



Metodología de investigación cuantitativa

.....

Matching y DiD

Pablo Geraldo Bastías

pdgerald@uc.cl

Estructura de la presentación

1. Cuasi-experimentos
2. Tipos de sesgo
3. (Propensity Score) Matching
4. Doble diferencia
5. Referencias

Cuasi-experimentos

Se denomina métodos cuasi-experimentales u observacionales al tipo de análisis que busca establecer relaciones causales en ausencia de asignación aleatoria:

- Aleatoriedad es la “base razonada para la inferencia” (Fisher).
- En ciertos contextos, randomizar es imposible.
- Imaginar: ¿Qué experimento sería necesario realizar para estimar el efecto causal de mi interés?
- Buscar el mejor acercamiento posible a un escenario experimental.

Estructura de la presentación

1. Cuasi-experimentos
2. Tipos de sesgo
3. (Propensity Score) Matching
4. Doble diferencia
5. Referencias

Tipos de sesgo

De acuerdo a Zeng y King (2006), el sesgo en la estimación de un efecto causal puede ser descompuesto en cuatro fuentes:

- **Variable omitida** (Δ_o): cuando no se ajusta por variables relevantes (“confounders”).
- **Post-tratamiento** (Δ_p): cuando se ajusta por variables que son afectadas por el tratamiento (“mecanismos”).
- **Interpolación** (Δ_i): cuando se ajusta por las variables adecuadas pero de forma inadecuada.
- **Extrapolación** (Δ_e): cuando se comparan casos fuera de la zona de soporte común.

Tipos de sesgo

Es importante notar que, en presencia de asignación aleatoria, las cuatro fuentes de sesgo recién descritas tienden a ser irrelevantes.

Por otro lado, aunque ajustar por ciertas variables sea necesario para evitar sesgos de interpolación y extrapolación, esto afecta también a la cantidad que está siendo estimada (ATE, ATT, ATC).

Estructura de la presentación

1. Cuasi-experimentos
2. Tipos de sesgo
3. (Propensity Score) Matching
4. Doble diferencia
5. Referencias

Matching

El matching consiste en “emparejar” observaciones, para comparar aquellas que sean más similares entre sí en sus características, pero que difieran sólo en su estatus de tratado o control. De este modo, se busca aproximar un experimento aleatorio.

A medida que el número de características a considerar aumenta, se hace más difícil encontrar sujetos similares. Por esta razón se recurre al *propensity score matching*: comparar sujetos que tienen probabilidades similares de haber recibido el tratamiento.

Propensity Score

El propensity score ($e(X)$) se define como la probabilidad de recibir el tratamiento ($D_i = 1$) condicional a las características observadas:

$$e(X) = Pr(D_i = 1 | X_i = x)$$

Supuestos del matching

Los siguientes supuestos son necesarios para estimar un efecto causal mediante el matching:

- **Balance:** Las características observadas están distribuidas igualmente entre los grupos, dado el propensity score.

$$D \perp\!\!\!\perp X | e(x)$$

- **Ignorabilidad (CIA):** La asignación al tratamiento es independiente de las características no observadas y de los resultados potenciales, dadas las características observadas (o el propensity score)

$$D \perp\!\!\!\perp (Y(1), Y(0), u) | X \implies D \perp\!\!\!\perp (Y(1), Y(0), u) | e(X)$$

Estructura de la presentación

1. Cuasi-experimentos
2. Tipos de sesgo
3. (Propensity Score) Matching
4. Doble diferencia
5. Referencias

Differences-in-differences

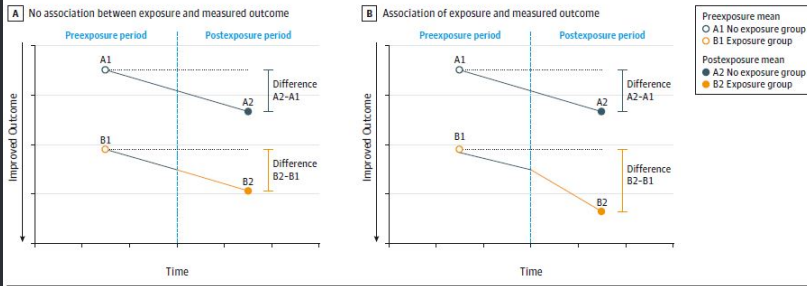
Una metodología cuasi-experimental que requiere de datos longitudinales es la doble diferencia o DiD.

Es similar a una metodología pre/post, con la diferencia de que en el primer momento ningún sujeto es tratado, y en un segundo momento sólo un grupo recibe el tratamiento.

De este modo, se busca controlar por las diferencias constantes entre los grupos, y al mismo tiempo por la tendencia temporal de los resultados.

Ilustración gráfica

Figure. Conceptual Illustration of a Difference-In-Differences Analysis for 2 Scenarios



Fuente: [acá](#)

Doble diferencia

El efecto de una intervención puede estimarse mediante una doble diferencia, de allí su nombre:

- **diferencia temporal**: dentro de cada grupo, se resta el resultado antes del tratamiento al resultados después del tratamiento. Así, tenemos $Y_{t1}(D = 1) - Y_{t0}(D = 1)$ para los tratados, y $Y_{t1}(D = 0) - Y_{t0}(D = 0)$ para los controles.
- **diferencia de grupos**: luego, restamos las diferencias obtenidas en el paso anterior. Así, tenemos $[Y_{t1}(D = 1) - Y_{t0}(D = 1)] - [Y_{t1}(D = 0) - Y_{t0}(D = 0)]$

Supuestos

El principal supuesto de esta metodología es conocido como **lineas paralelas**. Esto quiere decir que, de no existir la intervención, ambos grupos mantendrían la misma tendencia en sus resultados.

¿Matching o DiD?

Se suele afirmar que los supuestos del DiD son menos restrictivos (es decir, más débiles) que los del matching. Mientras éste último requiere que los grupos de tratamiento y control tengan **niveles** comparables en sus resultados potenciales, el DiD sólo requiere que sus **tendencias** sean comparables.

Estructura de la presentación

1. Cuasi-experimentos
2. Tipos de sesgo
3. (Propensity Score) Matching
4. Doble diferencia
5. Referencias

Referencias

- Imbens, G., y Rubin, D. (2015) Causal Inference for Statistics, Social, and Biomedical Sciences. Cambridge University Press.
- Maldonado, L. (2016) Inferencia Causal. Apuntes de Clase, Posgrado en Sociología UC
- Morgan, S., y Winship, C. (2015) Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge University Press (2nd Ed.)
- Pearl, J. (2009) Causality: Models, Reasoning, and Inference. Cambridge University Press (2nd. Ed.)
- Pearl, J. (2016) Causal Inference in Statistics: A Primer. Wiley & Sons.