Fundamentos Causalidad I

Alvaro J. Riascos Villegas

Julio de 2020

Contenido

Estadística y Modelos Causales

- Muchas preguntas que frecuentemente nos hacemos de un conjunto de datos: cómo (el mecanismo que causa un evento) y por qué (que sucedio que causo un evento), no es posible responderlas en el marco estadístico tradicional.
- Es necesario un marco conceptual adicional, una teoría de causalidad.

Contenido

- Edward Simpson (1922): Una relación estadística que se cumple para toda la población puede reversarse en cada subpoblación.
- Por ejemplo: en una población de estudiantes se podría encontrar que en promedio los fumadores tienen mejores notas. Sin embargo, entre cada grupo de edad los fumadores tienen peores notas y, entre cada grupo de edad y sexo los fumadores tiene mejores notas y asi puede reversarse la asociación anterior en cada subpoblación.

Paradoja de Simpson: Ejemplo

Example

Se les ofrece tomar de forma voluntaria una droga a 700 pacientes. 350 pacientes la toman y los demás no.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Debemos o no recomendar la droga?

Los datos sugieren que **si conocemos el sexo de las personas, debemos recomendar tomar la droga**. Pero si no lo conocemos, no!

Paradoja de Simpson: Ejemplo

Example

Se les ofrece tomar de forma voluntaria una droga a 700 pacientes. 350 pacientes la toman y los demás no.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Debemos o no recomendar la droga?
Los datos sugieren que **si conocemos el sexo de las personas, debemos recomendar tomar la droga**. Pero si no lo conocemos, no!

Paradoja de Simpson: Ejemplo

- En esta ejemplo tenemos 343 mujeres y 357 hombres.
- Es imposible racionalizar este fenómeno sin apelar a alguna teoría (i.e., hipótesis):
  Suponga que el estrógeno reduce la efectividad de la droga.
  Sin embargo, supongamos que esta es efectiva tanto en hombres como mujeres.

Paradoja de Simpson: Ejemplo

- Para evaluar el efecto de la droga sobre la población quisieramos eveluar que sucede con la población si todos toman la droga y compararla contra el resultado en el que nadie en la población toma la droga,
- Evidentemente esto no es lo que observamos en el experimento.
- Podríamos estimar ingenuamente el efecto si elegimos aleatoriamente una persona que tomo la droga y la comparamos contra una persona elegida aleatoriamente que no tomo la droga y repetimos varias veces esto y promediamos.
- Como es más probable elegir una mujer en el primer caso (i.e., son más mujeres las que toman la droga) y un hombre en el segundo caso, parece a nivel poblacional que la droga no funciona.
- Sin embargo esta forma de estimar el efecto poblaciónal esta fundamentalmente errada.

Paradoja de Simpson: Formalmente

- Formalmente queremos estimar (i.e., Teoría de resultados potenciales):
$$E(Y_t) - E(Y_c) \tag{1}$$
donde Y_t es el resultado sobre las personas de ser tratado y Y_c es el resultado sobre las personas si no son tratadas.
- **El valor esperado es sobre toda la población pero no se puede observar una persona simultáneamente tratada y no tratada**.
- Lo que observamos es (S, Y_S) donde S es una función indicadora de si la persona fue tratada o no. Luego observamos $E(Y \mid S)$.
- Pero en general $E(Y_t) \neq E(Y_t \mid S = t)$ y $E(Y_c) \neq E(Y_c \mid S = c)$. En nuestro caso, ser mujer y ser tratado no son independientes.
- Entonces:
$$E(Y_t) - E(Y_c) \neq \tag{2}$$
$$E(Y_t \mid S = t) - E(Y_c \mid S = c) \tag{3}$$

Hipótesis de Independencia

- Como mencionamos anteriormente $E(Y_t) \neq E(Y_t \mid S = t)$ y $E(Y_c) \neq E(Y_c \mid S = c)$.
- Hay por lo menos tres formas de resolver este problema:
  1. Estabilidad temporal y transitoriedad: el resultado de $Y_c(u)$, primero aplicar c a u y después observar el resultado $Y_c(u)$ no depende del momento en el que se haga y $Y_t(u)$ no depende de que anteriormente se haya expuesto u a c. El efecto causal se estima como $Y_t(u) - Y_c(u)$.
  2. Homogenedidad de las unidades: $Y_t(u_1) = Y_t(u_2)$ y $Y_c(u_1) = Y_c(u_2)$. El efecto casual se toma como $Y_t(u_1) - Y_c(u_2)$. Esta hipótesis es común en diseños experimentales.
  3. Independencia estadística: $E(Y_t) = E(Y_t \mid S = t)$ y $E(Y_c) = E(Y_c \mid S = c)$.
- Obérvese que en las dos primeras se estima el efecto causal por unidad.
- La última requiere que el resultado sea independiente de ser tratado o no (i.e., lo que no sucede en la Paradoja de Simpson porque existe un factor común, sexo, que afecta el resultado y el tratamiento).

Hipótesis de Independencia y Regresión: Lineal

- La hipótesis de independencia condicional se cumple cuando el tratamiento se asigna de forma aleatoria.
- En ese caso una forma de estimar el efecto es:

$$y = \beta_0 + \beta_1 T \tag{4}$$

  donde T es la dummy de ser tratado o no.
- En este caso el estimador de mínimos cuadrados ordinarios es un estimador no sesgado de el efecto promedio del tratamiento.

Paradoja de Simpson: Ejemplo

- Una versión continua de esta paradoja es la siguiente. Considere la relación entre el colesterol y el ejercicio de un grupo de personas.

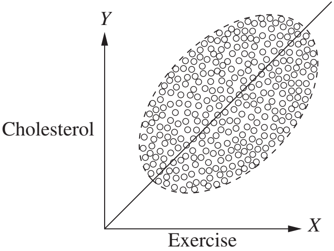

Figure 1.2 Results of the exercise–cholesterol study, unsegregated. The data points are identical to those of Figure 1.1, except the boundaries between the various age groups are not shown

Paradoja de Simpson: Ejemplo

- Ahora si condicionamos a la edad la historia es muy distinta.

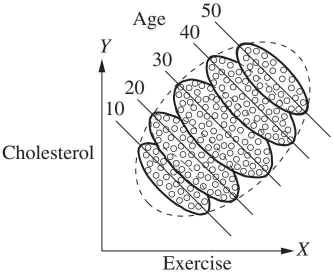

Figure 1.1 Results of the exercise–cholesterol study, segregated by age

Paradoja de Simpson: Ejemplo

- De nuevo, la unica forma de racionalizar esto es apelando a alguna teoría (historia) que va más allá de los datos disponibles.
- Suponga que los mayores hacen más ejercicio que los más jóvenes y tiene el colesterol más alto independientemente de si hacen ejercicio.
- Entonces la edad es un factor común del ejercicio y los niveles de colesterol.

Paradoja de Simpson: Desagregar los datos no es siempre la solución

- Suponga que en el primer ejemplo no se registra el sexo. Se registra la presión arterial después del experimento.
- Suponga que la droga afecta la recuperación reduciendo la presión pero también tiene un efecto tóxico. En cada subgrupo el efecto tóximo predomina.
- Los resultados del experimentos son como a continuación (esta gráfica es idéntica a la anteriores excepto que se intercambiaron los nombres de las columnas).

Paradoja de Simpson: Ejemplo

- Los resultados a nivel agregado sugieren que se debe usar la droga, pero no si se analizan los datos condicional a la presión arterial.
- Si la presión arterial se midiera antes del experimento y esta tuviera un efecto sobre las personas que acceden al tratamiento y no al contrario, encontrariamos lo contrario,

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No drug	Drug
Low BP	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High BP	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

- En este caso:

$$E(Y_t) = E(Y_t \mid S = t), E(Y_c) = E(Y_c \mid S = c) \qquad (5)$$

porque, por lo menos la otra variable obsevada la presion, no explica ser tratado (esta se mide después del tratamiento).

Paradoja de Simpson: Conclusiones

- En conclusion: En algunos casos la respuesta correcta puede estar en los datos desagregados, como en el primer ejemplo y en otros casos en la informacion agregada como en el segundo ejemplo.
- En la paradoja de Simpson, el problema está en asumir que la participación en el tratamiento y el resultado no son dependientes. Existe un factor común, el sexo, que induce una autoselección de los tratados.
- El análisis estadístico solo no permite responder estas preguntas, es necesario hipótesis adicionales y una teoría de como se generan los datos.

Correlación y Causalidad

- En los ejemplos anteriores hemos puesto a prueba nuestra intuición sobre la hipótesis de independencia. Sin embargo, la causalidad es un fenómeno aún más especial.
- Considere la correlación que eixste entre:
  1. Crimen y consumo de helados.
  2. Ocupación hotelera y precios.
  3. Incendios y número de bomberos.
  4. Personas que andan apuradas y llegada tarde a reuniones, etc. content...
- En las próximas secciones vamos a desarrollar una teoría de la dependencia y causalidad entre variables.

Contenido

- La paradoja del gato (Monty Hall).
- Esta paradoja tiene origen en que en ocasiones **nuestras creencias solo dependen de los datos observados y deconocen la forma como se generan esos datos**.
- Cuando el maestro de ceremonias abre la puerta donde no está el gato esto no es por si solo informativo hasta tanto no se cuestione el procesos de decisión que siguió este para abrir la puerta.

La paradoja del gato (Monty Hall) Formalmente

- Supongamos que la primera elección fue la tercera puerta.
- Sean A_1, A_2 y A_3 los eventos en los cuales el gato está detrás de la puerta $1, 2$ o 3 respectivamente.
- Sean B_1 y B_2 los eventos en los cuales el segundo jugador abre la puerta 1 o 2 respectivamente.
- Nuetro objetivo es calcular $P\left(A_i \mid B_j\right)$. Entonces dada la información del problema es natural suponer:

$$P(A_i) = \frac{1}{3}, P\left(B_1|A_1\right) = P\left(A_2|B_2\right) = 0$$

$$P(B_1|A_2) = P(B_2|A_1) = 1$$

y

$$P(B_1|A_3) = P(B_2|A_3) = \frac{1}{2}.$$

- Ahora, si la segunda persona abre la puerta 2 usando la regla de Bayes obtenemos $P\left(A_1|B_2\right) = \frac{2}{3}$

Contenido

Modelos Causales Estructurales: SCM

- Un modelo causal consiste de un conjunto de variables aleatorias U y V y un conjunto de funciones F que establecen una relación entre las variables de U y V.

Example (Salario, educación y experiencia)

Sea $U = \{X, Y\}, V = \{Z\}$ y $F = \{f_Z\}$, tal que f_Z es la distribución de $Z := 2X + 3Y$. El modelo representa la relación de causalidad entre X, años de escolaridad, Y experiencia profesional y Z salario.

Modelos Gráficos

- Grafos: Pariente, hijo, descendientes.
- Todo SCM define se puede representar como un modelo gráfico (grafos dirigidos acíclicos).

Example (Salario, educacion y experiencia)

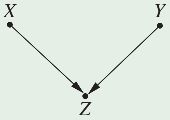

Figure 1.9 The graphical model of SCM 1.5.1, with X indicating years of schooling, Y indicating years of employment, and Z indicating salary

Modelos Gráficos: Regla de descomposición del producto

- Cuando un grafo es acíclico sobre variables aleatorias $\{X_1, ..., X_n\}$ la distribución de probabiliad conjunta de $(X_1, ..., X_n)$ se puede descomponer como:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i \mid pa_i)$$

donde pa_i son los parientes de X_i.

Contenido

Ejemplos SCM

Example (Financiación de la Escuela, SAT Promedio, Tasa Aceptación Universidad)

Sea $V = \{X, Y, Z\}$,$U = U_x, U_y, U_z$ y $F = \{f_x, f_y, f_z\}$:

$$X := U_X$$
$$Y := \frac{X}{3} + U_Y$$
$$z := \frac{Y}{16} + U_Z$$

Example (Horas de Trabajo, Entrenamiento, Tiempo de Carreras)

Sea $V = \{X, Y, Z\}$,$U = U_x, U_y, U_z$ y $F = \{f_x, f_y, f_z\}$ y:

$$X = U_X$$
$$Y = 84 - X + U_Y$$
$$Z = \frac{100}{Y} + U_Z$$

Modelos Gráficos Asociados: Cadenas

- Los dos ejemplos anteriores se pueden representar gáficamente como:

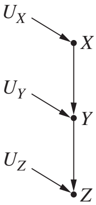

Figure 2.1 The graphical model of SCMs 2.2.1–2.2.3

- Este es un ejemplo de una **cadena**.
- Decimos que Z es potencialmente dependiente de Y, Y es potencialmente dependiente de X, Z es potencialmente dependiente de X; y Z e independiente de X condicional en Y.

Definition

Dos variables X y Z son condicionalmente independientes dado Y si existe solo un camino unidireccional entre X y Z, y Y es cualquier conjunto de variables que intercepta este camino.

Ejemplo SCM

Example (Temperatura, Helados y Crimen)

Sea $V = \{X, Y, Z\}$,$U = U_x, U_y, U_z$ y $F = \{f_x, f_y, f_z\}$ y:

$$\begin{aligned} x &:= U_x \\ y &:= 4x + U_y \\ z &:= \frac{x}{10} + U_z \end{aligned}$$

Modelos Gráficos Asociados: Bifurcaciones

- El ejemplo anterior se pueden representar gáficamente como:

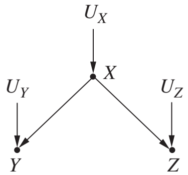

Figure 2.2 The graphical model of SCMs 2.2.5 and 2.2.6

- Este es un ejemplo de una **bifurcación**.
- Decimos que Z es potencialmente dependiente de X, Y es potencialmente dependiente de X, Z es potencialmente dependiente de Y, Z es independiente de Y condicional en X.

Modelos Gráficos: Independencia Condicional en Bifurcaciones

Definition

Si una variable X es una causa común de Y y Z y solo hay un camino entre Y y Z entonces Y y Z son independientes condicional a X.

Modelos Gráficos Asociados: Colisionadores

- Sea $Z = X + Y$ donde X y Y son independientes.
- Por ejemplo X, Y pueden ser el resultado de lanzar aleaoriamente una moneda al aire. Si por lo menos una cae cara Z se activa, suena una campana.
- Sea X la habilidad musical de un individuo y Y su desepeño académico. X, Y son posiblemente independientes. Ahora sea Z si la persona tiene una beca de estudio o no.

Modelos Gráficos Asociados: Colisionadores

- Lo ejemplos anteriores se pueden representar gáficamente como:

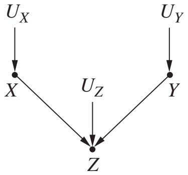

Figure 2.3 A simple collider

- X y Z, y Y y Z son dependientes. X y Y son independientes pero X y Y dado Z son dependientes.

Modelos Gráficos: Independencia Condicional en Colisionadores

Definition

Si una variable Z es una variable de colisión entre dos variables X y Y, y existe solo un camino entre X y Y, entonces X y Y son incondicionalmente independientes, pero son dependientes condionalmente en Z y cualquiera de sus descendietes.

Modelos Gráficos: Independencia Condicional en Colisionadores

- En el ejemplo de la paradoja del gato (i.e., Monty Hall problem) el nodo X es la puerta que el agente elige, el nodo Y es donde esta el gato y Z es la puerta que abre el maestro de ceremonia.
- Si X es distinto a Y, Z queda completamente determinado.
- Condicional a la elección de maestro (Z) la eleccón de la puerta (X) y donde esta el gato (Y) son dependientes.

Modelos Gráficos: Independencia Condicional en Colisionadores

Example (Z = X+Y)

X, Y Bernoulli independientes.

Table 2.2 Conditional probability distributions for the distribution in Table 2.2. (Top: Distribution conditional on $Z = 1$. Bottom: Distribution conditional on $Z = 0$)

X	Y	$P(X, Y \mid Z = 1)$
Heads	Heads	0.333
Heads	Tails	0.333
Tails	Heads	0.333
Tails	Tails	0

X	Y	$Pr(X, Y \mid Z = 0)$
Heads	Heads	0
Heads	Tails	0
Tails	Heads	0
Tails	Tails	1

Modelos Gráficos: Independencia Condicional en Colisionadores

- La condición de dependencia condicional a un colisionador es importante para determinar si un modelo causal generó un conjunto de datos, para descubrir el modelo a partir de los datos y para resolver la Paradoja de Simpson.
- Esta propiedad deja claro una nueva forma en la que dos variables pueden ser dependientes:
  1. Una variable causa la otra.
  2. Una tercera variable causa ambas.
  3. Dependencia condicional a Z. Obérvese que en esta última X y Y son independientes.

Modelos Gráficos: d-separación (separación direccional)

- El concepto de d-separación sirve como criterio para explorar las dependencias en cualquier modelo gráfico.
- Dos nodos que están d-separados son independientes. Si no están d-separados entonces son potencialmente dependientes.

Definition (d-separación)

Decimos que un camino p es bloqueado por un conjunto de nodos Z si se cumple alguna de estas:

1. p contiene un cadena $A \rightarrow B \rightarrow C$ donde $B \in Z$.
2. p contiene un bifurcación $A \leftarrow B \rightarrow C$ donde $B \in Z$.
3. p contiene un colisionador $A \rightarrow B \leftarrow C$ donde $B \notin Z$ y ningún descendiente de B está en Z.

Si Z bloquea todos los caminos de X a Y entonces X y Y están d-separados condicional en Z y por lo tanto X y Y son independientes condicional en Z.

Example (Z,Y d-separados condicional al conjunto vacio)

Solo hay un camino entre Z y Y y ese camino lo bloquea un colisionador: $Z \rightarrow W \leftarrow X$. Luego Z y Y son condicionalmente independientes (i.e., independientes condional al conjunto de nodos vacio).

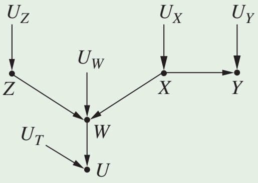

Figure 2.7 A graphical model containing a collider with child and a fork

Example (Z,Y d-conectados condicionales a W)

Solo hay un camino entre Z y Y. Si se condiciona a W queda una bifurcación X que no está en el conjunto W y el único colisionador sí está luego ese camino no está bloqueado.

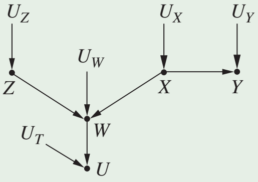

Figure 2.7 A graphical model containing a collider with child and a fork

Modelos Gráficos: d-separación

Example (Z,Y d-separados condicional a {W,X})

$\{W, X\}$ bloquea e único camino entre Z y Y.

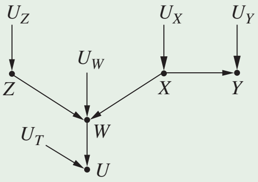

Figure 2.7 A graphical model containing a collider with child and a fork

Example (Z,Y incondicionalmente dependientes)

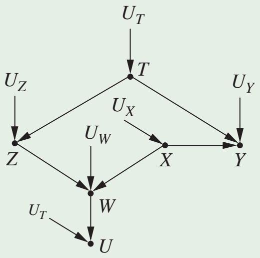

Figure 2.8 The model from Figure 2.7 with an additional forked path between Z and Y

Modelos Gráficos: d-separación

Example (Z,Y condicional a T independientes)

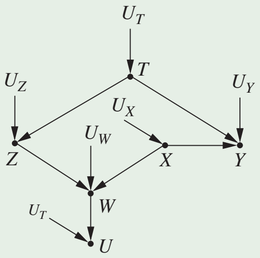

Figure 2.8 The model from Figure 2.7 with an additional forked path between Z and Y

Modelos Gráficos: d-separación

Example (Z,Y condicional a T, W d-conectados)

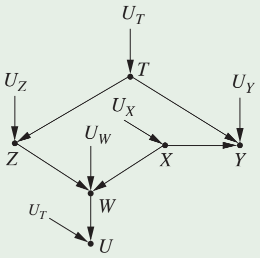

Figure 2.8 The model from Figure 2.7 with an additional forked path between Z and Y

Modelos Gráficos: d-separación

Example (Z,Y condicional a T, W, X independientes)

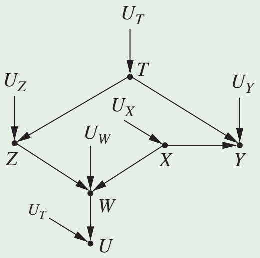

Figure 2.8 The model from Figure 2.7 with an additional forked path between Z and Y

Modelos Gráficos: d-separación

Example (Z,Y Resumen de Dependencias)

Z, Y son potencialmente dependientes condional a W, U,$\{W, U\}$,$\{W, T\}$,$\{U, T\}$,$\{W, U, T\}$,$\{W, X\}$,$\{U, X\}$ y $\{W, U, X\}$. Son d-separados o condicionalemnte independientes en T, $\{X, T\}$, $\{W, X, T\}$, $\{U, X, T\}$ y $\{W, U, X, T\}$.

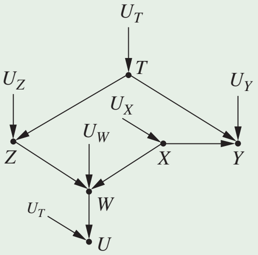

Figure 2.8 The model from Figure 2.7 with an additional forked path between *Z* and *Y*

Pruebas de Especificiacón de Modelos e Identificación Causalidad

- Supongamos que conjeturamos que el modelo porbabilístico que generó los datos es:

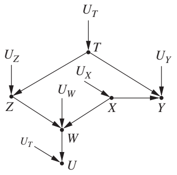

Figure 2.8 The model from Figure 2.7 with an additional forked path between Z and Y

- W, Z son indepenpiedntes condicional a Z_1. Esto se puede probar mediante una regresión: $w = r_X x + r_1 z_1$. Si r_1 es estadísticamente diferente a cero, entonces la hipótesis se rechaza, y el modelo gráfico no lo validan los datos (prueba local).