

Introducción a la Evaluación de Políticas Públicas

Rodolfo Stucchi

Escuela de Verano sobre Economías Latinoamericanas 2013
CEPAL, Santiago de Chile, Agosto 2013

El contenido y opiniones expresadas son responsabilidad del autor y no representan a las del Banco Interamericano de Desarrollo, su Directorio Ejecutivo o los países que representan.

Saber qué funciona en políticas públicas, así como también el por qué o por qué no funciona, es fundamental para una mejor asignación de los recursos públicos.

El objetivo de esta clase es presentar las principales herramientas usadas para estimar el efecto causal de las políticas públicas.

Referencias:

- Angrist, J. and J. Pischke (2009) *“Mostly harmless econometrics: An empiricist companion,”* Princeton University Press.
- Imbens, Guido W., y Jeffrey M. Wooldridge (2009) “Recent Developments in the Econometrics of Program Evaluation.” *Journal of Economic Literature*, 47(1): 5-86.

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 Métodos Cuasi-experimentales
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - Diferencias en Diferencias
 - Variables Instrumentales
 - Regresión Discontinua
- 4 Temas en evaluación de impacto
 - Validez
 - Dinámica
 - Dosis
 - Heterogeneidad
 - Spillovers

Objetivo de evaluación

- En estricto sentido, busca determinar la efectividad de un programa.
- El objetivo es identificar el impacto de una política (tratamiento) sobre una(s) variable(s) de resultado.
- Un concepto central en evaluación de impacto es el de **atribución o causalidad**, determinar si el cambio en el indicador de interés se debe a la intervención o a otros factores.
- Conceptualmente, se busca generar un **contrafactual**, es decir, determinar qué hubiera pasado sin el programa. Se contrasta la situación “con programa” y la situación “sin programa”, dónde todas las dimensiones relevantes para explicar los resultados (observables y no observables) son iguales

Mensajes esperados luego de la evaluación de impacto

- **Funciona?** Cuáles son los efectos atribuibles al programa ⇒ replicación, ampliación, eliminación
- **Por qué?** El programa produce efectos en la cadenas de indicadores que definen su modelo de intervención? ⇒ revisión del modelo de intervención
- **Para quién?** Para qué categorías de beneficiarios el programa es efectivo ⇒ focalización
- **Es eficiente?** Una vez estimados los efectos realmente atribuibles al programa puedo hacer una análisis costo-beneficio o costo-efectividad

Outline

- 1 Conceptos básicos
 - Causalidad
 - **Contrafactual**
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 Métodos Cuasi-experimentales
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - Diferencias en Diferencias
 - Variables Instrumentales
 - Regresión Discontinua
- 4 Temas en evaluación de impacto
 - Validez
 - Dinámica
 - Dosis
 - Heterogeneidad
 - Spillovers

El impacto del programa y el contrafactual

- El impacto de un programa es la diferencia entre:
 - Los resultados que los beneficiarios del programa obtienen después de participar en el programa y
 - Los resultados que esos mismos beneficiarios hubieran obtenido en ese mismo momento si no hubiesen participado en el programa (**contrafactual**)

- Para estimar el impacto del programa necesitamos estimar el contrafactual. **Qué hubiera pasado si?**

Contrafactual observable?



Sólo en cine . . .

En la práctica el contrafactual no es observable

- La evaluación de impacto intenta construir artificialmente (estadísticamente) un contrafactual
- Normalmente, esto se logra seleccionando un grupo de individuos/empresas/localidades que no participaron en el programa \Rightarrow grupo de control o grupo de comparación
- La validez de una evaluación depende de cuan buena es la estimación del contrafactual

Alternativas para construir un contrafactual

- Comparar la situación antes y después del programa \Rightarrow incorrecto: muchos otros factores que afectan el resultado pueden haber cambiado
- Comparar un grupo con programa y otro sin programa \Rightarrow supone que los grupos eran iguales antes del programa, que estuvieron expuestos a los mismos factores
- Para generar un grupo de referencia válido, hay dos opciones básicas:
 - Métodos experimentales
 - Métodos no-experimentales

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 **Metodos Experimentales**
 - **Metodos Experimentales**
- 3 Métodos Cuasi-experimentales
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - Diferencias en Diferencias
 - Variables Instrumentales
 - Regresión Discontinua
- 4 Temas en evaluación de impacto
 - Validez
 - Dinámica
 - Dosis
 - Heterogeneidad
 - Spillovers

- Experimentos aleatorios: personas/productores/empresas que aplican a un programa son asignados aleatoriamente a uno o más grupos de tratamiento y un grupo de control.
- Es la mejor manera de determinar si el programa funciona.
- Tiene limitaciones respecto a los experimentos de laboratorio: no hay placebo, hay desertores, externalidades, etc.
- Pero . . . se puede hacer evaluaciones experimentales en programas sociales?
- Puede haber cuestionamientos éticos y de eficiencia.
- Sin embargo, son una una manera transparente y justa de asignar beneficios cuando no hay presupuesto para todos y los potenciales beneficiarios son homogéneos.

Asignación aleatoria: Por qué funciona?

- El resultado potencial para el tratado si no participa es $Y_{1,0}$
- El resultado potencial para el tratado si participa es $Y_{1,1} = Y_{1,0} + \delta$
- El efecto causal del programa es $Y_{1,1} - Y_{1,0} = \delta$
- El resultado potencial para el no tratado cuando no participa es $Y_{0,0}$
- Tomando la diferencia en el promedio de los tratados y no tratados se obtiene

$$E(Y_{1,1}) - E(Y_{0,0}) = \delta + E(Y_{1,0}) - E(Y_{0,0})$$

- La aleatorización asegura que $E(Y_{1,0}) = E(Y_{0,0})$.
- Por lo tanto, si el tratamiento es asignado de manera aleatoria, la simple diferencia de medias estima el impacto del programa.

$$E(Y_{1,1}) - E(Y_{0,0}) = \delta$$

Asignación aleatoria: Se utiliza en la práctica?

- Muy común en medicina. Ej: nuevos tratamientos, medicinas, etc.
- Programas públicos de capacitación
 - Estos programas apuntan a aumentar el empleo y el ingreso
 - Estudios no experimentales mostraron que las personas capacitadas por estos programas tienen sueldos inferiores al grupo de comparación (Ashenfelter 1978; Ashenfelter and Card 1985, Lalonde 1995)
 - Problema de selección: Estos programas están diseñados para beneficiar a hombres y mujeres que tienen un potencial de ingresos bajos
 - Evaluaciones aleatorias han encontrado efecto positivo (Lalonde 1986, Orr et al 1996)
- Es una práctica que se está volviendo cada vez más común cuando el objetivo es aprender sobre qué es lo que realmente funciona en políticas públicas.

Asignación aleatoria: El experimento STAR, Tennessee

- Hay un vínculo causal entre el tamaño de la clase y el aprendizaje de los niños?
- El experimento STAR en Tennessee se diseñó para responder esta pregunta
- Estudios no-experimentales encontraron que no hay (o hay muy poca) evidencia de que los grupos pequeños favorecen el aprendizaje
- El programa costó aproximadamente \$12 millones y se implementó para una cohorte de niños de pre-escolar en 1985/86
- El estudio se hizo por cuatro años hasta que los niños llegaron a tercer grado
- Participaron aproximadamente 11,600 niños
- El tamaño medio de una clase regular en Tennessee en 1985/86 era de 22.3 niños por clase

- El experimento asignó a los estudiantes en uno de tres tratamientos
 - Clase pequeña (13-17 estudiantes)
 - Clase regular (22-25 estudiantes) y ayuda de maestro part-time
 - Clase regular (22-25 estudiantes) y ayuda de maestro full-time
- Krueger (1999) encontró que
 - (Como se espera luego de un experimento aleatorio) Las características de los estudiantes (edad, raza, genero) así como las características del maestro (raza, experiencia, nivel educativo, etc) estaban balanceadas entre los tres tratamientos
 - Los estudiantes en las clases pequeñas tuvieron mejor desempeño no que los estudiantes en otras clases
 - No encontró diferencias por la dedicación del maestro

PATCA: Un programa de desarrollo productivo con evaluación experimental

- PATCA II: segunda parte de un programa nacional de servicio de extensión agrícola en implementación en Rep. Dominicana.
- El programa provee un voucher para financiar el 80% del costo de adoptar una tecnología específica elegida por el productor agrícola.
- Criterios de elegibilidad:
 - Ciudadano Dominicano
 - Realiza actividades de agricultura
 - Pequeño o mediano productor
 - No haber sido beneficiado por PATCA I
 - Presentar evidencia suficiente que pruebe capacidad financiera para adquirir el 20% de la tecnología.

- El programa financia 8 tipos diferentes de tecnologías agrícolas:
 - Tecnología 1 = Nivelación de suelo
 - Tecnología 2 = Riego por Aspersión
 - Tecnología 3 = Riego por Micro-Aspersión
 - Tecnología 4 = Riego por Goteo
 - Tecnología 5 = Rehabilitación y Conservación de Pastos
 - Tecnología 6 = Cultivos Protegidos: Acolchados
 - Tecnología 7 = Cultivos Protegidos: Estructura Rústica
 - Tecnología 8 = Pos cosecha de Vegetales y Frutas
- El programa tiene una duración de 4 años (4 cohortes de tratamiento).
- Los productores se registraron para participar en el programa en las 8 oficinas regionales del PATCA.
- Restricciones de presupuesto
- Restricciones por capacidad limitada de oferta de tecnologías
- La randomización fue utilizada para manejar la sobre-demanda (las tecnologías 1, 3 y 6 no fueron aleatorizadas)

Aleatorización del PATCA



Actualmente el programa está en ejecución y se está haciendo la recolección de datos para la evaluación.

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 **Métodos Cuasi-experimentales**
 - **Métodos Cuasi-experimentales**
 - Propensity Score Matching
 - Diferencias en Diferencias
 - Variables Instrumentales
 - Regresión Discontinua
- 4 Temas en evaluación de impacto
 - Validez
 - Dinámica
 - Dosis
 - Heterogeneidad
 - Spillovers

Y si no es posible aleatorizar . . . Cómo construir el contrafactual?

- Los individuos no son asignados al tratamiento por un proceso aleatorio y, por lo tanto, no hay un grupo control experimental
- Es esencial entender y modelar el proceso por el cual se asigna el tratamiento
 - Autoselección (decisión individual)
 - Selección administrativa (individuos asignados según criterios específicos)
 - Combinación de ambas

El problema fundamental de la evaluación

- A = productores que reciben el programa
- B = productores que no reciben el programa
- Y = productividad
- El efecto del programa sobre los participantes es

$$ATT = E(Y_A | A \text{ participa}) - E(Y_A | A \text{ no participa})$$

- $E(Y_A | A \text{ no participa})$ es un contrafactual y, por lo tanto, no observado
- Si se usa el grupo B para el calculo, todos lo valores son observables y se obtiene lo siguiente

$$E(Y_A | A \text{ participa}) - E(Y_B | B \text{ no participa})$$

- Pero ... Es esto el efecto del programa?

El sesgo de selección

- Sumando y restando el contrafactual para el grupo A, se obtiene

$$\begin{aligned} & E(Y_A|A \text{ participa}) - E(Y_B|B \text{ no participa}) = \\ & E(Y_A|A \text{ participa}) - E(Y_A|A \text{ no participa}) + \\ & E(Y_A|A \text{ no participa}) - E(Y_B|B \text{ no participa}) \end{aligned}$$

- La primer linea es el efecto del programa buscado, ATT
- La segunda linea es la diferencia en productividad entre los dos grupos cuando ninguno participa (diferencias ex-ante). Se conoce como sesgo de selección.
- Entonces comparando las magnitudes observables, no se obtiene el efecto del programa. Sino que se obtiene el efecto del programa más el sesgo de selección:

$$E(Y_A|A \text{ participa}) - E(Y_B|B \text{ no participa}) = ATT + \text{Sesgo de Selección}$$

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 **Métodos Cuasi-experimentales**
 - Métodos Cuasi-experimentales
 - **Propensity Score Matching**
 - Diferencias en Diferencias
 - Variables Instrumentales
 - Regresión Discontinua
- 4 Temas en evaluación de impacto
 - Validez
 - Dinámica
 - Dosis
 - Heterogeneidad
 - Spillovers

Propensity Score Matching (PSM)

- PSM busca corregir el sesgo de selección utilizando características observables que afectan la asignación al tratamiento
 - Si se tiene información de A y B antes del programa, es posible estimar la probabilidad de participación, y luego usarla para comparar a cada beneficiario con un no beneficiario que tenía la misma probabilidad de participar, eliminando así sesgos por características observables.
- Supuesto de Independencia Condicional (CIA): Dado un conjunto de variables observables X , que no son afectadas por el tratamiento, los resultados potenciales son independientes del status de tratamiento.
- Si el status de tratamiento (selección) depende de características no observables, el estimador es sesgado.
- En prácticamente todos los programas hay selección por no observables.

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 **Métodos Cuasi-experimentales**
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - **Diferencias en Diferencias**
 - Variables Instrumentales
 - Regresión Discontinua
- 4 Temas en evaluación de impacto
 - Validez
 - Dinámica
 - Dosis
 - Heterogeneidad
 - Spillovers

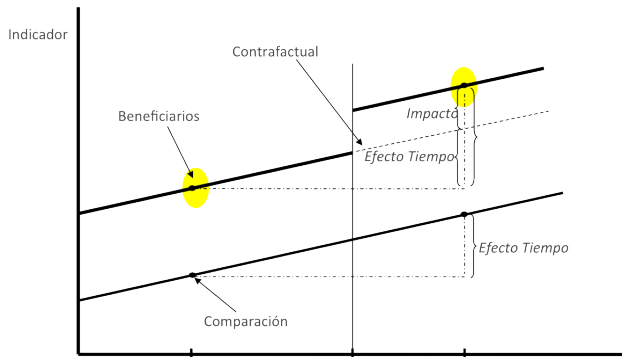
- Requiere el uso de datos de panel
- Corrige el sesgo de la estimación antes/después (AD) y con/sin (CS), usando un grupo de comparación como contrafactual:
- El estimador DD usa el AD en el grupo de comparación para quitar el efecto tiempo, y la diferencia ex-ante para quitar la diferencia entre grupos. Por ejemplo, si el programa ocurre en el el año 2000, y hay datos en 1999 y 2000, el estimador DD sería:

$$DD = E(Y_{A,2001} - Y_{A,1999}) - E(Y_{B,2001} - Y_{B,1999})$$

o

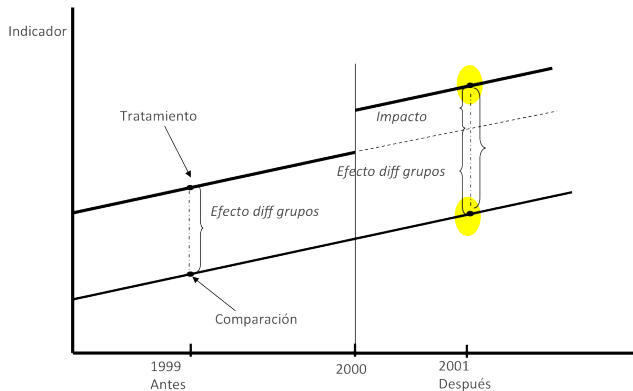
$$DD = E(Y_{A,2001} - Y_{B,2001}) - E(Y_{A,1999} - Y_{B,1999})$$

Estimador Antes/Después (AD)



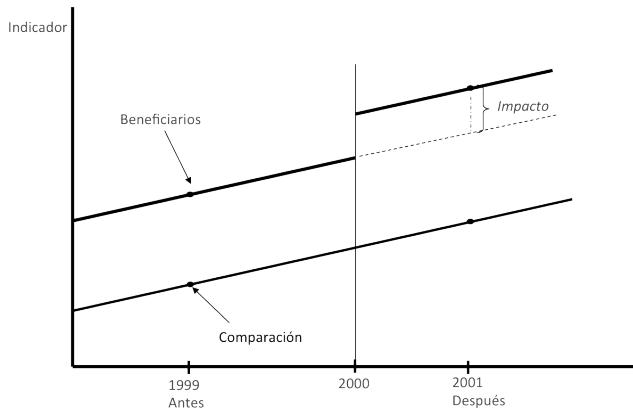
El estimador AD mide el impacto + efecto tiempo

Estimador Con/Sin (CS)



El estimador CS mide el impacto + diferencia entre grupos

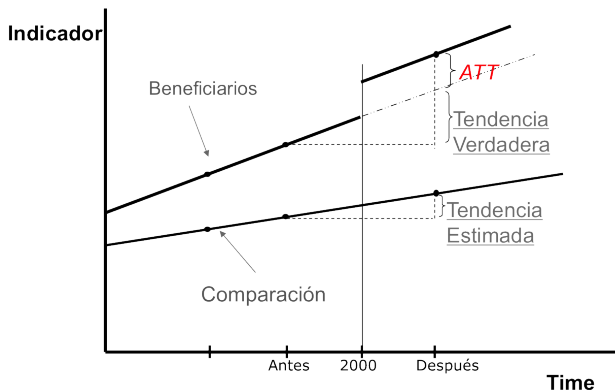
Estimador de Doble Diferencia (DD)



El supuesto fundamental del método DD es que las tendencias son similares (y que los no observables no cambian en el tiempo)

Estimador de Doble Diferencia (DD)

Qué ocurre si el supuesto de tendencias similares falla?



Los años anteriores son útiles para hacer más creíble el supuesto de que las tendencias hubiesen sido las mismas en ausencia del programa.

DD es un estimador muy usado en la evaluación de Programas de Desarrollo Productivo

- El diseño de los programas y la estructura de los datos existentes favorecen el uso de este estimador. En algunos casos se utiliza PSM para hacer el supuesto de tendencias iguales más creíble.
- Algunos ejemplos de evaluaciones que usaron este estimador:
 - **Servicios de extensión agropecuaria:** Programa Ganadero (Uruguay), PREDEG (Uruguay), PROSAP (Argentina), PROMSA (Ecuador), PROMOSA (Panama) y PATCA (Republica Dominicana)
 - **Innovación:** FONTAR (Argentina), PRE (Argentina), Innova (Chile), Fondos Sectoriales (Brasil), COLCIENCIAS (Colombia)
 - **Programas de Desarrollo de Proveedores:** PDP (Chile)
 - **Clusters y Cadenas de Valor:** TICs (Argentina), APL (Brasil)
 - **Crédito:** BANCOLDEX (Colombia) BNDES (Brasil), FNG (Colombia)

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 **Métodos Cuasi-experimentales**
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - Diferencias en Diferencias
 - **Variables Instrumentales**
 - Regresión Discontinua
- 4 Temas en evaluación de impacto
 - Validez
 - Dinámica
 - Dosis
 - Heterogeneidad
 - Spillovers

- En muchas ocasiones no podemos asignar el tratamiento, T , independientemente de Y , ni siquiera condicionalmente.
- En estos casos la estimación del efecto del programa, β , resultará inconsistente porque el término del error, u , estará correlacionado con T . Por qué?

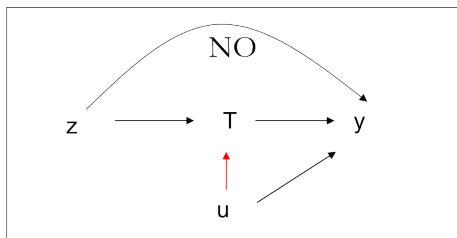
$$Y = \alpha + \beta T + \gamma X + u$$

- Pero quizás podemos afectar a T con otra variable que sea independiente de Y (que sea exógena al modelo de Y)
- Esa función la puede cumplir una variable Z a la que llamaremos variable instrumental

Variables Instrumentales

Una variable instrumental debe cumplir dos condiciones:

- 1 Relevancia: Estar relacionada con la participación en el programa
- 2 Exogeneidad: no afectar el indicador de ninguna otra manera



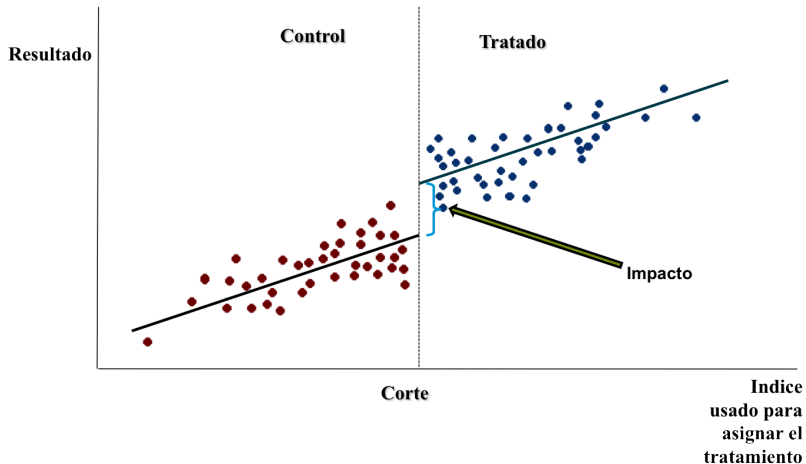
El mejor ejemplo es una lotería. En programas de desarrollo productivo puede pensarse en un esquema de incentivos a la participación en el programa.

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 **Métodos Cuasi-experimentales**
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - Diferencias en Diferencias
 - Variables Instrumentales
 - **Regresión Discontinua**
- 4 Temas en evaluación de impacto
 - Validez
 - Dinámica
 - Dosis
 - Heterogeneidad
 - Spillovers

- El tratamiento se asigna sobre la base de una variable (llamada “variable de asignación”)
- Si el valor de la variable cae en un lado del corte/umbral, la persona es asignada al grupo tratamiento
- Si el valor de la variable cae del otro lado del corte/umbral, la persona es asignada al grupo control
- Ejemplos: puntaje en un examen, ingreso familiar (medidos antes del programa), calificación de una empresa, etc.
- La variable de asignación es frecuentemente una medida de necesidad o de mérito
- Interpretación local: las estimaciones sólo se pueden generalizar a las personas cerca del corte/umbral

Regresión Discontinua



Ejemplos de evaluaciones con RD

- Benavente, Crespi, Figal Garone y Maffioli (2012) evalúan el impacto de FONDECYT en el número y calidad de las publicaciones de investigadores en Chile
 - Variable de asignación: posición en el ranking
 - Regresión discontinua difusa \Rightarrow La variable de asignación NO determina de manera determinística la participación en el programa. En estos casos, se llaman “Diseño Fuzzy” y el puesto en el ranking se usa como variable instrumental.
- En Programas de Desarrollo Productivo: El diseño de los **programas de emprendimiento** permite aplicar esta metodología. Todavía no hay evaluaciones de programas de emprendimiento con esta metodología.
- Programa de Vivienda de Interés Social (Honduras): El subsidio se entrega a hogares que más lo necesitan de acuerdo a un ranking que se construye utilizando la información de los postulantes.

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 Métodos Cuasi-experimentales
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - Diferencias en Diferencias
 - Variables Instrumentales
 - Regresión Discontinua
- 4 **Temas en evaluación de impacto**
 - **Validez**
 - Dinámica
 - Dosis
 - Heterogeneidad
 - Spillovers

Validez de las evaluaciones: 4 dimensiones

- **Interna:** El impacto estimado en la muestra se puede atribuir al programa? ⇒ Atribución
- **Externa:** Podemos generalizar los resultados de nuestro estudio a otros grupos, regiones, momentos en el tiempo? ⇒ Representatividad
- **Medición:** Estamos midiendo adecuadamente lo que nos interesa medir? ⇒ Especificidad
- **Estadística:** Estamos tomando en cuenta adecuadamente la incertidumbre estadística? ⇒ Significancia

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 Métodos Cuasi-experimentales
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - Diferencias en Diferencias
 - Variables Instrumentales
 - Regresión Discontinua
- 4 **Temas en evaluación de impacto**
 - Validez
 - **Dinámica**
 - Dosis
 - Heterogeneidad
 - Spillovers

- Los efectos de un programa pueden no ser instantáneos ni permanentes. Pueden tomar tiempo en aparecer y tener una duración variable
- La omisión de estos factores puede derivar en conclusiones erróneas sobre la efectividad de un programa
- Observar un período post-programa muy breve puede subestimar los beneficios
- Observar un período alejado del inicio de la intervención puede llevar a subestimar costos (por ej. de ajuste)

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 Métodos Cuasi-experimentales
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - Diferencias en Diferencias
 - Variables Instrumentales
 - Regresión Discontinua
- 4 **Temas en evaluación de impacto**
 - Validez
 - Dinámica
 - **Dosis**
 - Heterogeneidad
 - Spillovers

- Más allá del efecto de la participación como variable binaria (participa o no), es importante analizar el impacto de distintas “dosis”.
 - Ejemplo: Distintos montos de crédito o financiación
- Esto permite estudiar, entre otras cosas, cuál es la “intensidad óptima” del programa.
 - Ejemplo: Cuál es el monto de crédito que maximiza el impacto en productividad?
- Efectos marginales: cuando la variable de tratamiento es continua se puede estimar la elasticidad de los impactos al monto de tratamiento.
 - Binelli and Maffioli (2007) en programa de innovación.

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 Métodos Cuasi-experimentales
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - Diferencias en Diferencias
 - Variables Instrumentales
 - Regresión Discontinua
- 4 **Temas en evaluación de impacto**
 - Validez
 - Dinámica
 - Dosis
 - **Heterogeneidad**
 - Spillovers

- Un programa puede afectar de distinta manera a distintos subgrupos de la población
- El impacto promedio puede ser desagregado para obtener información más precisa sobre el efecto de la intervención
- Ejemplos:
 - Impactos diferenciales entre subgrupos: empresas grandes vs. pequeñas, diferente industria, diferente ubicación geográfica. Diferencias por género o por edad, etc.
 - Efectos en distintas partes de la distribución: el impacto en las empresas más productivas puede diferir del impacto en las menos productivas (efecto del tratamiento por cuantiles). Lo mismo podría ocurrir por nivel de ingreso.

Outline

- 1 Conceptos básicos
 - Causalidad
 - Contrafactual
- 2 Metodos Experimentales
 - Metodos Experimentales
- 3 Métodos Cuasi-experimentales
 - Métodos Cuasi-experimentales
 - Propensity Score Matching
 - Diferencias en Diferencias
 - Variables Instrumentales
 - Regresión Discontinua
- 4 **Temas en evaluación de impacto**
 - Validez
 - Dinámica
 - Dosis
 - Heterogeneidad
 - **Spillovers**

Spillovers

Muchos programas de desarrollo productivo se justifican en base a falta de apropiación (innovación) y externalidades por aglomeración (clusters). Es importante que la evaluación tenga en cuenta estos efectos. Ejemplo de clusters:

