



THE WORLD BANK



Medición de Impacto

Métodos de evaluación de impacto

Sebastián Martínez
Banco Mundial

Nota: diapositivas preparadas por Sebastián Martínez. El contenido de esta presentación refleja las opiniones del autor y no necesariamente las del Banco Mundial. Diciembre de 2007.

Medición de impacto

1) Inferencia causal

- Contrafactuales
- Contrafactuales falsos:
 - Antes y después (previo-posterior)
 - Inscrito-no inscrito (manzanas y naranjas)

2) Métodos de evaluación de impacto:

- Controles aleatorios
- Promoción aleatoria (IV)
- Diseño de discontinuidad/RDD (regression discontinuity design, *RDD* por su sigla en inglés)
- Diferencia en diferencias
- Pareamiento/matching (*propensity score matching*, *PSM* por su sigla en inglés)



Evaluación de impacto



- Marco lógico
 - Teoría
- **Medición de impacto**
 - **Estrategia de identificación**
- Datos
- Plan de operaciones
- Recursos

Medición de impacto

1) Inferencia causal

Contrafactuales

- Contrafactuales falsos:
 - Antes y después (previo-posterior)
 - Inscrito-no inscrito (manzanas y naranjas)

2) Métodos de evaluación de impacto:

- Controles aleatorios
- Promoción aleatoria (IV)
- Diseño de discontinuidad (RDD)
- Diferencia en diferencias
- Pareamiento/matching

Nuestro objetivo:

- Calcular el efecto **CAUSAL** (*impacto*) de
 - la intervención **P** (*programa o tratamiento*)
 - en
 - el resultado **Y** (*indicador, medida del éxito*)

- Ejemplo: ¿cuál es el efecto de un
 - **programa de transferencia monetaria (P)**
 - en
 - **el consumo del hogar (Y)**?

Inferencia causal

□ ¿Cuál es el efecto de **P** en **Y**?

□ Respuesta:

$$\alpha = (Y \mid P=1) - (Y \mid P=0)$$

□ ¿Listos? ¿Nos vamos a casa?

Problema de FALTA DE DATOS

$$\alpha = (Y | P=1) - (Y | P=0)$$

Para un beneficiario del programa:

- observamos ($Y | P=1$):
 - El nivel de consumo (Y) con un programa de transferencia monetaria (P)
- Pero no observamos ($Y | P=0$):
 - El nivel de consumo (Y) sin un programa de transferencia monetaria (P)

Solución

- ▣ Estimar lo que *hubiera* sucedido a **Y** en ausencia de **P**
- ▣ Esto se denomina...

CONTRAFACTUAL

La clave de una evaluación de impacto es tener un **contrafactual** válido



Cálculo del impacto de P en Y

$$\alpha = (Y \mid P=1) - (Y \mid P=0)$$

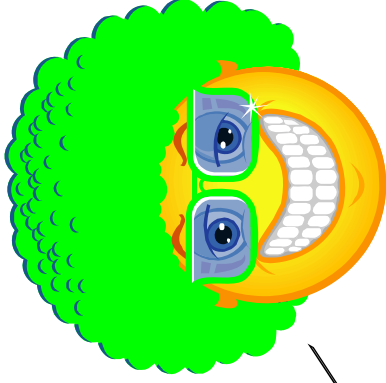
- ▣ OBSERVAR ($Y \mid P=1$)
- ▣ Intención de Tratar (**ITT**)
 - A quienes se ofreció tratamiento
- ▣ Tratamiento en Tratados (**TOT**) - Quienes están recibiendo tratamiento

- ▣ **CALCULAR** contrafactual para ($Y \mid P=0$)
 - ▣ Utilizar grupo de **comparación** o de **control**

IMPACTO = resultado con tratamiento -
contrafactual

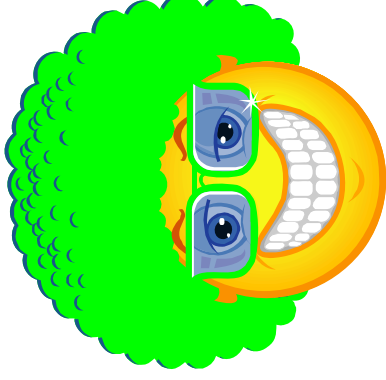
El “clon” perfecto

Beneficiario



6 dulces

Control

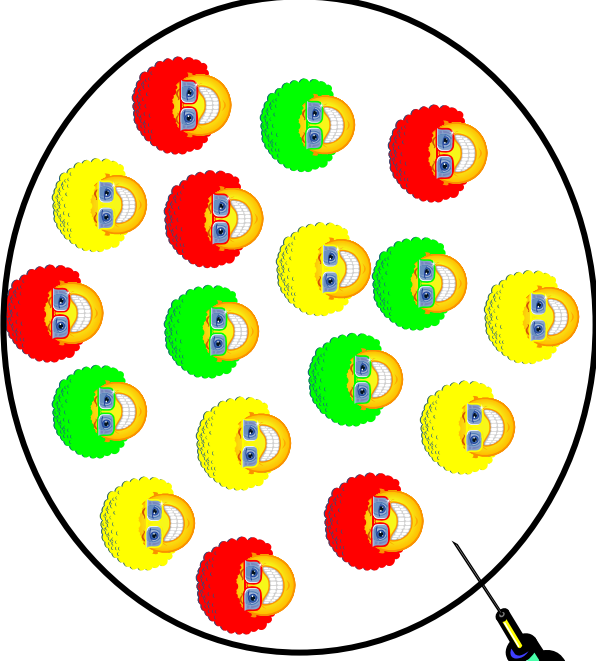


4 dulces

Impacto = 6 - 4 = 2 dulces

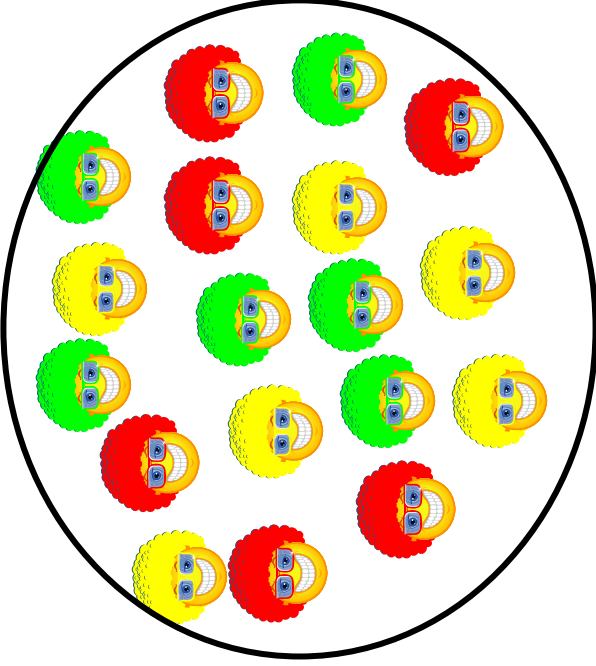
En realidad, utilizar estadística

Beneficiario



Promedio Y = 6 dulces

Control



Promedio Y = 4 dulces



Impacto = 6 - 4 = 2 dulces



Obtención de un buen contrafactual

- Entender el proceso de GENERACIÓN DE DATOS
 - Proceso conductual por el que se determina la participación en el programa (tratamiento)
 - ¿Cómo se asignan los beneficios?
 - ¿Cuáles son las reglas de elegibilidad?

- La observación tratada y el contrafactual:
 - tienen características idénticas, con excepción de los beneficios de la intervención

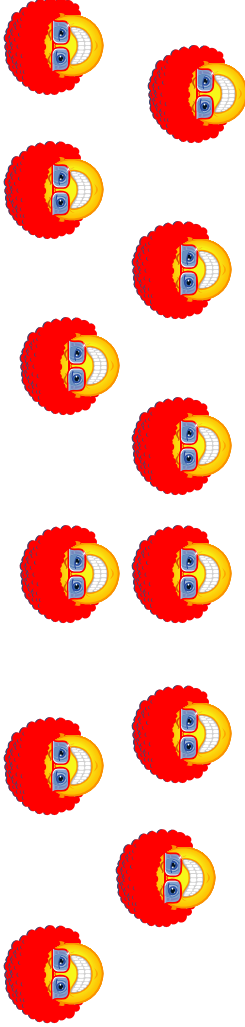
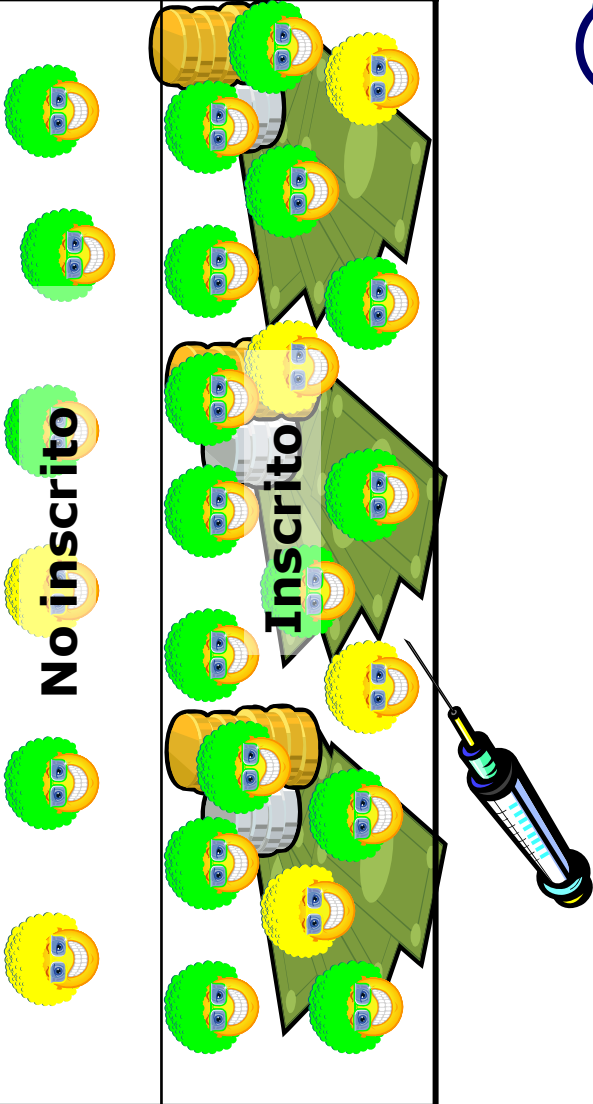
Con un buen contrafactual, el **único motivo** de la diferencia de resultados entre el grupo de tratamiento y el grupo de control es la **intervención (P)**

Estudio de caso

- ▣ ¿Cuál es el efecto de **un programa de transferencia monetaria (P)** en **el consumo del hogar (Y)**?
- ▣ Programa PROGRESA/OPORTUNIDADES
 - Programa nacional contra la pobreza en México
 - ▣ Comenzó en 1997
 - ▣ 5 millones de beneficiarios hacia 2004
 - ▣ Elegibilidad – basado en índice de pobreza
 - Transferencias monetarias
 - ▣ condicionadas a la asistencia escolar y visitas a centros de salud
 - Evaluación de impacto rigurosa con gran cantidad de datos
 - ▣ 506 comunidades, 24 mil hogares
 - ▣ Datos iniciales, 1997; seguimiento, 2008
 - Muchos resultados de interés. Aquí consideramos:
 - ▣ **Estándar de vida: consumo per cápita**

Estudio de casos

Elegibilidad e Inscripción

<p>No elegible (No pobre)</p>	
<p>Elegible (Pobre)</p>	<p>No inscrito</p> 

Medición de impacto

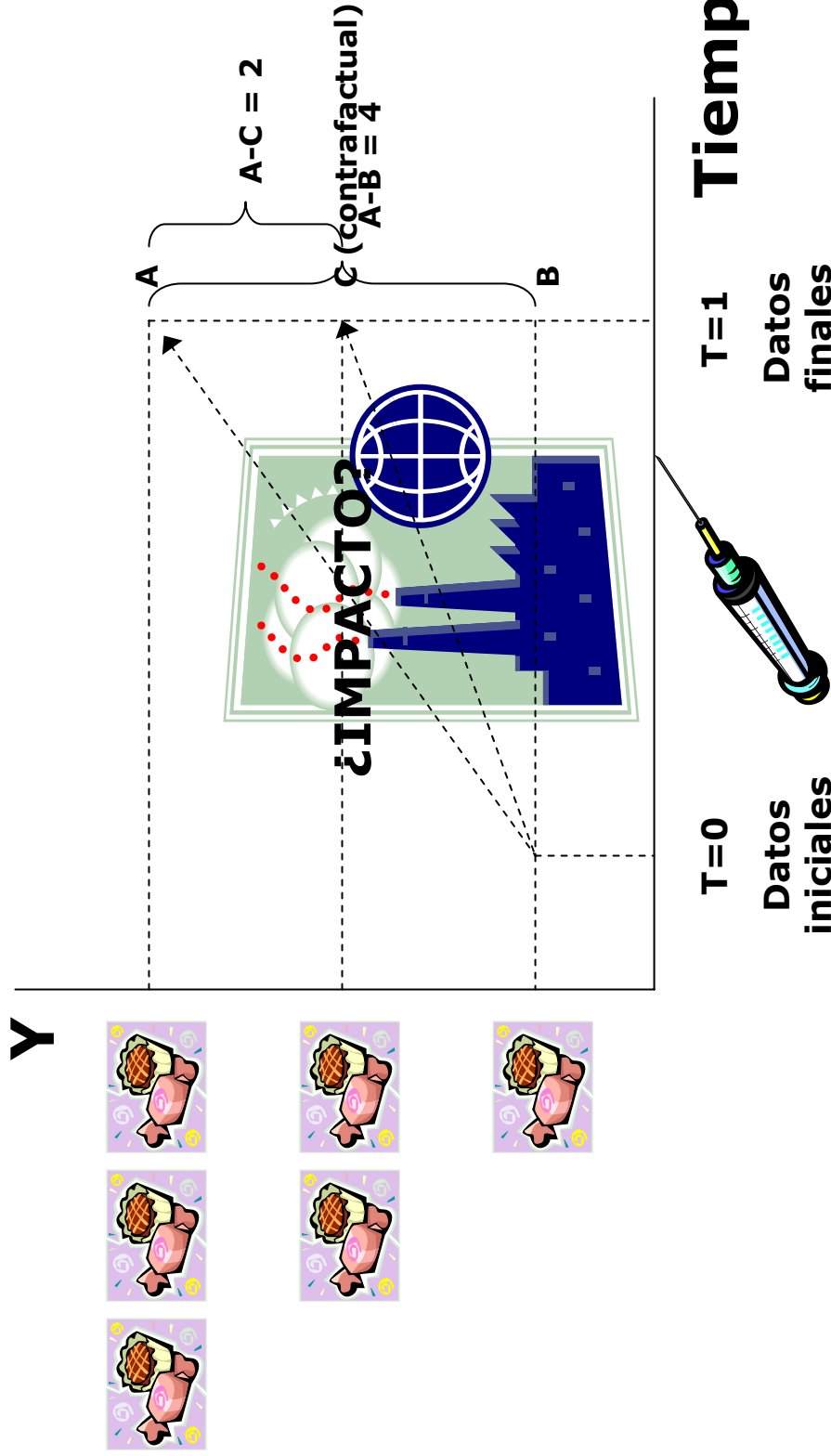
- 1) Inferencia causal
 - Contrafactuales
 - Contrafactuales falsos:**
 - Antes y después (previo-posterior)**
 - Inscrito-no inscrito (manzanas y naranjas)
- 2) Métodos de evaluación de impacto:
 - Controles aleatorios
 - Promoción aleatoria (IV)
 - Diseño de discontinuidad (RDD)
 - Diferencia en diferencias
 - Pareamiento/matching

Contrafactuales falsos

- ❑ Dos contrafactuales comunes que deben evitarse:
 - ❑ Antes y después (previo-posterior)
 - ❑ Datos sobre los mismos individuos antes y después de la intervención
 - ❑ Inscritos-no inscritos (manzanas y naranjas)
 - ❑ Datos sobre un grupo de individuos inscrito en el programa, y otro grupo no inscrito
 - ❑ No conocemos la causa
- ❑ Ambos contrafactuales pueden llevar a resultados sesgados

Contrafactual falso n.º 1

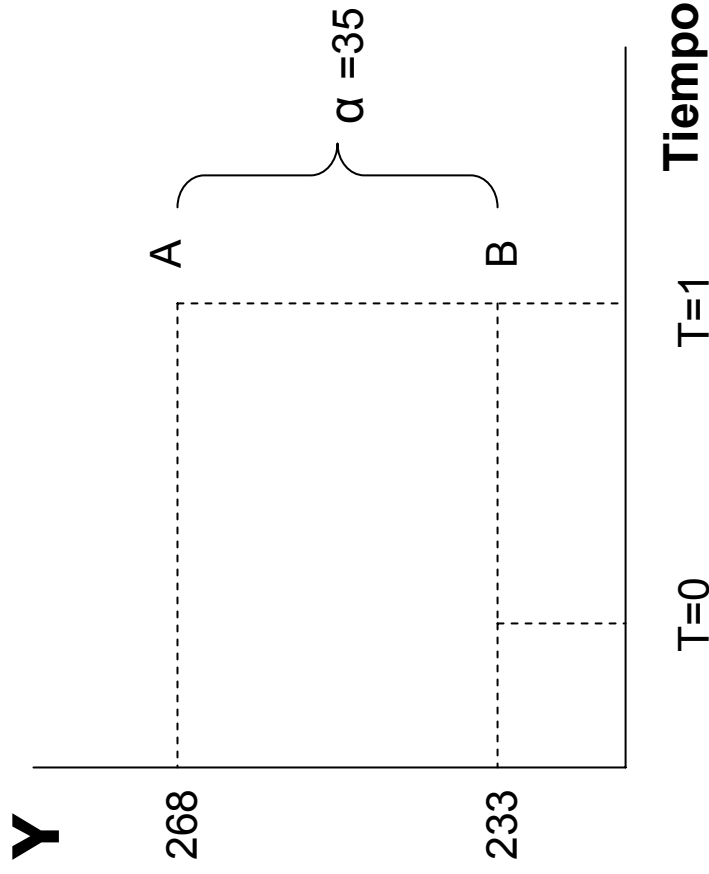
Antes y Después



Caso 1: Antes y Después

¿Cuál es el efecto de **un programa de transferencia monetaria (P)** en el **consumo del hogar (Y)**?

- 2 Puntos en el Tiempo
- Se mide a beneficiarios en:
 - Consumo en T=0
 - Consumo en T=1
- Cálculo del contrafactual
($Y_{i,t} | P=0$) = ($Y_{i,t-1} | P=0$)
- "Impacto" = $A - B = 35$



Caso 1: Antes y Después

Caso 1 - Antes y Después

	Antes	Despues	t-stat
Consumo Promedio	233.47	268.75	16.3

Caso 1 - Antes y Después

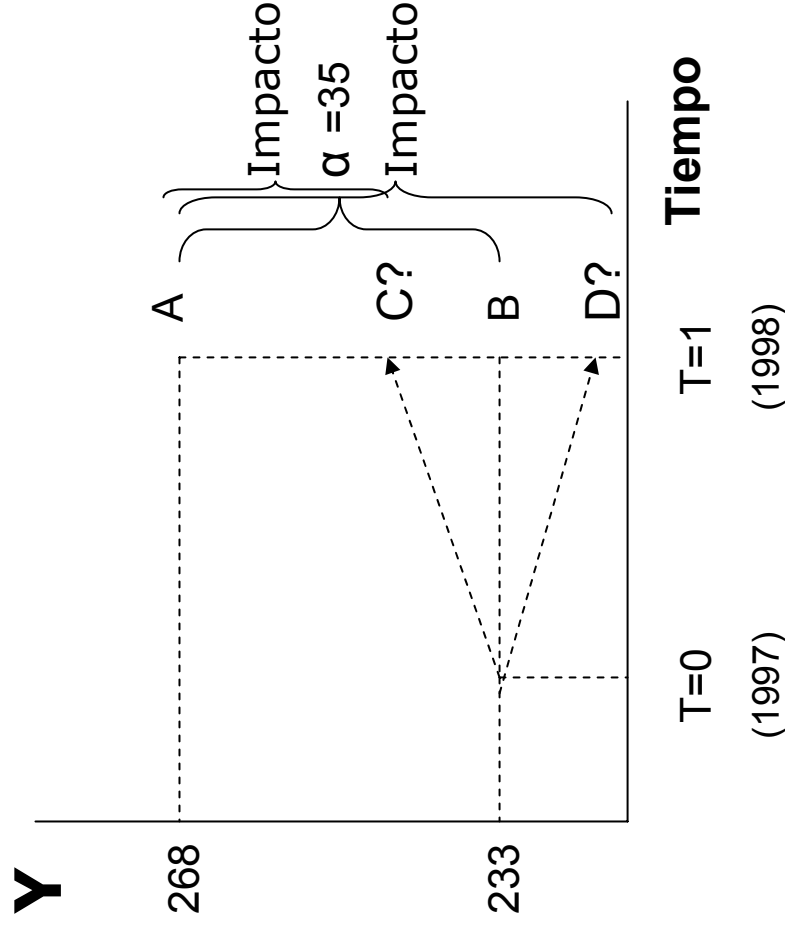
	Regresión Lineal	Regresión Lineal Multivariable
Impacto Estimado	35.27** (2.16)	34.28** (2.11)

** Significativo al 1%

Caso 1: Antes y Después

¿Cuál es el problema?

- 2 Puntos en el Tiempo
- Sólo mide beneficiarios en:
 - Consumo en $T=0$
 - Consumo en $T=1$
- Cálculo del contrafactual
 $(Y_{i,t} | P=0) = (Y_{i,t-1} | P=0)$
- "Impacto" = $A-B = 35$
- **No controla por factores que varían en el tiempo**
 - **Boom:** Impacto = $A-C$
 - $A-B =$ sobreestimación
 - **Recesión:** Impacto = $A-D$
 - $A-B =$ subestimación



1) Inferencia causal

- ❑ Contrafactuales

❑ **Contrafactuales falsos:**

- ❑ Antes y después (previo-posterior)
- ❑ **Inscrito-no inscrito (manzanas y naranjas)**

2) Métodos de evaluación de impacto:

- ❑ Controles aleatorios
- ❑ Promoción aleatoria (IV)
- ❑ Diseño de discontinuidad (RDD)
- ❑ Diferencia en diferencias
- ❑ Pareamiento/matching

Contrafactual falso n.º 2

Inscrito-no inscrito

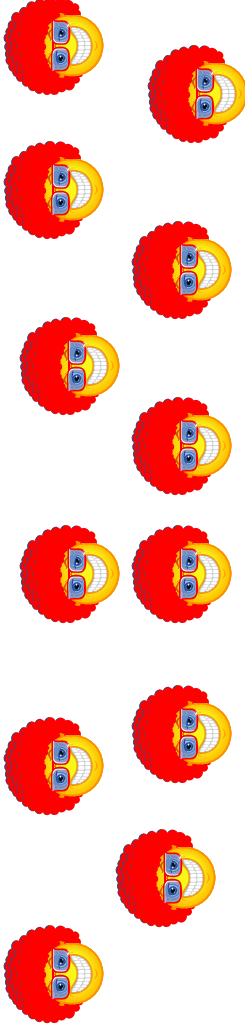


- Datos posteriores al tratamiento en 2 grupos
 - Inscrito: grupo de tratamiento
 - No inscrito: grupo de “control” (contrafactual)
 - Los **no elegibles** para participar
 - Los que **optan por NO** participar



- Sesgo de selección
 - El motivo de la no inscripción puede estar correlacionado con el resultado (Y)
 - Posible controlar las características observables
 - ¡Pero no controlar las no observables!
 - El impacto estimado se confunde con otros factores

Caso 2: Inscrito – no inscrito

Resultados en período posterior al tratamiento (1998)

<p>No elegible (No pobre)</p>	
<p>Elegible (Pobre)</p>	<div style="display: flex; justify-content: space-between;"> <div style="text-align: center;">  <p>No inscrito Y = 290</p> </div> <div style="text-align: center;">  <p>Inscrito Y = 268</p> </div> </div>

¿En qué pueden diferenciarse el grupo inscrito y el no inscrito, fuera de su participación en el programa?

Caso 2: Inscrito – no inscrito

Caso 2 - Inscrito/No Inscrito

	No Inscrito	Inscrito	t-stat
Consumo Promedio	290.16	268.75	5.6

Caso 2 - Inscrito/No Inscrito

	Regresión Lineal	Regresión Lineal Multivariable
Impacto Estimado	-22.7** (3.78)	-4.15 (4.05)

** Significativo al 1%

Estudio de casos

- Considere los resultados.....

Impacto Estimado	Caso 1 - Antes y Después		Caso 2 - Inscrito/No Inscrito	
	Regresión Lineal	Regresión Lineal Multivariable	Regresión Lineal	Regresión Lineal Multivariable
	35.27** (2.16)	34.28** (2.11)	-22.7** (3.78)	-4.15 (4.05)

** Significativo al 1%

- ¿Cuál resultado es el mas cierto?
- Problema con Antes-Después:
 - *No se consideran otros factores que varían en el tiempo*
- Problema con Inscrito-no inscrito:
 - *No sabemos si otros factores, fuera de la intervención, inciden en el resultado*

Medición de impacto

1) Inferencia causal

- Contrafactuales
- Contrafactuales falsos:
 - Antes y después (previo-posterior)
 - Inscrito-no inscrito (manzanas y naranjas)

2) Métodos de evaluación de impacto:

- Controles aleatorios
- Promoción aleatoria (IV)
- Diseño de discontinuidad (RDD)
- Diferencia en diferencias
- Pareamiento / matching



Selección del método...

- Para identificar un método de evaluación de impacto apropiado para su programa, considere:
 - Prospectivo/retrospectivo
 - Reglas de elegibilidad
 - Plan de implementación
- ¿El universo de elegibles es mayor que los recursos disponibles en un momento dado?
 - Presupuesto y capacidad de implementación
 - Exceso de demanda para el programa
 - Criterios de elegibilidad
 - Focalización geográfica, etc.

Seleccione la estrategia más sólida de acuerdo con el contexto operacional

Selección del método

- Identificar el “mejor” diseño posible de acuerdo con el contexto operacional
 - Mejor diseño = menor riesgo de contaminación
 - ¿Se controla por “todo”?
 - Validez interna
 - ¿El resultado es válido para “todos”?
 - Validez externa
 - Efecto local y efecto global del tratamiento



Medición de impacto

1) Inferencia causal

- ❑ Contrafactuales
- ❑ Contrafactuales falsos:
 - ❑ **Antes y después (previo-posterior)**
 - ❑ **Inscrito-no inscrito (manzanas y naranjas)**

2) Métodos de evaluación de impacto:

❑ **Controles aleatorios**

- ❑ Promoción aleatoria (IV)
- ❑ Diseño de discontinuidad (RDD)
- ❑ Diferencia en diferencias
- ❑ Pareamiento/matching

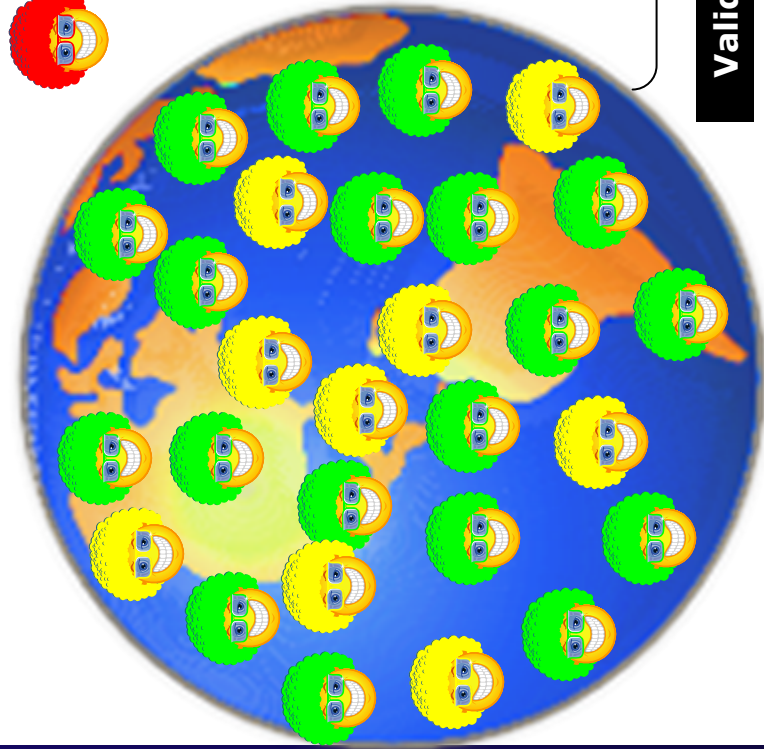


Controles aleatorios

- Cuando universo de elegibles > que n.º beneficios:
 - Realizar una selección aleatoria
 - Asignar por sorteo a quiénes se le ofrecen los beneficios
 - Modo equitativo, transparente y ético de asignar beneficios entre quienes tienen el mismo derecho a recibirlos
- Exceso de Demanda:
 - Dar a cada unidad elegible las mismas oportunidades de recibir tratamiento
 - Comparar el grupo al que se le ofreció tratamiento con el grupo al que no se le ofreció (controles)
- Aleatorización por etapas:
 - Dar a cada unidad elegible las mismas oportunidades de recibir tratamiento en primer lugar, segundo, tercero...
 - Comparar el grupo al que se le ofreció tratamiento primero con el grupo al que se le ofreció más adelante (controles)

Aleatorización

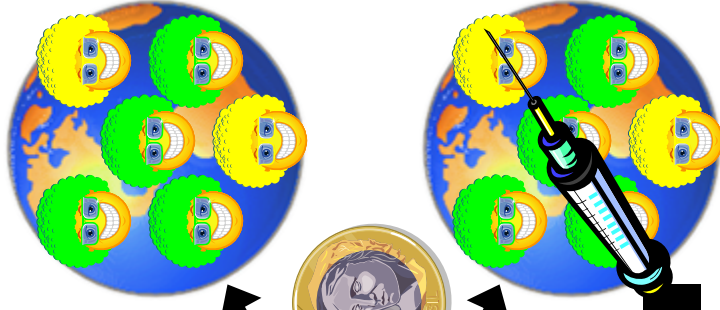
1. Universo



2. Muestra aleatoria de elegibles



3. Tratamiento aleatorio



Validez externa

Validez interna

No elegible =



Elegible =



Unidad de aleatorización

- Seleccionarla según el tipo de programa:
 - Individual/Hogar
 - Escuela/Centro salud/zona captación
 - Manzana/Pueblo/Comunidad
 - Distrito municipal/Departamento/Región
- Tener en cuenta los siguientes puntos:
 - Es necesario un número “suficientemente grande” de unidades para detectar el impacto mínimo deseado
 - *Spillovers*/contaminación
 - Costos operativos y de encuestas



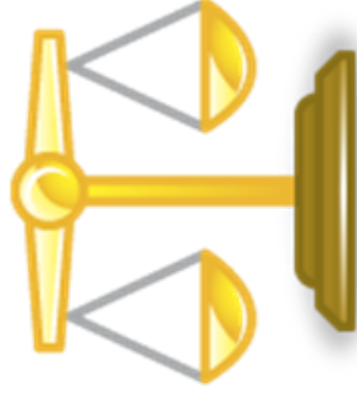
Como regla general, opte por aleatorizar en la mínima unidad de implementación viable.

Caso 3: Aleatorización

- Muestra de la evaluación de *Oportunidades*
- Unidad de aleatorización:
 - comunidad
- Aleatorización por etapas :
 - 320 comunidades de tratamiento (14.446 hogares)
 - Primeras transferencias distribuidas en abril 1998
 - 186 comunidades de control (9.630 hogares)
 - Primeras transferencias en noviembre 1999

Caso 3: Equilibrio inicial

Variables	RANDOMIZATION		
	Treatment (4,670)	Control (2727)	t-stats
Consumption per capita	233.47	233.4	-0.04
Head's age	1.02	1.3	1.2
Head's education	2.95	2.81	-2.16
Spouse's age	37.02	36.96	-0.38
Spouse's education	2.76	2.76	-0.006
Speaks an indigenous language	41.69	41.95	0.21
Head is female	0.007	0.009	0.66
Household at baseline	5.76	5.7	-1.21
Bathroom at baseline	0.57	0.56	-1.04
Total hectares of land	1.63	1.72	1.35
Min. Distance loc-urban	109.28	106.59	-1.02
	0.6	0.81	



Caso 3: Aleatorización

Caso 3 - Aleatorización

	Control	Treatamiento	t-stat
Promedio Basal	233.40	233.47	0.04
Promedio Seguimiento	239.5	268.75	9.6

Caso 3 - Aleatorización

Regresión Lineal Regresión Lineal Multivariable

Impacto Estimado	29.25**	29.79**
	(3.03)	(3.00)

** Significativo al 1%

Estudio de casos

	Caso 1 - Antes y Después Regresión Lineal Multivariable	Caso 2 - Inscrito/No Inscrito Regresión Lineal Multivariable	Caso 3 - Aleatorización Regresión Lineal Multivariable
Impacto Estimado	34.28** (2.11)	-4.15 (4.05)	29.79** (3.00)

** Significativo al 1%

Medición de impacto

1) Inferencia causal

- Contrafactuales
- Contrafactuales falsos:
 - Antes y después (previo-posterior)**
 - Inscrito-no inscrito (manzanas y naranjas)**

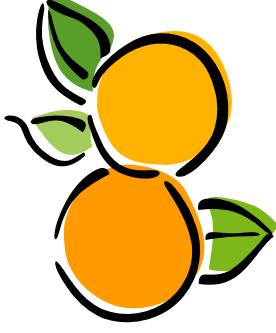
2) Métodos de evaluación de impacto:

- Controles aleatorios
- Promoción aleatoria (IV)**
- Diseño de discontinuidad (RDD)
- Diferencia en diferencias
- Pareamiento/matching



Promoción aleatoria (VI)

- Escenarios comunes:
 - Programa nacional con elegibilidad universal
 - Inscripción voluntaria al programa
- ¿Se puede comparar inscritos y no inscritos?
 - ¡Sesgo de selección!



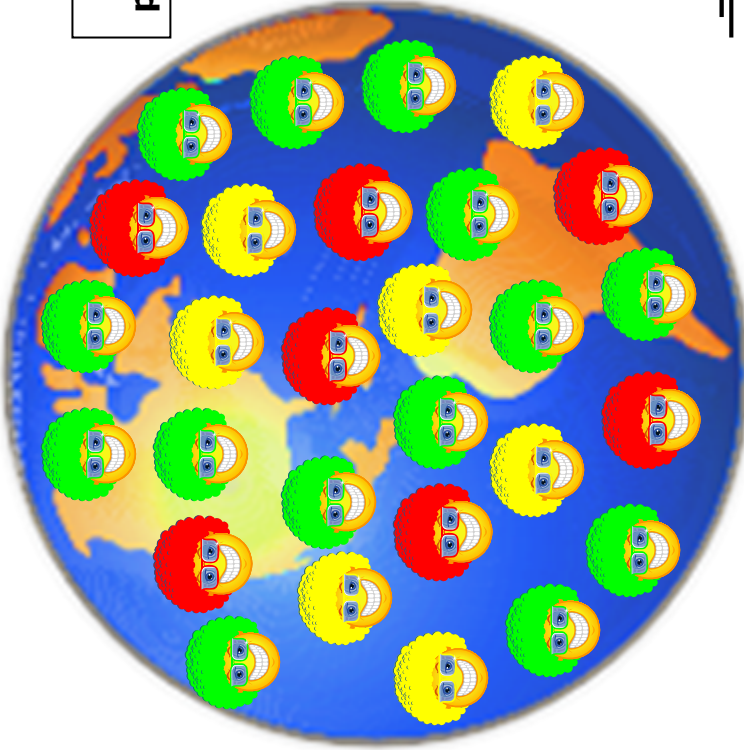
Promoción aleatoria (IV)

- Posible solución: proporcionar promoción adicional, motivaciones o incentivos a una submuestra aleatoria:
 - Información
 - Motivación (pequeño premio o beneficio)
 - Transporte
 - Otra ayuda/otros incentivos

- Condiciones necesarias:
 1. Grupos con promoción y sin promoción son comparables:
 - La promoción no está correlacionada con las características de la población
 - Garantizado por la aleatorización
 2. Inscripción más alta al programa en el grupo con promoción
 3. La promoción no afecta los resultados de modo directo

Promoción aleatoria

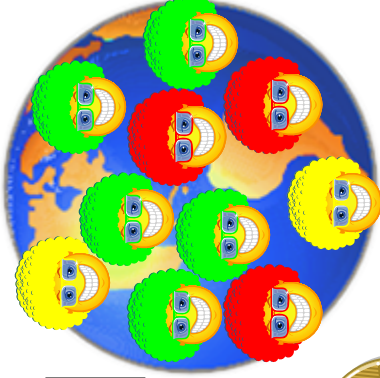
Elegibilidad
universal



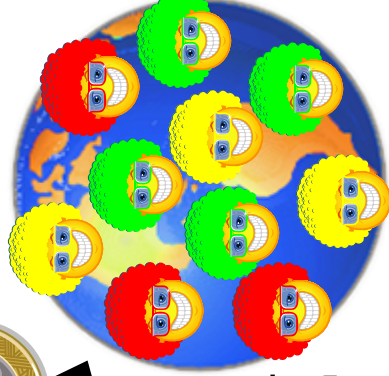
Elegible = **Nunca** **Siempre**

Sin
promoción

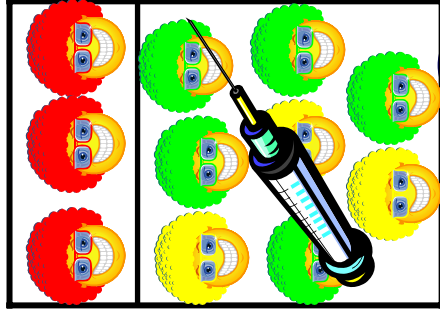
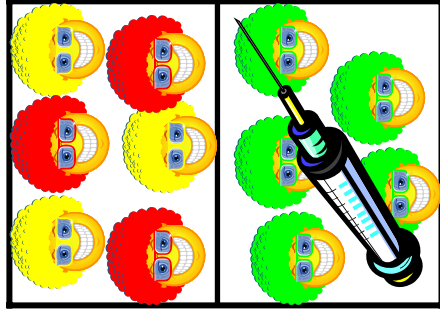
Promoción
aleatoria



Promoción



Inscripción





Promoción aleatoria

	Con promoción	SIN promoción	IMPACTO
	Inscritos= 80% Y = 100	Inscritos= 30% Y = 80	Δ Inscritos= 0.5 Δ Y=20 Impacto = 40
Nunca se inscribe			
Se inscribe si se lo incentiva			
Se inscribe siempre			

Ejemplos



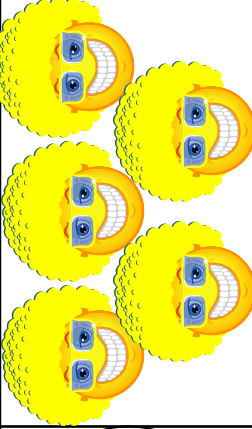
- Seguro de salud materno infantil en Argentina
 - Campañas intensivas de información
- Programa de empleos en Argentina
 - Tarjetas de transporte
- Administración escolar comunitaria en Nepal
 - Apoyo de ONG
- Fondos para prevención de riesgos de salud en India
 - Apoyo de equipos de recursos comunitarios

Promoción aleatoria

- Prueba piloto rigurosa de la estrategia de promoción
- Se obtiene información adicional de interés:
 - Modos de aumentar la inscripción
- No se tiene que “excluir” a nadie, pero...
 - La estrategia depende del éxito y la validez de la promoción
 - Se obtiene un promedio local del efecto del tratamiento
- La promoción aleatoria es una variable instrumental (VI)
 - Una variable correlacionada únicamente con el tratamiento (por ejemplo, Promoción aleatoria)
 - Puede encontrarse más información en el apéndice

Caso 4: VI



	Tratamiento aleatorio (Con promoción)	Control aleatorio	IMPACTO
Nunca se inscribe	Inscritos= 92% Y = 268	Inscritos= 0% Y = 239	Δ Inscritos= 0.92 Δ Y=29 Impacto TT= 31
Se inscribe si se lo incentiva			



Caso 4: VI - TOT

- Calcule el efecto TOT de *Oportunidades* sobre consumo
- Ejecute la regresión 2SLS

Caso 4 - IV

Regresión Lineal

Regresión Lineal Multivariable

Impacto Estimado

29.88**

(3.09)

30.44**

(3.07)

** Significativo al 1%

Medición de impacto

1) Inferencia causal

- Contrafactuales
- Contrafactuales falsos:
 - Antes y después (previo-posterior)**
 - Inscrito-no inscrito (manzanas y naranjas)**

2) Métodos de evaluación de impacto:

- Controles aleatorios
- Promoción aleatoria (IV)
- Diseño de discontinuidad (RDD)**
- Diferencia en diferencias
- Pareamiento/matching



Discontinuidades en la elegibilidad

Los programas sociales muchas veces focalizan el tratamiento mediante un índice de elegibilidad:

- Programas de lucha contra la pobreza:
 - → destinados a hogares por debajo de un determinado índice de pobreza
- Programas de pensiones:
 - → destinados a una población de más de cierta edad
- Becas escolares:
 - → destinadas a estudiantes con puntajes altos en pruebas estandarizadas

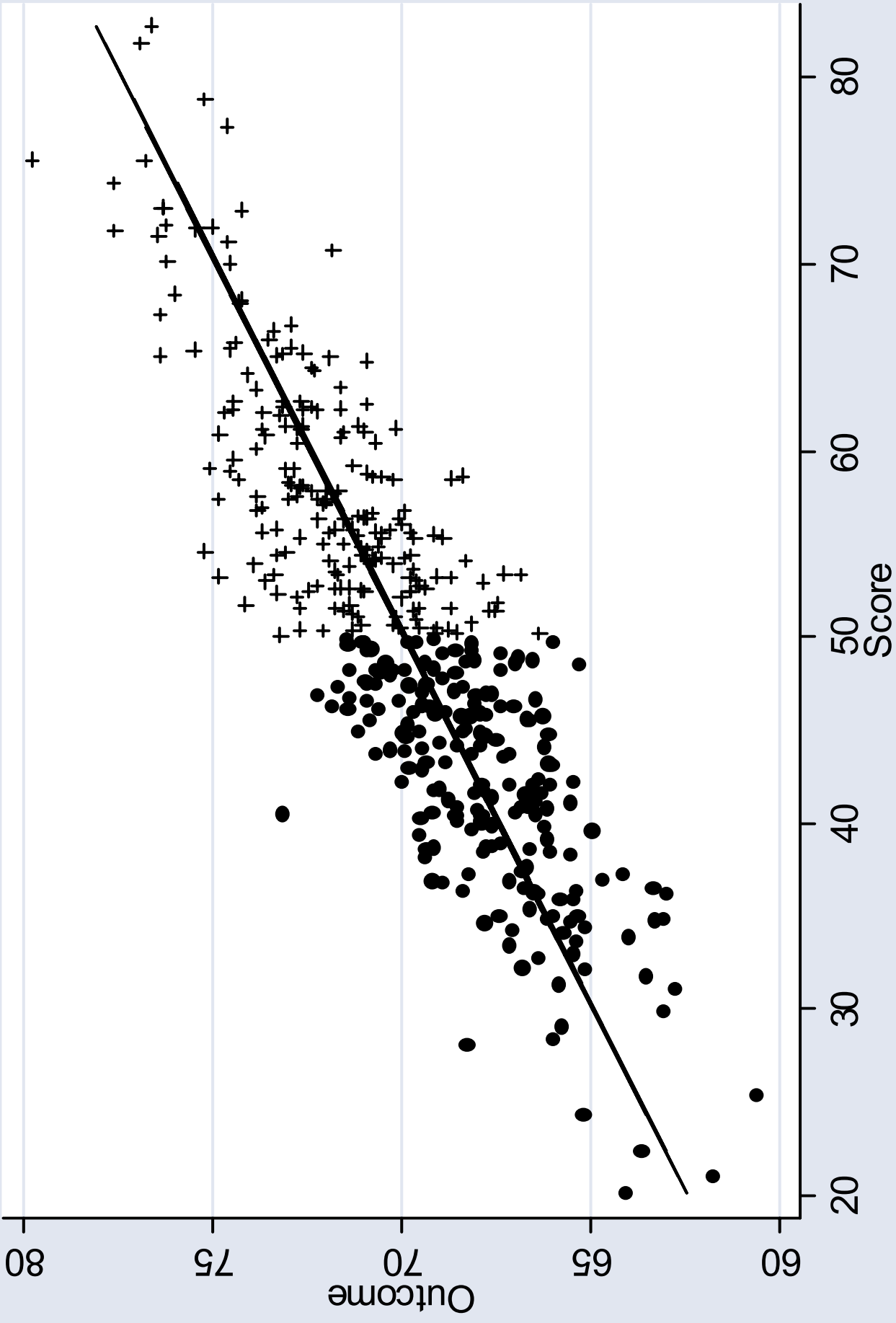
Para un diseño de discontinuidad se necesita:

- Índice continuo de elegibilidad
- Punto de corte claramente definido

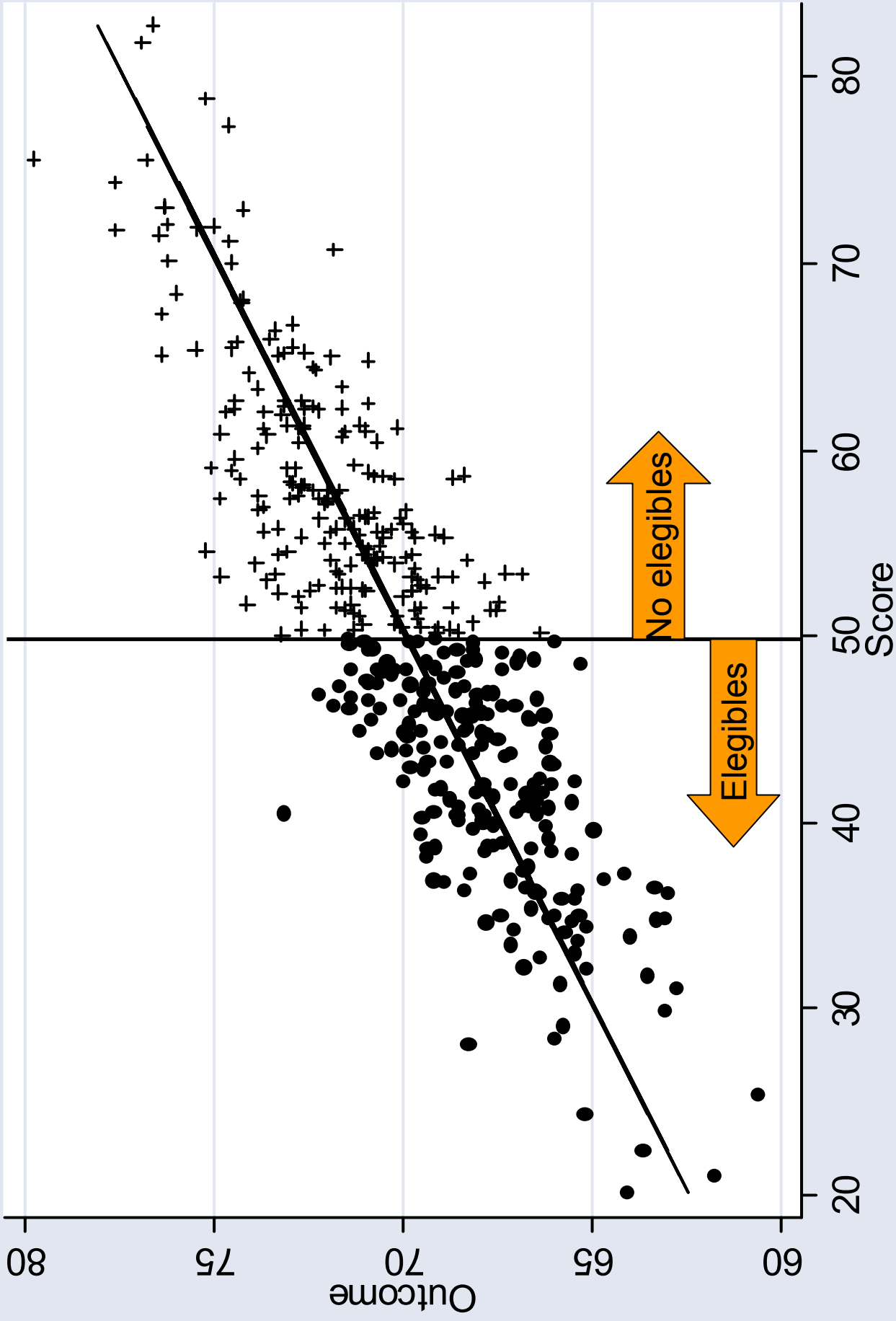
Ejemplo:

- Índice de elegibilidad (puntaje) de 1 a 100
 - Sobre la base de características anteriores a la intervención
- Puntaje ≤ 50 son elegibles
- Puntaje > 50 no son elegibles
- Se ofrece tratamiento a los elegibles

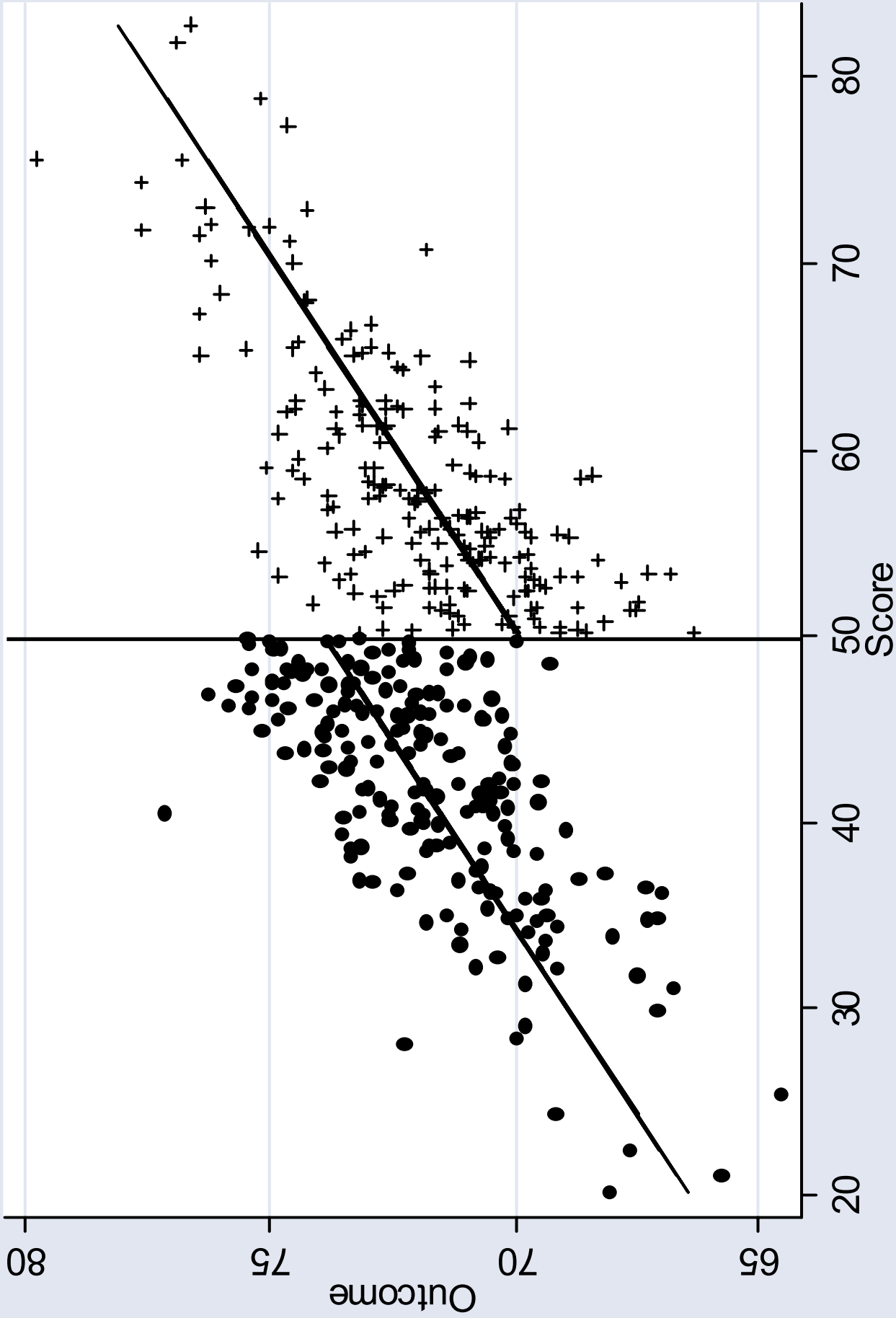
Regression Discontinuity Design - Baseline



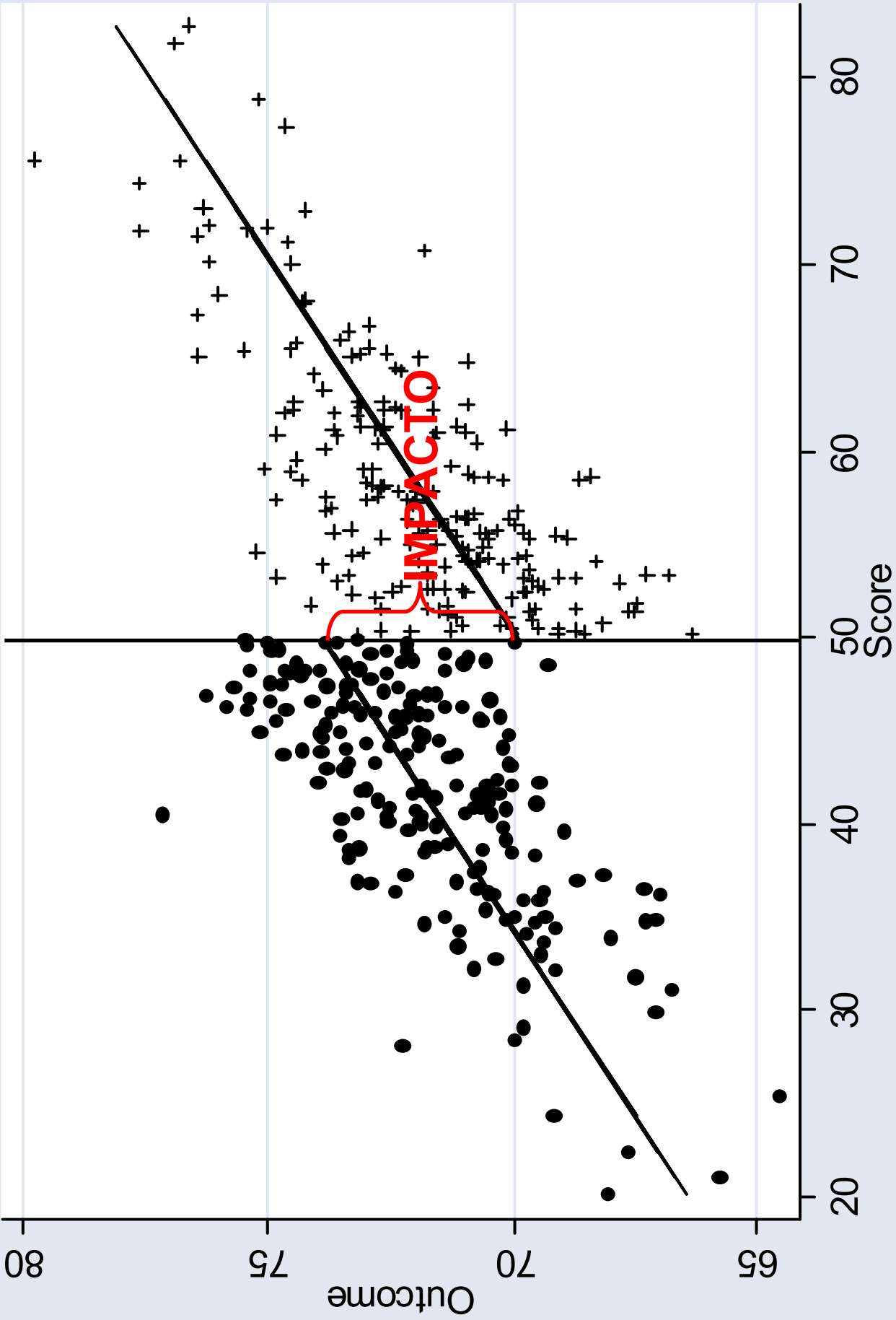
Regression Discontinuity Design - Baseline



Regression Discontinuity Design - Post Intervention



Regression Discontinuity Design - Post Intervention

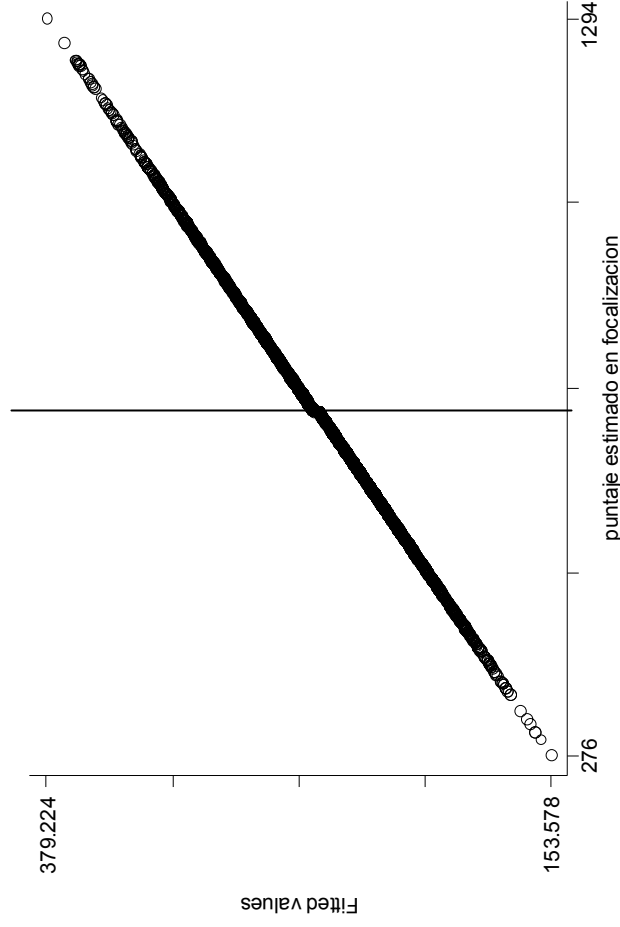


Caso 5: Diseño de discontinuidad

- *Oportunidades* asignó beneficios sobre la base de un índice de pobreza
- Donde
- *Tratamiento* = 1 si *puntaje* ≤ 750
- *Tratamiento* = 0 si *puntaje* > 750

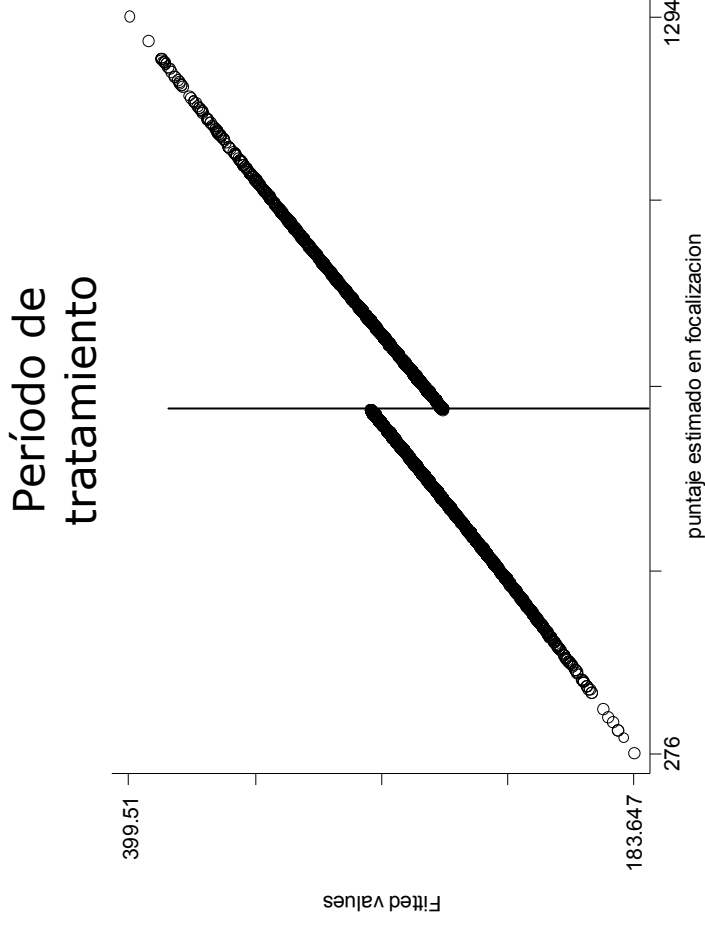
Caso 5: Diseño de discontinuidad

Datos iniciales –
Sin tratamiento



$$y_i = \beta_0 + \beta_1 Treatment_i + \delta(score) + \varepsilon_i$$

Caso 5: Diseño de discontinuidad



Caso 5 - Regression Discontinuity
 Regresión Lineal Multivariable

Impacto Estimado

30.58**
(5.93)

** Significativo al 1%

Posibles desventajas de RDD

- Efectos locales promedio; no siempre generalizables
- Alcance: el efecto se calcula en la discontinuidad, por lo que suelen obtenerse menos observaciones que en un experimento aleatorio con una muestra del mismo tamaño
- La especificación puede ser sensible a la forma funcional: verifique que la relación entre la variable de asignación y la variable de resultado esté correctamente diseñada; incluidas:
 - Relaciones no lineales
 - Interacciones



Ventajas de RDD para la evaluación

- Con RDD se obtiene una estimación sin sesgo del efecto del tratamiento en la discontinuidad
- Muchas veces puede valerse de una regla conocida para la asignación de beneficios que son comunes en los diseños de políticas sociales
 - No es necesario “excluir” del tratamiento a un grupo de hogares/individuos elegibles

Medición de impacto

1) Inferencia causal

- Contrafactuales
- Contrafactuales falsos:
 - Antes y después (previo-posterior)**
 - Inscrito-no inscrito (manzanas y naranjas)**

2) Métodos de evaluación de impacto:

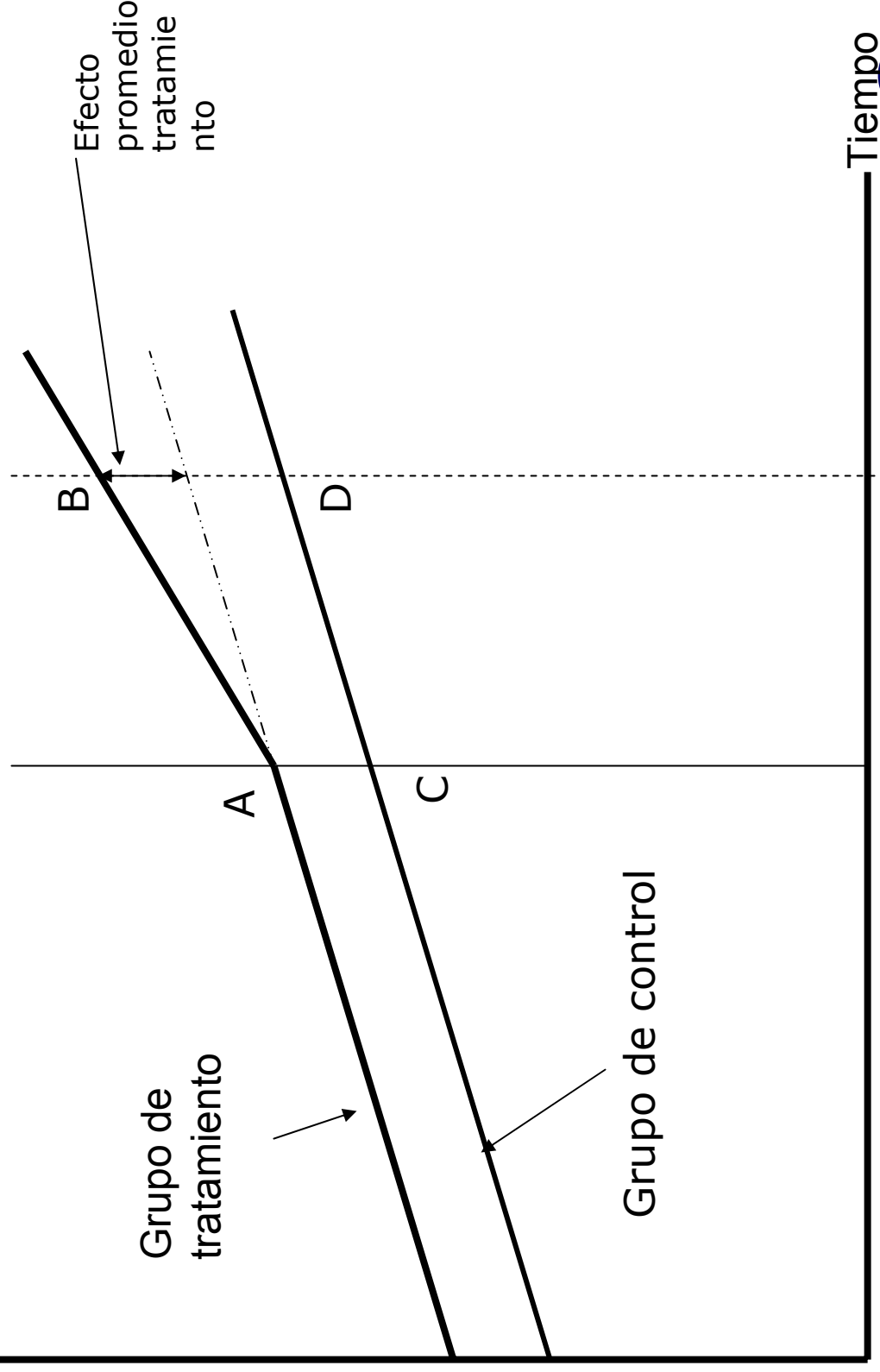
- Controles aleatorios
- Promoción aleatoria (IV)
- Diseño de discontinuidad (RDD)
- Diferencia en diferencias**
- Pareamiento/matching

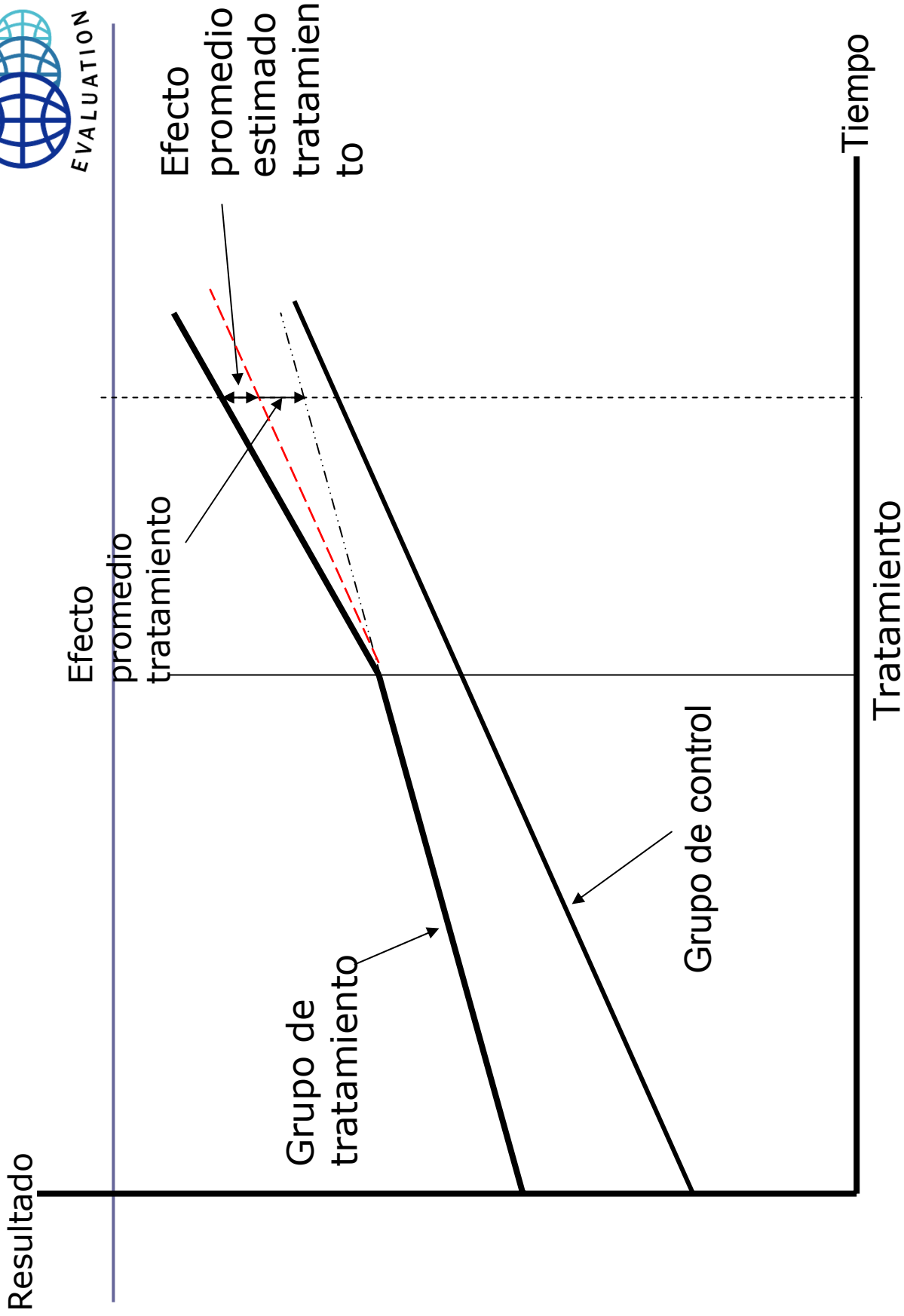


Diferencia en diferencias

- Se compara la modificación en los resultados entre grupos con tratamiento y sin tratamiento
 - El impacto es la diferencia en la modificación de los resultados
- Impacto = $(Y_{t1} - Y_{t0}) - (Y_{c1} - Y_{c0})$
- Controla por todos los factores que NO varían en el tiempo
 - Observables y no-observables

Resultado





Diferencia en diferencias

- Supuesto fundamental de que las tendencias (pendientes) son las mismas en tratamientos y controles antes de la intervención
- Se necesita un mínimo de dos puntos en el tiempo antes de la intervención para verificar el supuesto
 - Mayor confianza que no existen otros factores que si varían en el tiempo

Caso 6: Diferencia en diferencias

Caso 6 - Dif en Dif

	No Inscrito	Inscrito	t-stat
Mean Δ CPC	8.26	35.92	10.31

Caso 6 - Dif en Dif

	Regresión Lineal	Regresión Lineal Multivariable
Impacto Estimado	27.66** (2.68)	25.53** (2.77)

** Significativo al 1%

Estudio de casos

	Case 1 - Before and After	Case 2 - Enrolled/Not Enrolled	Case 3 - Randomization	Case 4 - IV (TOT)	Case 5 - Regression Discontinuity	Case 6 - Diff in Diff
	Multivariate Linear Regression	Multivariate Linear Regression	Multivariate Linear Regression	2SLS	Multivariate Linear Regression	Multivariate Linear Regression
Estimated Impact on CPC	34.28** (2.11)	-4.15 (4.05)	29.79** (3.00)	30.44** (3.07)	30.58** (5.93)	25.53** (2.77)

** Significant at 1% level

Medición de impactos

1) Inferencia causal

- Contrafactuales
- Contrafactuales falsos:
 - Antes y después (previo-posterior)**
 - Inscrito-no inscrito (manzanas y naranjas)**

2) Métodos de evaluación de impacto:

- Controles aleatorios
- Promoción aleatoria (IV)
- Diseño de discontinuidad (RDD)
- Diferencia en diferencias

Pareamiento / matching



Pareamiento/matching

- Se selecciona la comparación ideal que corresponda al grupo de tratamiento de un estudio más amplio.
- Las correspondencias son seleccionadas sobre la base de similitudes en características observadas
- Se asume el supuesto de que no existe sesgo de selección en características no observables.

Fuente: Martin Ravallion

Propensity Score Matching (PSM)

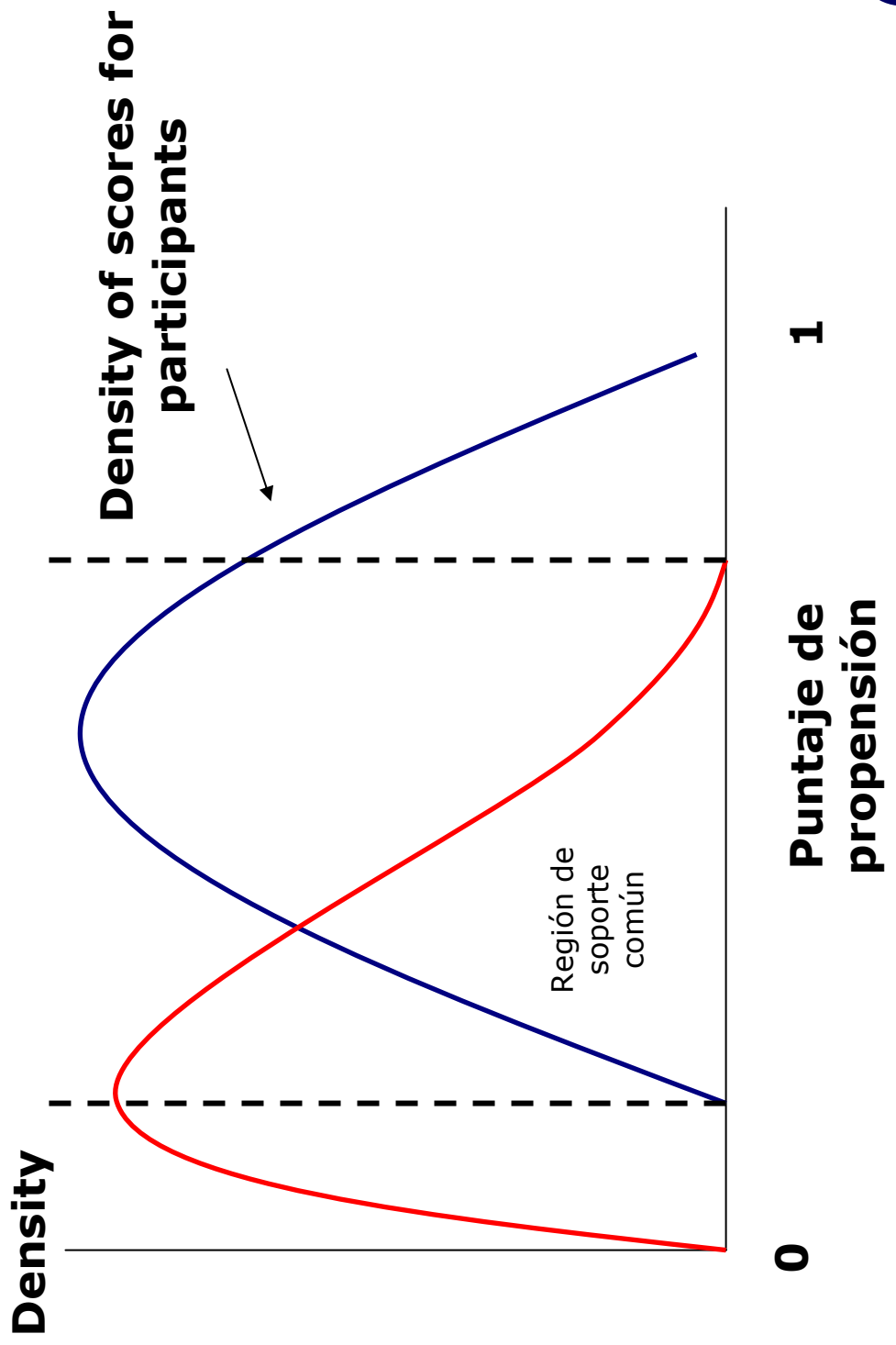
- Controles: no participantes con las mismas características que participantes
 - En la práctica, es muy difícil. El vector completo de características observadas de X podría ser enorme.
- Rosenbaum y Rubin: buscar correspondencia sobre la base del puntaje de propensión=

$$P(X_i) = Pr (D_i=1|X)$$

- En lugar de intentar asegurar que el puntaje de control de cada participante **tenga exactamente el mismo valor que X**, se puede alcanzar el mismo resultado buscando la correspondencia en la **probabilidad de participación**.
- Esto supone que la participación es independiente de los resultados dado X.

Pasos en PSM

1. Encuestas altamente representativas y comparables de no participantes y participantes.
2. Unir las dos muestras y estimar un modelo Logit (o Probit) de participación en el programa.
3. Restringir las muestras para asegurar un **soporte común** (fuente importante de sesgo en estudios de observación).
4. Buscar para cada participante una muestra de no participantes con puntajes de propensión similares.
5. Comparar los indicadores de resultados. La diferencia es el **cálculo de la ganancia** debida al programa para dicha observación.
6. Calcular la media de estas ganancias individuales para obtener el promedio de la ganancia general.



PSM no es un simple experimento

- Un experimento, en su forma pura, no requiere el supuesto de independencia condicional no comprobable de características observables
- PSM requiere muestras grandes de datos confiables

Lecciones sobre métodos de pareamiento

- Se suelen utilizar cuando no es posible aplicar aleatorización, RDD u otras opciones cuasi experimentales (por ejemplo, porque no se tienen datos iniciales)
 - Se debe tener cautela en el pareo *ex-post*
 - Comparación sobre variables endógenas
- El pareo ayuda a tener un control adecuado de la heterogeneidad OBSERVABLE
- Parear los datos iniciales puede ser muy útil:
 - Estimación:
 - Se combina con otras técnicas (por ejemplo, diferencia en diferencias)
 - Se conoce la regla de asignación (se compara sobre la base de esta regla)
 - Muestreo:
 - Selección de muestras de evaluación no aleatorias
- Se necesitan datos confiables
 - El soporte común puede representar un problema

Caso 7: Pareamiento

Case 7 - PROPENSITY SCORE: Pr(treatment=1)

Variable	Coef.	Std. Err.
Age Head	-0.0282433	0.0024553
Educ Head	-0.054722	0.0086369
Age Spouse	-0.0171695	0.0028683
Educ Spouse	-0.0643569	0.0093801
Ethnicity	0.4166998	0.0397539
Female Head	-0.2260407	0.0714199
_cons	1.6048	0.1013011

P-score Quintiles

xi	Quintile 1			Quintile 2			Quintile 3			Quintile 4			Quintile 5		
	T	C	t-score	T	C	t-score	T	C	t-score	T	C	t-score	T	C	t-score
Age Head	68.04	67.45	-1.2	53.61	53.38	-0.51	44.16	44.68	1.34	37.67	38.2	1.72	32.48	32.14	-1.18
Educ Head	1.54	1.97	3.13	2.39	2.69	1.67	3.25	3.26	-0.04	3.53	3.43	-0.98	2.98	3.12	1.96
Age Spouse	55.95	55.05	-1.43	46.5	46.41	0.66	39.54	40.01	1.86	34.2	34.8	1.84	29.6	29.19	-1.44
Educ Spouse	1.89	2.19	2.47	2.61	2.64	0.31	3.17	3.19	0.23	3.34	3.26	-0.78	2.37	2.72	1.99
Ethnicity	0.16	0.11	-2.81	0.24	0.27	-1.73	0.3	0.32	1.04	0.14	0.13	-0.11	0.7	0.66	-2.3
Female Head	0.19	0.21	0.92	0.42	0.16	-1.4	0.092	0.088	-0.35	0.35	0.32	-0.34	0.008	0.008	0.83

Caso 7: Pareamiento

Caso 7 - Pareamiento

Regresión Lineal

Regresión Lineal Multivariable

Impacto Estimado

1.16
(3.59)

7.06+
(3.65)

** Significativo al 1%, + Significativo al 10%

Estudio de casos: Síntesis de resultados

	Caso 1 - Antes y Después	Caso 2 - Inscrito/No Inscrito	Caso 3 - Aleatorización	Caso 4 - IV (TOT)	Caso 5 - RDD	Caso 6 - Dif en Dif	Caso 7 - Matching
	Regresión Lineal Multivariable	Regresión Lineal Multivariable	Regresión Lineal Multivariable	2SLS	Regresión Lineal Multivariable	Regresión Lineal Multivariable	Regresión Lineal Multivariable
Impacto Estimado	34.28** (2.11)	-4.15 (4.05)	29.79** (3.00)	30.44** (3.07)	30.58** (5.93)	25.53** (2.77)	7.06+ (3.65)

** Significativo al 1%, + Significativo al 10%

Síntesis de métodos

	Aleatorización	Promoción aleatoria VI	Diseño de discontinuidad	Diferencia en diferencias	Pareo
Validez interna					
Validez externa					
Riesgos					

Medición de impactos

- 1) Inferencia causal
 - Contrafactuales
 - Contrafactuales falsos:
 - Antes y después (previo-posterior)
 - Inscrito-no inscrito (manzanas y naranjas)

2) Métodos de evaluación de impacto:

- Controles aleatorios**
- Promoción aleatoria (IV)**
- Diseño de discontinuidad/RDD**
- Diferencia en diferencias**
- Pareamiento/matching (*propensity score matching, PSM* por su sigla en inglés)**
- Combinaciones de estos métodos



Recuerde...

- El objetivo de la evaluación de impacto es estimar el efecto *CAUSAL* de un programa en los resultados de interés
- Al diseñar el programa debemos entender el proceso de generación de datos
 - el proceso conductual que genera los datos
 - el modo en que se asignan los beneficios
- Se debe adecuar el mejor diseño de evaluación al contexto operacional

Ahora vamos a evaluar!

GRACIAS!

Apéndice 1: Mínimos Cuadrados en Dos Etapas (2SLS, por su sigla en inglés)

- Modelo con *Tratamiento* endógeno (T):

$$y = \alpha + \beta_1 T + \beta_2 x + \varepsilon$$

- Etapa 1: Realizar la regresión de variable endógena sobre VI (Z) y otros regresores exógenos

$$T = \delta_0 + \delta_1 x + \theta_1 Z + \tau$$

- Calcular el valor previsto para cada observación: T gorro

Apéndice 1: Mínimos Cuadrados en Dos Etapas (2SLS, por su sigla en inglés)

- Etapa 2: Realizar la regresión de resultado y sobre la variable prevista (y otras variables exógenas)

$$y = \alpha + \beta_1(T) + \beta_2x + \varepsilon$$

- Es necesario corregir Errores Estándar (se basan en T gorro más que en T)
- En la práctica se utiliza STATA - ivreg
- Intuición: T ha quedado “limpia” de su correlación con ε .