

- 1 Los mecanismos son los caminos a través de los cuales X causa la variable de resultado Y
- 2 Aunque no *necesitamos* conocer el mecanismo para concluir que X causa Y, hay varias razones por las que *queremos* conocerlo
- 3 Pero es muy difícil identificar los mecanismos causales porque los propios mecanismos no se asignan al azar...
- 4 ...y porque los efectos del tratamiento rara vez son homogéneos
- 5 Muchos estudios intentan descomponer el efecto total del tratamiento en sus efectos “directos” e “indirectos”
- 6 Pero tenga cuidado al utilizar el análisis de regresión para descomponer efectos
- 7 A veces, el análisis de subgrupos puede proporcionar pruebas a favor o en contra de un mecanismo
- 8 También podemos buscar pruebas sugerentes al observar los efectos de nuestro tratamiento en ciertas variables de resultado
- 9 El diseño de tratamientos complejos puede ayudarnos a comprender mejor qué parte del tratamiento está “haciendo el trabajo”
- 10 A pesar de las dificultades para medir empíricamente los mecanismos, vale la pena prestarles mucha atención, pero también siendo cautos en nuestro lenguaje

A los científicos sociales nos encantan las preguntas causales. Tan pronto nos enteramos de que X causa Y, queremos entender *por qué* X causa Y. Esta guía explora el rol de los “mecanismos” en el análisis causal, y además le ayudará a entender qué tipo de conclusiones puede sacar sobre estos.

1 Los mecanismos son los caminos a través de los cuales X causa la variable de resultado Y

Desde hace ya mucho tiempo, los mecanismos han sido parte principal de la medicina. Cada vez que un médico prescribe un tratamiento, lo hace a partir de la comprensión de los factores químicos o físicos que causan una enfermedad, y prescribe un tratamiento que es eficaz porque interrumpe estos factores. Por ejemplo, muchos psicólogos clínicos recomiendan hacer ejercicio a los pacientes que sufren depresión. El ejercicio aumenta las endorfinas en la química del cuerpo, lo cual desencadena sensaciones positivas y actúan como analgésicos, que reducen el dolor. Las endorfinas, por lo tanto, son un mecanismo por el cual el ejercicio ayuda a reducir la depresión. El ejercicio puede tener efectos positivos sobre otras variables dependientes (por ejemplo, las enfermedades cardíacas) a través de otros mecanismos (por ejemplo, elevando la frecuencia cardíaca), pero el mecanismo que hace que afecte a la depresión en particular son las endorfinas. También podríamos concluir que otro tratamiento, como un fármaco que eleve las endorfinas, pueda tener efectos similares sobre la depresión.

Los mecanismos son igual de importantes para las ciencias sociales. Consideremos como ejemplo las investigaciones recientes que han relacionado el cambio climático con un aumento de los conflictos civiles. Un estudio¹ busca identificar el efecto causal de las perturbaciones climáticas en los conflictos violentos al estudiar la tasa de conflictos civiles en los países afectados por El Niño; comparando los años en que estos países sufrieron El Niño frente a los años en los que no los sufrieron. Supongamos que este estudio es correcto. ¿Por qué una perturbación climática puede provocar un aumento de los conflictos en un país? Un mecanismo podría ser la pobreza: las perturbaciones climáticas perjudican a la economía y, al reducirse los costos de oportunidad, los individuos son más propensos a unirse a grupos armados. Un mecanismo alternativo es el fisiológico: las personas están físicamente predispuestas a ser más agresivas con

temperaturas más altas. Tal vez el mecanismo sea la migración: las perturbaciones climáticas desplazan a las personas que habitan regiones costeras, lo que produce un conflicto social entre los migrantes y los nativos. En realidad todos estos mecanismos (así como otros no enumerados aquí) podrían estar operando simultáneamente ¡Incluso en el mismo caso! En muchas de las preguntas más interesantes de las ciencias sociales, hay varios canales (“M”) que podrían transmitir el efecto total de X sobre Y.

2 Aunque no *necesitamos* conocer el mecanismo para concluir que X causa Y, hay varias razones por las que *queremos* conocerlo

En el ejemplo anterior sobre clima y conflicto, podemos confiar plenamente en la capacidad de los investigadores para identificar de forma causal que las perturbaciones climáticas provocan conflictos, y sin embargo no tenemos evidencia de cuáles mecanismo(s) está(n) operando. Pero a los científicos sociales les interesa conocer dichos mecanismos porque están estrechamente relacionados con las teorías de las ciencias sociales. Por ejemplo, el mecanismo de la “pobreza” mencionado anteriormente está estrechamente relacionado con la teoría de Gurr² de que los individuos se rebelan cuando sus costos de oportunidad del conflicto son bajos, mientras que el mecanismo de la “migración” podría apoyar una teoría del conflicto basada en los descontentos entre grupos sociales. No es de extrañar que al enterarse de que X causa Y, los científicos sociales se pregunten inmediatamente cuál es el mecanismo: ¡quieren relacionar este hallazgo con la teoría!

Entender los mecanismos tiene ventajas tanto teóricas como prácticas. En primer lugar, conocer M nos permite adivinar para qué poblaciones X conducirá a Y. Si el mecanismo para el clima/conflicto es la respuesta fisiológica al calor, entonces las perturbaciones climáticas podrían causar conflictos sólo cuando la temperatura es bastante cálida. En segundo lugar, conocer M nos ayuda a considerar otras variables de resultado que pueden verse afectadas por X. Si el mecanismo del clima/conflicto es la migración, entonces también podríamos esperar que las perturbaciones climáticas provoquen un uso excesivo de los bienes públicos en zonas urbanas. En tercer lugar, conocer M nos ayuda a considerar otras formas de causar o evitar cambios en Y. Si el mecanismo para el clima/conflicto es la pobreza, entonces los programas de desarrollo podrían disminuir el conflicto al reducir la sensibilidad de ingresos a las perturbaciones climáticas, aunque no puedan cambiar dichas perturbaciones.

3 Pero es muy difícil identificar los mecanismos causales porque los propios mecanismos no se asignan al azar...

Consideremos un el siguiente experimento. Chong et al. (2015)³ utilizaron un experimento de campo para estudiar el efecto de la información sobre corrupción en la participación electoral. Para eso, asignaron aleatoriamente algunos distritos electorales en México para que recibieran información sobre el uso corrupto de fondos dentro de ese municipio. Sorprendentemente, descubrieron que los distritos tratados

acudieron a votar en menor proporción que los distritos de control. Sugieren el siguiente mecanismo: la información sobre corrupción convence a los votantes de que el municipio es tan corrupto, que elegir un político honesto no generará ningún cambio, lo que hace que las personas creen que su voto vale menos.

Básicamente su argumento es:⁴

Recibir información sobre corrupción (X) $\xrightarrow{+}$ Cree que la corrupción es demasiado grave (M) $\xrightarrow{+}$ Quedarse en casa (Y)

Chong et al. se enfrentan a un obstáculo común a la hora de interpretar sus resultados: su mecanismo propuesto no se asignó al azar. Algunas personas están más inclinadas a creer que “todos los políticos son unos sinvergüenzas”, mientras que otras tienen tendencia a impulsar “un cambio en el que podamos creer”. Desafortunadamente sólo podemos observar el tratamiento aleatorio que recibió un individuo y su convicción no-aleatoria sobre la corrupción; no podemos saber qué convicciones sobre la corrupción *habría tenido* si hubiera recibido la otra condición de tratamiento. Esto nos impide determinar hasta qué punto su decisión de acudir a votar fue causada por el mecanismo propuesto versus otros mecanismos.

Algunos investigadores intentan resolver este problema al estimar el efecto *promedio* del tratamiento sobre el mecanismo, para luego estimar el efecto *promedio* del mecanismo sobre la variable de resultado. Una de las razones por las que esto es problemático es que podemos imaginar varios factores distintos al tratamiento que podrían estar causando tanto M como Y. Supongamos que el nivel de apatía -llamémoslo Q- varía entre los ciudadanos de nuestro estudio, y que Q tiene un efecto muy fuerte tanto en M como en Y. Los individuos muy apáticos podrían ser más propensos a creer que los problemas no tienen solución, y también podrían ser más propensos a quedarse en casa el día de las elecciones. Por lo tanto, es probable que observemos una fuerte correlación entre M y Y, impulsada por el factor de confusión de Q, no por nuestro tratamiento X. Mecánicamente, nuestros resultados estarán sesgados a favor de encontrar pruebas del efecto de X sobre Y a través de M, simplemente porque Q ha producido una relación entre M y Y.

4 ...y porque los efectos del tratamiento rara vez son homogéneos

Otro problema al intentar descomponer los efectos promedios de X en M y luego de M en Y es que este enfoque supone que cada sujeto responde al tratamiento de manera idéntica. Si recordamos nuestro ejemplo en el que X es el tratamiento de información, M es la convicción de que la corrupción es demasiado grave, y Y es quedarse en casa, podemos imaginar dos tipos diferentes de encuestados. El Tipo A pensaba que la corrupción era demasiado grave como para poder solucionarla, hasta que recibió una tarjeta postal con información sobre la corrupción en su distrito. Se sorprendió al ver que el problema no era tan grave como imaginaba. Formalmente, para el Tipo A, $M(X = 0) = 1$ y $M(X = 1) = 0$, por lo que X tiene un efecto *negativo* sobre M. El Tipo B pensaba que la corrupción era un problema manejable, hasta que recibió una información por correo sobre la corrupción en su distrito. Se sorprendió tanto de la magnitud del problema, que perdió la esperanza de resolverlo. Formalmente, para el Tipo B, $M(X = 0) = 0$ y $M(X = 1) = 1$, por lo que X tiene un efecto *positivo* sobre M. Si promediáramos los efectos de estos dos tipos, no veríamos ninguna relación entre X y M.

Type	X (Tratamiento de información)	Condicional M en X=0 (sin observar)	Condicional M en X=1 (observado)	Efecto de X en M	Efecto de M en Y	Y (Quedarse en casa)
A	1	1	0	negativo	negativo	1
B	1	0	1	positivo	positivo	1

Estimar el rol de M puede complicarse aún más cuando la relación entre M y Y es también heterogénea. Imaginemos que el Tipo A sólo vota cuando está enfadado (en otras palabras, M tiene un efecto *negativo* sobre Y). El Tipo A planeaba votar para expresar su enfado por la presencia prevalente de la corrupción en su distrito, aunque sabía que no cambiaría nada, hasta que se enteró de que la corrupción no era tan mala como esperaba. Al desaparecer su ferviente convicción, decide quedarse en casa el día de las elecciones. Sin embargo, el Tipo B sólo vota cuando cree que su voto puede marcar la diferencia (en otras palabras, M tiene un efecto *positivo* sobre Y). El Tipo B iba a votar por los políticos no corruptos de su distrito, hasta que se enteró de que todos eran corruptos. Sin ninguna esperanza de cambiar la situación, también decidió quedarse en casa el día de las elecciones. Tanto para el Tipo A como para el Tipo B, existe un “efecto indirecto” de M (en otras palabras, X afecta a Y a través de M). Pero esta relación se nos escapará en el agregado, porque no podremos obtener estimaciones sin sesgo del efecto promedio de X sobre M.⁵

Podemos imaginar muchos más “tipos” distintos aparte de A y B - el punto aquí es demostrar de manera intuitiva que debido a que M no se asigna al azar (y porque es poco probable que los efectos de X en M y M en Y sean idénticos para todos), será muy difícil caracterizar con precisión cuánto de nuestro efecto está mediado a través de M.

5 Muchos estudios intentan descomponer el efecto total del tratamiento en sus efectos “directos” e “indirectos”

Dado que aprender sobre mecanismos es tan prometedor desde el punto de vista teórico, a los investigadores les encantaría cuantificar qué parte del efecto de X sobre Y opera a través de M. A veces los investigadores intentan hacer esto mediante una técnica llamada “descomposición de efectos”.

Un análisis de descomposición de efectos trata de descomponer un efecto *total* de X sobre Y en el efecto que X tiene sobre Y *directamente*, y el efecto de X sobre Y que se produce *indirectamente* a través de M. El “efecto total” se refiere al Efecto Promedio del Tratamiento (ATE, por sus siglas en inglés), que es simplemente el efecto promedio que X tiene sobre Y. Cualquier experimento que asigne de manera aleatoria un tratamiento para observar sus efectos sobre alguna variable de resultado, está estimando el ATE. A continuación el investigador trata de cuantificar el tamaño del efecto que X tiene sobre Y a través del mecanismo M. Esto se conoce como el “efecto indirecto” -porque X está afectando a Y indirectamente a través de M- o también se conoce como el Efecto Causal Mediado Promedio (ACME, por sus siglas en inglés). Por último, el investigador tratará de estimar el efecto de X sobre Y que no pasa por M. Esto se conoce como el “efecto directo” de X sobre Y o como el Efecto Directo Controlado Promedio (ACDE, por sus siglas en inglés), porque es el efecto de X sobre Y cuando controlamos el trabajo que está haciendo M.

6 Pero tenga cuidado al utilizar el análisis de regresión para descomponer efectos

Aunque es utilizado comúnmente, el uso del análisis de regresión de mediación asume algunas suposiciones fuertes y poco realistas. Utilizaremos un código para exponer lo que implica este método y demostrar las condiciones en las que puede producir estimaciones sesgadas.

La idea básica es que si tenemos datos sobre el tratamiento que recibió un individuo (X), si se presenta el mecanismo propuesto (M), y cuál es la variable de resultado (Y), entonces podemos diferenciar estos efectos utilizando las siguientes tres regresiones.

1.
$$M_i = \alpha_1 + aX_i + e_{1i}$$
2.
$$Y_i = \alpha_2 + cX_i + e_{2i}$$
3.
$$Y_i = \alpha_3 + dX_i + bM_i + e_{3i}$$

¿Cómo podríamos hacer esto? Al utilizar la ecuación 1, podemos hacer una regresión de M sobre X para obtener el efecto directo de X sobre M, que es el coeficiente a . Luego pasamos a la ecuación 3, en la que hacemos regresión de Y sobre M y X. En esta regresión, el coeficiente b representa el efecto directo de M sobre Y cuando controlamos por X. Un análisis de descomposición de efectos multiplicaría $a * b$ para revelar el efecto indirecto de X sobre Y a través de M. Para encontrar el efecto directo de X sobre Y podemos hallarlo en d , que es el coeficiente de X en la ecuación 3 cuando controlamos por M. En otras palabras, d es el efecto de X sobre Y que no pasa por M. Si sumamos el efecto indirecto y el efecto directo, obtendremos el “efecto total” de X sobre Y, que es igual a c . En resumen, el análisis de descomposición de efectos desagrega el efecto total en el efecto que está mediado a través de M y en el efecto que no está mediado a través de M, lo que permite al investigador concluir cuan importante es M para explicar la relación entre X y Y.⁶

El problema es que este cálculo sólo funciona bajo ciertos supuestos muy fuertes. Uno de estos supuestos es que los términos de error en las regresiones 1 y 3 no están relacionados entre sí, es decir, M no puede ser predicho por los factores no observables que también predicen Y. Describimos este problema de forma intuitiva en el punto 3 cuando introducimos Q, una variable de confusión que contribuye tanto a M como a Y, que por lo tanto genera una relación muy fuerte entre ellas, incluso si el efecto de X sobre Y no está operando a través de M. Ahora describiremos este problema mediante una simulación.

En el siguiente código, comenzamos creando la variable Q para cada individuo, y definimos los efectos “verdaderos” de X sobre M, M sobre Y y X sobre Y. A continuación creamos variables de resultados potenciales e hipotéticas para M. Es decir, para cada individuo, definimos qué valor de M revelarían si fueran tratados, y qué valor de M revelarían si no fueran tratados. Estos valores están relacionados no sólo con el “verdadero” efecto de X sobre M, sino también con Q. Así mismo podemos definir variables de resultados potenciales e hipotéticas para Y. Hacemos esto para cuatro escenarios, todos los cuales suponen efectos constantes, una suposición que relajaremos más adelante. Dos de ellos son variables de resultado potenciales simples de Y: el Y del individuo cuando no es tratado y revela su variable de resultado potencial M sin tratar, y el Y del individuo cuando es tratado y revela su variable de resultado potencial M, que ha sido tratado. Sin embargo, también definimos dos variables de resultado potenciales complejas de Y: el Y del individuo cuando no es tratado pero revela su variable de resultado potencial M, que ha sido tratado, y el Y del individuo que es tratado pero revela sus potenciales variables de resultado M, que no han sido

tratados. Aunque estas variables de resultado potenciales nos puedan confundir un poco, es importante definirlos en el sentido hipotético para poder calcular los efectos directos e indirectos “verdaderos” (pero intrínsecamente inobservables) con los que podamos comparar nuestro análisis de descomposición.

En la segunda mitad del código, realizamos una asignación aleatoria del tratamiento y procedemos al análisis de descomposición de efectos descrito anteriormente, utilizando los datos que “observamos”. Bajo los (fuertes) supuestos de que los términos de error no están correlacionados y los efectos son constantes entre los sujetos; $a * b = ACME$, $d = ACDE$ y $c = ATE$. Sin embargo, la simulación revela que $a * b > ACME$ y $d < ACDE$; es decir, sobrestimamos el efecto promedio indirecto o mediado (ACME) y subestimamos el efecto promedio medio directo (ACDE). Nuestro análisis de descomposición de efectos estaba sesgado porque el primer supuesto -términos de error no correlacionados- no se cumplía: la variable no observada Q predecía tanto M como Y , y esto nos llevó a sobreestimar el papel del mecanismo M .

Hide

```
rm(list = ls())

set.seed(20160301)

N <- 1000000

# Simular Datos, Crear Variables de Resultado Potenciales, Estimar Efectos "Verdaderos" -----

# construir una característica idiosincrática sin observar
Q_i <- rnorm(N)

# crear un "modelo verdadero" al definir nuestros efectos de tratamiento (tau)
tau_X_on_M <- 0.2 # efecto de X sobre M
tau_M_on_Y <- 0.1 # efecto de M sobre Y
tau_X_on_Y <- 0.5 # efecto total de X sobre Y (ATE), ambos a través de M y sin atravesar M

# construir las variables de resultado potenciales (POs) para el mediador
# individuo muestra M_1 si es tratado; M_0 si no es tratado
# M es una función del tratamiento y la característica sin observar
M_0 <- 0 * tau_X_on_M + Q_i
M_1 <- 1 * tau_X_on_M + Q_i

# podemos estimar el Efecto Promedio del Tratamiento sin sesgo (ATE) de X sobre M
ATE_M <- mean(M_1 - M_0)
ATE_M
```

```
[1] 0.2
```

Hide

```

# construir POs para la variable de resultado
Y_M0_X0 <- tau_M_on_Y * (M_0) + tau_X_on_Y * 0 + Q_i
Y_M1_X1 <- tau_M_on_Y * (M_1) + tau_X_on_Y * 1 + Q_i
Y_M0_X1 <- tau_M_on_Y * (M_0) + tau_X_on_Y * 1 + Q_i # esto es un PO "complejo"
Y_M1_X0 <- tau_M_on_Y * (M_1) + tau_X_on_Y * 0 + Q_i # esto es un PO "complejo"
# algunos POs son "complejos" porque imaginamos que Y nosotros
# observaríamos si asignáramos el tratamiento,
# pero observamos el resultado potencial de M sin tratar o
# el valor de Y si asignamos control pero observamos
# el PO de M tratada
# Para construir esto POs complejos es necesario para estimar los efectos "verdaderos, directos e indirectos

# podemos estimar el Efecto Causal Mediado Promedio (ACME) sin sesgo
# estimamos los efectos de M manteniendo la constante X
# son lo mismo
# este es el "efecto indirecto"
ACME_X0 <- mean(Y_M1_X0 - Y_M0_X0)
ACME_X1 <- mean(Y_M1_X1 - Y_M0_X1)
ACME <- mean(((Y_M1_X1 - Y_M0_X1) + (Y_M1_X0 - Y_M0_X0)) / 2)

# podemos estimar el Efecto Directo Controlado Promedio (ACDE) sin sesgo
# estimamos los efectos de X manteniendo la constante M
# son lo mismo
# este es el "efecto directo"
ACDE_M0 <- mean(Y_M0_X1 - Y_M0_X0)
ACDE_M1 <- mean(Y_M1_X1 - Y_M1_X0)
ACDE <- mean(((Y_M0_X1 - Y_M0_X0) + (Y_M1_X1 - Y_M1_X0)) / 2)

# ahora construimos las PO simples para Y
Y_1 <- tau_M_on_Y * (M_1) + tau_X_on_Y * 1 + Q_i
Y_0 <- tau_M_on_Y * (M_0) + tau_X_on_Y * 0 + Q_i

# estimamos la ATE verdadera de X sobre Y
# este es el "efecto total"
ATE <- mean(Y_1 - Y_0)

ATE

```

```
[1] 0.52
```

[Hide](#)

```
ACDE + ACME # notese que los efectos directos e indirectos suman el total
```

```
[1] 0.52
```

Hide

ACDE

[1] 0.5

Hide

ACME

[1] 0.02

Hide

ATE_M

[1] 0.2

Hide

```
# Asignación Aleatoria, Revelación de POs, Intento de Descomponer Efectos -----
# asignamos mitad de nuestra muestra al tratamiento y la otra mitad a control
X <- sample(c(rep(1, (N / 2)), rep(0, (N / 2))))
# revelamos las POs para M y Y basados en la asignación de tratamiento
M <- X * M_1 + (1 - X) * M_0
Y <- X * Y_1 + (1 - X) * Y_0

modell <- lm(M ~ X)
a <- coef(modell)[2] # extraer el coeficiente para encontrar el efecto de X sobre
M
a
```

X
0.2001291

Hide

```
model2 <- lm(Y ~ X)
c <- coef(model2)[2] # extraer el coeficiente para encontrar el efecto total de X
sobre Y
c
```


X
0.520142

Hide

```
model3 <- lm(Y ~ X + M)
d <- coef(model3)[2] # extraer este coeficiente para encontrar el efecto de X sobre Y controlando M
b <- coef(model3)[3] # extraer este coeficiente para encontrar el efecto de M sobre Y controlando X

# se multiplicaría el efecto promedio de X sobre M y el efecto promedio de M sobre Y para hallar el efecto promedio indirecto/mediado de X sobre Y a través de M (ACME)
a * b
```

X
0.220142

Hide

pero cuando comparamos esto con el ACME verdadero, veremos que es sesgado
ACME

[1] 0.02

Hide

algunos también interpretarían el efecto promedio de X sobre Y controlando para M, como el efecto directo promedio controlado (ACDE)
d

X
0.3

Hide

pero cuando comparamos esto con el ACDE verdadero, veremos que es sesgado
ACDE

[1] 0.5

Hide

```
# nótese que hemos SOBReestimado el efecto indirecto promedio y hemos SUBestimado
el efecto directo promedio

# los estimados que no son sesgados son los efectos promedio de X sobre Y y X sobre M
porque X es asignado aleatoriamente
```

```
      X
0.2001291
```

Hide

```
ATE_M
```

```
[1] 0.2
```

Hide

```
c
```

```
      X
0.520142
```

Hide

```
ATE
```

```
[1] 0.52
```

Volvamos a relacionar este ejercicio con el problema planteado en el punto 3. Esta simulación demostró que la cuantificación del efecto mediado resulta difícil cuando las variables predictivas de fondo confunden la relación entre M y Y. Dado que M no se asigna de manera aleatoria, es importante que pensemos en la probabilidad de que tanto M como Y se vean afectados por variables no observadas. En principio, si no hay variables de confusión en esta relación, entonces el análisis de descomposición de efectos puede no tener sesgo, pero esta suposición es fuerte y difícil de probar.

Aunque no fue verificado en esta simulación, también es posible mostrar que la descomposición de efectos no funciona bien cuando los efectos del tratamiento son heterogéneos (introducimos la intuición de esto en el punto 4). La razón técnica para ello proviene de nuestra ley de expectativas, que es: $E[a * b] = E[a]E[b] + cov(a, b)$. Si tenemos efectos de tratamiento constantes, entonces a y b no covarían, el término de covarianza desaparece, y podemos simplemente multiplicar $a * b$ para obtener el ACME. Sin embargo, si el término de covarianza es distinto de cero, entonces no podemos estimar este efecto indirecto a partir de estos dos coeficientes que han sido obtenidos de regresiones separadas. Construimos efectos de tratamiento constantes para poder demostrar el proceso de descomposición de efectos, pero si volviéramos a hacer la

simulación con efectos de tratamiento heterogéneos que covarían, entonces ni siquiera podríamos calcular el ACME o el ACDE utilizando el enfoque de potenciales variables de resultado al principio de nuestro código.

Lo que puede hacer... Antes de embarcarse en un análisis de descomposición de efectos, pregúntese:

- ¿Puedo imaginar alguna variable no observada que prediga tanto M como Y?
- ¿Es posible que mis sujetos respondan a los efectos del tratamiento de maneras diferentes?

Si la respuesta a cualquiera de estas preguntas es afirmativa, le recomendamos que proceda con precaución. Piense cuidadosamente en cómo las variables no observadas y los efectos heterogéneos del tratamiento afectarían a su estrategia de estimación.

7 A veces, el análisis de subgrupos puede proporcionar pruebas a favor o en contra de un mecanismo

En los puntos 3 a 6, hemos advertido a los investigadores de que no intenten cuantificar con certeza la proporción de un efecto que está mediado por un mecanismo particular, pero puede haber otras formas de aprender más sobre los mecanismos en funcionamiento de un estudio concreto. En el punto 1 subrayamos la estrecha relación entre los mecanismos y la teoría. El hecho de que sea un reto cuantificar directamente las evidencias de un mecanismo, no significa que no podamos explorar las predicciones comprobables de la teoría en la que aparece nuestro mecanismo.

Una estrategia consiste en utilizar el análisis de subgrupos, o las interacciones entre tratamientos y covariables, para ver si las distintas poblaciones responden al tratamiento de forma diferente según nuestras teorías. Por ejemplo, supongamos que queremos saber más sobre el rol de los ingresos en la mediación de la relación clima/conflicto. Una de las implicaciones comprobables de una teoría en la que los ingresos desempeñan un papel mediador, es que esperaríamos que las perturbaciones climáticas estuvieran asociadas a los conflictos en zonas en las que los ingresos son sensibles a las perturbaciones climáticas, pero no en las que los ingresos son independientes de las perturbaciones climáticas. Sarsons (2015)⁷ hace exactamente esto. Aprovechando el hecho de que los ingresos de los distritos situados agua abajo de las represas de riego no dependen de las precipitaciones de lluvia, pero los distritos situados río arriba de las represas de riego sí lo hacen, se explora el mecanismo de los ingresos al probar si la lluvia pueden predecir la incidencia de los disturbios sociales en los distritos situados río abajo, pero no en los distritos situados aguas arriba. Formalmente, se pone a prueba estas hipótesis:

- $X \rightarrow Y$ en lugares en los que se sabe que X afecta a M [La precipitación de lluvia aumentarán los disturbios sociales en zonas en las que la lluvia afectará negativamente los ingresos económicos (río arriba de la represa)].
- X no tiene ningún efecto sobre Y en lugares donde X no tiene ningún efecto sobre M [Las precipitaciones de lluvia no tendrán ningún efecto sobre los disturbios sociales en zonas donde los ingresos no son sensibles a las precipitaciones (río abajo de la represa)].

Sin embargo se descubrió que la relación entre las lluvias y la incidencia de los disturbios sociales se mantenía con la misma intensidad en los distritos situados en la parte baja del río, donde los ingresos no eran sensibles a las lluvias. Sarsons interpreta este resultado como una prueba “sugestiva” en contra del

mecanismo de los ingresos. Para ser claros, Sarsons no llevó a cabo ningún análisis de mediación: no midió los ingresos de cada pueblo ni cuantificó el efecto directo de las perturbaciones de las lluvias en los disturbios sociales, y el efecto indirecto de las perturbaciones de las lluvias en los disturbios sociales a través de los ingresos. En su lugar buscó los efectos de tratamiento heterogéneo que la teoría habría insinuado y, al no encontrar pruebas de ellos, concluyó que el canal del ingreso puede ser menos importante de lo que se pensaba.

Lo que puede hacer... En futuros proyectos, pregúntese: Si el mecanismo es M, ¿qué grupos o unidades esperaríamos que mostraran un efecto del tratamiento y qué grupos o unidades esperaríamos que no respondieran al tratamiento? A continuación, compruebe si estas predicciones están respaldadas por sus datos e interpréte los como pruebas sugerentes a favor o en contra del mecanismo M propuesto. Tenga en cuenta que estas pruebas no son decisivas porque los grupos podrían diferir en otros aspectos que afectarían a la capacidad de respuesta al tratamiento.

8 También podemos buscar pruebas sugerentes al observar los efectos de nuestro tratamiento en ciertas variables de resultado

Aunque es difícil cuantificar pruebas de un mecanismo, siempre podemos explorar las implicaciones comprobables de la teoría en la que figura nuestro mecanismo. En el punto 7, hicimos esto al revisar si el tratamiento afectaba a subgrupos concretos para los que nuestra teoría implicaba un efecto del tratamiento. Otro enfoque puede ser el explorar si el tratamiento afecta sólo a las variables de resultado insinuadas por nuestra teoría.

Por ejemplo, muchos científicos sociales están interesados en cómo la educación de masas influye en la democracia. Varias teorías de democratización esperan que diferentes mecanismos conectasen educación y democracia. En primer lugar, según la teoría de la modernización, la educación podría facilitar el buen funcionamiento de la democracia al socavar los vínculos de grupo (como la etnia o la religión) en favor del mérito⁸ En segundo lugar, según los teóricos sociales que estudian opresión, la educación podría debilitar a la democracia al reforzar la obediencia a la autoridad, algo que es inherente a la estructura de los centros educativos.⁹ En tercer lugar, según muchos politólogos y psicólogos, la educación puede fomentar la participación democrática al empoderar a los individuos con la capacidad de adquirir conocimiento y actuar en consecuencia¹⁰ Friedman et al. (2011)¹¹ deciden trabajar estos mecanismos investigando los resultados de un experimento de campo en el que se asignó al azar a niñas kenianas para que recibieran un subsidio educativo. Hicieron un seguimiento de las estudiantes cinco años después del programa y les plantearon varias preguntas para comprobar cuál de estos tres mecanismos estaba en funcionamiento: ¿Aceptaban las niñas el derecho del marido a pegar a su mujer? ¿Participaron sus padres en la elección de cónyuge? ¿En qué medida se identificaba la niña con su grupo religioso o étnico? ¿Leía la niña las noticias de manera regular?

La siguiente tabla resume la dirección de los efectos que sugeriría cada teoría. Obsérvese que los distintos mecanismos que se ensayan aquí son el resultado de teorías con predicciones divergentes sobre algunos de estas variables de resultado. Las predicciones de cada uno de los tres mecanismos se indican en las filas, seguidas de los resultados reales. Podemos ver que dos de las variables de resultado recogidas apoyan la teoría de la modernización. Sin embargo, la teoría de la modernización habría predicho una disminución de

la asociación con grupos religiosos o étnicos (en realidad, no hubo ningún efecto) y no tuvo ninguna predicción sobre el número de lectores de periódicos (en realidad, el número de lectores aumentó). Ninguna de las predicciones sobre el mecanismo de obediencia a la autoridad se vio respaldada por los datos. Sin embargo, los datos apoyaron las cuatro predicciones de la teoría del empoderamiento individual. Los autores concluyen que es más probable que $X \rightarrow M3 \rightarrow Y$ que $X \rightarrow M1, M2 \rightarrow Y$.¹²

Mecanismo | Aceptación del derecho del marido a maltratar a la mujer (Y1) | Participación de los padres en la selección del cónyuge (Y2) | Asociación con religión o identidad étnica (Y3) | Lectura de noticias (Y4)
 ----- | ----- | ----- | ----- | ----- (M1) Modernización | ↓ | ↓ | ↓ | No hay efecto (M2)
 Obediencia a la autoridad | ↑ | ↑ | ↑ | No hay efecto (M3) Empoderamiento individual | ↓ | ↓ | No hay
 efecto | ↑ Efecto actual | ↓ | ↓ | No hay efecto | ↑

Este estudio, al igual que el de Sarsons, no trata de cuantificar qué parte del efecto de X sobre Y se transmite a través de M. Sin embargo, a través de una investigación minuciosa de diversas variables de resultado, los autores son capaces de proporcionar evidencias sugestivas de cuales mecanismos parecen más plausibles.

****Lo que puedes hacer...** Para futuros proyectos hágase esta pregunta: Si el mecanismo es M, ¿cuáles variables de resultado esperaría que se vieran afectadas por mi tratamiento, y cuales esperaría que no se vieran afectados por mi tratamiento? Luego compruebe si estas predicciones se ven respaldadas por sus datos e interpréte los como evidencias sugestivas a favor o en contra de su mecanismo M propuesto.

9 El diseño de tratamientos complejos puede ayudarnos a comprender mejor qué parte del tratamiento está “haciendo el trabajo”

A veces, los investigadores experimentales intentan comprender mejor los mecanismos al sumar o restar elementos del tratamiento que se cree que desencadenan diferentes mecanismos. Este enfoque se denomina “análisis de mediación implícita” porque se cree que los diferentes componentes X manipulan *implícitamente* ciertos mecanismos. Esto, por supuesto, es una suposición: como no estamos midiendo M directamente, nos basamos en una afirmación teórica de que el componente A desencadenará M, mientras que el componente B no lo hará.

Por ejemplo, muchos gobiernos, como los de México, Brasil, Tanzania y Uganda, han creado programas de transferencias monetarias condicionadas para hacer frente a la pobreza. Estos programas proporcionan dinero en efectivo a individuos pobres, pero a menudo vienen con condiciones como la asistencia a la escuela o a un programa de formación laboral. Hasta hace poco, sólo sabíamos que estos programas (X) reducían con éxito la pobreza (Y) y que X causaba Y, bien a través del dinero en efectivo o a través de la asistencia requerida a la escuela o a los programas de empleo. Para diferenciar estos mecanismos, Baird et al. (2011)¹³ llevaron a cabo un experimento en Malawi: asignaron a un grupo de familias a recibir una transferencia de efectivo *condicionada* a que sus hijas asistieran regularmente a la escuela, a otro grupo de familias a recibir el efectivo *incondicionalmente*, y a un grupo de control a no recibir ninguna transferencia. Este diseño manipulaba “implícitamente” a M: aunque las niñas del grupo de transferencia incondicional también podían buscar la educación, la asistencia a la escuela (la condición objeto de estudio) sería probablemente mayor en el grupo que estaba obligado a buscarla. No es de extrañar que la asistencia a la escuela y el rendimiento en los exámenes fueran mejores para el grupo que recibió las transferencias

monetarias condicionadas. Sin embargo, sus mediciones de Y -la tasa de embarazo o de matrimonio de las niñas- fueron en realidad mejores (más bajas) en el grupo que recibió las transferencias monetarias incondicionales. Los autores concluyeron que los requisitos de asistencia a la escuela asociados a las transferencias monetarias condicionadas no eran probablemente el mecanismo responsable del éxito de estos programas en la reducción de los síntomas de la pobreza.

Estudios como éste no sólo ayudan a los científicos sociales a aprender más sobre los canales a través de los cuales X causa Y, sino también ayudan a los legisladores a explorar y descubrir nuevos tratamientos. Después de que otros estudios se unieran a los de Baird et al. en la demostración de los notables efectos de las transferencias monetarias incondicionales, muchos gobiernos y organizaciones han comenzado a aplicar programas de transferencias monetarias incondicionales.

Lo que puede hacer... Para futuros proyectos hágase esta pregunta: ¿Puede mi tratamiento ser “descompuesto” en múltiples brazos de tratamiento, algunos que implícitamente manipulan M, y otros que no? Considere la posibilidad de utilizar un diseño factorial para identificar los efectos de los diferentes brazos de tratamiento. Si tiene suficiente poder, la comparación de los distintos brazos de tratamiento le proporcionará pruebas sugestivas a favor o en contra de M.

10 A pesar de las dificultades para medir empíricamente los mecanismos, vale la pena prestarles mucha atención, pero también siendo cautos en nuestro lenguaje

El intentar identificar los mecanismos causales es un esfuerzo noble. Articular los mecanismos causales es lo que nos permite extraer los tratamientos de “caja negra” y comprender por qué y cómo funcionan determinados tratamientos. Aunque se pueden hacer (y a menudo se hacen) afirmaciones causales sin pruebas de un mecanismo causal, explorar los mecanismos causales es lo que nos permite ampliar la frontera de la investigación y reevaluar cómo nuestras evidencias se corresponden con nuestras teorías. Por estas razones, el público (ya sea el público en general o los revisores académicos) suele estar comprensiblemente ansioso de que usted exponga los mecanismos causales después de demostrar las pruebas de una provocativa afirmación causal. Para anticiparse a esto, vale la pena considerar si es posible diseñar una forma de probar los mecanismos causales antes de realizar un experimento. Si no es así, considere si ciertas medidas de variable de resultado o interacciones entre tratamientos y covariables podrían proporcionar algún apoyo a un mecanismo causal concreto, y sea explícito sobre las limitaciones de este tipo de análisis en su redacción. Los mecanismos son un campo de investigación apasionante y deben tenerse en cuenta tanto en el diseño como en el análisis de un experimento, pero debemos asegurarnos de hablar de los mecanismos con la precaución adecuada a nuestra capacidad para identificar un mecanismo concreto y evitar exagerar con los argumentos.

-
1. Solomon M. Hsiang, Kyle C. Meng y Mark A. Cane, “Civil Conflicts Are Associated with the Global Climate”, *Nature* 476.7361 (2011): 438-441. ↩
 2. Ted Gurr, *Why Men Rebel*, Princeton University Press, 1970. ↩

3. Alberto Chong, Ana L. de la O, Dean Karlan y Leonard Wantchekon, “Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice, and Party Identification”, *The Journal of Politics* 77.1 (2015): 55-71. ↩
4. Nótese que Chong et al. prueban su argumento utilizando datos a nivel de distrito electoral, no a nivel individual, pero hemos adaptado su argumento al nivel individual para facilitar la explicación. ↩
5. Para una discusión más rigurosa sobre estas falacias, léase Adam N. Glynn, “The Product and Difference Fallacies for Indirect Effects”, *American Journal of Political Science* 56.1 (2012): 257-269. ↩
6. Explicación adaptada de Alan Gerber y Donald Green, *Field Experiments*, W.W. Norton and Company, 2012, capítulo 10. ↩
7. Heather Sarsons, “Rainfall and Conflict: A Cautionary Tale”. *Journal of Development Economics* 115 (2015): 62-72. ↩
8. Marion Joseph Levy, *Modernization and the Structure of Societies*, Princeton University Press, 1966. ↩
9. Frantz Fanon, *The Wretched of the Earth*, Grove Press, 1964. John Lott, Jr., “Public Schooling, Indoctrination and Totalitarianism,” *Journal of Political Economy* 107(6), 1999. ↩
10. Gabriel Almond and Sidney Verba, *The Civic Culture: Political Attitudes and Democracy in Five Nations*, Sage Publications, 1963. Robert Mattes and Michael Bratton, “Learning about Democracy in Africa: Awareness, Performance, and Experience,” *American Journal of Political Science*, 51(1), 2007. ↩
11. Willa Friedman, Michael Kremer, Edward Miguel, and Rebecca Thornton, “Education as Liberation?” NBER Working Paper 16939, 2011. ↩
12. En el estudio actual, los autores se sorprendieron al descubrir evidencias de que la educación también aumentaba la aceptación de la violencia política por parte de los individuos. Aunque siguen sosteniendo que el empoderamiento individual es el responsable de la relación entre educación y democracia, advierten que la educación no siempre conduce a la democratización (es decir, $M_3 \rightarrow Y$ pero también es posible que $M_3 \rightarrow \text{NO } Y$). No obstante, su enfoque es una útil demostración de cómo múltiples variables de resultado pueden aclarar cosas sobre los mecanismos. ↩
13. Sarah Baird, Craig McIntosh, Berk Ozler, “Cash or Condition? Evidence from a Cash Transfer Experiment,” *Quarterly Journal of Economics* 126, 2011. ↩