

LA REVOLUCIÓN DE LA INFERENCIA CAUSAL

Noventa años después de los primeros EAC/RCT se ha logrado formalizar matemáticamente una propuesta que aborda el problema de los sesgos introducidos por las variables de confusión en los estudios observacionales. Esta formalización se encuentra desarrollada en múltiples fuentes, Miguel Hernán (Hernán & Robins, 2020, # 3-10) propone diferenciar la definición de efecto causal en dos dimensiones:

- efecto causal a nivel individual y
- efecto causal en el grupo (población)

Hernan y Rubin, señalan:

Seguimos siendo agnósticos sobre conceptos metafísicos como causalidad y causa. En cambio, nos enfocamos en la identificación y estimación de efectos causales en poblaciones, es decir, cantidades numéricas que miden cambios en la distribución de un resultado bajo diferentes intervenciones. (Hernán & Robins, 2020, # vii)

Al declararse agnósticos, cuando dicen que su objeto de trabajo último es la “medición de cambios en la distribución de un resultado bajo diferentes intervenciones”, hacen una definición que ineludiblemente tiene una clara implicación causal y que necesita un contexto metafísico.

En una escala individual, una unidad de observación que previamente no presenta la consecuencia/resultado, al ser expuesta/intervenida a la “causa” tiene dos potenciales resultados (potential outcomes, resultados contrafácticos), que la consecuencia/resultado se presente o no se presente. En la escala individual solo se puede observar uno de los dos potenciales resultados, la presencia o la ausencia de la consecuencia, el otro resultado potencial permanecerá inobservado en cada caso individual.

La notación para los resultados potenciales es la siguiente:

T = tratamiento, Y = resultado

T=1 tratamiento presente, T=0 tratamiento ausente

Y=1 resultado presente, Y=0 resultado ausente

Y_i se refiere al resultado en una unidad de observación específica, Y es el resultado en el grupo de observaciones

Para una unidad de observación específica expuesta el potencial resultado está presente: do (T=1), (Y=1), efecto causal bajo tratamiento, Y_i=(1)

Para la misma unidad de observación específica no expuesta el resultado está ausente: do (T=0), (Y=0), efecto causal sin tratamiento, Y_i=(0)

El problema fundamental de la inferencia causal es que no se pueden observar los dos resultados en la misma unidad de observación, solo se puede observar uno de los dos resultados potenciales: que el efecto aparezca luego de la intervención o que el efecto no aparezca si no se ha producido la intervención. Cualquiera de los dos resultados potenciales que se observen en una unidad específica de observación se constituye en un hecho fáctico y su contrario no observado sería entonces algo que podría haber sucedido pero que no ha sido observado, se constituye de esta forma en un resultado “contrafáctico”.

No se puede establecer la relación causal en la misma unidad de observación específica entre el resultado potencial con y sin tratamiento: $Y_i(1) - Y_i(0) = 1$. Esta notación se puede replicar cuando se estudia un cierto número de unidades de observación; se espera ($E = \text{expectativa}$) que si la relación de causalidad es absoluta, la misma relación lineal que se observa a escala individual se observe a escala grupal. $E[Y_i(1) - Y_i(0)] = E[Y(1) - Y(0)]$. A esto se refieren las filósofas que han propuesto considerar a la teoría contrafáctica de la causalidad como una forma de regularidad.

En el enfoque frecuentista tradicional de la epidemiología, el tratamiento estadístico de esta situación que se manifiesta en varios tipos de investigación se resolvió con una tabla doble entrada, de 2×2 , con lo que se calcula la razón matemática entre la incidencia en expuestos dividida para la incidencia en no expuestos, en estudios prospectivos, esta razón aritmética es el riesgo relativo. Para estudios transversales, la variante es la razón de disparidad (odds ratio) y para los estudios de tiempo hasta el evento la regresión de Cox.

El cambio con la notación expresada en forma de probabilidades condicionales es evidente y modifica sustancialmente la interpretación de los supuestos en los que se basa la relación de causalidad y la interpretación de los resultados de las pruebas estadísticas.

Esta “expectativa” de que los resultados en cada unidad de observación sea adecuadamente representada en el valor promedio del grupo total de observaciones, supone que la relación entre el “tratamiento” y el resultado (efecto) es lineal, si en un grupo de unidades de observación todas son expuestas y todas desarrollan la consecuencia y si en otro grupo ninguna de las unidades de observación han sido expuestas y ninguna de ellas desarrolla la consecuencia, entonces se considera que hay total consistencia entre la exposición y la consecuencia. La probabilidad de estar expuesta y desarrollar la consecuencia es igual a 1 y la probabilidad de no estar expuesta y no desarrollar la consecuencia es también igual a 1. La diferencia entre las dos probabilidades es 0 (cero) y la razón entre estas dos probabilidades es igual a 1, de la misma manera que la “razón de razones” de estas probabilidades es igual a 1. Existe una congruencia total entre la exposición y la consecuencia, tanto para las expuestas como para las no expuestas, si bien no son resultados contrafácticos en la misma unidad de observación, son resultados en un grupo expuesto y otro no expuesto.

$$(i) \Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 1] = 0$$

$$(ii) \frac{\Pr[Y^{a=1} = 1]}{\Pr[Y^{a=0} = 1]} = 1$$

$$(iii) \frac{\Pr[Y^{a=1} = 1] / \Pr[Y^{a=1} = 0]}{\Pr[Y^{a=0} = 1] / \Pr[Y^{a=0} = 0]} = 1$$

Tomado de (Hernán & Robins, 2020, # 7)

(i) La probabilidad de que el resultado en las unidades de observación que han recibido la intervención es igual a 1, menos la probabilidad de que el resultado en las unidades de observación que NO han recibido la intervención es igual a 1 da como resultado cero (0).

(ii) La probabilidad de que el resultado en las unidades de observación que han recibido la intervención es igual a 1, dividido para la

probabilidad de que el resultado en las unidades de observación que NO han recibido la intervención es igual a 1 da como resultado uno (1).
 (iii) La probabilidad de que el resultado en las unidades de observación que han recibido la intervención es igual a 1, dividido para la probabilidad de que el resultado en las unidades de observación que NO han recibido la intervención es igual a 0, todo esto dividido La probabilidad de que el resultado en las unidades de observación que han recibido la intervención es igual a 1, dividido para la probabilidad de que el resultado en las unidades de observación que NO han recibido la intervención es igual a 0, da como resultado uno (1).

La notación corresponde a variables categóricas y se aplica con las precisiones correspondientes a variables continuas.

Pero en otros casos se puede dar la situación en la que algunas de las unidades expuestas no presenten la consecuencia y por otra parte, algunas de las unidades no expuestas si presenten la consecuencia. Esta situación pone en evidencia que la relación causal no alcanza una total consistencia sino que existe una parte de la relación que presenta una discrepancia que conduce a la noción de **expectativas condicionales** (conditional expectations=E).

$$E [Y (1) - Y (0)] \neq E [Y | X=1] - E [Y | X=0].$$

La consistencia total es poco probable en el trabajo científico en las ciencias de la salud. Una alternativa genial para evitar que la consistencia se reduzca demasiado por los sesgos al elegir las unidades de observación, recolectar los datos y proceder a su análisis estadístico es formar dos grupos de unidades de observación asignando aleatoriamente las unidades a cada grupo e introduciendo una intervención (tratamiento) en uno de los grupos. La intervención, que matemáticamente se formaliza como una probabilidad condicional, puede introducirse bien sea en el grupo 1 o en el grupo 2, puesto que el conjunto de las unidades de observación en cualquiera de los dos grupos tiene características similares al conjunto de las unidades del otro grupo, son intercambiables (**exchangeability**). La fortaleza de los ECA/RCT reside tanto en la intercambiabilidad de los grupos expuesto y no expuesto, cuanto en el hecho de que la relación entre la intervención y el resultado queda aislada del resto de características (variables) de las unidades de observación que podrían tener una “relación causal con la mencionada intervención”. Si alguna duda queda sobre el papel de las terceras variables en la relación causal, la estadística desarrolló durante el siglo pasado un arsenal impresionante de posibles análisis de esta “incertidumbre”, tanto de los valores promedio de las variables en cada grupo de estudio, cuanto de las distribuciones de cada variable o de las distribuciones de la cuantificación matemática de las discrepancias. Unos ejemplos clásicos sobre el análisis de la incertidumbre introducida por las terceras variables y su correspondiente análisis estadístico para variables categóricas los tenemos en el Chi cuadrado de Mantel-Hanzel (para análisis estratificado por una tercera variable) y para las variables continuas la regresión logística (para análisis estratificado por un modelo de asociación de terceras variables) y los modelos de regresión lineal y multivariada. Dentro de los ECA/RCT estos procedimientos simplemente debían verificar que los supuestos declarados en el diseño de los experimentos se cumplieran.

Tanto en los estudios experimentales como en los estudios observacionales, en los que se compara uno o varios grupos expuestos y no expuestos es altamente probable que se presente una discrepancia matemática entre la expectativa de total consistencia entre la exposición y el resultado y lo que efectivamente se consigue

observar. La diferencia está en el hecho de que en los estudios experimentales los grupos son intercambiables (se puede elegir cualquiera de los dos para ser el grupo experimental) y que la intervención ha sido aislada de las posibles variables que podrían tener una relación de causalidad con ella. Mientras que en los estudios observacionales, los grupos se forman por la influencia de variadas circunstancias que distribuyen las unidades de observación en cada grupo ejerciendo múltiples tipos de influencias, lo que resulta en grupos no intercambiables y adicionalmente las relaciones de causalidad con la intervención de las variables que perfilan a cada unidad de observación no han sido aisladas, sino que ejercen su influjo con escasa restricción.

Cabe preguntarse entonces:

¿Existe discrepancia entre la probabilidad de que se produzca el resultado/consecuencia DADO que se ha observado y medido la presencia concomitante y antecedente de la exposición/intervención en la población general (primer escalón) con la probabilidad de que se produzca el resultado/consecuencia DADO que se ha inducido intencionalmente o “naturalmente” la exposición/intervención/tratamiento (segundo escalón) en grupos seleccionados aleatoriamente?

Además de los sesgos de selección, de recolección y de análisis, una tercera variable o grupo de variables debe estar presente para explicar la discrepancia. Estas terceras variables no estarán distribuidas homogéneamente en el grupo (os) expuesto (s) y no expuesto (s) y en algunos segmentos de las unidades de observación ejercerán un influjo causal sobre la intervención con lo que se introducirán sesgos adicionales que perturbarán la relación causal hasta (en un caso extremo) prácticamente anularla. Las conclusiones que se obtengan, incluso haciendo el correspondiente ajuste por las terceras variables han sido consideradas como incapaces de presentar evidencia suficiente sobre una posible relación causal. Esto es así en especial para los estudios observacionales, pero también para los experimentales, puesto que las unidades de observación seleccionadas para el experimento no suelen provenir de muestras representativas de la población, sino de participantes autorreferidos o a quienes se solicita su participación. La aleatorización de los grupos experimental y de control solo garantiza que no se introduzcan más sesgos en la distribución de las variables de confusión entre los dos grupos (intercambiabilidad).

En alguna medida esto ha conducido a la calificación de los estudios observacionales como débiles e inconsistentes para revelar o comprobar una relación causal. Adicionalmente, las unidades de observación no son intercambiables, puesto que estuvieron expuestos o no expuestos de acuerdo con su condición individual, no fueron asignadas aleatoriamente a cada grupo. El desarrollo de la estadística tanto para su uso en los ECA/RCT cuanto en los estudios observacionales responde a la necesidad de mejorar la capacidad de unos y otros a disipar la incertidumbre introducida por las terceras variables, en especial aquellas que actúan como causa común de la variable de exposición y de la variable de resultado; en los últimos tiempos también se han desarrollado pruebas estadísticas para procesar el papel de las terceras variables como mediadoras en la relación entre la exposición y el resultado.

Sin embargo, la definición clásica de confusión de la epidemiología ha sido objeto de cuestionamientos. Esta definición afirma que:

“Un elemento que produce “confusión” en X (el tratamiento) y Y (el resultado) es una variable Z que está (1) asociada con X en la población en su conjunto, y (2) asociada con Y entre la gente que no ha estado expuesta al tratamiento X. En años recientes, esto ha sido complementado con una tercera condición: (3) Z no debería estar en la trayectoria causal entre X y Y” (Pearl & Mackenzie, 2018, #153).

Según algunos autores, en especial Pearl, por mucho que se amplié la capacidad de las pruebas estadísticas, las relaciones de causalidad evidenciadas por los estudios observacionales difícilmente alcanzarán un nivel de aceptación suficiente como evidencia científica de las relaciones de causalidad, a menos que se produzca una revolución de la inferencia causal, cambiando los paradigmas filosóficos, encontrando un mecanismo preciso para poner en claro los supuestos teóricos en los que se basa la asociación de variables y su papel en la relación causal y desarrollando un nuevo lenguaje matemático (engendrado por los dos elementos anteriores) que diseñe un andamiaje axiomático y de formalización de las relaciones entre variables. Una imagen objetivo ideal sería lograr este sistema combinando estudios experimentales con estudios observacionales de tal manera que permitan aprovechar las ventajas de cada uno, minimizando sus deficiencias. Esto no solo es cierto para las ciencias de la salud, se ha identificado su necesidad en múltiples dominios del conocimiento como se argumenta a continuación:

“El peldaño asociativo (inferior) esencialmente no requiere información adicional más allá de lo que se proporciona en los datos de observación. Para inferir el efecto de las intervenciones necesitamos conocimiento causal estructural. Y para responder preguntas contrafácticas necesitamos conocimiento causal estructural y ecuaciones estructurales”.
(Berrevoets et al., 2023, #)

Podría decirse que aprender sobre causa y efecto es el objetivo principal de la econometría aplicada. En la práctica, la validez de estas inferencias causales depende de una serie de supuestos críticos relacionados con el tipo de datos que se han recopilado y el conocimiento sustantivo disponible sobre el fenómeno que se investiga. Por ejemplo, los factores de confusión no observados amenazan la validez interna de las estimaciones; la disponibilidad de datos a menudo se limita a muestras no aleatorias con sesgo de selección; los efectos causales deben aprenderse de experimentos sustitutos con cumplimiento imperfecto; y el conocimiento causal debe extrapolarse entre poblaciones estructuralmente heterogéneas. Se requiere un marco de inferencia causal potente y flexible para abordar todos estos desafíos, que plagan esencialmente cualquier análisis de datos en diversos grados. Sobre la base de la perspectiva estructural de la causalidad introducida por Haavelmo (1943) y el enfoque teórico de grafos propuesto por Pearl (1995), la literatura sobre inteligencia artificial (IA) ha desarrollado una amplia gama de técnicas para la inferencia causal que permiten aprovechar la información de varios fuentes de datos imperfectas, heterogéneas y sesgadas (Bareinboim y Pearl, 2016). (Hünermund & Bareinboim, 2023, #)

Como se ha mencionado anteriormente, el enfoque contrafáctico de la causalidad está contenido en la misma definición de causalidad de David Hume, pero ha tomado un par de siglos pasar esta noción metafísica al campo de la ciencia. El RCT/ECA representó un crucial avance hacia la posibilidad de tener información tanto del efecto de la intervención, cuánto del efecto de la no intervención, lo que amplía la capacidad de la ciencia para identificar las relaciones de causalidad pero no consigue

desentrañar relaciones de causalidad que escapan a la aleatorización, la intercambiabilidad entre grupo expuesto y no expuesto y la generalización de los resultados a una población de referencia o su aplicación mecánica a cada unidad de observación. Entonces resulta indispensable examinar el enfoque contrafáctico desde la perspectiva de su incorporación a la metodología de la ciencia.

Conversaciones con ChatGPT y Bard sobre el enfoque contrafáctico, dentro de la revolución de la causalidad:

PREGUNTA: Por favor elabore una síntesis basada en los conceptos desarrollados por Judea Pearl y otros autores sobre la revolución de la causalidad en relación con los principios del análisis contrafáctico (counterfactual analysis)

ChatGPT 3.5:

La revolución de la causalidad, según Judea Pearl y otros autores, se refiere a un cambio fundamental en la forma en que entendemos y analizamos las relaciones causales entre eventos. En contraste con enfoques tradicionales que se centran en la correlación entre variables, la revolución de la causalidad propone que entender la causalidad implica examinar cómo cambiar un factor afecta directamente a otro, incluso en situaciones hipotéticas.

El análisis contrafáctico es un pilar fundamental de esta revolución. Se basa en la idea de "contrafactuals" o escenarios alternativos, es decir, explorar qué habría sucedido si un evento o condición hubiera sido diferente. Esto permite evaluar el impacto causal de una variable al comparar lo que realmente ocurrió con lo que podría haber ocurrido en un mundo hipotético.

BARD:

Según la revolución de la causalidad liderada por Judea Pearl, el análisis contrafáctico es un método para evaluar el impacto de un cambio en una variable causal sobre otra variable. Este método se basa en la idea de que las causas afectan a sus efectos de manera condicional, es decir, que el efecto de una causa depende del contexto en el que se produce.

En el contexto de la revolución de la causalidad, el análisis contrafáctico se puede utilizar para responder a preguntas como:

- ¿Qué hubiera pasado si...?*
- ¿Cuál es el impacto de cambiar X en Y?*
- ¿Cuál es la probabilidad de que Y ocurra si X cambia?*

Para responder a estas preguntas, el análisis contrafáctico utiliza modelos causales para representar la relación entre las variables involucradas. Estos modelos se pueden utilizar para generar escenarios contrafácticos, es decir, escenarios en los que las variables causales se modifican de manera hipotética.

El análisis contrafáctico tiene una amplia gama de aplicaciones en campos como la ciencia, la medicina, la política y la economía. Por ejemplo, se puede utilizar para evaluar el impacto de nuevas políticas públicas, para desarrollar nuevos tratamientos médicos o para predecir el comportamiento del mercado. Una definición precisa de análisis contrafáctico en el contexto de la revolución de la causalidad es la siguiente:

El análisis contrafáctico es un método para evaluar el impacto de un cambio en una variable causal sobre otra variable, utilizando modelos causales para representar la relación entre las variables involucradas.

Esta definición enfatiza los siguientes aspectos del análisis contrafáctico:

- *Se basa en la idea de que las causas afectan a sus efectos de manera condicional.*
- *Utiliza modelos causales para representar la relación entre las variables involucradas.*
- *Puede utilizarse para responder a preguntas sobre el impacto de los cambios en las variables causales.*

El análisis contrafáctico es una herramienta poderosa que puede utilizarse para comprender la causalidad y sus efectos. La revolución de la causalidad liderada por Judea Pearl ha contribuido a popularizar este método y a ampliar su alcance a nuevas áreas de aplicación.

¿Cuáles son los procedimientos matemáticos para procesar el análisis contrafactual?

ChatGPT

El análisis contrafactual implica el uso de procedimientos matemáticos y técnicas estadísticas para estimar y analizar los efectos causales de variables en situaciones hipotéticas. Algunos de los enfoques matemáticos y estadísticos utilizados en el análisis contrafactual incluyen:

Potencial Causal: Este enfoque se basa en la idea de "potenciales causales" y utiliza el concepto de diferencias entre los resultados observados y los resultados hipotéticos que habrían ocurrido en ausencia de una variable específica. La idea central es estimar la diferencia entre el resultado observado y el resultado contrafactual para medir el efecto causal.

Emparejamiento y Propensity Score Matching: Estos métodos se utilizan para crear grupos de tratamiento y control que sean comparables en términos de características observables. El "propensity score" es la probabilidad de recibir el tratamiento condicionado a las características observables. Al emparejar individuos que tienen valores de "propensity score" similares, se busca crear grupos comparables para estimar el efecto causal.

Modelos de Regresión y Econometría: Los modelos de regresión permiten modelar la relación entre una variable dependiente y una o más variables independientes, controlando el efecto de otras variables. Los modelos econométricos, como el modelo de regresión lineal, pueden ser utilizados para estimar los efectos causales después de controlar por otras variables relevantes.

Experimentos Controlados: En situaciones donde es posible llevar a cabo experimentos controlados, se asignan aleatoriamente los tratamientos a grupos de estudio. La aleatorización ayuda a controlar factores no observados y permite atribuir cambios en los resultados a la manipulación de la variable de interés.

Estimación de Efectos Marginales y Condicionales: Al estimar los efectos marginales, se analiza cómo un cambio en una variable afecta directamente el resultado, manteniendo constantes otras variables. Los efectos condicionales se centran en cómo cambia el resultado cuando múltiples variables varían simultáneamente.

Grafos Causales y Modelado Estructural: Los grafos causales son representaciones visuales de las relaciones causales entre variables. El modelado estructural utiliza estos grafos para establecer relaciones causales y derivar ecuaciones que describen cómo las variables interactúan.

Estos son solo algunos de los enfoques matemáticos y estadísticos utilizados en el análisis contrafactual. La elección de la técnica dependerá de la naturaleza de los datos disponibles, la complejidad del problema y los supuestos que se pueden hacer sobre las relaciones causales.

BARD

Los principales procedimientos matemáticos que se utilizan para el análisis contrafáctico son los siguientes:

- *Modelos causales basados en redes bayesianas: Las redes bayesianas son una representación gráfica de las relaciones causales entre las variables. Se pueden utilizar para generar escenarios contrafácticos, simulando los cambios en las variables causales y calculando los efectos resultantes.*
- *Modelos causales basados en regresiones: Las regresiones son un tipo de modelo estadístico que se utiliza para predecir el valor de una variable en función de los valores de otras variables. Se pueden utilizar para estimar el impacto de un cambio en una variable causal sobre otra variable, incluso en ausencia de un modelo causal explícito.*

Modelos causales basados en aprendizaje automático: El aprendizaje automático es una disciplina de la inteligencia artificial que permite construir modelos a partir de datos. Se puede utilizar para desarrollar modelos causales que puedan aprender las relaciones entre las variables a partir de datos.

Estos procedimientos matemáticos se pueden utilizar de forma independiente o en combinación para realizar análisis contrafácticos. La elección del procedimiento más adecuado depende de las características del problema específico que se desea abordar.

A continuación, se ofrece una breve descripción de cada uno de estos procedimientos:

Modelos causales basados en redes bayesianas: Las redes bayesianas son una representación gráfica de las relaciones causales entre las variables. Una red bayesiana consta de nodos, que representan variables, y arcos, que representan las relaciones causales entre las variables.

Para generar escenarios contrafácticos, se puede modificar el valor de una variable causal en una red bayesiana y luego calcular las probabilidades de los valores de las demás variables.

Las redes bayesianas son una herramienta poderosa para el análisis contrafáctico, pero pueden ser difíciles de construir y mantener.

Modelos causales basados en regresiones: Las regresiones son un tipo de modelo estadístico que se utiliza para predecir el valor de una variable en función de los valores de otras variables.

Para estimar el impacto de un cambio en una variable causal sobre otra variable, se puede utilizar una regresión para predecir el valor de la variable afectada en función del valor de la variable causal. Las regresiones son una

herramienta relativamente simple de usar, pero pueden ser menos precisas que los modelos causales basados en redes bayesianas.

Modelos causales basados en aprendizaje automático: El aprendizaje automático es una disciplina de la inteligencia artificial que permite construir modelos a partir de datos. Se puede utilizar el aprendizaje automático para desarrollar modelos causales que puedan aprender las relaciones entre las variables a partir de datos. Los modelos causales basados en aprendizaje automático pueden ser más precisos que los modelos causales basados en redes bayesianas o regresiones, pero requieren más datos para su entrenamiento.

En conclusión, los principales procedimientos matemáticos que se utilizan para el análisis contrafáctico son los modelos causales basados en redes bayesianas, los modelos causales basados en regresiones y los modelos causales basados en aprendizaje automático.