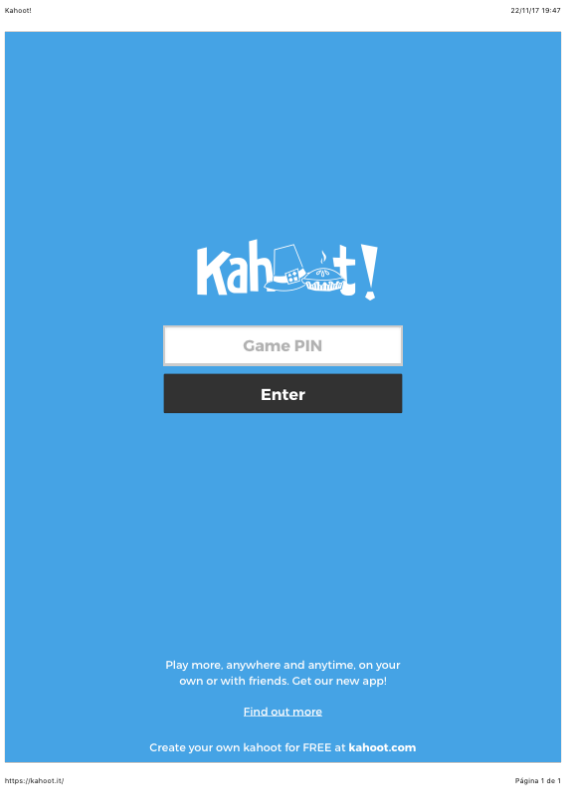
# La Evaluación de Impacto

Ana Balsa
Universidad de Montevideo

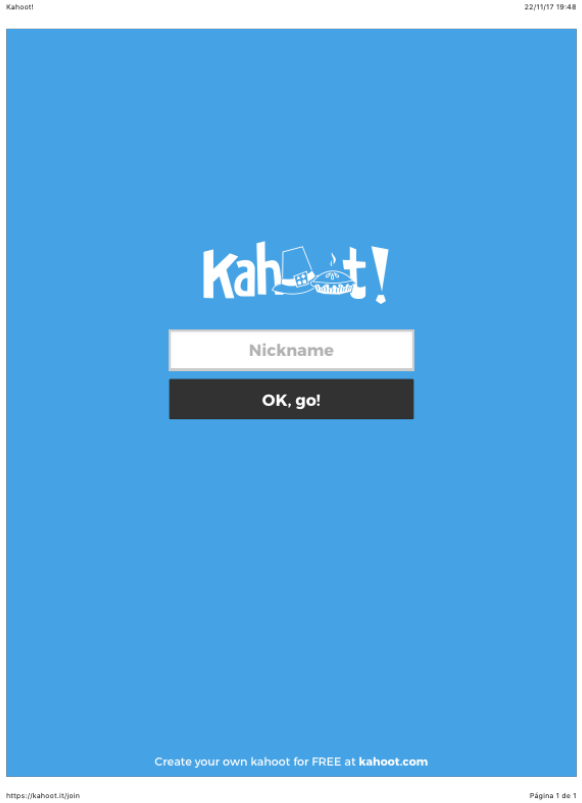
Argentina SEMIDE
Noviembre 2017

# Kahoot.it

Ingresar PIN




Ingresar apodo



# ¿Qué es una evaluación de impacto?

- Enfoque tradicional: basado en insumos, procesos y monitoreo financiero

- Política basada en la evidencia: **foco en resultados**
  - Asignación eficiente de los recursos (eficacia, costo-efectividad, costo-beneficio)
  - Rendición de cuentas
  - Innovación
  - Aprendizaje
  - Discusión informada

# Ejemplo

- Programas de transferencia condicionada
  - Bolsa Familia (Brasil)
  - Progresa Oportunidades (México)
  - Asignación Universal por Hijo (Argentina)
  - Asignación Universal por Embarazo (Argentina)
- Impactos en:
  - Pobreza
  - Capital humano
  - Salud

# Componentes de una evaluación

# Componentes de una evaluación

- Evaluación de necesidades
- Evaluación teórica
- Evaluación de procesos
- Evaluación de impacto
- Evaluación de eficiencia

# Componentes de una evaluación

Evaluación de necesidades

Evaluación teórica

Evaluación de procesos

Evaluación de impacto

Evaluación de eficiencia

***El problema a atender***. *Población, necesidades, causas y cómo se está atacando.*

# Componentes de una evaluación

Evaluación de necesidades

***El problema a atender**. Población, necesidades, causas y cómo se está atacando.*

Evaluación teórica

*La **teoría del cambio**. ¿Cómo resolver el problema? Cadena causal: objetivos, insumos, productos, resultados Inter-medios, impacto*

Evaluación de procesos

Evaluación de impacto

Evaluación de eficiencia

# Componentes de una evaluación

Evaluación de necesidades

***El problema a atender**. Población, necesidades, causas y cómo se está atacando.*

Evaluación teórica

*La **teoría del cambio**. ¿Cómo resolver el problema? Cadena causal: objetivos, insumos, productos, res. Intermedios, impacto*

Evaluación de procesos

*La **implemen-tación.** Fidelidad al plan original y ajustes necesarios.*

Evaluación de impacto

Evaluación de eficiencia

# Componentes de una evaluación

**Evaluación de necesidades**

***El problema a atender**. Población, necesidades, causas y cómo se está atacando.*

**Evaluación teórica**

*La **teoría del cambio**. ¿Cómo resolver el problema? Cadena causal: objetivos, insumos, productos, res. Intermedios, impacto*

**Evaluación de procesos**

*La **implemen-tación.** Fidelidad al plan original y ajustes necesarios.*

**Evaluación de impacto**

*Los **resultados** atribuibles a la intervención. Cumplimiento y magnitud de los objetivos*

**Evaluación de eficiencia**

# Componentes de una evaluación

**Evaluación de necesidades**

***El problema a atender***. *Población, necesidades, causas y cómo se está atacando.*

**Evaluación teórica**

*La **teoría del cambio**. ¿Cómo resolver el problema? Cadena causal: objetivos, insumos, productos, res. Intermedios, impacto*

**Evaluación de procesos**

*La **implemen-tación.** Fidelidad al plan original y ajustes necesarios.*

**Evaluación de impacto**

*Los **resultados** atribuibles a la intervención. Cumplimiento y magnitud de los objetivos*

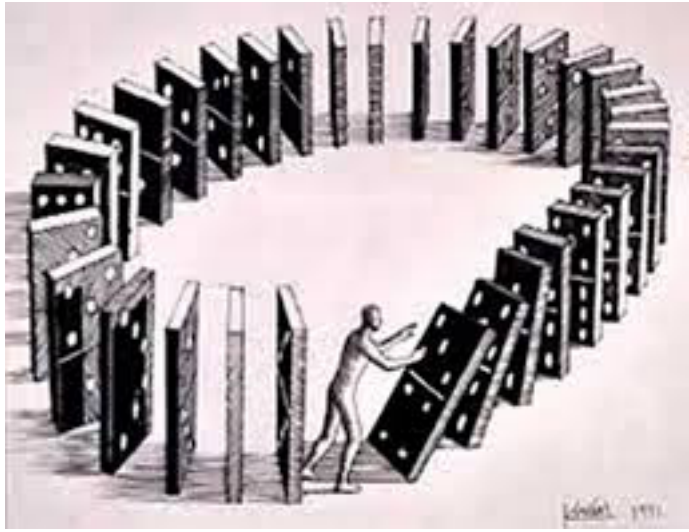
**Evaluación de eficiencia**

***Comparación con alternativas*** *Costo-efectividad Costo-beneficio.*

# Evaluación de impacto y Causalidad

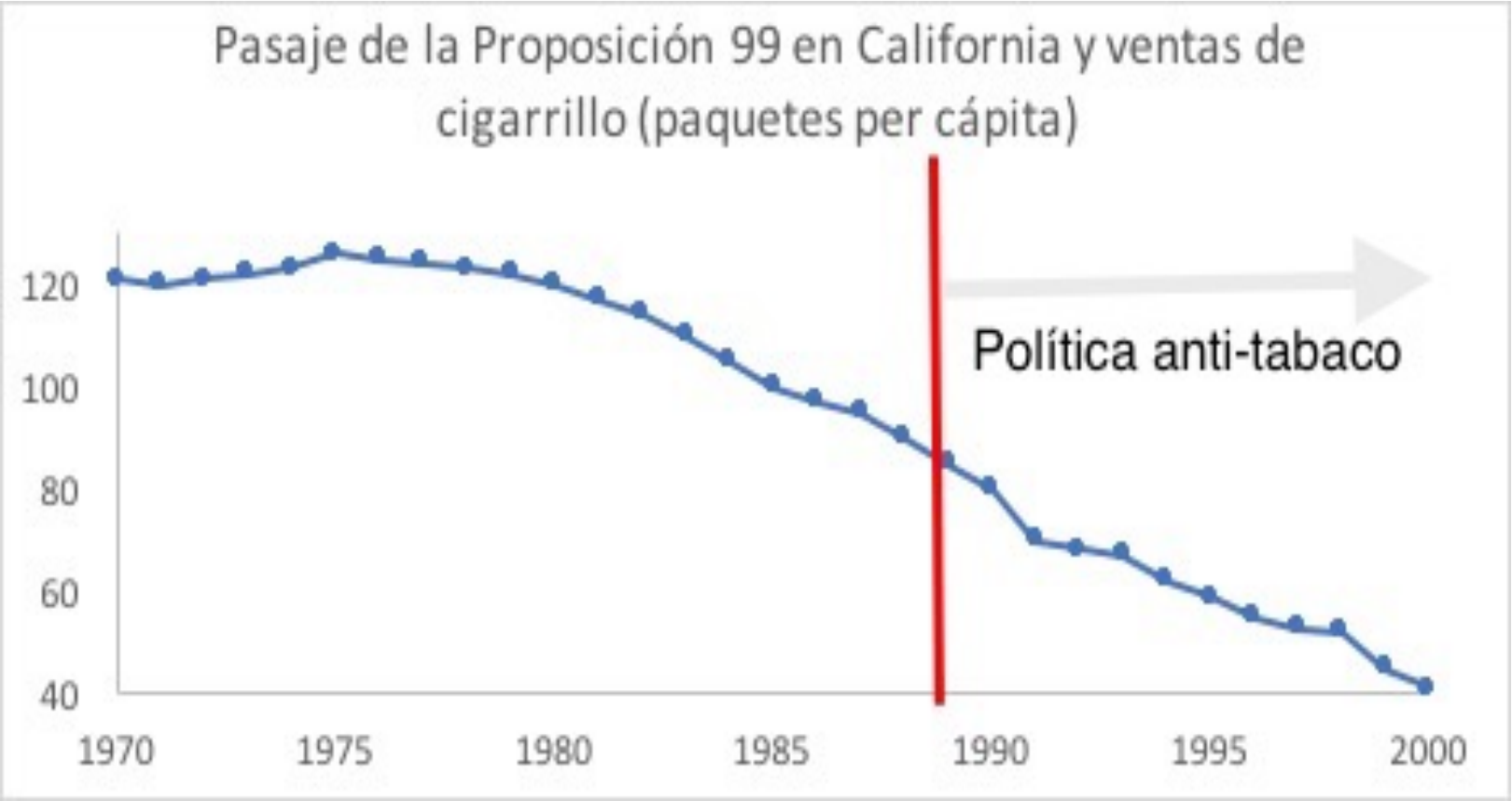
*“La afirmación paradigmática en el estudio de las relaciones causales es que la manipulación de una causa va a resultar en la manipulación de un efecto...La causalidad implica que, al variar un factor, puedo hacer que otro varíe.”*

Cook & Campbell, 1979



# Kahoot.it

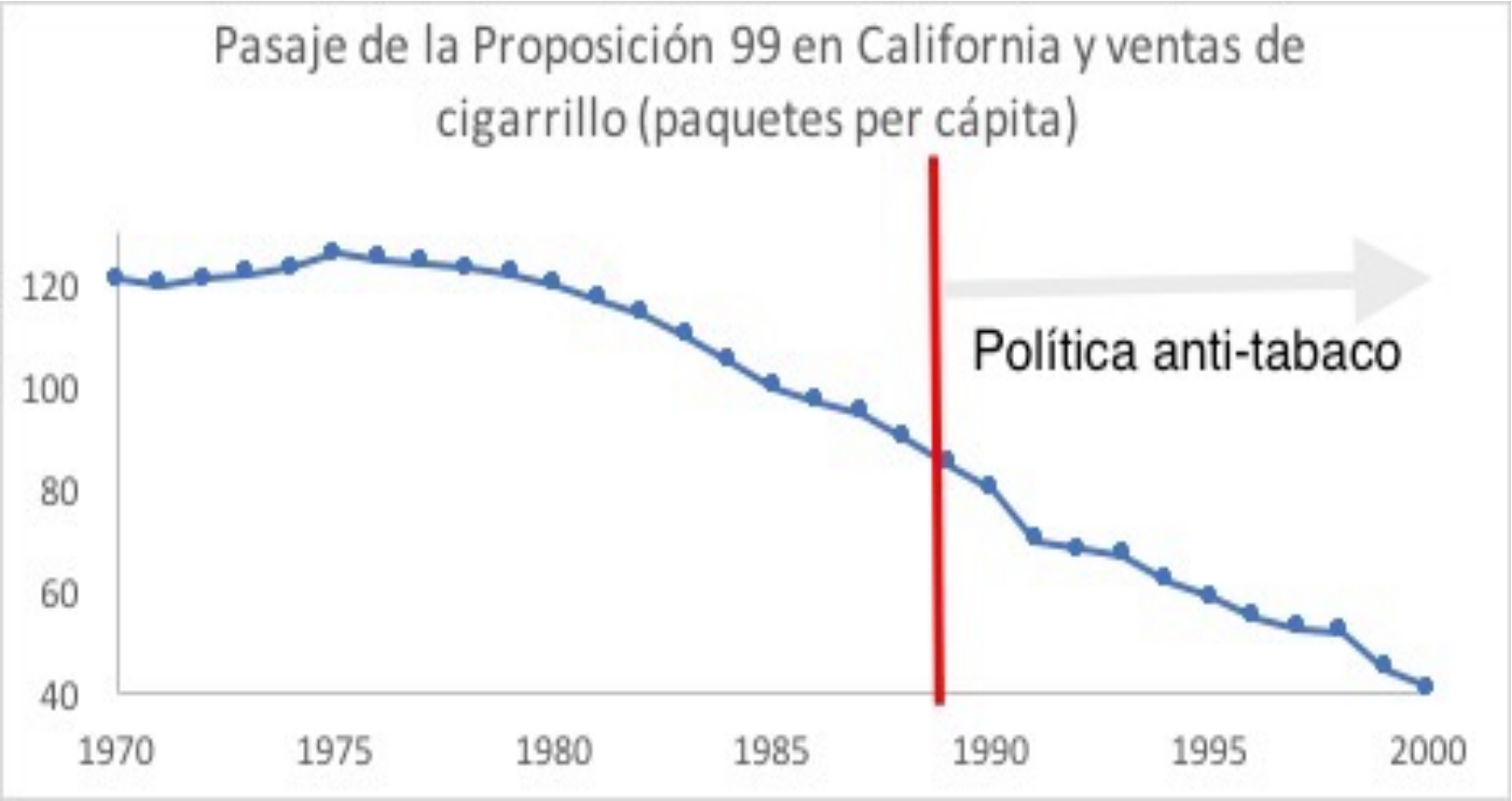
- Política anti-tabaco en California en 1988




¿**Causó** esta política la caída de las ventas?

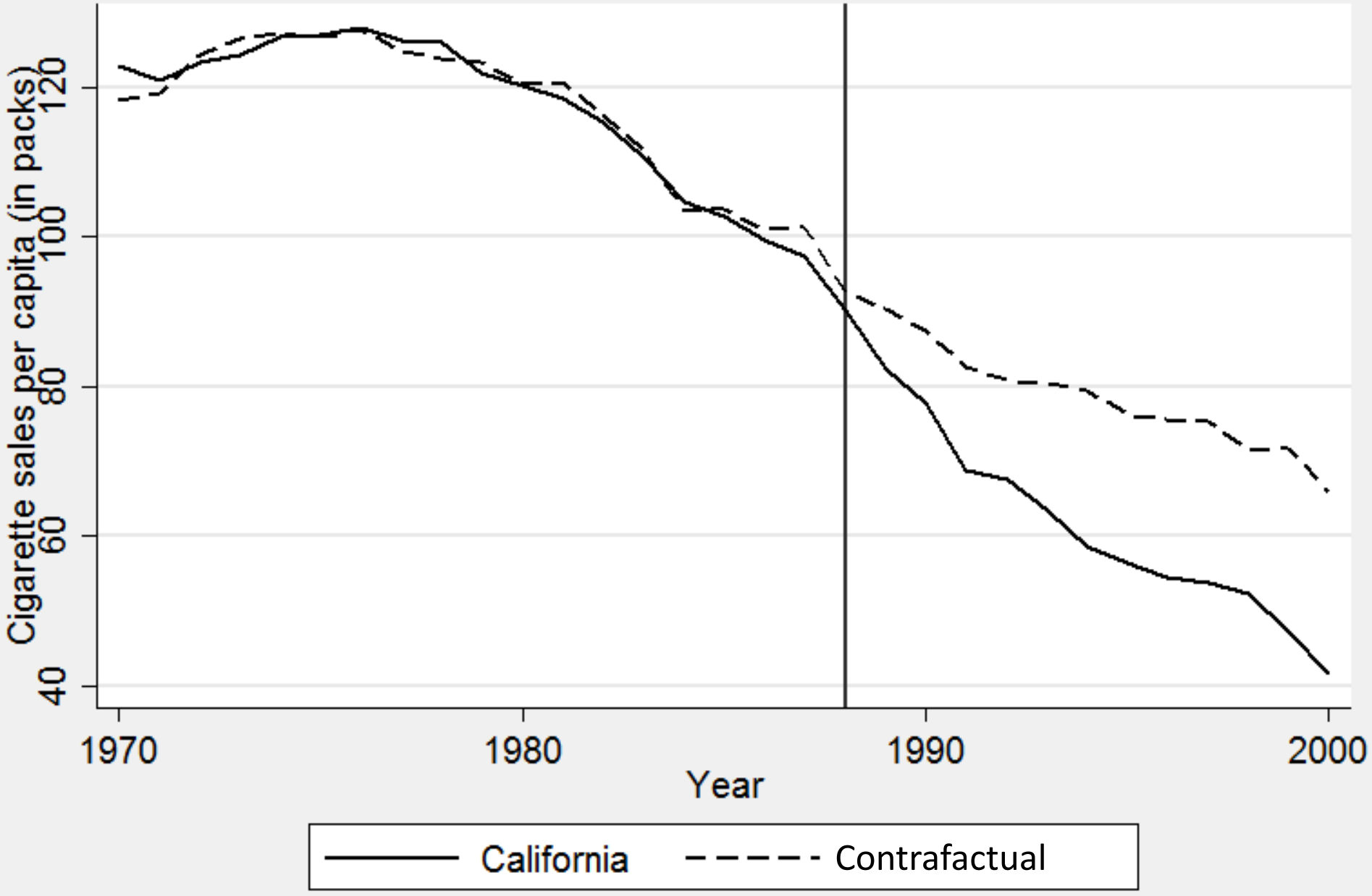
# Kahoot.it

- Política anti-tabaco en California en 1988

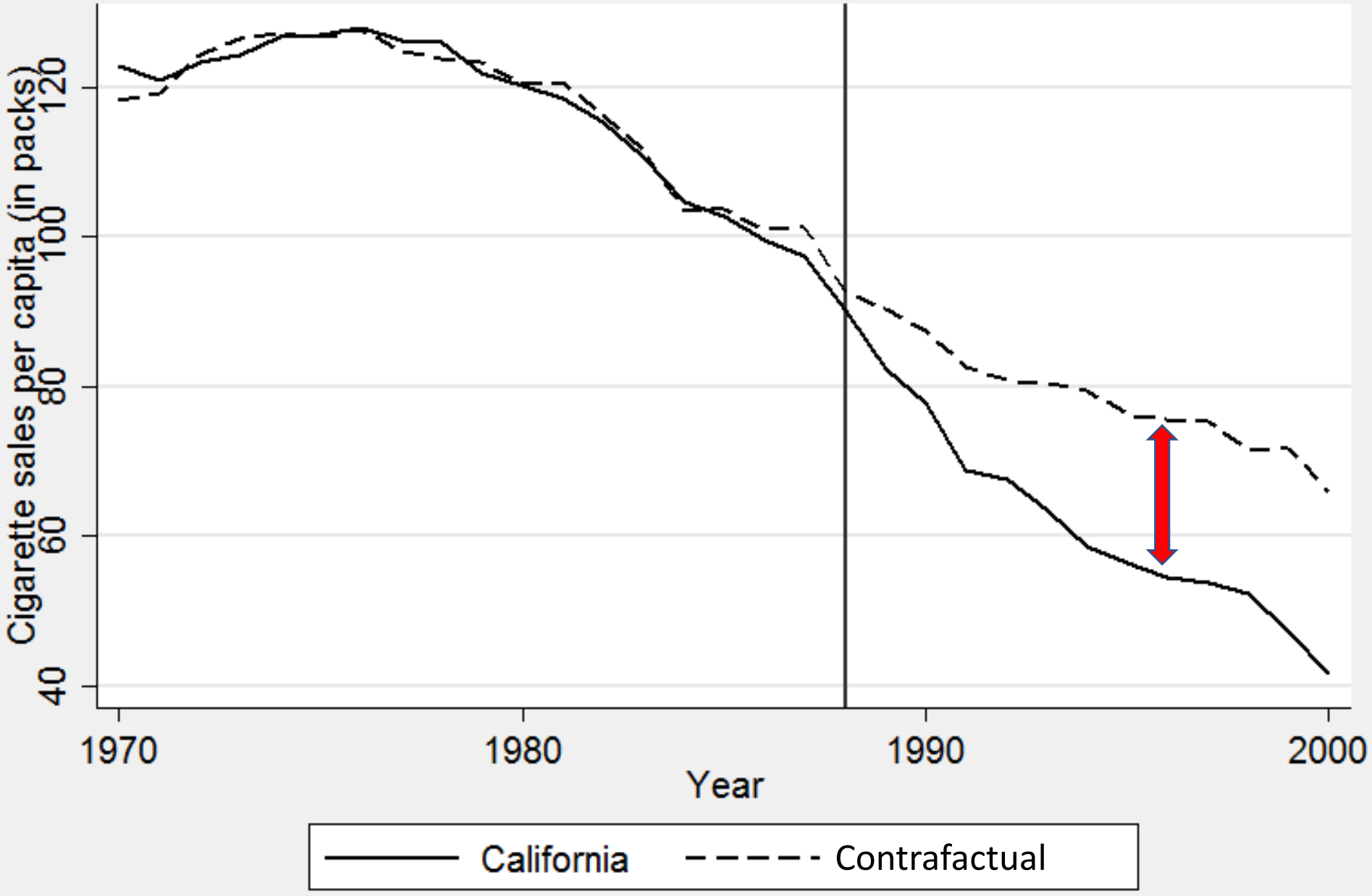



¿**Causó** esta política la caída de las ventas?

# Proposición 99 en California y venta de cigarrillos



# Proposición 99 en California y venta de cigarrillos



# Impacto y Contrafactual

- El **contrafactual** es el resultado que habrían tenido los participantes del programa en caso de que el programa no se hubiese implementado

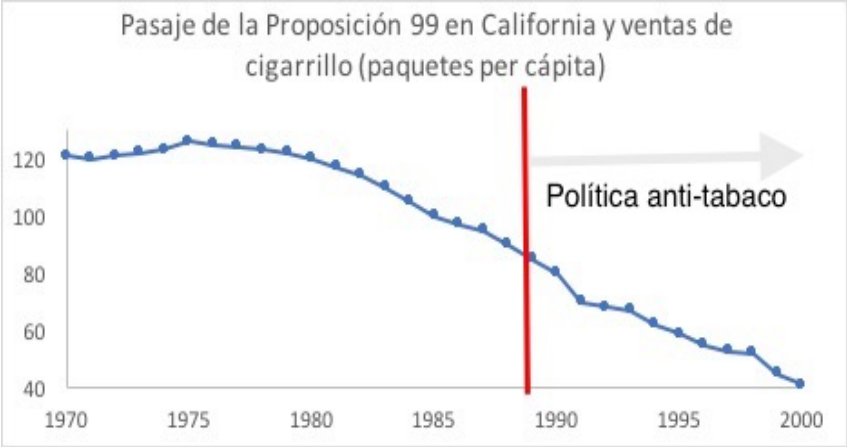
- Impacto = Cambio causal en los resultados o cambios en los resultados atribuibles al programa
  - Resultado de los participantes luego de pasar por el programa – Contrafactual

# Grupos de tratamiento y control

- El problema de las evaluaciones de impacto es que **no podemos observar al mismo tiempo el resultado para un sujeto con y sin programa**.
- *Grupo de tratamiento*: participantes de la intervención, programa o política
- *Grupo de control*: sujetos que son idénticos en sus características promedio al grupo de tratados pero que no participan de la intervención

# Diseños incorrectos de evaluación

1) Comparación antes y después del grupo de participantes



# Diseños incorrectos de evaluación

1) Comparación antes y después del grupo de participantes
2) Comparación entre los participantes y los no participantes
   - California vs. estados que no implementaron

# Kahoot.it

# El Sesgo de Selección

- El sesgo de selección ocurre cuando las razones por las cuales un individuo participa en el tratamiento están correlacionadas con los resultados.
    - Hospitalización y salud autoreportada (estado de salud general)
    - Participación en olimpíadas matemáticas y aprendizaje en matemática (habilidad)
    - Participación en programas de capacitación laboral (potencial de generar ingresos)
    - Tamaño de clase y aprendizajes (niños con más dificultad se asignan a clases más chicas)

- Asegurarse que el impacto estimado esté libre de sesgo de selección es uno de los objetivos y desafíos más importantes de cualquier evaluación de impacto

# Metodologías de evaluación de impacto

CLAVE

**Encontrar un grupo de control que sea un buen contrafactual**

# Metodologías de evaluación de impacto

## Experimentales

- La asignación al tratamiento se realiza en forma aleatoria

## Cuasi-experimentales

- No hay asignación aleatoria ex-ante al tratamiento; se construye un grupo de control a través de diseños no experimentales
  - *Técnicas de emparejamiento*
  - *Diferencias en diferencias*
  - *Variables instrumentales*
  - *Regresión discontinua*

# Diseños experimentales

# Kahoot.it

- ¿Por qué constituye la evaluación experimental un buen diseño de evaluación de impacto?

# Diseños de evaluación experimentales

- Asignación aleatoria (no sistemática) de sujetos a grupo de tratamiento y grupo de control
- Asegura que en promedio las características de los grupos sean iguales en la línea base (antes de implementar la intervención)

- Cualquier diferencia entre los grupos al finalizar la intervención puede atribuirse al programa

# Pasos de un experimento aleatorio

- Diseño del experimento (definir sub-población, necesidades, pertinencia, insumos y procesos)
- Aleatorización
- Recoger datos de línea de base y chequear balance
- Recoger datos de seguimiento
- Análisis de desgaste de muestra y de integridad de tratamiento y control
- Recoger datos del proceso
- Evaluar impacto

# Métodos de aleatorización

- <u>Según unidad de aleatorización</u>
  - A nivel individual
  - A nivel de grupo o cluster (escuelas, hospitales, barrios)

# Kahoot.it

- Aleatorización grupal vs. individual

# Métodos de aleatorización

- <u>Según restricciones del programa</u>
  - Sobre-suscripción
    - Lotería simple
    - Problema de desgaste diferencial
  - Ingreso al programa en fases
    - Ventajas
      - Más justo
      - Incentivo a participar en el futuro evita desvinculación (attrition) del control
    - Desventajas
      - Puede impedir estimación de efectos de mediano y largo plazo
      - Grupo de comparación puede verse afectado por expectativa de tratamiento futuro
  - Aleatorización en la burbuja
  - Aleatorización de los incentivos para participar

# Métodos de aleatorización

- <u>Según tipo de aleatorización</u>
  - Aleatorización simple
    - De probabilidad fija. Todos tienen la misma probabilidad de ser elegidos sin importar las asignaciones previas
    - Restrictiva
  - Estratificada
    - Construcción de celdas que combinan variables iniciales correlacionadas con resultados y se aleatoriza al interior de las celdas

# Consideraciones éticas de la evaluación experimental

- ¿Es ético dejar gente fuera del programa?
- Son muy caras
- Privacidad de datos
- Que nadie quede peor

# Consideraciones éticas

- Puede ser más caro y menos ético asignar presupuestos a programas cuya efectividad no está comprobada
- No debería negarse tratamiento por el solo hecho de evaluar. Se trata de aprovechar las limitantes administrativas y financieras del propio programa, para generar grupos de tratamiento y control satisfactorios
- Privacidad y “que no quede nadie peor” son consideraciones que aplican a cualquier programa nuevo

# Kahoot.it

- Validez interna y externa

# Validez interna y externa

- Una evaluación es **internamente válida** si genera un estimativo preciso del contrafactual a través de un grupo de comparación válido

- Una evaluación es **externamente válida** si la muestra evaluada representa en forma precisa a la población de los sujetos elegibles. En este caso los resultados de la evaluación pueden ser generalizados a la población de sujetos elegibles

# Cálculo de poder

- Es muy costoso evaluar a toda la población –> muestreo
- Error estándar (error de estimación): aumenta con la varianza del resultado y disminuye con el tamaño de la muestra
- Intervalo de confianza al 95%:  Media +/- 1,96*error estándar
  - Confianza: en 100 muestras, 95 de ellas reflejan el verdadero valor poblacional

# Kahoot.it

- Significancia estadística

# Kahoot.it

- Significancia estadística: Error tipo I = Probabilidad de encontrar efecto dado que NO lo hay.
    - 5% es lo habitual

- Poder estadístico: 1-error tipo II = Probabilidad de encontrar impacto dado que SI lo HAY
    - 80% a 90% es lo habitual
    - El poder depende de
        - La magnitud del efecto (+)
        - el tamaño de la muestra (+)
        - La varianza del resultado (-)
        - El % asignado a tratamiento vs. control )

# Kahoot.it

- Efecto mínimo detectable

$$MDE = (t_{(1-\kappa)} + t_{\alpha}) * \sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}}$$

# Kahoot.it

- Efecto mínimo detectable

$$MDE = (t_{(1-\kappa)} + t_{\alpha}) * \sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}}$$

- El cálculo de poder es clave para saber si vale la pena embarcarse en la evaluación y para calcular el tamaño muestral mínimo para mantener cierto poder

# Problemas de seguimiento en un RCT

- Contaminación
- Desgaste
- Cumplimiento imperfecto

# Diseños cuasi-experimentales

# Diseños cuasi-experimentales

- Selección basada en características observables
  - Técnicas de emparejamiento

- Selección basada en características inobservables
  - Diferencias en diferencias
  - Variables instrumentales
  - Regresión discontinua

# Técnicas de emparejamiento o “matching”

- Consisten en el use de bases de datos grandes y técnicas estadísticas para construir el mejor grupo de comparación basado en características observables

- Emparejamiento exacto
- Emparejamiento en base al propensity score
- Emparejamiento en diferencias en diferencias
- Control sintético

# Emparejamiento exacto

| Treated units | | | |
|---|---|---|---|
| Age | Gender | Months unemployed | Secondary diploma |
| 19 | 1 | 3 | 0 |
| 35 | 1 | 12 | 1 |
| 41 | 0 | 17 | 1 |
| 23 | 1 | 6 | 0 |
| 55 | 0 | 21 | 1 |
| 27 | 0 | 4 | 1 |
| 24 | 1 | 8 | 1 |
| 46 | 0 | 3 | 0 |
| 33 | 0 | 12 | 1 |
| 40 | 1 | 2 | 0 |

| Untreated units | | | |
|---|---|---|---|
| Age | Gender | Months unemployed | Secondary diploma |
| 24 | 1 | 8 | 1 |
| 38 | 0 | 1 | 0 |
| 58 | 1 | 7 | 1 |
| 21 | 0 | 2 | 1 |
| 34 | 1 | 20 | 0 |
| 41 | 0 | 17 | 1 |
| 46 | 0 | 9 | 0 |
| 41 | 0 | 11 | 1 |
| 19 | 1 | 3 | 0 |
| 27 | 0 | 4 | 0 |

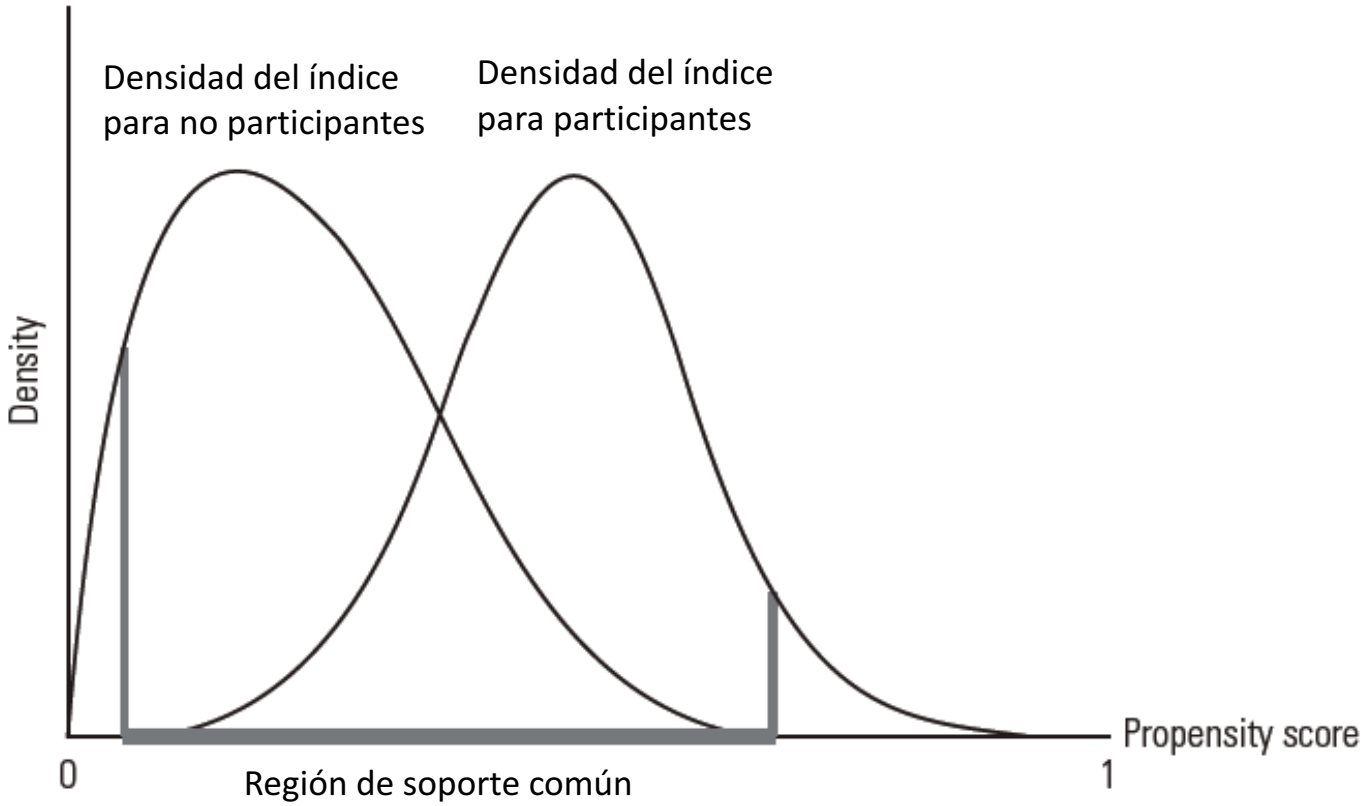
Emparejamiento exacto. Gertler et al 2016.

# Emparejamiento en base a propensity score

- Construimos un índice de participación basado en características observables del individuo, el propensity score
  - Nos quedamos con observaciones tratadas y no tratadas con un soporte común del propensity score
  - Se asignan al grupo de tratamiento a individuos tratados y al grupo de control a individuos similares en observables pero que no han sido tratados

# Emparejamiento y soporte común

## Ejemplo de soporte común




Fuente: Banco Mundial Manual de Evaluación de Impacto

# Emparejamiento en base a propensity score

- Construimos un índice de participación basado en características observables del individuo, el propensity score
  - Nos quedamos con observaciones tratadas y no tratadas con un soporte común del propensity score
  - Se asignan al grupo de tratamiento a individuos tratados y al grupo de control a individuos similares en observables pero que no han sido tratados
- Supuestos claves:
  - Kahoot.it

# Emparejamiento en base a propensity score

- Supuestos claves:
  - Independencia condicionada: una vez que controlo por todas las características que observo, elimino el sesgo de selección (no hay correlación entre la participación y características del individuo que lo lleven a obtener un resultado)
  - Soporte común

# Diferencias en diferencias

- Compara los CAMBIOS en resultados en el tiempo entre unidades que participaron del programa (tratamiento) y unidades que no (control). Esto permite corregir por diferencias entre tratamiento y control que son constantes en el tiempo

- Pischke (2007): efectos de la longitud del año escolar en el aprendizaje
  - Hasta 1960 todos los Estados alemanes empezaban el año escolar en la primavera, menos Bavaria
  - 1966-67: se decidió que todos empezarían en otoño
  - transición: 2 años cortos para cohortes afectados (24 en lugar de 37 semanas)

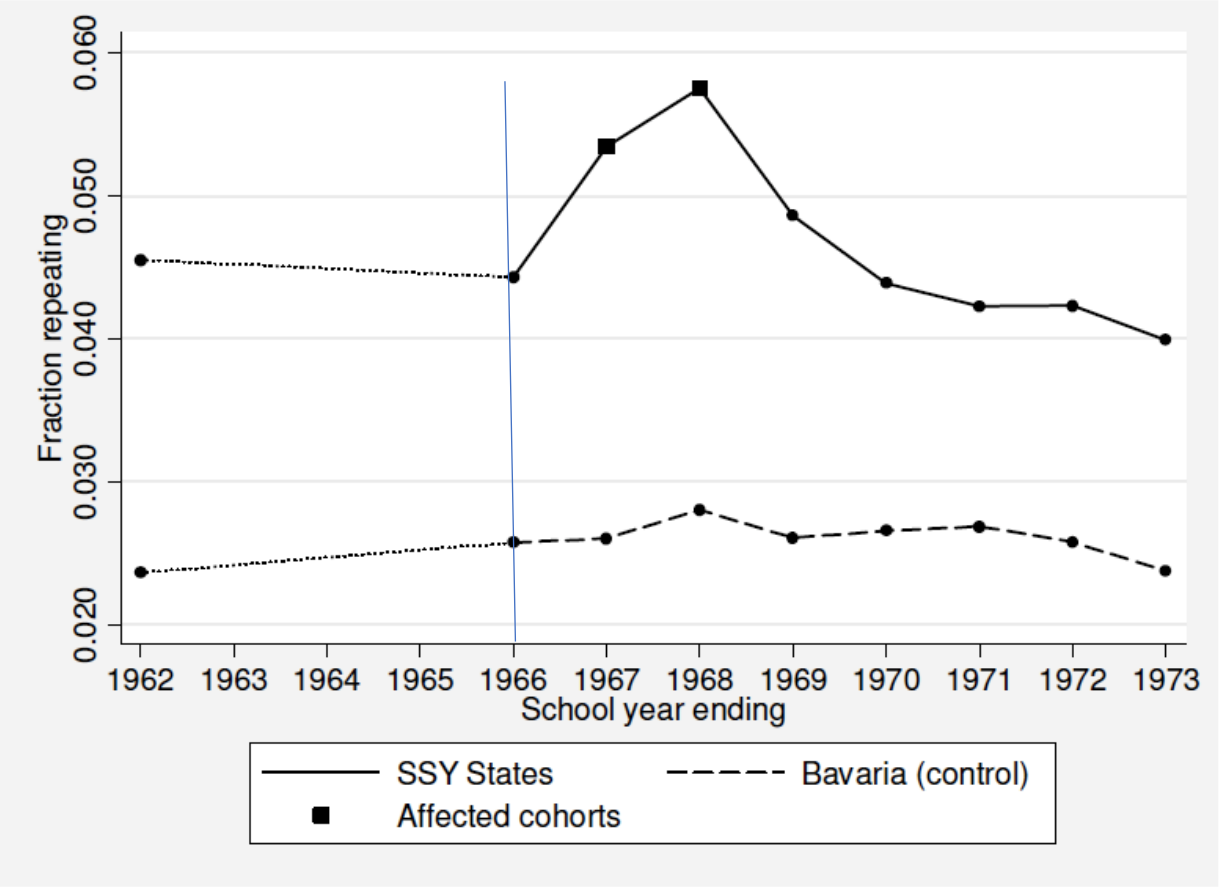
# Diferencias en diferencias

| | Antes | Después | Diferencia |
|---|---|---|---|
| Otros Estados | Resultado OA | Restultado OD | OD - OA |
| Bavaria | Resultado BA | Resultado BD | BD - BA |
| | OA - BA | OD - BD | (OD-OA)-(BD-BA) |

# Diferencias en diferencias

- Kahoot.it
- Supuesto clave

Supuesto: Bavaria es un buen contrafactual para el resto de los Estados. No tiene por qué ser idéntico en niveles, pero sí debe tener una tendencia similar antes de la política



# Variables instrumentales

- Asignación al tratamiento depende exógenamente de ciertas variables no correlacionadas con los resultados

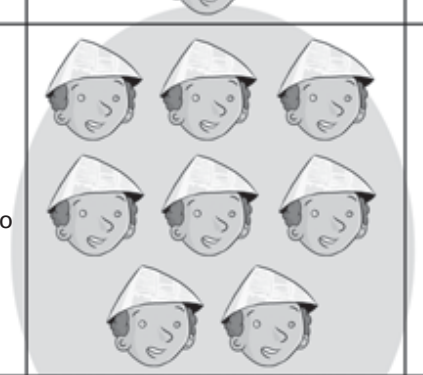
- Ejemplos:
  - distancia a la clínica
  - Paros de transporte
  - Incentivos aleatorios a participar de un programa
  - Mes de nacimiento
  - Presencia de una universidad en la zona donde vive el individuo

# Variables instrumentales

- Aprovecho la variación en la participación que surge a partir de esa variable o variables exógenas para construir un grupo de control
    - Grupo de tratamiento = los que participan porque tienen “activada” esa variable exógena
    - Grupo de control = los que no participan porque la tienen desactivada

- Supuesto clave: no hay correlación entre esta variable exógena y las características del individuo
- Limitaciones: calcula solo un efecto de tratamiento local sobre aquellas personas que participarían del programa cuando la variable instrumental se activa pero que no participarían o reducirían su participación si la variable está desactivada

# Variables instrumentales: LATE

| | Group assigned to treatment | Group not assigned to treatment | Impact |
|---|---|---|---|
| | Percent enrolled = 90%<br>Average $Y$ for those assigned to treatment = 110 | Percent enrolled = 10%<br>Average $Y$ for those not assigned to treatment = 70 | $\Delta$% enrolled = 80%<br>$\Delta Y$ = ITT = 40<br>LATE = 40/80% = 50 |
| Never enroll | | | — |
| Only enroll if assigned to treatment |  | | |
| Always enroll | | | — |

# Regresión discontinua

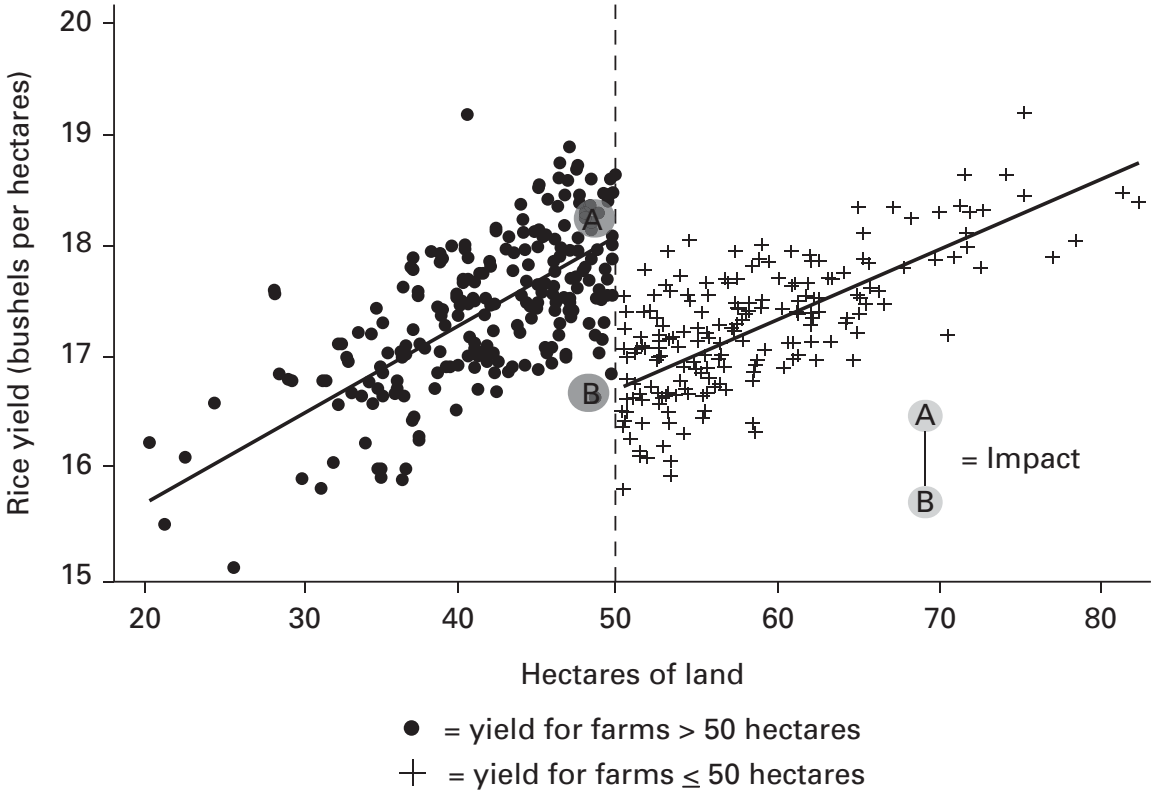
- Programas que usan un índice para asignar tratamiento y que tienen un punto de corte que determina si los individuos son elegibles al programa
  - Asignaciones familiares
  - Ingresos a la universidad

- Se puede usar técnicas de regresión discontinua cuando el índice
  - Ordena individuos en forma continua
  - Tiene un corte claramente definido para asignar tratamiento
  - Solo se usa para asignar tratamiento
  - No se puede manipular el “puntaje” de cada individuo

# Regresión discontinua

- Si la regulación implica que valores del índice a la derecha del punto de corte reciben tratamiento: la regresión discontinua define como
  - Grupo de tratamiento a los individuos con un valor del índice cercano pero por encima del punto de corte
  - Grupo de control a los individuos con un valor del índice cercano pero por debajo del punto de corte

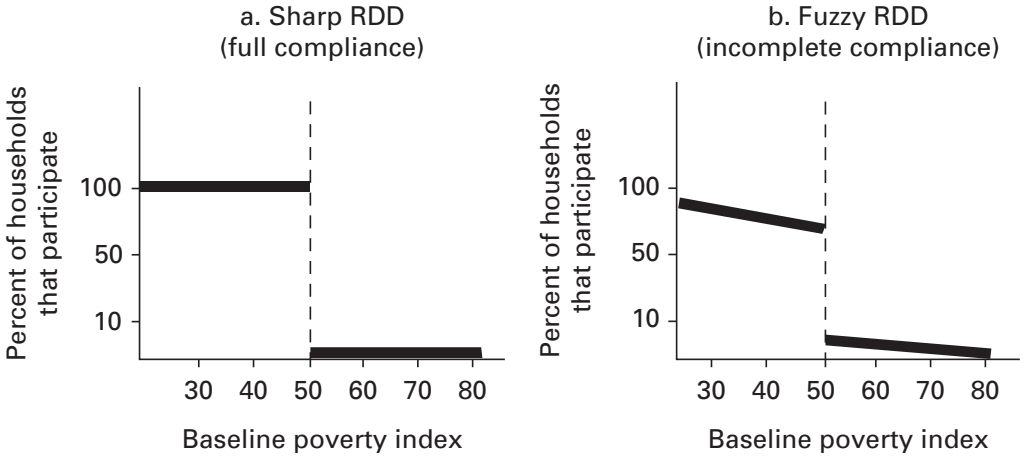
# Ejemplo: subsidios a fertilizantes para agricultores con menos de 50 hectáreas




Fuente: Gertler et al 2016

# Regresión discontinua

- 2 tipos:
  - Sharp o determinística: el tratamiento se activa con probabilidad 1 cuando se pasa el punto de corte
  - Fuzzy o probabilística: pasar el punto de corte aumenta la probabilidad de tratamiento pero no es 1 (discontinuidad en la asignación opera como variable instrumental)




Fuente: Gertler et al 2016

# Resumen métodos cuasi-experimentales

| | SUPUESTOS | Grupo tratado | Grupo control |
|---|---|---|---|
| Emparejamiento | Elimino el sesgo de selección condicionando por características observables | Participantes con soporte común | No participantes emparejados a participantes |
| Diferencias en diferencias | Previo al programa, el grupo de control tiene idéntica tendencia en resultados que el tratado | Participantes | No participantes con misma tendencia previa |
| Variables instrumentales | La variable instrumental solo está correlacionada con el resultado a través de la participación en el programa | Sujetos que participan porque está “activada” la variable instrumental | Sujetos que no participan porque está “desactivada” la variable instrumental |
| Regresión discontinua | No hay manipulación en torno al punto de corte. Puedo estimar bien la forma funcional entre resultados y el índice de asignación en torno al punto de corte | Sujetos en torno al punto de corte que participan porque clasifican según el índice | Sujetos en torno al punto de corte que no participan porque no clasifican según el índice |

Muchas gracias

abalsa@um.edu.uy