


THE WORLD BANK



IMPACT  
EVALUATION

# Sesión Técnica I Inferencia Causal

Lima, 2009

Departamento de  
Desarrollo Humano

Fondo Español para  
Evaluación de Impacto

## Justificación

---

- Las preguntas de investigación que motivan la mayoría de los estudios en las ciencias de la salud son de naturaleza causal.
- Por ejemplo:
  - ¿Cuál es la eficacia de un medicamento en una población dada?
  - ¿Qué porcentaje de muertes por una enfermedad dada podrían haberse evitado si se hubiera aplicado un tratamiento o una política adecuada?

## Justificación

---

- Las preguntas empíricas más difíciles en economía también involucran relaciones causa-efecto:
  - ¿Mejora la descentralización de las escuelas la calidad de la educación?
  - ¿Un año de capacitación *causa* mayores ingresos?
  - ¿Mejoran las transferencias condicionadas la salud de niños?



## Justificación

---

El interés en estas preguntas está motivado por:

- Preocupaciones de política
  - Ej. ¿ Los programas públicos reducen la pobreza?
- Consideraciones teóricas
- Problemas que enfrentan los tomadores de decisiones.



## Intuición del problema: un ejemplo hipotético

- Un programa de transferencias condicionadas de ingresos para mujeres embarazadas
  - Se someten a controles médicos periódicos
  - Se refuerza la alimentación
  - Se realizan reuniones informativas sobre riesgos de fumar y consumir alcohol durante el embarazo
  
- La pregunta es: el programa tiene algún impacto sobre el peso del niño al nacer?
  - Cuáles son los canales de impacto? Discusión



## Transferencias y peso al nacer

- Asuma que tiene datos sobre mujeres en un pueblo con TMC, y sobre mujeres en un pueblo vecino sin TMC

Y = promedio del peso al nacer, en gramos				
	Mujeres con TMC	Mujeres sin TMC	Diferencia	Doble Difference
	Pueblo con TMC	Pueblo vecino		
Antes del inicio de las TMC	?	?		
Después de que se iniciaron las TMC	3,250	3,100	150	?
Asuma: las diferencias son significativas				



## Transferencias y peso al nacer

- Datos: personas en el programa y personas sin programa, pueblo vecino, y datos antes del programa

Y = promedio del peso al nacer, en gramos

	Mujeres con TMC	Mujeres sin TMC	Diferencia	Doble Diferencia
	Pueblo con TMC	Pueblo vecino		
Antes del inicio de las TMC	3,025	2,840	185	
Después de que se iniciaron las TMC	3,250	3,100	150	?

Asuma: las diferencias son significativas



## Transferencias y peso al nacer

- Asuma que el programa fue asignado a través de una lotería

Y = promedio del peso al nacer, en gramos

	Mujeres con TMC	Mujeres sin TMC	Diferencia	Doble Diferencia
	Pueblo con TMC	Pueblo vecino		
Antes del inicio de las TMC	3,025	2,840	185	
Después que se iniciaron las TMC	3,250	3,100	150	-35

	Mujeres con TMC por lotería	Mujeres sin TMC por lotería	Diferencia	Doble Diferencia
Antes del inicio de las TMC	3,028	3,028	0	
Después de que se iniciaron las TMC	3,250	3,105	145	145

Asuma: las diferencias son significativas



## Análisis Estadístico Estandar

---

- ❑ **Herramientas:** probabilidad y otras técnicas de estimación
- ❑ **Objetivo:** inferir parámetros de una distribución, a partir de muestras obtenidas de esa distribución.
- ❑ **Uso:** Con la ayuda de tales parámetros, uno puede:
  - Inferir asociación entre variables,
  - Estimar la probabilidad de eventos pasados y futuros,
  - Así como actualizar la probabilidad de eventos a la luz de nueva evidencia o nuevas mediciones.
- ❑ **Condición:** las condiciones experimentales no pueden cambiarse.



## Análisis Causal

---

- ❑ El análisis causal va un paso adelante del análisis estadístico estandar:
- ❑ Su objetivo es inferir aspectos del proceso de generación de datos.
- ❑ Con la ayuda de tales aspectos, uno puede
  - deducir la probabilidad de eventos bajo condiciones estáticas (como en análisis estadístico estandar)
  - y también la dinámica de eventos bajo condiciones cambiantes.



## Análisis Causal

---

- La *dinámica de eventos bajo condiciones cambiantes* incluye:
  1. Predecir los efectos de intervenciones
  2. Predecir los efectos de cambios espontáneos
  3. Identificar las causas de eventos reportados
  
- Esta distinción implica que los conceptos causales y de asociación no se mezclan.

11

## Análisis Causal

---

- Análisis estadístico / teoría de probabilidad estandar:
  - La palabra *causa* no está en su vocabulario
  - Todo lo que nos permite decir es que dos eventos están mutuamente *correlacionados*, o son dependientes
  - Es decir: si encontramos uno, podemos esperar encontrar el otro.
  
- Eso no es suficiente para decisiones de política:
  - Necesitamos fundamentos para la toma de decisiones: ej. Si hacemos XXX, lograremos YYY?
  - Tenemos que suplir el término de probabilidad con un vocabulario para *causalidad*.

12

## Análisis Causal

---

Nuestro vocabulario de causalidad

El Modelo de Resultados Potenciales  
de Neyman-Rubin

Ing.: The Neyman-Rubin potential outcome model

(Neyman, 1923; Rubin, 1974)

13

## Modelo Causal de Rubin

---

- Define a la población por la letra  $U$ .  
Cada unidad en  $U$  es denotada por  $u$ .
- Para cada  $u \in U$ , hay asociado un valor  $Y(u)$  de la variable de interés  $Y$ , la cual llamamos variable de respuesta.
- Sea  $A$  una segunda variable definida en  $U$ .  
Llamamos a  $A$  un atributo de las unidades en  $U$ .

14

- ❑ La idea clave es el *potencial* para exponer o no cada unidad a la acción de una causa:
- ❑ Cada unidad tiene que ser potencialmente expuesta a cualquiera de las causas.
- ❑ Así, Rubin **toma** la posición de que las causas son acciones que, en términos hipotéticos, podrían ser tratamientos.
- ❑ Un atributo no puede ser una causa en un experimento, porque la idea de *exposición potencial* no aplica.

15

- ❑ Por simplicidad, asumimos que hay sólo 2 causas o niveles de tratamiento.
- ❑ Sea  $D$  una variable que indica la causa a la cual cada unidad en  $U$  es expuesta:

$$D = \begin{cases} t & \text{si la unidad } u \text{ es expuesta a tratamiento} \\ c & \text{si la unidad } u \text{ es expuesta a control} \end{cases}$$

- ❑ En un estudio controlado,  $D$  es construido por el experimentador.
- ❑ En un estudio sin control,  $D$  es determinado por factores que están más allá del control del experimentador.

16



- ❑ Los valores de  $Y$  son potencialmente afectados por la causa particular,  $t$  o  $c$ , a la cual la unidad es expuesta.
- ❑ Así, necesitamos dos variables de respuesta:

$$Y_t(u), Y_c(u)$$

- ❑  $Y_t(u)$  es el valor de la respuesta que sería observada si la unidad  $u$  fuera expuesta a  $t$ .
- ❑  $Y_c(u)$  es el valor que sería observado en la misma unidad  $u$  si ésta fuera expuesta a  $c$ .



Alternativamente,  $D$  se puede expresar como una variable binaria:

$$D = \begin{cases} 0 & \text{si la unidad } u \text{ es expuesta a tratamiento} \\ 1 & \text{si la unidad } u \text{ es expuesta a control} \end{cases}$$

Entonces,

$Y_1(u)$  es el el resultado si la unidad  $u$  es expuesta a tratamiento

$Y_0(u)$  es el el resultado si la unidad  $u$  es expuesta a control

Entonces, el resultado de cada individuo puede ser escrito como:

$$Y(u) = DY_1(u) + (1 - D)Y_0(u)$$

Nota: Asumamos que el estado de tratamiento de un individuo no afecta los resultados potenciales de otros individuos.



Definition: Para cada unidad tratada  $u$ , el tratamiento causa

$$\delta_u = Y_1(u) - Y_0(u)$$

❑ **Problema Fundamental de Inferencia Causal:**

Es imposible observar el valor de  $Y_1(u)$  y  $Y_0(u)$  en la misma unidad  $u$ ,

=> es imposible observar el efecto de  $t$  en  $u$ .

❑ Es decir: No podemos inferir el efecto de tratamiento por que no tenemos la evidencia *contrafactual*.

*¿Qué hubiera sucedido en ausencia de tratamiento?*



- ❑ Dado que el efecto causal para una sola unidad  $u$  no puede ser observada,
- ❑ Intentamos identificar el **efecto causal promedio** para la población entera o para las sub-poblaciones.
- ❑ Computamos el **efecto de tratamiento promedio (ATE)** de  $t$  (relativo a  $c$ ) sobre  $U$  (o una sub-población de  $U$ ):

$$TE_u = \delta_u = Y_1(u) - Y_0(u)$$

↓

$$\begin{aligned} ATE_U &= E_U [Y_1(u) - Y_0(u)] \\ &= E_U [Y_1(u)] - E_U [Y_0(u)] \\ &= \bar{Y}_1 - \bar{Y}_0 \\ &= \bar{\delta} \end{aligned} \quad (1)$$



- ❑ La solución estadística reemplaza el efecto causal imposible de observar de  $t$  en una unidad  $u$  específica con la posibilidad de estimar el *efecto causal* promedio de  $t$  sobre una población  $U$  de unidades.
- ❑  $E_U(Y_1)$  y  $E_U(Y_0)$  no pueden ser calculados, pero pueden ser estimados.
- ❑ La mayoría de los métodos econométricos intentan construir a partir de datos observados, estimaciones consistentes de

$$E_U(Y_1) = \bar{Y}_1 \quad \text{y} \quad E_U(Y_0) = \bar{Y}_0$$

21

- ❑ Entonces tratamos de estimar:

$$\begin{aligned} ATE_U &= E_U[Y_1(u)] - E_U[Y_0(u)] \\ &= \bar{Y}_1 - \bar{Y}_0 \end{aligned} \quad (1)$$

- ❑ Considere el siguiente estimador simple de ATE (“después de tratamiento”):

$$\hat{\delta} = [\hat{Y}_1 | D = 1] - [\hat{Y}_0 | D = 0] \quad (2)$$

- ❑ La ecuación (1) está definida para la población completa,
- ❑ La ecuación (2) representa un estimado a ser evaluado en una muestra obtenida de esa población.

22

Asuma que  $\pi$  es igual a la proporción de la población que sería asignada al grupo de tratamiento.

Se puede descomponer  $ATE$  de manera siguiente:

$$\begin{aligned}\bar{\delta} &= \pi \bar{\delta}_{\{D=1\}} + (1-\pi) \bar{\delta}_{\{D=0\}} \\ &= \pi \left[ (\bar{Y}_1 - \bar{Y}_0) | D=1 \right] + (1-\pi) \left[ (\bar{Y}_1 - \bar{Y}_0) | D=0 \right] \\ &= \left[ \pi [\bar{Y}_1 | D=1] + (1-\pi) [\bar{Y}_1 | D=0] \right] - \\ &\quad \left[ \pi [\bar{Y}_0 | D=1] + (1-\pi) [\bar{Y}_0 | D=0] \right] \\ &= \bar{Y}_1 - \bar{Y}_0\end{aligned}$$

23

Si se supone que

$$[\bar{Y}_1 | D=1] = [\bar{Y}_1 | D=0] \text{ and } [\bar{Y}_0 | D=1] = [\bar{Y}_0 | D=0]$$

entonces

$$\begin{aligned}\bar{\delta} &= \left[ \pi [\bar{Y}_1 | D=1] + (1-\pi) [\bar{Y}_1 | D=1] \right] - \\ &\quad \left[ \pi [\bar{Y}_0 | D=0] + (1-\pi) [\bar{Y}_0 | D=0] \right]\end{aligned}$$

$$\bar{\delta} = [\bar{Y}_1 | D=1] - [\bar{Y}_0 | D=0]$$

Lo cual puede ser estimado de manera consistente por:

$$\hat{\delta} = [\hat{Y}_1 | D=1] - [\hat{Y}_0 | D=0]$$

24

- Así, una condición suficiente para el estimador estándar para estimar consistentemente el ATE verdadero es que:

$$[\bar{Y}_1|D = 1] = [\bar{Y}_1|D = 0] \quad \text{y} \quad [\bar{Y}_0|D = 1] = [\bar{Y}_0|D = 0]$$

- *El resultado promedio bajo el tratamiento no difiere entre los grupos de tratamiento y de control.*
- *y el resultado promedio bajo control no difiere entre los grupos de tratamiento y de control.*
- A fin de satisfacer estas condiciones, es suficiente que la asignación del tratamiento  $D$  no esté correlacionado con las potenciales distribuciones resultantes de  $Y_1$  e  $Y_0$ .
- La forma principal de conseguir esta no-correlación es a través de la asignación aleatoria del tratamiento.

25

- En la mayoría de los casos, simplemente no hay información disponible sobre cómo aquellos individuos en el grupo control habrían reaccionado si hubieran recibido el tratamiento.
- Ésta es la base para comprender los potenciales sesgos del estimador estándar (2).
- Usando un poco de álgebra, se puede mostrar que:

$$\begin{aligned} \hat{\delta} &= \bar{\delta} + \underbrace{\left([\bar{Y}_0 | D = 1] - [\bar{Y}_0 | D = 0]\right)}_{\text{Diferencia en la línea base}} \\ &+ (1 - \pi) \underbrace{\left(\bar{\delta}_{\{D=1\}} - \bar{\delta}_{\{D=0\}}\right)}_{\text{Heterogeneidad del tratamiento}} \end{aligned}$$

26

$$\hat{\delta} = \bar{\delta} + \underbrace{([\bar{Y}_0 | D = 1] - [\bar{Y}_0 | D = 0])}_{\text{Diferencia en la línea base}} + (1 - \pi) \underbrace{(\bar{\delta}_{\{D=1\}} - \bar{\delta}_{\{D=0\}})}_{\text{Heterogeneidad del tratamiento}}$$

- Esta ecuación especifica las dos fuentes de sesgo que necesitan ser eliminadas de las estimaciones de efectos causales de estudios observacionales.
  1. Sesgo de selección: diferencias en la línea de base.
  2. Heterogeneidad del tratamiento
- La mayoría de los métodos disponibles solamente se enfocan en el sesgo de selección, asumiendo que el efecto del tratamiento es constante en la población o redefiniendo los parámetros de interés en la población.

27

## Tratamiento en los tratados

- *ATE* no siempre es el parámetro de interés.
- En una variedad de contextos de política, el efecto de tratamiento promedio *para la persona bajo intervención* es de sumo interés:

$$\begin{aligned} TOT &= E [Y_1(u) - Y_0(u) | D = 1] \\ &= E [Y_1(u) | D = 1] - E [Y_0(u) | D = 1] \end{aligned}$$

28

## Tratamiento en la persona tratada

---

- Cuando queremos estimar  $TOT$  ...

$$TOT = E [Y_1(u) | D = 1] - E [Y_0(u) | D = 1]$$

- ... el estimador estándar (2) ...

$$\hat{\delta} = [\hat{Y}_1 | D = 1] - [\hat{Y}_0 | D = 0]$$

- ... estima consistentemente  $TOT$  a condición de que:

$$[\bar{Y}_0 | D = 1] = [\bar{Y}_0 | D = 0]$$



## Referencias

---

- Judea Pearl (2000): Causality: Models, Reasoning and Inference, CUP. Chapters 1, 5 and 7.
- Trygve Haavelmo (1944): "The probability approach in econometrics," *Econometrica* 12, pp. iii-vi+1-115.
- Arthur Goldberger (1972): "Structural Equations Methods in the Social Sciences," *Econometrica* 40, pp. 979-1002.
- Donald B. Rubin (1974): "Estimating causal effects of treatments in randomized and nonrandomized experiments," *Journal of Educational Psychology* 66, pp. 688-701.
- Paul W. Holland (1986): "Statistics and Causal Inference," *Journal of the American Statistical Association* 81, pp. 945-70, with discussion.

