



THE WORLD BANK



Sesión Técnica VI: Métodos de Apareamiento

Karen Macours

Managua, 5 Marzo 2008

En el caso de asignación aleatoria...

Supuesto asignación aleatoria al tratamiento:

ambos resultados potenciales son independientes del tratamiento:

$$Y_1, Y_0 \perp D$$

En ese caso, no importa que no podamos observar el contrafactual:

$$E_i \{Y_0(u_i) | D_i=0\} = E_i \{Y_0(u_i) | D_i=1\} = E_i \{Y_0(u_i)\}$$

$$E_i \{Y_1(u_i) | D_i=0\} = E_i \{Y_1(u_i) | D_i=1\} = E_i \{Y_1(u_i)\}$$

Entonces d = Efecto promedio de tratamiento en los tratados

$$= E_i \{\Delta_i | D_i = 1\}$$

$$= E_i \{Y_1(u_i) - Y_0(u_i) | D_i = 1\}$$

$$= \underbrace{E_i \{Y_1(u_i) | D_i = 1\}}_{\text{se puede medir en el análogo muestral}} - \underbrace{E_i \{Y_0(u_i) | D_i = 1\}}_{\text{no se puede medir en el análogo muestral}}$$

$$= E_i \{Y_1(u_i) | D_i = 1\} - \underbrace{E_i \{Y_0(u_i) | D_i = 0\}}_{\text{se puede medir en el análogo muestral}}$$

Cuándo utilizamos métodos de apareamiento?

- La aleatorización asegura que el sesgo de selección sea zero:

$$E_i \{Y_0(u_i) | D_i = 1\} - E_i \{Y_0(u_i) | D_i = 0\} = 0$$

- Ahora, que pasa si la aleatorización no es factible y la selección se hace en base a (un conjunto de) variables observables ?
 - *Ahí es cuando utilizamos los métodos de pareo: Los métodos de pareo son técnicas para construir grupos de comparación cuando la asignación al tratamiento no es aleatoria pero se hace en base a los observables*
- **Cuidado:** los métodos de pareo NO dejan construir grupos de comparación cuando la asignación al tratamiento se hace en base a los no-observables!!!
- **Intuición:** El grupo de comparación debe ser lo mas similar posible al grupo de tratamiento en término de los elementos observables antes de que el tratamiento se desarrolle (asumiendo no hay diferencias en los no observables).

Pregunta clave....

Cual es
el efecto de tratamiento en los tratados,
cuando la selección al tratamiento
está basada en observables?

Unconfoundedness y selección en base a observables

- Suponemos que X denota una matriz en la cual cada fila es un vector de variables observables pre-tratamiento, para cada individuo i .

- **Definición** “Unconfoundedness”

La asignación al tratamiento es “inconfundida” dadas las variables pre-tratamiento X si

$$Y_1, Y_0 \perp D \mid X$$

- Note: “unconfoundedness” es equivalente a:
 - Dentro de cada celda de observaciones definido por X : la asignación al tratamiento es aleatoria.
 - La asignación al tratamiento depende solamente de las variables observables X .

Cual es el efecto de tratamiento en los tratados, bajo selección en observables?

Asumiendo "unconfoundedness":

$$E_i \{Y_0(u_i) | D_i=0, X\} = E_i \{Y_0(u_i) | D_i=1, X\} = E_i \{Y_0(u_i) | X\}$$

$$E_i \{Y_1(u_i) | D_i=0, X\} = E_i \{Y_1(u_i) | D_i=1, X\} = E_i \{Y_1(u_i) | X\}$$

Ahora, para cada celda definida por X , podemos definir

d_X = efecto promedio de tratamiento en los tratados de la celda definida por X

$$= E_i \{\Delta_i | D_i=1, X\}$$

$$= E_i \{Y_1(u_i) - Y_0(u_i) | D_i=1, X\}$$

$$= \underbrace{E_i \{Y_1(u_i) | D_i=1, X\}}_{\text{se puede medir en el análogo muestra}} - \underbrace{E \{Y_0(u_i) | D_i=1, X\}}_{\text{no se puede medir en el análogo muestra}}$$

$$= E_i \{Y_1(u_i) | D_i=1, X\} - \underbrace{E_i \{Y_0(u_i) | D_i=0, X\}}_{\text{se puede medir en el análogo muestra}}$$

Efecto promedio de tratamiento en los tratados bajo “unconfoundedness”

Ahora cuál es la relación entre...

d "efecto promedio de tratamiento en los tratados"... y....

d_x "efecto promedio de tratam. en los tratados en la celda definida por X "?

d = efecto promedio de tratamiento en los tratados

$$= E_i \{ \Delta_i \mid D_i = 1 \}$$

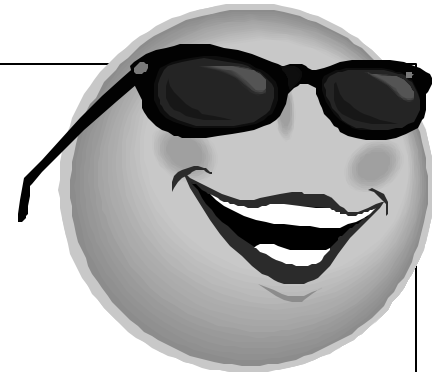
↓ por la ley de las expectativas iteradas

$$= E_i \{ E_x [\Delta_i \mid D_i=1, X] \}$$

$$= E_x \{ E_i [\Delta_i \mid D_i=1, X] \}$$

$$= E_x \{ d_x \}$$

$$= E_x \{ \text{efecto promedio de tratam. en los tratados, en la celda definida por } X \}$$



Estrategia de estimación del efecto promedio del tratamiento en los tratados – selección en observables

Eso sugiere la estrategia siguiente para estimar el efecto promedio de tratamiento en los tratados d :

- Estratificar los datos dentro de cada celda definida para cada valor de X
- Dentro de cada celda (i.e. condicionado en X) calcular la diferencia en la variable de resultado promedio entre el grupo de tratamiento y el de control;
- El promedio de estas diferencias con respecto a la distribución de X , en la población de unidades tratadas

Preguntas:

- Es factible esa estrategia?
- Esa estrategia es diferente de una regresión lineal de Y en D , controlando de manera no-paramétrica por el conjunto completo de efectos principales y interacciones entre los covariados?

Es factible el pareo? El problema de la dimensionalidad

- En la regresión, tenemos que incluir un conjunto completo de interacciones nonparametricas entre todos las variables observables.

- Puede ser que eso no sea factible cuando...
 - La muestra es pequeña,
 - Hay muchos covariados / variables observables
 - Los covariados / variables observables tienen varios valores, o son continuos

- Eso es lo que se llama... el problema de la dimensionalidad

El problema de la dimensionalidad

- Ejemplos:
 - Cuantas celdas tenemos cuando hay 2 variables X que son binarias (“dummies”)? Y con 3 variables X binarias? Y con K variables X binarias?
 - Y cuantas celdas tenemos cuando hay 2 variables que pueden tomar 7 valores cada una?

- A medida que crece el número de celdas, vamos a encontrar una falta de soporte común, es decir
 - Celdas que contienen solo tratamientos
 - Celdas que contienen solo controles



Una alternativa para solucionar el problema de la dimensionalidad

- Rosenbaum y Rubin (1983) proponen una estrategia alternativa equivalente pero factible basada en el “propensity score”.
 - El propensity score convierte el problema multi-dimensional en un problema uni-dimensional
 - Y así reduce el problema de la multi-dimensionalidad

Una alternativa para solucionar el problema de la dimensionalidad

- **Definición:** El “propensity score” es la probabilidad condicional de recibir el tratamiento dadas las variables observadas X antes del tratamiento:

$$p(X) = \Pr\{D = 1|X\} = E_X\{D|X\}$$

- **Lema 1:** Si $p(X)$ es el “propensity score,” $\Rightarrow D \perp X | p(X)$
“Dado el “propensity score,” las variables observables pre-tratamiento X están balanceadas entre los beneficiarios y los no-beneficiarios.”
- **Lema 2 :** $Y_1, Y_0 \perp D | X \Rightarrow Y_1, Y_0 \perp D | p(X)$
“Suponiendo que existe independencia condicional entre el resultado y la asignación al tratamiento, condicional en X , entonces existe también independencia condicional en $p(X)$.”

El propensity score soluciona el problema de multi-dimensionalidad?

- Si!
- El propensity score tiene la propiedad de *balancear características*: (Lemma 1) asegura que:
 - Observaciones con el mismo propensity score tienen la misma distribución de covariados observables, independientemente que sean tratamientos o controles.
 - Para un propensity score dado: la asignación al tratamiento es “aleatoria”. Por eso, en promedio, los tratamientos y los controles son idénticos del punto de vista observacional.

Efecto del tratamiento promedio en los tratados usando “Propensity Score”

Ahora podemos parear los tratamientos y los controles en base al propensity score $p(X)$ en vez de X .

$$E_i \{Y_0(u_i) | D_i=0, p(X_i)\} = E_i \{Y_0(u_i) | D_i=1, p(X_i)\} = E_i \{Y_0(u_i) | p(X_i)\}$$

$$E_i \{Y_1(u_i) | D_i=0, p(X_i)\} = E_i \{Y_1(u_i) | D_i=1, p(X_i)\} = E_i \{Y_1(u_i) | p(X_i)\}$$

Usando estas expresiones, podemos definir para cada celda definida por $p(X)$

$$\begin{aligned} d_{p(X)} &= \text{efecto promedio de tratamiento en los tratados en la celda definida por } p(X) \\ &= E_i \{ \Delta_i | D_i = 1, p(X) \} \\ &= E_i \{ Y_1(u_i) - Y_0(u_i) | D_i = 1, p(X) \} \\ &= \underbrace{E_i \{ Y_1(u_i) | D_i = 1, p(X) \}}_{\text{se puede medir en el análogo muestral}} - \underbrace{E_i \{ Y_0(u_i) | D_i = 1, p(X) \}}_{\text{no se puede medir en el análogo muestral}} \\ &= E_i \{ Y_1(u_i) | D_i = 1, p(X) \} - \underbrace{E_i \{ Y_0(u_i) | D_i = 0, p(X) \}}_{\text{se puede medir en el análogo muestral}} \end{aligned}$$

Average effects of treatment and the propensity score

Ahora cuál es la relación entre

d "efecto promedio de tratamiento en los tratados"... y....

$d_{p(X)}$ "efecto promedio de tratamiento en los tratados dentro de la celda definida por $p(X)$ "?

d = efecto promedio de tratamiento en los tratados

$$= E_i \{ \Delta_i \mid D_i = 1 \}$$

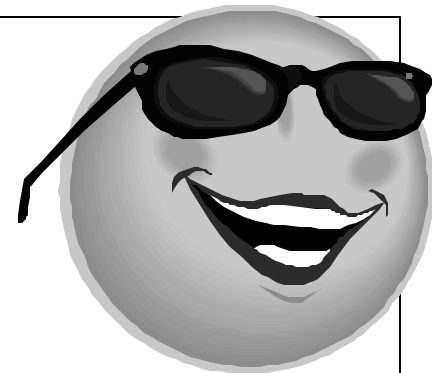
⇓ por la ley de las expectativas iteradas

$$= E_i \{ E_{p(X)} [\Delta_i \mid D_i = 1, p(X)] \}$$

$$= E_{p(X)} \{ E_i [\Delta_i \mid D_i = 1, p(X)] \}$$

$$= E_{p(X)} \{ d_{p(X)} \}$$

$$= E_{p(X)} \{ \text{efecto promedio de tratamiento en los tratados en la celda definida por } p(X) \}$$



Implementación de la estrategia de estimación

Eso sugiere que sigamos la estrategia siguiente para estimar el efecto promedio del tratamiento en los tratados d :

- i. **Primera etapa:** estimar el “propensity score”
- ii. **Segunda etapa:** estimar el efecto promedio de tratamiento dado el “propensity score”

Primera etapa: Estimación del propensity score

- Los modelos probabilísticos estándares pueden ser utilizados para estimar el propensity score. Por ejemplo, un modelo logit sería:

$$Pr\{D_i | X_i\} = \frac{e^{Ih(X_i)}}{1 + e^{Ih(X_i)}}$$

donde $h(X_i)$ es una función de los covariados X con términos lineares, interacciones y potencias.

- Cuales son las interacciones y potencias que tenemos que incluir en $h(X_i)$?
 - Eso está determinado solamente por la necesidad de obtener un estimado del propensity score que satisfaga la propiedad de balance.

Primera etapa: Estimación del propensity score

□ Recuerda:
$$Pr\{D_i | X_i\} = \frac{e^{Ih(X_i)}}{1 + e^{Ih(X_i)}}$$

La especificación de $h(X_i)$

- Es más parsimonia que el conjunto completo de interacciones entre los observables X .
 - Pero no puede ser demasiado parsimonia: porque tiene que satisfacer la propiedad de balance.
- Nota: la estimación del propensity score no requiere ninguna interpretación relacionada con la conducta de los beneficiarios.

Primera etapa: Un algoritmo para estimar el propensity score

1. Empiece con un modelo logit o probit parsimonioso para estimar el propensity score.
2. Ordene los datos según su propensity score estimado (del más bajo al más alto).
3. Estratifique todas las observaciones en bloques, de tal manera que, dentro de cada bloque, el propensity score no sea estadísticamente diferente entre los tratamientos y los controles:
 - a. Empiece con cinco bloques de magnitud igual, por ejemplo *{0 hasta 0.2, 0.2 hasta 0.4, 0.4 hasta 0.6, 0.6 hasta 0.8, 0.8 hasta 1}*
 - b. Dentro de cada uno de esos bloques, compruebe que los promedios de los scores para los tratamientos y los controles son estadísticamente diferentes.
 - c. En caso que sí, aumente el número de bloques, reitere paso b.
 - d. En caso que no, pase al paso siguiente.

Primera etapa: Un algoritmo para estimar el

~~propensity score (continuación)~~

4. Compruebe que la propiedad de balance esté satisfecha en cada uno de los bloques, para cada uno de los covariantes en X :
 - a) Para cada covariante, compruebe que los promedios (y posiblemente los otros momentos) para los tratamientos y los controles son estadísticamente diferentes. Haga eso en cada bloque.
 - b) En caso que uno de los covariados no sea balanceado en uno de los bloques, divida el bloque en dos partes, y compruebe de nuevo en esos bloques más pequeños.
 - c) En caso que uno de los covariados no sea balanceado en una mayoría de los bloques, retome el modelo logit y añada más interacciones y potencias. Reitere las pruebas.

Utilice el programa STATA “**pscore.ado**” (y `psmatch2.ado`)

Descargable en <http://www.iue.it/Personal/Ichino/Welcome.html>



Cuando está apropiado el método de propensity score?

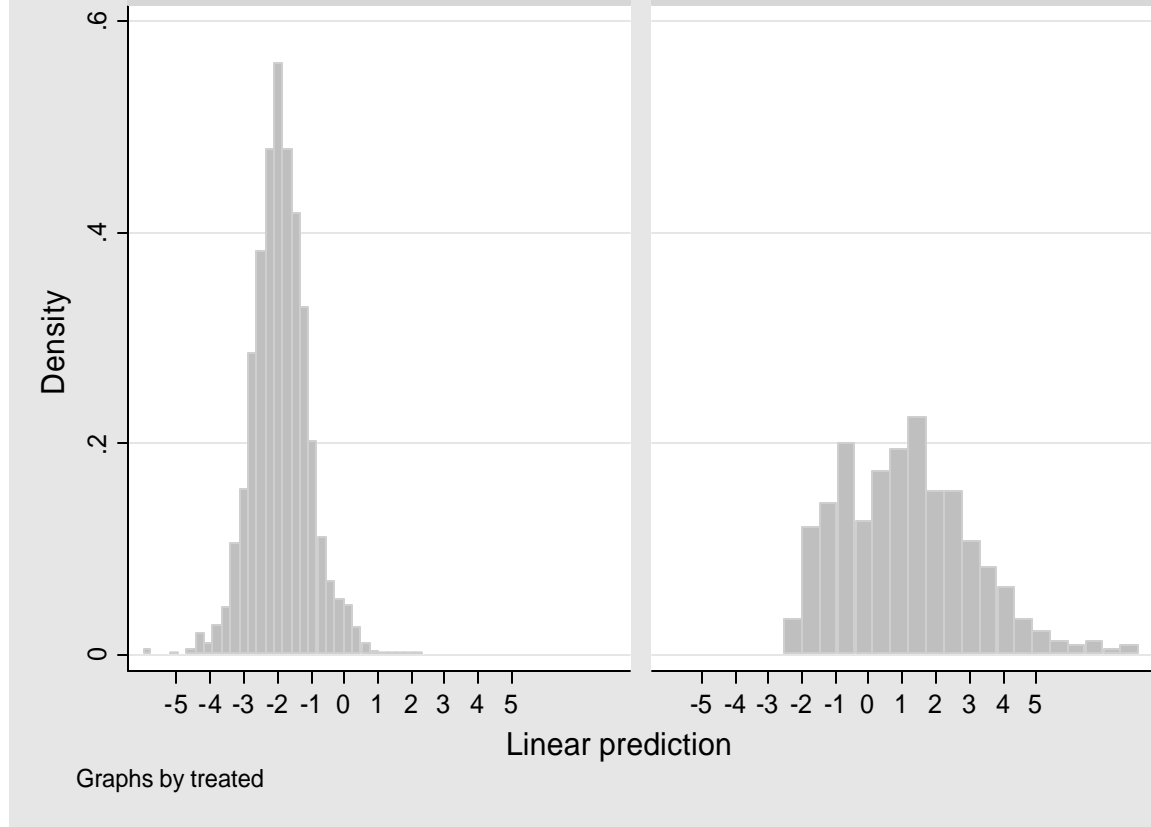
- ❑ Los métodos de propensity score están basados en la idea que la estimación de los efectos de tratamiento requieren un apareamiento cuidadoso entre los tratamientos y los controles.
- ❑ En caso que los tratamientos y los controles sean muy diferentes en términos de características observables, ese apareamiento no estará lo suficientemente dependable, hasta puede ser imposible de realizar.
- ❑ La comparación de los propensity score entre los tratamientos y los controles puede ser una herramienta de diagnóstico útil cuando uno quiere evaluar que tanto similares son los tratamientos y los controles, y que tan confiable es la estrategia de estimación.

Lo que quiere usted es que los propensity scores de los tratamientos y los controles sean “similares”...

- ❑ Pero que es “similar” ???
- ❑ El rango de variación del propensity score debería ser lo mismo para los tratamientos y los controles.
 - Cuente cuantos controles tienen un p.s. que sea abajo del mimino, o arriba del máximo de los p.s. de los tratamientos.
 - Y vice versa.
- ❑ La frecuencia de los propensity scores debería ser la misma para los tratamientos y los controles.
 - Traze histogramas de los p.s. para los tratamientos y los controles.
 - Las celdas de los histogramas corresponden la los bloques utilizados en la estimación de los propensity scores.

Un ejemplo de problema con el soporte común

Figure A1:
Propensity Scores For EIC Phase 1 and non-EiC schools 1



Implementación de la estrategia de estimación

Acuerdense que estamos discutiendo de la estrategia que va utilizar para estimar el efecto promedio del tratamiento en los tratados, o sea d

- **Primera etapa: estimamos el propensity score (hecho!)**
- **Segunda etapa: estimamos el efecto promedio del tratamiento en los tratados, dado el propensity score.**
 - Aparee los tratamientos y los controles que tienen exactamente el mismo p.s. estimado
 - Compute el efecto del tratamiento para cada uno de los valores estimados del p.s.
 - Compute el promedio de estos efectos condicionados.

Segunda etapa: estimamos el efecto promedio del tratamiento, dado el propensity score

- ❑ En la práctica, eso es imposible porque es poco común de encontrar dos observaciones con exactamente el mismo p.s.
- ❑ La solución práctica es de aparear cada unidad tratada, a la unidad control “más cercana”, en cuanto a su p.s.
- ❑ **“Más cercana”** se puede definir de varias maneras en estadística. Cada manera corresponde a una manera de hacer el apareamiento. Por ejemplo.
 - Estratificación
 - Vecino mas cercano (pareo uno a uno o mas de uno) con o sin reemplazo;
 - Apareamiento por Radius
 - Apareamiento “kernel”
 - Weighting on the basis of the Score.

Referencias

- ❑ Dehejia, R.H. and S. Wahba (1999), “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs”, *Journal of the American Statistical Association*, 94, 448, 1053-1062.
- ❑ Dehejia, R.H. and S. Wahba (1996), “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs”, Harvard University, Mimeo.
- ❑ Hahn, Jinyong (1998), “ ON the role of the propensity score in efficient semiparametric estimation of average treatment effects”, *Econometrica*, 66,2,315-331.
- ❑ Heckman, James J. H. Ichimura, and P. Todd (1998), “ Matching as an econometric evaluation estimator ”, *Review of Economic Studies*, 65, 261-294.
- ❑ Hirano, K., G.W. Imbens and G. Ridder (2000), “Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score”, mimeo.
- ❑ Machin, McNally, Meghir. “Excellence in cities: Evaluation of an education policy in disadvantaged areas”
- ❑ Rosenbaum, P.R. and D.B. Rubin (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects”, *Biometrika* 70, 1, 41–55.
- ❑ Vinha, K. (2006) A primer on Propensity Score Matching Estimators” Documento CEDE 2006-13, Universidad de los Andes
- ❑ Imbens, G. (2004), “Semiparametric Estimation of Average Treatment Effects under Exogeneity: A review”, *Review of Economics and Statistics*, 86(1): 4-29.