

Propensity Score Matching

PSM es una técnica común que nos ayuda a medir el efecto de algún tratamiento cuando no hemos podido realizar algún experimento aleatorio controlado (RCT por sus siglas en inglés). Usamos PSM cuando tenemos datos observacionales que incluyen características definidas previas al tratamiento y que determinan si los individuos recibirán o no el tratamiento.

ej. Supongamos que una ONG construye clínicas en varias localidades y sabemos que estas localidades no fueron elegidas de manera aleatoria, sino con base a sus necesidades. Nos interesa saber si el programa afectó la mortalidad infantil. Tenemos los siguientes datos:

	T	imrate
1.	1	10
2.	1	15
3.	1	22
4.	1	19
5.	0	25
6.	0	19
7.	0	4
8.	0	8
9.	0	6

Si comparamos las medias de imrate:

$$0.25(10+15+22+19) - 0.2(25+19+4+8+6) = +4.1$$

¡las clínicas matan! 😞

En realidad, no es que las clínicas ↑ imrate, si no que las clínicas fueron asignadas a las localidades más necesitadas: ↑ pobreza ↓ docs.

	T	imrate	povrate	pdcoos
1.	1	10	.5	.01
2.	1	15	.6	.02
3.	1	22	.7	.01
4.	1	19	.6	.02
5.	0	25	.6	.01
6.	0	19	.5	.02
7.	0	4	.1	.04
8.	0	8	.3	.05
9.	0	6	.2	.04

el grupo T es, en promedio, más pobre y tiene menos docs per cápita.

¿cómo resolvemos este problema?

Para cada una de las obs. en el grupo T vamos a seleccionar una obs. en el grupo C que se le parezca mucho y compararemos el resultado promedio entre grupos. Hagamos el emparejamiento solo con povrate por el momento. Las obs 1 y 6 tienen 50% de povrate, así que las emparejaremos. Ahora lo mismo con las obs. 2 y 5 que tienen povrate = 0.6 (las obs. 3 y 4 también las emparejamos con 5. Notemos que una sola observación puede emparejarse con varias otras y que

algunas se quedarán sin pareja porque son muy distintas y no representarían un buen control o "contrafactual".

El emparejamiento lo podemos hacer con múltiples características, aunque se vuelve un poco más complicado de hacer a mano.

El ejemplo anterior ilustra la esencia del PSM: vemos los datos y analizamos qué variable predice cuál localidad recibe el tratamiento y cuál no. La manera en que lo hacemos es a través de un Probit/Logit en el que la dependiente toma el valor de 1 cuando se recibe el tratamiento y 0 en otro caso y nuestras var. indep. son todas aquellas características que creemos que están relacionadas con el tratamiento; en nuestro ejemplo, será povrate y pdcoos.

	T	imrate	povrate	pdcoos	ps1	Predicted prob of treatment aka the propensity score
1.	1	10	.5	.01	.4165713	
2.	1	15	.6	.02	.7358171	
3.	1	22	.7	.01	.9268033	
4.	1	19	.6	.02	.7358171	
5.	0	25	.6	.01	.752714	
6.	0	19	.5	.02	.395162	
7.	0	4	.1	.04	.0016534	
8.	0	8	.3	.05	.0268033	
9.	0	6	.2	.04	.0070107	

no son muy útiles

Ya que obtuvimos nuestro propensity score, vamos a hacer el emparejamiento entre grupo T y grupo C. Para la obs. 1 tomamos el ps1 = 0.41 y lo vamos a emparejar con obs. 6 que tiene ps1 = 0.39. Las obs. 2-4 las vamos a emparejar con la obs 5 que tiene el ps1 más cercano. En la vida real, quisiéramos emparejar al grupo T con más de 2 observaciones en grupo C.

Ahora sí, después de hacer el PSM podemos estimar el efecto del tratamiento en la mortalidad infantil:

$$(0.25)(10+15+22+19) - (0.25)(19+25+25+25) = -7$$

¡ya tiene sentido!

Propensity Score Theorem

Veámoslo ahora de manera más formal. El teorema de propensity score dice que si se cumple el supuesto de independencia condicional (CIA), entonces la variable de respuesta es independiente a la asignación del tratamiento condicional al PSM. Es decir:

$$Si: y_i, y_{i0} \perp\!\!\!\perp D_i | X_i \Rightarrow y_i, y_{i0} \perp\!\!\!\perp D_i | P(X_i)$$

Este teorema es de mucha utilidad porque, en la práctica, puede darte el problema de que X_i es multidimensional y esto puede derivar en X_i insuficiente para cada sub-muestra [de X_i] en grupo T y grupo C. Por otro lado $P(X_i)$ es un escalar, es decir, es uni-dimensional, entonces el problema se vuelve mucho más sencillo.

Proof. (De mostly harmless econometrics)

$$P.D. \quad P(D_i = 1 | y_{ji}, P(X_i)) \neq F(y_{ji}) \quad (1)$$

Es decir, queremos demostrar que la probabilidad de que un individuo reciba el tratamiento, dado su nivel potencial de y_j y dado el Propensity Score, no es una función de $y_j \rightarrow$ son independientes.

De (1), sabemos que:

$$E[D_i | y_{ji}, P(X_i)] = (1) P(D_i = 1 | y_{ji}, P(X_i)) + (10) P(D_i = 0 | y_{ji}, P(X_i))$$

$$\rightarrow P(D_i = 1 | y_{ji}, P(X_i)) = E[D_i | y_{ji}, P(X_i)], \text{ reescribiendo} \\ = E[E[D_i | y_{ji}, P(X_i), X_i] | y_{ji}, P(X_i)]$$

Notemos que al tener X_i en la expresión, ya tenemos también $P(X_i)$, mientras que el inverso no necesariamente se cumple porque el PS no es único para un set particular de co-variables. entonces, podemos simplificar:

$$P(D_i = 1 | y_{ji}, P(X_i)) = E[E[D_i | y_{ji}, X_i] | y_{ji}, P(X_i)]$$

Ahora, el término sombreado viene de CIA y, como asumimos que se cumple CIA, podemos eliminar y_{ji} de la expresión:

$$P(D_i = 1 | y_{ji}, P(X_i)) = E[E[D_i | X_i] | y_{ji}, P(X_i)]$$

El término ahora sombreado es equivalente a la probabilidad de que $D_i = 1$, que es justo el PS, entonces podemos reescribir:

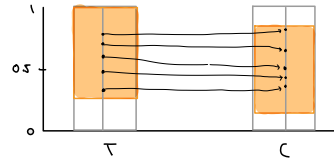
$$P(D_i = 1 | y_{ji}, P(X_i)) = E[P(X_i) | y_{ji}, P(X_i)] \\ = P(X_i) \neq F(y_{ji})$$

Entonces, si se cumple CIA, sucede que $y_{ji} \perp\!\!\!\perp D_i | P(X_i)$. Este teorema nos permite asumir que el tratamiento ha sido aleatoriamente asignado, condicional al PSM.

Tipos de Emparejamiento

Vecino más cercano (nearest neighbour)

- Para cada observación i , seleccionamos la observación que tiene el $p(x_i)$ más cercano. $\min \|p_i - p_j\|$



Emparejamiento por radio (radius matching)

- Cada observación i es emparejada con las observaciones de control j que estén contenidas en un radio especificado

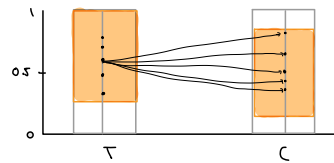
$$\|p_i - p_j\| < r$$

Emparejamiento de Kernel

- Cada observación i es emparejada con varias observaciones de control, en peso inversamente proporcional a la distancia entre las obs. T y obs. C.
- Si el emparejamiento lo hacemos con PS, los pesos se definen como:

$$w(i,j) = \frac{\kappa\left(\frac{p_i - p_j}{h}\right)}{\sum_{j=1}^n \kappa\left(\frac{p_j - p_i}{h}\right)}$$

con h un parámetro de ancho de banda.



Emparejamiento por intervalos

- Comparamos los PS entre bloques o intervalos. ej. 0.0-0.1, 0.11-0.2, 0.21-0.3, ...

Soporte común

- restringimos el emparejamiento sólo a range común de PS.

