i) = E(Y_{1i} - Y_{0i})$$

- Efecto promedio del programa sobre los tratados (ATET)

$$\Delta^{ATET} = E(\Delta_i | D = 1) = E(Y_{1i} - Y_{0i} | D = 1)$$

- Efecto promedio del programa sobre los no tratados (ATEN)

$$\Delta^{ATEN} = E(\Delta_i | D = 0) = E(Y_{1i} - Y_{0i} | D = 0)$$

Todos estos parámetros pueden responder preguntas claves de política públicas

- Por ejemplo, consideremos un gobierno que quiere entregar bonos para que las madres lleven a sus hijas al control de la niña sana. Pero no todas las beneficiarias toman el programa:

→ ¿Es esto común? ¿Personas elegibles a un programa no toman el programa? ¿Por qué? ¿Son distintas estas personas de las que si toman el programa?
- Sea la población de interés todas las niñas y niños que pertenecen al 50% más pobre de los hogares de Chile. Sea $D_i = 1$ un indicador de si la madre i recibe el bono y los resultados potenciales Y_{1i}, Y_{0i} el IMC de la niña:
 - ¿Cuándo quisiéramos estimar Δ^{ATET} ?
 - ¿Cuándo quisiéramos estimar Δ^{ATE} ?
 - ¿Cuándo quisiéramos estimar Δ^{ATEN} ?
- ¿Esperaríamos que estos parámetros difieran?

Todos estos parámetros pueden responder preguntas claves de política públicas

- Si es que los impactos son los mismos para todas las personas entonces se dice que existen efectos homogéneos y todos los parámetros que vimos son equivalentes
- Pero en general estos parámetros difieren entonces los efectos del programa son heterogéneos:

$$\rightarrow \Delta^{ATET} \neq \Delta^{ATE} \neq \Delta^{ATEN}$$

- Esto simplemente dice que el impacto del bono condicional puede ser distinto para personas que eligen tomarlo que para personas que no eligen tomarlo.
- De todas formas estos parámetros se relacionan entre sí:

$$\rightarrow \Delta^{ATE} = Pr(D_i = 1)\Delta^{ATET} + Pr(D_i = 0)\Delta^{ATEN}$$

- ¿De donde viene esto?

El efecto promedio sobre los tratados

- En este curso nos enfocaremos en el ATET:

$$\Delta^{ATET} = E(\Delta_i | D = 1) = E(Y_{1i} - Y_{0i} | D = 1)$$

- Este parámetro consiste en dos promedios. El primero, $E(Y_{1i} | D = 1)$ lo podemos estimar de los datos observados ya que:

$$E(Y_{1i} | D = 1) \approx \bar{Y}_{1,D=1}$$

- ¿Qué nos falta? La segunda parte del parámetro, el contrafactual (estado sin tratamiento) para la población que tomó el programa:

$$E(Y_{0i} | D = 1) \approx ?$$

- Problema: No observamos esto para la persona i tratada: **Es el contrafactual que falta**

Sesgo de selección en una asignación de tratamiento no aleatoria

- Sigamos definiendo "sesgo de selección" (auto selección, selección en no observables, etc...)



Sesgo de selección en una asignación de tratamiento no aleatoria

- Sigamos definiendo "sesgo de selección" (auto selección, selección en no observables, etc...)
- Vamos a nuestro estimador ingenuo: ¿Cómo lo armamos?

$$\begin{aligned}\Delta^N &= E(Y_{1i}|D = 1) - E(Y_{0i}|D = 0) \\ &= E(Y_{1i}|D = 1) - E(Y_{0i}|D = 0) + \underbrace{E(Y_{0i}|D = 1) - E(Y_{0i}|D = 1)}_{\text{agrego un } 0} \\ &= E(Y_{1i}|D = 1) - E(Y_{0i}|D = 1) - E(Y_{0i}|D = 0) + E(Y_{0i}|D = 1) \\ &= \underbrace{E(Y_{1i} - Y_{0i}|D = 1)}_{ATET} + \underbrace{E(Y_{0i}|D = 1) - E(Y_{0i}|D = 0)}_{\text{sesgo de selección}}\end{aligned}$$

- $E(Y_{0i}|D = 1) - E(Y_{0i}|D = 0)$
 - ¿Cómo interpretamos esto?
 - ¿Es igual a cero este término?

Selección en el modelo de MCO

- Podemos escribir el valor observado (realizado) de Y_i en términos de los resultados potenciales:

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

- Esto viene de el hecho que:

$$Y_i = Y_{1i} \text{ si } D_i = 1$$

$$Y_i = Y_{0i} \text{ si } D_i = 0$$

- Ahora suponga que tenemos efectos de tratamiento que son homogéneos (es idéntico para toda la población):

$$\Delta_i = Y_{1i} - Y_{0i} = \delta$$

Selección en el modelo de MCO

- Podemos escribir

$$\begin{aligned} Y_i &= D_i Y_{1i} + (1 - D_i) Y_{0i} \\ &= D_i Y_{1i} + Y_{0i} D_i Y_{0i} \\ &= Y_{0i} + (Y_{1i} - Y_{0i}) D_i + [E(Y_{0i}) - E(Y_{0i})] \\ &= E(Y_{0i}) + (Y_{1i} - Y_{0i}) D_i + [Y_{0i} - E(Y_{0i})] \\ &= \alpha + \delta D_i + \varepsilon_i \end{aligned}$$

- Dónde:

- $\alpha = E(Y_{0i})$: el promedio de Y_{0i}
- $\delta = Y_{1i} - Y_{0i}$: el efecto promedio constante
- $\varepsilon_i = Y_{0i} - E(Y_{0i})$: el componente aleatorio de Y_{0i}

Selección en el modelo de MCO

- Ahora tomemos el valor esperado de Y_i condicional en $D_i = 1$ y condicional en $D_i = 0$:

$$E[Y_i|D_i = 1] = \alpha + \delta + E[\varepsilon_i|D_i = 1]$$

$$E[Y_i|D_i = 0] = \alpha + E[\varepsilon_i|D_i = 0]$$

- Podemos escribir:

$$\begin{aligned}\Delta^N &= E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\ &= \delta + E[\varepsilon_i|D_i = 1] - E[\varepsilon_i|D_i = 0]\end{aligned}$$

- ¿Es esta expresión familiar? ¿Lo podemos interpretar?

Selección en el modelo de MCO

- Recuerden nuestro término de sesgo de selección:

$$E(Y_0 i | D_i = 1) - E(Y_0 i | D_i = 0)$$

- La nueva expresión $E[\varepsilon_i | D_i = 1] - E[\varepsilon_i | D_i = 0]$ es la misma selección pero expresada en términos del error en una regresión lineal
- Si queremos obtener un estimador sin selección, necesitamos asumir entonces que el error es independiente del estado de tratamiento ¿Qué significa esto?
- En el lenguaje de regresión: El error no esta correlacionado con la variable de tratamiento, D_i

MCO y los parámetros de tratamiento

- Empecemos de un modelo muy general en donde los resultados potenciales son una función de las características observadas, X , y las características no observadas capturadas en el término de error, ε :

$$Y_{1i} = g_1(X_i, \varepsilon_i)$$
$$Y_{0i} = g_0(X_i, \varepsilon_i)$$

- El primer supuesto es que estos dos términos son **separables**

$$Y_{1i} = g_1(X_i, \varepsilon_i) = g_1(X_i) + \varepsilon_{1i}$$
$$Y_{0i} = g_0(X_i, \varepsilon_i) = g_0(X_i) + \varepsilon_{0i}$$

MCO y los parámetros de tratamiento

- Asumamos que la función $g(\cdot)$ es \textcolor{blue}{lineal} en los parámetros:

$$Y_{1i} = \beta_1 X_i + \varepsilon_{1i}$$

$$Y_{0i} = \beta_0 X_i + \varepsilon_{0i}$$

- También asumiremos que estas funciones de regresión tienen parámetros idénticos y sólo difieren por una constante según el estado de tratamiento D_i :

$$Y_{1i} = Y_{0i} + \delta$$

- Esto nos lleva a la regresión de MCO típica:

$$Y_i = X_i \beta + \delta D_i + \varepsilon_i$$

MCO y los parámetros de tratamiento

- ¿Qué supuestos necesitamos imponer en el error para poder estimar un efecto causal del tratamiento, D_i , sobre nuestra variable de resultado Y_i ?
 - $E(\varepsilon) = 0$
 - $E(\varepsilon|X_i) = E(\varepsilon)$
 - $E(\varepsilon|D_i) = E(\varepsilon)$
- O, combinándolos todos: $E(\varepsilon|X_i, D_i) = E(\varepsilon) = 0$
- ¿Qué nos dice este supuesto? $E(\varepsilon|X_i, D_i) = 0$ implica que puede haber selección en observables X , pero que no hay selección en no observables, representados por ε

Resumen

- A modo de resumen, la regresión: $Y_i = X_i\beta + \delta D_i + \varepsilon_i$
 - Impone la restricción de que: $\delta = \Delta_i$. El tratamiento es el mismo para todos
 - Asume que no hay selección en no observables
- ¿Qué pasa si es que estos supuestos no se cumplen?
 - Si los efectos son heterogéneos, el parámetro δ es un promedio ponderado de los parámetros individuales
 - Si hay selección en no observables, el parámetro δ nos entrega un estimador sesgado del impacto del programa.

Repaso MCO

- Consideremos la regresión lineal de MCO:

$$Y_i = X_i\beta + \delta D_i + \varepsilon_i$$

- Bajo los supuestos anteriores y en particular bajo $E(\varepsilon|D, X) = 0$, sabemos que el estimador de MCO nos entrega una estimación insesgada tal que $E(\hat{\delta}) = \delta$.
- $E(\varepsilon|D, X) = 0$ implica que podría haber selección en observables, X , pero no selección en no observables, ε .

¿Cuándo incluir las X en la regresión?

Algunos tips para pensar en cuando incluir X en la regresión:

1. Si las Xs son afectadas por el tratamiento no hay que incluirlas pues estas deberían ser analizadas como variables de resultado (Y_i)
2. Si las Xs pueden ser variables omitidas de la regresión deben ser incluidas porque de lo contrario podrían generar sesgo de variable omitida.
3. Incluir a las Xs también puede mejorar los errores estándar de los coeficientes (si mejoran el R^2).

5. Experimentos

Lo que hay detrás de los experimentos

Cuando el tratamiento es asignado de manera aleatoria, tenemos que:

$$F(X, \varepsilon | D = 1) = F(X, \varepsilon | D = 0) = F(X, \varepsilon)$$

¿Qué significa esto?

--

La distribución de características observables y no observables para la población tratada y no tratada son iguales a la distribución de la población

No hay diferencias sistemáticas entre un grupo y otro.

No hay **sesgo de selección** por construcción

Lo que hay detrás de los experimentos

- Esto nos permite entonces al menos cumplir con algo muy importante para el análisis de impacto con regresión lineal

$$E(Y_{1i}|D=1) = E(Y_{1i}|D=0) = E(Y_{1i})$$

$$E(Y_{0i}|D=0) = E(Y_{0i}|D=1) = E(Y_{0i})$$

- ¿Qué nos permite esta condición? \pause{} estimar ATE

$$\begin{aligned}\Delta^{ATE} &= E(Y_{1i}) - E(Y_{0i}) \\ &= E(Y_{1i}|D=1) - E(Y_{0i}|D=0)\end{aligned}$$

- ¡Este es el estimador más sencillo que nos podemos imaginar!
- El impacto es simplemente la diferencia entre promedios grupales

Lo que se hace generalmente

- Verificar si la aleatorización funcionó
- Revisar cuanta gente asignada al tratamiento efectivamente toma el tratamiento
- Estimar el ATE (u otros parámetros)
- Estimar efectos para distintas poblaciones

King et al. (2009)

- Los sistemas de salud que promueven calidad en la atención no llegan a los sectores más vulnerables de la población
- México implementa el Seguro popular en el año 2003 para garantizar cobertura financiera y acceso a servicios de salud
- Objetivo principal es el de reducir el gasto catastrófico de bolsillo. Este es el gasto que incurrimos por sobre lo que el seguro paga
- El SP paga a proveedores de salud acreditados → Hospitales tienen incentivos a invertir para la acreditación

Antes de ver el detalle

- Pensemos junt@s la regresión que King debería correr
 - ¿Cuál es la pregunta de impacto aquí? Principalmente, ¿Cuál es el efecto del SP en el gasto de bolsillo en salud?
 - ¿Qué datos necesitamos?
 - Con estos datos, ¿cómo escribimos la regresión para el análisis de impacto?
 - ¿Cómo estimamos esta regresión? ¿Qué supuestos hay detrás?
- $y_i = \alpha + \delta D_i + \varepsilon_i$
 - ¿Qué es y_i ?
 - ¿Qué es i ?

Diseño experimental

- ¿Porque aleatorizar este programa?
- Implementación del SP
 1. Problemas de presupuesto permiten dejar grupos fuera/esperando
 2. Seguro popular se implementó como un "phased roll-out". ¿Qué significa esto?
 3. ¿Cómo elegir cuáles comunidades entran primero?

Diseño experimental

1. "Matched pair cluster randomised trial" **¡¿QUÉ?!**
2. Definen geográficamente 12284 clusters ("catchment areas").
3. 13 Estados del país aceptan ser parte de la implementación: (5439 rurales y 1639 urbanos)
4. "Matched pair" en términos de tamaño y algunas características observables. **¿Qué están tratando de hacer con esto?**
5. Luego obtienen información de 74 clusters en 7 Estados

Estados que participan de estudio

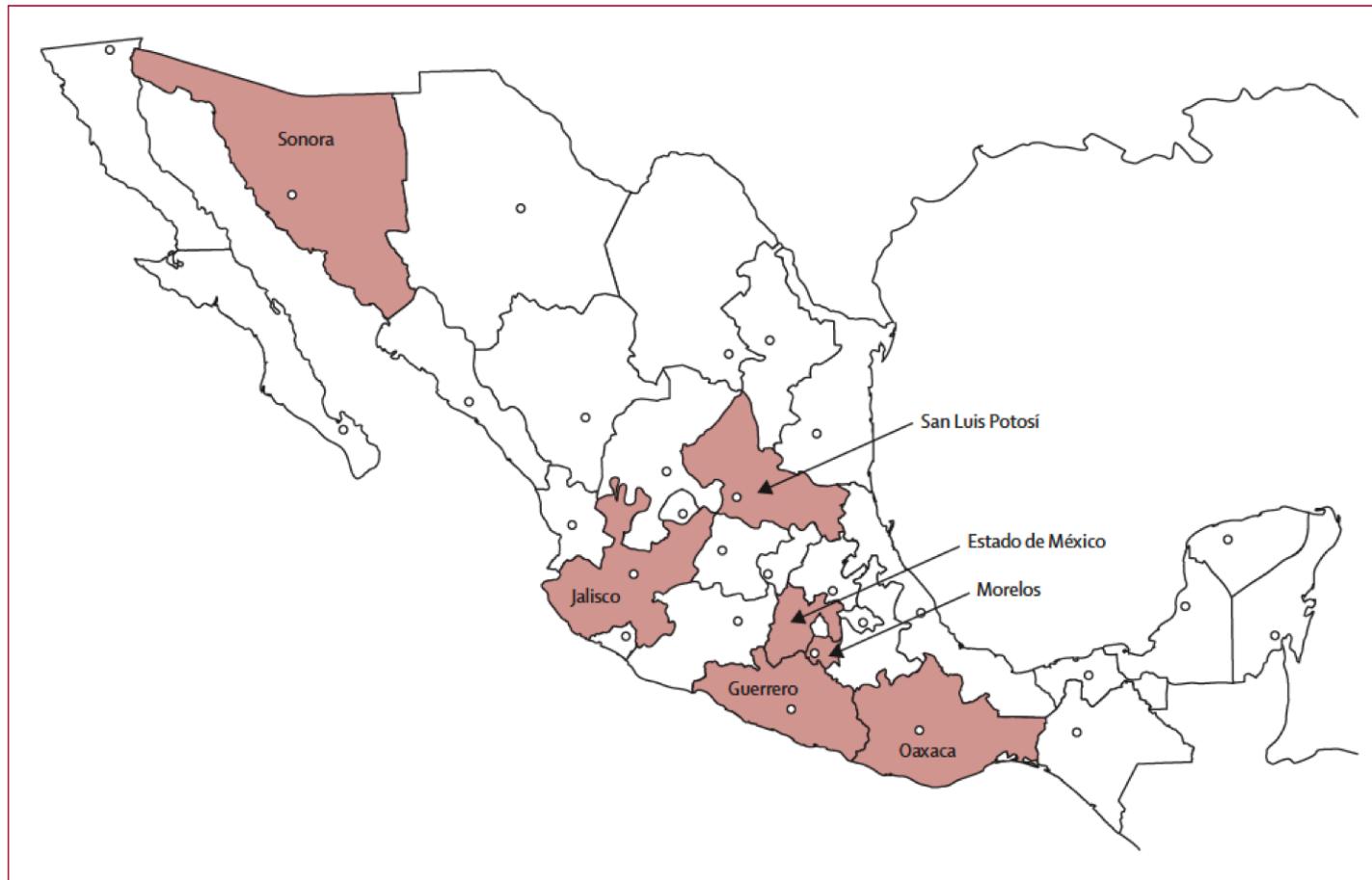


Figure 1: Mexican states participating in the Seguro Popular assessment

Diseño experimental

- Diseño
 1. Con estos 74 clusters seleccionan uno al grupo de tratamiento y otro al grupo de control en cada par formado ¿Cómo lo haríamos?
- Tratamiento
 1. Una campaña publicitaria masiva en comunidades de tratamiento para fomentar la inscripción en el SP
 2. Personas ya inscritas en otros programas sociales (OPORTUNIDADES) se inscriben automáticamente
 3. Por el lado de la oferta de salud, se inician los procedimientos para mejorar la calidad de la atención y el inventario de medicamentos
- ¿A qué nivel es el tratamiento entonces? Tratamiento a nivel de comunidad
- Grupo de control
 1. "Business as usual". Recibe el mismo programa que siempre se ha recibido en Mexico pero espera un año para el SP.

¿Qué tan representativo el estudio de la población de México? (Validez externa)

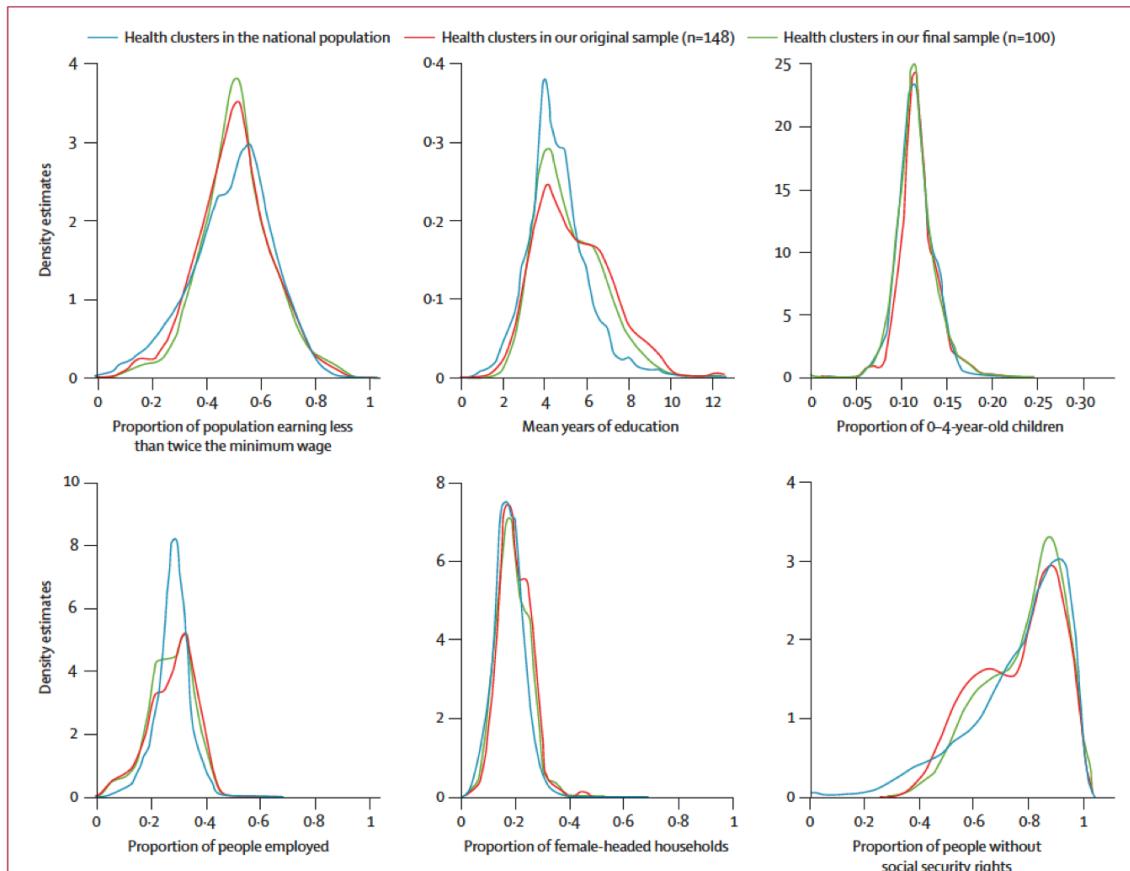


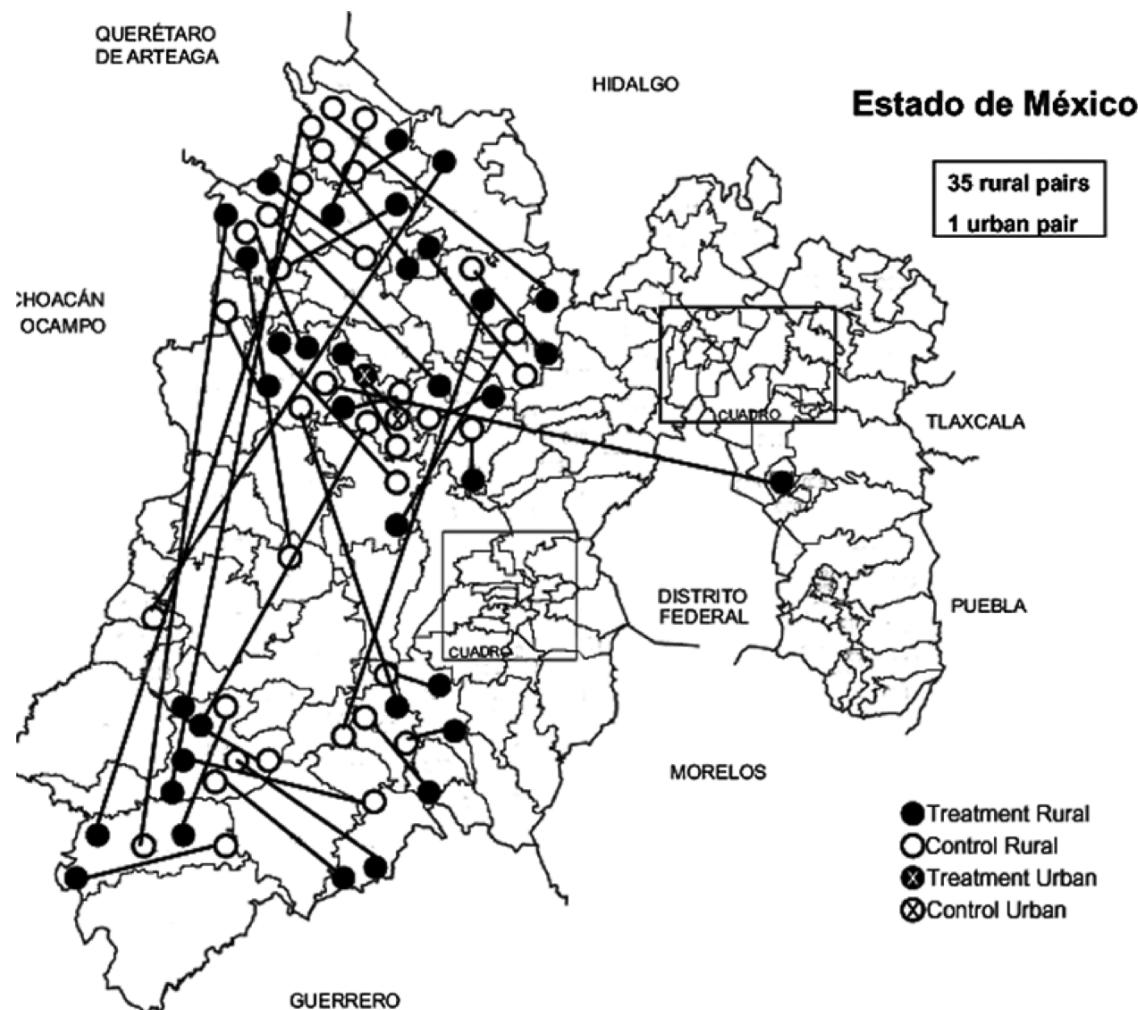
Figure 2: Comparisons between selected and all health clusters

Each histogram (density estimates) of health clusters indicates the national population (blue lines), our original sample of 148 (red lines), and our final sample of 100 (green lines). These histograms display the distribution of eligible cluster statistic, not the distribution for Mexican people or families.

Diseño experimental

- Diseño
 - 1. Muestra final de 100 clusters
 - Se reduce principalmente porque algunos Estados o lugares querían implementar más rápido el programa o porque algunos gobiernos locales no querían cooperar con el estudio.
 - 2. Guerrero (8), Jalisco (2), Estado de México (54), Morelos (26), Oaxaca (4), San Luis Potosí (4), and Sonora (2).
- Datos

Estados que participan en el estudio



Diseño experimental

- Si la aleatorización funcionó, las características del grupo de control y tratamiento deberían ser similares
- Esto, ¿puede *testearse*? Al menos para características observables se puede. Esto se conoce como "revisar balance" de la muestra
- En el caso del SP, es complicado pues el tratamiento fue estratificado por Estado
- Fue hecho al nivel de comunidad pero los datos son recolectados a nivel individual

Diseño experimental

- Recuerde que la aleatorización hace que la distribución condicional de Y_0 y de Y_1 condicional en D_i sea igual a la distribución incondicional

$$E(Y_{1i}|D=1) = E(Y_{1i}|D=0) = E(Y_{1i})$$

$$E(Y_{0i}|D=0) = E(Y_{0i}|D=1) = E(Y_{0i})$$

- ¿Qué nos permite esta condición? estimar ATE
- Lo mismo se sostiene para cualquier X

$$E(Y_{1i}|X) = E(Y_{1i})$$

$$E(Y_{0i}|X) = E(Y_{0i})$$

- Por esta razón se reporta el "balance en las Xs". Por ejemplo mostrar $E[\hat{X}_i]$ para cada grupo

Datos

Balance entre grupos

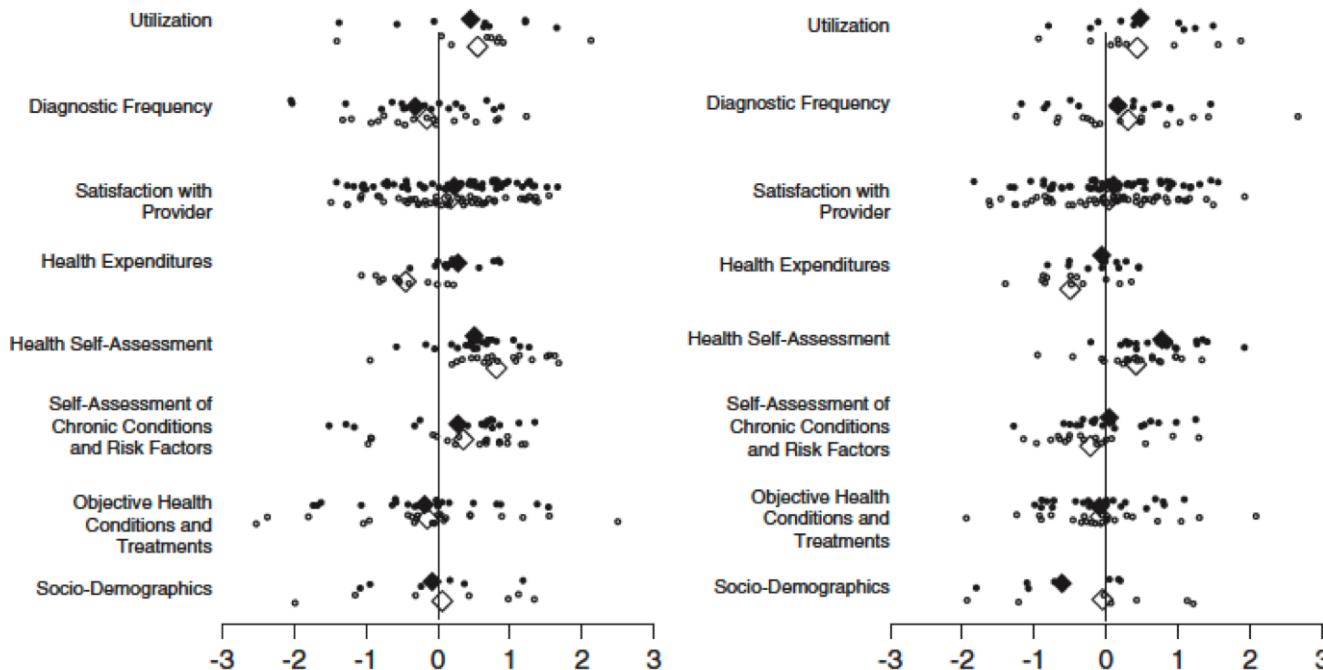


Figure 5. Effects of random assignment on outcome measures at baseline, for all families (left graph) and poor families, in Oportunidades (right graph). If the experiment were implemented properly, we would see zero effect (near the vertical line) plus or minus random error. The horizontal axis is in standard deviation units, and so we expect relatively few estimates outside the $[-2, 2]$ interval, for example, which appears to be the case. Estimates appear without (in open circles) and with (closed disks) covariate adjustment; corresponding diamonds represent the average for each category.

Estimando el ATE

- Podemos estimar el ATE con una simple regresión lineal:

$$Y_i = \alpha + \delta D_i + \varepsilon_i$$

- Donde

$$\begin{aligned}\delta &= E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0) \\ &= E(Y_{1i}) - E(Y_{0i}) \\ &= \Delta^{ATE}\end{aligned}$$

- La segunda igualdad viene de la aleatorización de D_i

Resultados

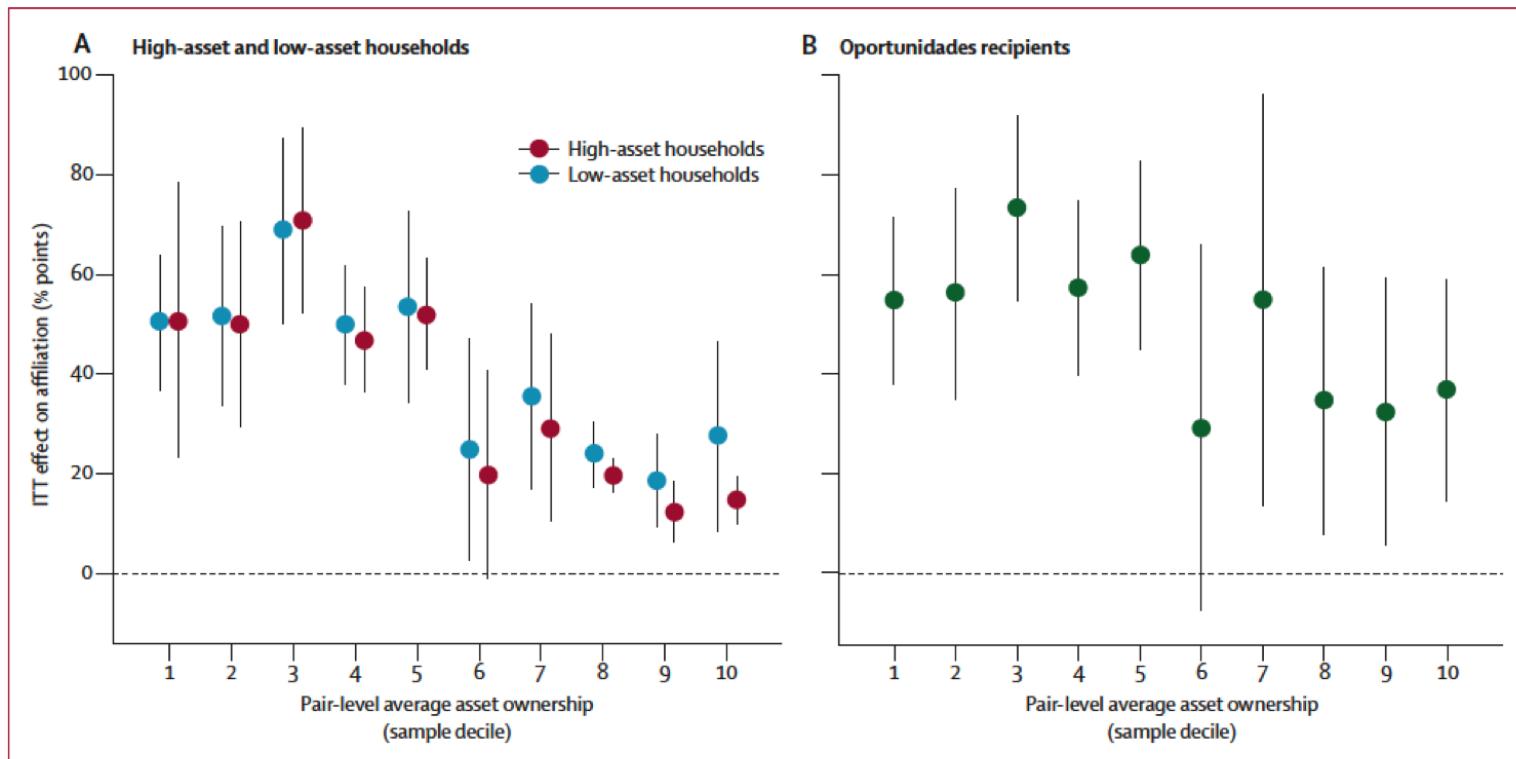


Figure 4: Effect of cluster-level assignment on affiliation to Seguro Popular for all households (A) and for those enrolled in the Oportunidades anti-poverty programme (B)

The horizontal axis in (A) and (B) indicates asset per person ownership deciles within our sample of matched pairs. The vertical axis in (A) and (B) represents the percentage point difference in self-reported Seguro Popular affiliation resulting from encouragement to affiliate. Points represent point estimates and the vertical bars are 95% CIs. ITT=intention to treat.

Resultados

	All participants				Experimental compliers only			
	Control group*	ITT	SE	95% CI	Control group*	CACE	SE	95% CI
All	8·4	1·9†	0·9	0·2 to 3·7	9·5	5·2†	2·3	0·8 to 9·6
Low asset	9·9	3·0†	1·3	0·5 to 5·5	11·0	6·5†	2·5	1·6 to 11·3
High asset	7·1	0·9	0·8	-0·7 to 2·5	7·9	3·0	2·7	-2·3 to 8·4
Female-headed	8·5	1·4	1·1	0·7 to 3·5	9·6	3·8	3·0	-2·1 to 9·7

ITT=intention to treat. SE=standard error. CACE=complier average causal effect. * Average value of the variable, as a baseline. † $p \leq 0.05$ (one-tailed t test). Positive values correspond to favourable outcomes. Catastrophic expenditures are defined as out-of-pocket health expenses greater than 30% of post-subsistence income.

Table 1: Causal effect of rolling out Seguro Popular on the reduction in the number of households suffering catastrophic health expenditures

Resultados

	All participants			Experimental compliers only		
	Control group*	ITT	SE	Control group*	CACE	SE
Overall						
All	1631·3	258·0	175	1712·7	689·7	453
Low asset	1360·2	425·6†	197	1502·6	915·3†	392
High asset	1867·9	128·4	201	1933·2	428·2	669
Female-headed	1509·1	156·5	207	1535·0	428·6	566
Inpatient care						
All	532·5	96·9†	44	557·1	259·1†	112
Low asset	527·1	188·2†	73	579·0	404·8†	142
High asset	537·2	31·1	52	536·2	103·6	173
Female-headed	452·5	115·1†	68	463·3	315·2†	182
Outpatient care						
All	448·3	116·7†	63	499·1	312·0†	161
Low asset	412·3	176·7†	73	466·3	380·0†	147
High asset	479·7	81·9	69	533·0	272·9	230
Female-headed	416·3	110·4	75	451·3	302·4	202
Medicine						
All	521·1	20·0	41	534·5	53·3	109
Low asset	427·3	17·8	46	444·7	38·3	100
High asset	603·0	29·4	47	627·5	98·1	157
Female-headed	625·6	53·6	55	671·2	146·8	151
Medical devices						
All	139·7	-8·8	23	117·8	-23·4	62
Low asset	72·0	-0·2	20	72·8	-0·5	43
High asset	198·8	-16·5	29	165·6	-55·1	98
Female-headed	155·5	10·9	34	147·9	30·0	94

ITT=intention to treat. SE=standard error. CACE=complier average causal effect. *Average value of the variable, as a baseline. † $p \leq 0.05$ (one-tailed t test). Values are expressed in MXN\$. Positive values correspond to favourable outcomes. Inpatient care includes self-reported spending on staying overnight in a hospital or health clinic, whereas outpatient care includes spending on medical care that did not require an overnight stay. Medicine includes spending on traditional and non-traditional medications, whereas medical devices include glasses, prostheses, hearing aids, and others.

Resultados

	All participants			Experimental compliers only		
	Control group*	ITT	SE	Control group*	CACE	SE
Utilisation (procedures)						
Used outpatient services	62.6%	-1.5%	1.9%	64.8%	-4.0%	5.2%
Outpatient visits (count)	1.6	-0.03	0.09	1.7	-0.08	0.23
Hospitalised	7.6%	-0.2%	0.5%	7.9%	-0.5%	1.3%
Hospitalisations (count)	0.1	-0.003	0.006	0.1	-0.01	0.02
Satisfaction with provider	68.0%	-1.0%	1.6%	69.8%	-2.6%	4.5%
Utilisation (preventative)						
Eye exam (past year)	10.0%	-0.7%	0.7%	9.8%	-1.8%	1.9%
Flu vaccine	25.7%	-1.8%	1.4%	27.2%	-4.9%	3.7%
Mammogram (past year)	5.1%	-0.9%	0.6%	5.2%	-2.3%	1.6%
Cervical (past year)	21.8%	-1.3%	2.0%	22.2%	-3.2	4.8%
Pap test (past year)	31.9%	-2.3%	2.1%	33.2%	-5.8%	5.0%

*Average value of the variable, as a baseline. Data are percentages, unless otherwise indicated. ITT=intention to treat. SE=standard error. CACE=complier average causal effect.

Table 3: Causal effect of Seguro Popular on utilisation of medical procedures and preventive care

Conclusiones

- Discusión final
 1. El tratamiento tiene un impacto en afiliar a casi el 44% de la población elegible
 2. Los efectos muestran que el SP reduce GBS para consultas ambulatorias y hospitalarias
 3. Efectos mucho más notorios en población más pobre
 4. No hay efectos en gasto en medicamentos

Diseño de experimentos sociales

- Hay alternativas en la forma de aleatoizar
- Pueden haber externalidades
- Puede variar el nivel de aleatorización
- Estratificación
- Cálculos de poder estadístico

Externalidades

- Hasta ahora asumimos que no hay "contaminación" entre individuos
 - Stable Unit Treatment value Assumption (SUTVA)
 - Asume que el hecho de una persona sea tratada no afecta los resultados de otra persona
¿Qué pasa si hay redes sociales fuertes?
- Es importante considerar esto y ver cómo se puede controlar.
- Por ejemplo, ¿qué implica este supuesto para el caso de King et al. (2009)?
 - Recuerde que el SP fue aleatorizado a nivel de comunidad
 - ¿Por qué SUTVA sería complicado si el tratamiento fuese asignado nivel individual en este caso?

Nivel de aleatorización

- En algunos casos se aleatoriza a nivel individual y en otros casos a un nivel más agregado (colegios, comunidades, consultorios)
- ¿Porqué hay diferencia en niveles de aleatorización?
 - Reducir externalidades o contaminación: ej., SUTVA es más probable si aleatorizamos a nivel de comunidad que a nivel individual
 - Aspectos éticos: Es más fácil tener grupos controlados a nivel de comunidad que separar tratamiento y control a nivel de comunidad.

Resumen: Experimentos I

- Los datos experimentales resuelven el problema de selección que encontramos para tener estimadores validos de los efectos del tratamiento.
- Los experimentos permiten a los tomadores de decisiones políticas (públicas) de observar efectos de nuevos tratamientos que no han sido revelados anteriormente.
- Los resultados son convincentes y fáciles de explicar ya que son diferencias medias.
- El diseño es crucial en un experimento: nivel de aleatorización, unidad de análisis, etc.

6. Variables Instrumentales

Variables instrumentales

La visión tradicional

- Repasemos el modelo de regresión lineal estándar:

$$Y_i = \delta D_i + X_i \beta + \epsilon_i$$

- Dónde Y_i es la variable de resultado, D_i es la variable de interés, y X_i es un vector de controles que incluye una constante y dónde $\text{cov}(X, \epsilon) = 0$
- Si $\text{Cov}(D, \epsilon) \neq 0$ entonces decimos que ... D es endogeno y estimadores MCO... ¿es sesgado e inconsistente.
- Esto puede ocurrir por distintas razones: variable omitida, error de medición, simultaneidad, sesgo de selección, etc.

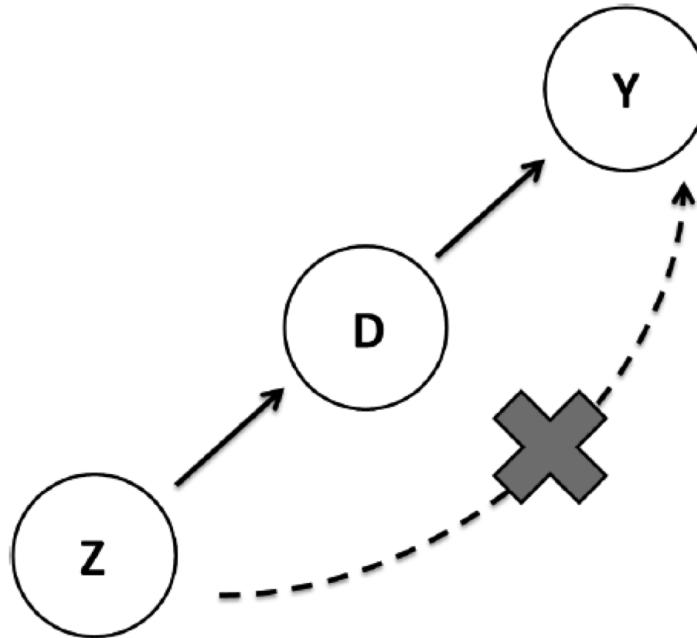
Variables instrumentales

- Podemos pensar en la variable D como:
- $D_i = B_i \varepsilon_i + C_i$
- Dónde $\text{cov}(C, \varepsilon) = 0$ Reemplazando en la regresión anterior tenemos que:
- $Y_i = \delta C_i + X_i \beta + \underbrace{(1 - \delta B_i) \varepsilon_i}_{v_i}$
- Si observáramos los componentes de D_i podríamos correr una regresión de Y sobre C y obtener δ
- Pero en la realidad no observamos los componentes de D_i . Lo mejor que podemos hacer es buscar una variable instrumental.

Variables instrumentales

- La idea detrás de IV es la de encontrar una variable que este correlacionada con C (la parte exógena de D) pero no correlacionada con ε
- Entonces Z es un instrumento para D , cuando:
 - Restricción de exclusión: $Cov(Z, \varepsilon) = 0$. Condicionando en D y X , Z no está correlacionada con ε
 - Condición de instrumento: $Cov(Z, D) \neq 0$. Z está correlacionada con D
- Con estas dos condiciones podemos usar MCO en dos etapas para estimar δ

Variables instrumentales



- La restricción de exclusión afirma que Z afecta a Y sólo a través de D . Z no puede afectar Y a través de otros canales y por eso se "excluye" de la regresión de Y sobre D .

MCO en 2 etapas

- En la primera etapa estimamos una regresión de la variable endógena D con todas las variables exógenas X incluyendo Z :

$$1E \quad D_i = \alpha Z_i + X_i \lambda + v_i$$

- Predecimos los valores de D_i en esta regresión, \hat{D}_i .
- En la segunda etapa, estimamos una regresión de la variable de resultado sobre el valor predicho \hat{D}_i . y las otras X

$$2E \quad Y_i = \delta \hat{D}_i + X_i \beta + \varepsilon_i$$

- $\hat{\delta}$ es el estimador de IV

Asignación aleatoria como IV

- En clase

Ejemplo

Angrist (1990)

Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records

By JOSHUA D. ANGRIST*

The randomly assigned risk of induction generated by the draft lottery is used to construct estimates of the effect of veteran status on civilian earnings. These estimates are not biased by the fact that certain types of men are more likely than others to service in the military. Social Security administrative records indicate that in the early 1980s, long after their service in Vietnam was ended, the earnings of white veterans were approximately 15 percent less than the earnings of comparable nonveterans. (JEL 824)

Ejemplo: Efecto de Servicio Militar sobre Salud y Salarios

- Muchos programas en EE.UU ofrecen servicios de salud, educación, capacitación a veteranos de guerra en parte para compensar por las desventajas a las que estuvieron expuestos por ir a la guerra o al servicio militar
- Por ir al servicio, tienen menos años de experiencia laboral y no reciben la misma compensación que otros no veteranos(as) de la misma edad.
- La **pregunta de investigación** de esos estudios es: ¿Cuál es el efecto del servicio militar sobre la salud y salarios de los veteranos?
- La **pregunta de política pública** es: ¿Son los veteranos compensados de manera adecuada?

Discusión en clase: ¿Por qué creen que sería difícil estudiar estas preguntas mediante un simple MCO?

Usando la selección mediante lotería como instrumento

- El Departamento de Defensa de EE.UU implementó cinco loterías para la guerra de Vietnam en 1970 a 1975 enfocadas en hombres de 19-20 años.
- La de 1970 se enfoco en hombres nacidos en 1950; la de 1971 en hombres de 1951,...
- En las loterías, se asignaron Número aleatorios mediante un muestreo aleatorio de cumpleaños y enumerándolos del 1 al 365
 - Se seleccionaron puntos de corte para cada año y todo aquél por debajo del corte fue seleccionado para ir a la guerra
 - 1970: RSN 195, 1971: RSN 125, 1972: RSN 95
- Angrist construye el siguiente instrumento Z : $ELIG_i = 1$ si $RSN < c$, y $= 0$ en caso contrario.
¿Que quiere instrumentar Angrist? El hecho de que un hombre haya ido o no al servicio militar

Primera etapa

$$(2) \quad VET_i = X'_i \pi_0 + \pi_1 ELIG_i + \eta_i,$$

- VET: Vietnam-era veteran status
- ELIG: the draft-eligibility instrument

Table 2. First-Stage Estimates

Source	Pooled		By Year of Birth	
	1950-52	1950	1951	1952
(1)	(2)	(3)	(4)	
A. White				
SSA	0.184 (0.004)	0.171 (0.007)	0.162 (0.007)	0.225 (0.008)
Census	0.140 (0.001)	0.129 (0.002)	0.134 (0.002)	0.164 (0.002)
B. Nonwhite				
SSA	0.085 (0.007)	0.077 (0.012)	0.085 (0.012)	0.086 (0.013)
Census	0.064 (0.002)	0.062 (0.004)	0.064 (0.004)	0.065 (0.004)

Segunda etapa

$$(1) \quad Y_{it} = X'_i \gamma_t + \beta_t VET_i + \epsilon_{it}$$

- Y: annual FICA taxable earnings, employment status, or disability status

Table 3. OLS and 2SLS Estimates for White Men Born 1950-52

Years	Earnings			Working		
	Nonvet mean	2SLS	OLS	Nonvet mean	2SLS	OLS
(1)	(2)	(3)	(4)	(5)	(6)	
1970-79	6932	-0.153 (0.032)	0.015 (0.006)	0.801	0.043 (0.014)	0.137 (0.002)
1980-89	21204	-0.100 (0.039)	-0.096 (0.008)	0.788	0.004 (0.017)	0.079 (0.003)
1990-99	34652	-0.023 (0.042)	-0.083 (0.009)	0.763	0.030 (0.019)	0.070 (0.003)
2000-07	47246	-0.068 (0.048)	-0.070 (0.010)	0.717	0.012 (0.021)	0.054 (0.004)

Efectos de largo plazo

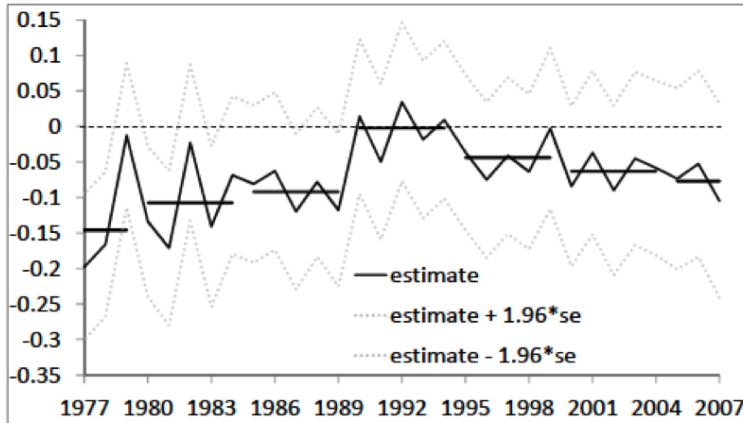


Figure 1. Draft-lottery Estimates of Vietnam-era Service Effects on In(Earnings) for White Men Born 1950-52

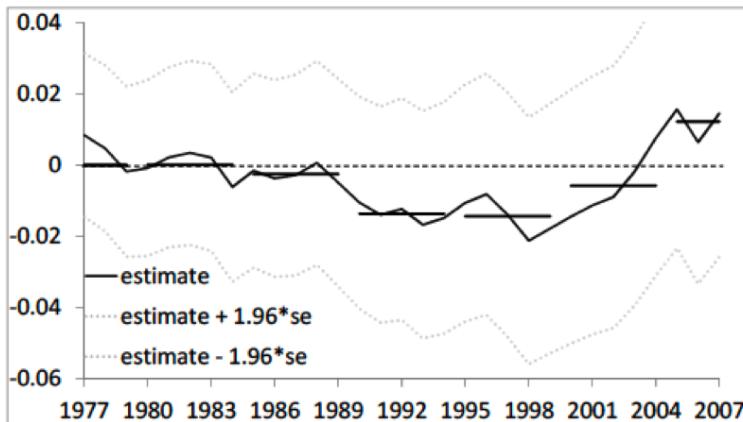


Figure 2. Estimates of the Veteran Effects on Application (Ever) for Disability for White Men Born 1950-52

Efectos en salud

Table 2. Data on Civilian Mortality for White Men Born in 1950 and 1951

Year	Draft eligibility ^a	Number of deaths ^b	Number of suicides ^c	Probability of death ^d	Probability of suicide	Probability of military service ^e
1950	Yes	2,601	436	.0204 (.0004)	.0034 (.0002)	.3527 (.0325)
	No	2,169	352	.0195 (.0004)	.0032 (.0002)	.1934 (.0233)
	<i>Difference (Yes minus No)</i>			.0009 (.0006)	.0002 (.0002)	.1593 (.0401)
	<i>IV estimates^f</i>			.0056 (.0040)	.0013 (.0013)	
1951	Yes	1,494	279	.0170 (.0004)	.0032 (.0002)	.2831 (.0390)
	No	2,823	480	.0168 (.0003)	.0029 (.0001)	.1468 (.0180)
	<i>Difference (Yes minus No)</i>			.0002 (.0005)	.0003 (.0002)	.1362 (.0429)
	<i>IV estimates</i>			.0015 (.0037)	.0022 (.0016)	

^a Determined by lottery number cutoff: RSN 195 for men born in 1950, and RSN 125 for men born in 1951.

- Hombres blancos nacidos en 1950 que fueron a Vietnam por la lotería tuvieron una probabilidad de 0.56% mayor de morir luego de la guerra

Un check list para la estimación de IV

- ¿Es el supuesto de asignación aleatoria válido para el caso de Vietnam?
 - Si, las loterías fueron esencialmente aleatorias
- ¿Cómo es la restricción de exclusión en este caso?
 - Se invalidaría si es que salarios o salud de los hombres con bajo RSN se vieran afectados de alguna otra forma distinta a su participación en la guerra. ¿Hay algún ejemplo en que esto pase? **Difícil**

Un check list para la estimación de IV

- ¿Cómo es la condición de instrumento en este caso? ¿Hay relación entre la lotería e ir a la guerra?
 - Si, los datos muestran que aquellos con RSN bajo el corte tienen mayor probabilidad de ir a la guerra
- ¿Cómo es la restricción de exclusión en este caso?
 - Se invalidaría si es que salarios o salud de los hombres con bajo RSN se vieran afectados de alguna otra forma distinta a su participación en la guerra. ¿Hay algún ejemplo en que esto pase? **Difícil**



Semana 3

Introducción a Inferencia Causal

07 de octubre, 2022

 **Pablo A. Celhay** |  pacelhay@uc.cl

Diseño y formato de la presentación:  José Daniel Conejeros |  jdconejeros@uc.cl |  JDConjeros