

Metodología de investigación cuantitativa

.....

Modelo contrafactual de causalidad

Pablo Geraldo Bastías

pdgerald@uc.cl

Estructura de la presentación

1. Identificación y validez
2. Resultados potenciales
3. Gráficos causales (DAGs)
4. Controversias
5. Referencias

Propósito de la sesión

- Introducir a los estudiantes al modelo contrafactual de causalidad, identificando el denominado “problema fundamental de la causalidad”
- Comprender la utilidad de la asignación aleatoria para identificar relaciones causales
- Comprender la diferencia entre validez interna y validez externa en los estudios causales
- Comprender la equivalencia de distintos lenguajes para expresar supuestos y relaciones causales: resultados potenciales y gráficos causales

Identificación y causalidad

- Inferir relaciones causales a partir de un conjunto de datos siempre implica invocar supuestos: Los datos por sí mismos sólo ofrecen “correlaciones”, no causalidad.
- Explicitar y discutir la plausibilidad de los supuestos necesarios para que una determinada correlación se considere causal se denomina “Estrategia de identificación”.
- La técnica estadística utilizada, o método de estimación, tiene un rol secundario respecto de la estrategia de identificación.

Tipos de validez

La **validez interna** refiere a la existencia, o no, de un efecto tratamiento en determinada situación.

Se suele afirmar que los experimentos ofrecen el más alto grado de validez interna.

La **validez externa** refiere a la posibilidad de extrapolar un resultado obtenido en una muestra particular a una población mayor.

Se suele afirmar que los experimentos ofrecen muy poca validez externa.

Experimentos

Existen diversos tipos de experimentos. Lo que tienen en común es la asignación aleatoria del tratamiento, lo que permite atribuir causalidad al factor que está siendo “manipulado” por el investigador.

- Experimentos de laboratorio
- Experimentos de campo
- Experimentos en encuestas
- Experimentos de lista
- Experimentos naturales

Estructura de la presentación

1. Identificación y validez
2. Resultados potenciales
3. Gráficos causales (DAGs)
4. Controversias
5. Referencias

Modelo de Neyman-Rubin

Una de las formalizaciones más influyentes del concepto contrafactual de causalidad fue desarrollada por Jerzy Neyman (1894-1981), y luego extendida por Donald Rubin (1943-).

En la actualidad, se habla del modelo causal de Neyman-Rubin, o el modelo de resultados potenciales. La utilidad del modelo consiste en su capacidad de expresar de manera transparente los supuestos necesarios para identificar relaciones causales, tanto en contextos experimentales como observacionales.

Resultados potenciales

Tenemos una población (U) compuesta por individuos indexados por i . Nos interesa estudiar el efecto de determinado tratamiento (exposición, causa, etc.) sobre estos individuos.

- Por simplicidad, consideraremos que un individuo puede ser expuesto a un tratamiento (t) o ser asignado a un grupo de control (c).
- La situación a la que es asignada cada unidad la denotaremos D_i . Así, $D_i = t$ indica que un individuo es tratado, mientras $D_i = c$ indica que es control.

Resultados potenciales

- La respuesta de cada unidad al tratamiento recibido se denota Y_i .
 - $Y_i(t)$ corresponde a la respuesta que observaríamos si la unidad i recibiera el tratamiento.
 - $Y_i(c)$ corresponde a la respuesta que observaríamos si la unidad i estuviese en situación de control.
- Es **fundamental** que, para cada unidad, el valor de D_i pudo ser distinto (situación contrafactual).
- De este modo, mientras D_i corresponde a una variable que **puede** ser asignada aleatoriamente, la respuesta $Y_i(D)$ es fijado el tratamiento recibido.

Problema fundamental

El efecto del tratamiento lo denotamos como τ_i , y está definido por:

$$\tau_i = Y_i(t) - Y_i(c) \quad (1)$$

¿Qué consecuencias se pueden desprender de esta definición?

Problema fundamental de la inferencia causal

Los resultados potenciales $Y_i(t)$ e $Y_i(c)$ no pueden ser observados simultáneamente para la misma unidad. Por ende, τ_i tampoco puede ser observado.

Efecto causal promedio (ATE)

Si consideramos que $D_i = t = 1$ y que $D_i = c = 0$, tenemos que la relación entre los resultados potenciales, el resultado observado, y la asignación al tratamiento es la siguiente:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) \quad (2)$$

Una solución al problema fundamental es hacer uso de los valores poblacionales, para obtener el valor esperado del efecto tratamiento:

$$\mathbb{E}(\tau_i) = \mathbb{E}(Y_i(t) - Y_i(c)) = \mathbb{E}(Y_i(t)) - \mathbb{E}(Y_i(c)) \quad (3)$$

Esta fórmula corresponde al **Average Treatment Effect** o ATE.

Supuestos

Para identificar el ATE a partir de una muestra, es necesario considerar tres supuestos:

- **Independencia:** La asignación al tratamiento no guarda relación con los resultados potenciales. En términos formales:

$$Y_i(t), Y_i(c) \perp\!\!\!\perp D_i$$

Lo que implica que:

$$\mathbb{E}(Y_i(obs)|D_i = t) = \mathbb{E}(Y_i(t)|D_i = t)$$

$$\mathbb{E}(Y_i(obs)|D_i = c) = \mathbb{E}(Y_i(c)|D_i = c)$$

¿Qué relación guarda la asignación aleatoria con el supuesto de independencia?

Supuestos

- **Restricción de exclusión:** Los resultados potenciales dependen únicamente del tratamiento recibido, y no del tratamiento asignado.
- **Consistencia:** También conocida como SUTVA (Stable Unit-Treatment-Value Assumption), implica que los resultados potenciales están bien definidos, no existe interferencia entre unidades, y no existen versiones del tratamiento.

Estructura de la presentación

1. Identificación y validez
2. Resultados potenciales
3. Gráficos causales (DAGs)
4. Controversias
5. Referencias

¿Qué es un DAG?

El tipo de gráficos causales del que nos ocuparemos en esta presentación corresponde a los *directed acyclic graphs* o simplemente DAGs.

- **Gráfico:** corresponde a una representación visual de una estructura de datos (“modelo de realidad”).
- **Dirigido:** los DAGs indican direccionalidad en la relación entre variables (“supuestos causales”).
- **Acíclico:** en la relación entre variables, no es posible recorrer ningún camino que lleve de vuelta al punto de partida (“temporalidad”).

¿Qué es un DAG?

Para representar las relaciones causales de interés, los DAGs hacen uso fundamentalmente de dos componentes:

- **Nodos:** representan variables. En general, se usan puntos negros para referir a variables observadas, y puntos “vacíos” para referir a variables no observadas. Complementariamente, pueden denotarse con letras o nombres.
- **Flechas:** representan las relaciones entre variables. Así, la presencia de dos nodos conectados por una flecha implica la existencia de una influencia causal desde una variable a otra. Una flecha bidireccional (entrecortada) sólo se usa para indicar la presencia de una causa común no observada.

Fuentes de asociación

Existen diversos caminos que pueden llevar de una variable a otra. Estos caminos implican relaciones causales y, en términos probabilísticos, “asociaciones”. Es decir, conocer el valor de una de estas variables nos entrega información acerca del valor más probable de la otra.

Es importante notar que, así como se suele indicar que correlación no implica causalidad, en el contexto de los gráficos causales no hay causalidad sin correlación.

Causación y mediación

Supongamos la siguiente relación: el origen social de una familia (O) determina la calidad de la educación que recibirá su hijo (E), y ésta determinará a su vez su futuro desempeño académico, por ejemplo, en la PSU. En este escenario, tendríamos el siguiente gráfico causal:

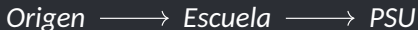


Figura: Relación de mediación

- La variable PSU se denomina “hijo” de E y “descendiente” de O. Inversamente, la variable O se denomina “padre” de E, y “ascendiente” de PSU.
- O tiene un efecto causal sobre E, mientras E tienen un efecto causal sobre PSU. Por ende, también existe un efecto de O sobre PSU que se encuentra mediada por el efecto de la Escuela.
- En este gráfico se excluye la existencia de un efecto directo de O sobre PSU, pero eventualmente podría ser incorporado.

Dependencia mutua

Por otro lado, es posible que el origen social de una familia (O) afecte tanto la calidad de la educación que reciben sus hijos (E), como su futuro logro educativo (L), por ejemplo, que cursen o no educación superior. Entonces tendríamos el siguiente camino causal:

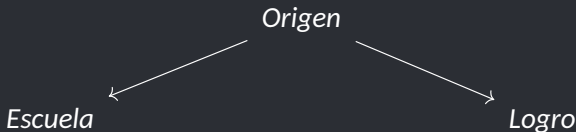


Figura: Relación de mutua dependencia

- Existe una asociación entre E y L sin que ninguna afecte causalmente a la otra. La asociación se produce porque comparten una causal común, que es el origen social. Es decir, un confounder.
- Es necesario *controlar* por la causa común, es decir, el origen social, para revelar que no existe asociación entre E y L.
- En este gráfico se excluye la existencia de un efecto directo de la escuela sobre el logro educativo, pero eventualmente podría ser incorporado.

Causación mutua

Imaginemos que, gracias a una serie de reformas, el origen social ya no influya más en la educación que recibimos. De todas formas el origen social podría seguir influyendo en la PSU (v.g. capital cultural). Por otro lado, la calidad de la escuela influirá sobre los resultados de la PSU. Entonces tendríamos el siguiente gráfico causal:



Figura: Relación de causación mutua

- NO existe una asociación entre origen social y calidad de la educación (¡la reforma!). La variable PSU, con dos flechas apuntando hacia ella, “bloquea” cualquier asociación entre ambas.
- Sin embargo, si condicionamos por la variable PSU, surgirá una asociación entre origen social y educación: un estudiante con menos de 500 puntos en una buena escuela, tiene probablemente bajo capital cultural. Un estudiante con 700 puntos y bajo capital cultural, proviene seguramente de una excelente escuela.
- En este gráfico se excluye la existencia de un efecto directo de O sobre E sólo con fines explicativos.

Ajuste y sesgo

A partir de los tres tipos de relaciones entre variables ilustradas previamente, es posible comprender cuando resulta pertinente ajustar o controlar por determinadas variables y cuando no lo es. Es decir, a partir de determinado DAG, es posible emprender un estudio de identificación de un efecto causal de interés.

Cadenas

- **Mediación:** en términos gráficos, corresponde a la “cadena” ($A \rightarrow C \rightarrow B$).
- Una relación de mediación ocurre cuando el efecto causal es transmitido por una variable intermedia. En nuestro ejemplo, para identificar el efecto (total) de O sobre PSU se requiere *no ajustar* por la Escuela, pues se encuentra en el camino entre dos variables conectadas causalmente.
- Controlar por un mediador equivale a sobre-controlar, induciendo sesgo pues elimina una proporción del efecto causal que se busca estimar.

Bifurcaciones

- **Confounder:** en términos gráficos, corresponde a una “bifurcación” ($A \leftarrow C \rightarrow B$).
- La dependencia mutua produce asociación en ausencia de relaciones causales. De allí que un verdadero mantra de la investigación social sea que “correlación no implica causalidad”.
- Para identificar el efecto causal de interés (en nuestro ejemplo, la influencia inexistente entre E y L) es necesario *ajustar* por la causa común.

Bifurcación invertida

- **Collider:** en términos gráficos, corresponde a una “bifurcación invertida” ($A \rightarrow C \leftarrow B$).
- Un peligro menos considerando en la práctica es la posibilidad de generar asociaciones (espúreas) al ajustar por determinadas variables. En términos gráficos, estas variables son colliders, es decir, puntos de choque entre dos caminos causales.
- Para identificar el efecto causal de interés es necesario *no condicionar* por un collider.

Ejemplo: logro educativo

Supongamos que el efecto de interés que buscamos capturar es el efecto de la escuela sobre el logro educativo. Sin embargo, no basta con conocer la asociación entre ambos, porque dicha asociación oculta tanto relaciones causales (las que nos interesan!) como no causales (las que queremos descartar).

Debemos, por lo tanto, emprender un análisis de identificación con los siguientes pasos:

1. Construir el modelo causal que expresa nuestros supuestos sobre el proceso.
2. Determinar si es posible aislar el efecto de interés a partir de la información disponible.
3. Determinar si haría falta recoger información adicional para aislar el efecto de interés.
4. Discutir la plausibilidad del modelo causal construido.

Ejemplo: logro educativo

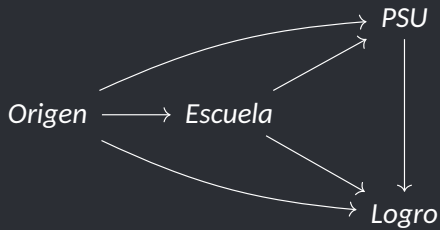


Figura: Modelo de logro educativo

Identificación

A partir de dicho modelo, las fuentes de asociación entre la escuela (nuestra variable de interés, tratamiento o exposición), y el logro educativo (nuestra variable resultado) son las siguientes:

- *Escuela* \rightarrow *Logro*: efecto directo
- *Escuela* \rightarrow *PSU* \rightarrow *Logro*: efecto mediado
- *Escuela* \leftarrow *Origen* \rightarrow *Logro*: camino trasero (confounder)
- *Escuela* \leftarrow *Origen* \rightarrow *PSU* \rightarrow *Logro*: camino trasero (confounder)

Identificación

Si consideramos que el DAG presentado refleja adecuadamente nuestra concepción acerca del proceso de logro educativo (incluyendo la investigación previa e hipótesis plausibles), entonces podemos concluir lo siguiente:

- Ajustar por PSU no es de utilidad e induce sesgo. Esto, pues si bien permite bloquear (parte) del efecto confundente del origen social, también bloquea (parte) del efecto de interés que es mediado por la PSU.
- Ajustar por origen social es suficiente para identificar el efecto de interés. Se eliminan los caminos traseros que van desde la escuela al logro educativo, sin interceptar el efecto directo e indirecto de la escuela sobre el logro educativo.

Estructura de la presentación

1. Identificación y validez
2. Resultados potenciales
3. Gráficos causales (DAGs)
4. Controversias
5. Referencias

Contribuciones

Entre los usuarios de DAGs, crecientes en sociología, ciencia política y epidemiología, se reconocen importantes aportes de los gráficos causales a la comprensión de la identificación causal.

- El peligro de los modelos de “todas las causas”
- La falacia de la Tabla 2
- Comprensión del sesgo de selección endógena
- Clarificación de supuestos en datos panel

Controversias

Sin embargo, existen todavía importantes controversias en torno a la utilidad de los DAGs, especialmente en campos como la economía.

- “No decimos que nuestras inferencias estén justificadas porque dependen de supuestos, sino que distinguimos entre inferencias sustentadas o no (...) Un supuesto no es una base para la inferencia a menos que el supuesto esté garantizado” (Rosenbaum, 1995).
- “Creemos que los métodos de Pearl, aunque son herramientas formidables para manipular DAGs, pueden fácilmente adormecer a los investigadores en una falsa confianza acerca de sus conclusiones causales” (Imbens y Rubin, 1995).

Controversias

- “El trabajo de Pearl es interesante, y muchos investigadores consideran convincente su argumento de que los diagramas son una manera natural y conveniente de expresar supuestos sobre estructuras causales. En nuestro propio trabajo, sin embargo, quizás por el tipo de ejemplos que surgen en las ciencias sociales y médicas, no hemos encontrado que este enfoque ayude a establecer inferencias causales” (Imbens y Rubin, 2015).

Estructura de la presentación

1. Identificación y validez
2. Resultados potenciales
3. Gráficos causales (DAGs)
4. Controversias
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.