

# Cuadernos Metodológicos

---

---

## Métodos de análisis causal

Juan Díez  
Medrano

## Causa y medición

### La noción de causa

Desde hace mucho tiempo humanistas y científicos sociales han tratado de desvelar las relaciones entre acontecimientos sociales, de manera a comprender, manipular y predecir. Una de las estrategias que han utilizado es la búsqueda de causas y efectos. Ahora bien, la definición de qué es lo que constituye una causa y la determinación del tipo de evidencia que necesita ser recogida para establecer causalidad ha sido fuente de innumerables debates que los párrafos siguientes tratan de resumir.

La definición clásica del concepto de causalidad nos la ofrece HUME, que en su *Tratado sobre la Naturaleza Humana*, señala tres criterios fundamentales: (1) Contigüidad entre Causa y Efecto, (2) la precedencia temporal de la Causa sobre el Efecto, y (3) la conjunción constante entre Causa y Efecto; es decir, que siempre que la Causa esté presente se observe el Efecto y que siempre que la Causa esté ausente no se observe el Efecto. Tanto para HUME como para la mayoría de los positivistas éstos son los tres criterios necesarios para poder decir que una determinada relación es causal. Así, RUSSELL (1913) define la causalidad de la manera siguiente: «Dado un acontecimiento  $e_1$ , existen un acontecimiento  $e_2$  y un intervalo temporal  $T$ , de tal manera que cada vez que  $e_1$  se da,  $e_2$  le sigue tras un intervalo  $T$ .

Desde la perspectiva positivista, por tanto, la distinción entre meras regularidades o correlaciones y las relaciones causales, que implican la existencia de un argumento teórico previo que ligue Causa y Efecto, no es posible y no se plantea. Para aquellos, sin embargo, que piensan que dicha distinción es posible y necesaria, el criterio enunciado por John Stuart MILL según el cual para que podamos definir una relación como causal debemos haber excluido todo mecanismo causal alternativo, se convierte en determinante a la hora de determinar la existencia de una relación de Causa y Efecto. En otras palabras, desde el momento en que aceptamos la posibilidad de distinguir entre relaciones causales y relaciones no causales, el criterio principal sobre el que se apoyará cualquier tesis sobre la existencia de una relación causal determinada es este principio introducido por MILL.

Ello nos conduce hacia la búsqueda de métodos adecuados que nos permitan eliminar, del análisis de una determinada relación causal, causas alternativas potenciales. La mayor o menor capacidad que nuestro diseño de investigación tenga a la hora de eliminar causas alternativas es la que determina la llamada Validez Interna de nuestro diseño de investigación. Esta validez interna es máxima en los experimentos, en los cuales los individuos u objetos que constituyen nuestra unidad de análisis son asignados de manera aleatoria a un grupo experimental y a un grupo de control. En muchas ocasiones, sin embargo, razones de diversa índole (éticas, financieras, etc.) impiden la utilización del método experimental, en cuyo caso el investigador trata de diseñar investigaciones que permitan eliminar un máximo de causas alternativas potenciales. Por ejemplo, la introducción de variables de control en análisis estadísticos, o la utilización del método comparado, en base al método de la concordancia y el método de las diferencias, propuestos por MILL. El método de la concordancia implica comparar dos casos que difieran en todo, menos en el efecto, para de esta manera constreñir al máximo el número de causas potenciales, que son aquellas pocas variables causales potenciales en las que los dos casos coincidan. El método de las diferencias, por el contrario, implica comparar dos casos que no difieran en casi nada, menos en el efecto estudiado, de manera a constreñir el número de causas potenciales, que son aquellas pocas variables causales potenciales en las que los dos casos difieren.

Si bien el criterio para determinar la existencia de una relación causal propuesto por MILL ha conllevado el desarrollo de diseños de investigación que permitan eliminar un máximo de causas alternativas potenciales, el criterio de la conjunción constante entre Causa y Efecto ha conllevado el desarrollo de métodos para determinar los límites contextuales de dicha conjunción constante. Se tiende a aceptar que toda conjunción constante puede ser probada únicamente dentro de determinados límites impuestos por lo exhaustivo de nuestra observación empírica. Incluso aquellas relaciones que nos parecen más constantes podrían no darse en el futuro, al que no tenemos acceso. Para determinar el grado de regularidad con el que se producen determinadas relaciones, los metodólogos han propuesto diversos métodos de mayor o menor complejidad.

En general, se puede decir que cuanto mayor sea el número de ocasiones en que se observa una determinada relación causal, mayor es nuestra confianza en su generabilidad. Sin embargo, no es lo mismo observar dicha relación causal en varios contextos similares que observarla en contextos que difieran tanto como sea posible entre sí. Nuestra capacidad para decidir la magnitud del ámbito contextual en el que se da tal relación causal es mayor en este segundo caso. Sin embargo, el método más adecuado para determinar la generabilidad de una relación causal es el muestreo probabilístico. Los métodos de muestreo nos ayudan a determinar en qué contextos se produce una determinada relación causal y nos ayudan a deter-

minar con un cierto grado de confianza si la relación causal observada en una muestra se da en la población de la que se extrajo dicha muestra.

Llegados a este punto, merece la pena volver a Bertrand RUSSELL para reconsiderar el concepto de causa. RUSSELL señala con acierto que encontrar procesos causales que cumplan el requisito de la conjunción constante es altamente improbable y que, incluso si ésta se da en el presente y en el pasado, no sabemos qué ocurrirá en el futuro. Además un análisis de la práctica científica demuestra que el investigador raramente se dedica exclusivamente a la búsqueda de tales procesos causales. RUSSELL propone abandonar de una vez por todas el concepto de causa y que, en lugar de ello, nos dediquemos a formular leyes científicas, es decir, relaciones funcionales del tipo  $e = mc^2$ . Ateniéndose, en sentido amplio a esta recomendación, los científicos se han dedicado a especificar el sentido y la magnitud de relaciones entre tipos de acontecimientos y a circunscribir el ámbito en el que se aplican determinadas relaciones. Por ejemplo, en lugar de intentar determinar si el descenso de los tipos de interés va siempre seguido de un aumento en la tasa de crecimiento económico, tratamos de determinar cuál es la probabilidad de que un descenso en los tipos de interés se traduzca en un mayor crecimiento económico, o tratamos de estimar cuál es el impacto sobre el crecimiento económico de cambios en los tipos de interés. Además, intentamos medir cómo varía esta relación a través del tiempo y el espacio.

Si siguiéramos utilizando la definición tradicional de causa, sólo podríamos utilizar este término cuando una relación cumple los cuatro criterios citados en esta sección. En la práctica, sin embargo, los investigadores hablan de causa cuando existe una relación entre dos variables y si se puede descartar la idea de que tal relación sea espúrea. En lugar de distinguir entre relaciones causales y relaciones no causales, distinguimos entre relaciones causales fuertes, relaciones causales débiles y relaciones no causales.

## El problema de la medición

La mayoría de nuestras descripciones, explicaciones o predicciones sobre la sociedad se basan en la medición consciente o inconsciente. Ello no significa que todo conocimiento deba basarse en la medición de características sociales. Ciertamente algunas corrientes filosóficas creen en la existencia de otras fuentes de conocimiento: revelaciones, intuiciones, razonamientos deductivos. Hay que tener en cuenta, sin embargo, que incluso el razonamiento más deductivo se basa en ciertos supuestos previos que uno contrasta con la realidad. De este modo, incluso la lógica deductiva se basa en la medición previa de la realidad objetiva. La lógica inductiva se convirtió en

la rama dominante dentro de la ciencia únicamente en tiempos recientes. Hoy, las ciencias sociales tienden a basar sus conocimientos en la medición sistemática de la realidad.

La medición consiste en dos etapas principales: la primera es la definición de lo que se va a medir. La segunda es la operacionalización de lo que va a medirse, en base a esa definición. Respecto al primer punto, hay que notar que uno de los problemas más serios con los que se enfrentan las ciencias sociales es la falta de consenso respecto a la definición de los conceptos utilizados, así como la incapacidad para darse cuenta de que la falta de concordancia entre los resultados de diversos tests para determinadas hipótesis se debe a que los investigadores están utilizando definiciones y operacionalizaciones distintas de un mismo concepto. Por tanto, el primer paso en toda investigación consiste en la definición de los conceptos utilizados. Para BOLLEN (1989):

Un concepto es una idea que une distintos fenómenos (i.e. actitudes, comportamientos, rasgos) bajo una misma etiqueta. El enfado, por ejemplo, es el elemento común que une distintas manifestaciones individuales como el elevar la voz, el lanzamiento de objetos, el enrojecimiento o el comportamiento descontrolado. El concepto de enfado resume una serie de características exhibidas por los individuos. Otros conceptos juegan un papel similar [...] ¿Podemos acaso hablar de la existencia de conceptos? Los conceptos tienen la misma realidad o falta de realidad que otras ideas. Los crea la gente que cree que determinados fenómenos comparten algo en común. El concepto identifica aquellos rasgos compartidos por dichos fenómenos.

Una vez que un concepto ha sido seleccionado, los cuatro pasos siguientes son: (1) dar un significado al concepto, (2) identificar las dimensiones que lo representan, (3) crear indicadores, y (4) especificar la relación entre los indicadores y las variables latentes (o conceptos). El primer paso conlleva la formulación de una definición teórica. Una definición teórica explica en términos lo más simple posibles el significado de un concepto. Esta definición teórica cumple determinadas funciones. Una de ellas es la de unir una etiqueta al concepto. En segundo lugar, la definición teórica enumera las dimensiones del concepto. Cada concepto, efectivamente, consta de varias dimensiones. Las dimensiones de un concepto no pueden ser divididas fácilmente en componentes adicionales... En tercer lugar, la definición teórica provee información respecto al tipo de indicadores que uno debe seleccionar.

El concepto de terrorismo puede servir de ejemplo. La CIA lo define como sigue: «La amenaza o el uso de la violencia por motivos políticos por individuos o grupos, en favor o en contra de la autoridad gubernamental establecida, cuando dichas acciones tienen como fin el connotacionar o intimidar a un grupo mayor que el de sus víctimas inmediatas.» Esta definición tiene, como se puede apreciar, tres dimensiones fundamentales. La primera

Se que  
no define  
razones  
están en

es cualquier amenaza o uso de violencia por motivos políticos; la segunda es la existencia de un objetivo que trasciende al individuo o grupo objeto de tal violencia; la tercera es el que la acción sea en favor o en contra de la autoridad gubernamental establecida.

El paso siguiente en el proceso de medición, la selección de indicadores, depende directamente de la definición teórica. Este paso es denominado fase de definición operacional. La definición operacional describe los procedimientos a seguir para seleccionar indicadores de la variable latente o variables latentes (si el concepto tiene varias dimensiones) que representan a un determinado concepto. En determinadas situaciones la variable o variables latentes son operacionalizadas en base a respuestas individuales a un cuestionario. Otras veces dichos indicadores provienen de estadísticas recogidas por la administración pública, ya sean datos del censo o de los registros civiles, etc... Un indicador determinado es apropiado en la medida en que provea al investigador con una variable empírica que corresponda al significado asignado al concepto. Muchas veces, sin embargo, la información proporcionada por un indicador no es completa, y una cantidad considerable de sentido común entra en juego a la hora de asignar indicadores a un determinado concepto. La diversidad de las fuentes utilizadas, la profesionalidad de los codificadores, son factores que influyen sobre la calidad de los indicadores. En el caso del terrorismo, la calidad de su medición dependerá de la diversidad de fuentes de información utilizadas, del cuidado tomado por los reporteros, e incluso de la capacidad de acceso a determinadas áreas geográficas. Por ejemplo, las fuentes de información occidentales tienden más a hablar de actos terroristas cuando estas acciones son tomadas por países o grupos hostiles al mundo capitalista occidental.

Prácticamente todas las medidas que utilizamos contienen error de medición. Por eso el cuarto paso en el proceso de medición consiste en formalizar ese tipo de errores. No hay criterios definitivos respecto a la definición de un concepto; tal como indiqué anteriormente, las ciencias sociales se caracterizan por su falta de consenso respecto a conceptos importantes. Sin embargo, significaría un importante avance el empezar a crear consenso tratando de (1) especificar claramente las definiciones utilizadas, (2) utilizar en las investigaciones las definiciones utilizadas más frecuentemente, y (3) siempre que sea necesario y posible, evaluar la calidad representativa de la realidad de las diferentes definiciones de los conceptos utilizados. En cuanto a la operacionalización de los conceptos, los científicos sociales han desarrollado técnicas para juzgar la bondad de distintos indicadores a la hora de medir los conceptos utilizados. Los dos criterios principales utilizados son la validez y la fiabilidad.

La validez de un concepto se refiere al grado en que una variable determinada mide lo que se supone debe medir. ¿Hasta qué punto, por ejemplo, podemos decir que la información sobre terrorismo incluida en los medios de comunicación mide realmente el terrorismo? ¿Hasta qué punto los tests

de inteligencia miden la inteligencia? ¿Hasta qué punto el Producto Nacional Bruto mide el valor real de los bienes y servicios producidos en un país? Estas cuestiones se refieren a la validez de distintos indicadores, que nunca se puede determinar de modo absoluto. Sin embargo, aunque nunca podemos probar la validez de un concepto, sí que podemos obtener evidencia que determine el grado mayor o menor de validez de un indicador.

Existen medios teóricos y empíricos de determinar la validez de un indicador. Los primeros definen lo que se denomina validez de contenido. La validez de contenido se refiere al grado en que los indicadores de un concepto cubren todas sus dimensiones. En tanto en cuanto lo hagan podemos hablar de la validez de contenido de los indicadores. La pregunta crucial que nos podemos hacer entonces es ¿cómo sabemos cuáles son las dimensiones de un concepto? Para responder a esta pregunta debemos volver al primer paso dentro del proceso de medición del que hablamos en la primera parte de este capítulo. Es decir, que para poder tomar en cuenta todas las dimensiones de un concepto es necesario tener una definición teórica previa de ese concepto. En determinadas ocasiones, sin embargo, nuestros instrumentos de medida no nos permiten considerar todas las dimensiones de un determinado concepto. Lo mejor en tales ocasiones es reconocer el carácter parcial de nuestros resultados.

La limitación principal del criterio de la validez de contenido de un indicador es que depende de la definición teórica del concepto. Para la mayoría de los conceptos utilizados en las ciencias sociales no existe un consenso absoluto sobre su definición teórica. El investigador debe en estas situaciones no sólo proveer una definición teórica aceptada por los demás colegas sino también obtener indicadores que cubran completamente todas las dimensiones del concepto. Se han sugerido distintos métodos empíricos para determinar la validez de los indicadores para un concepto determinado. Ninguno de ellos es enteramente apropiado. Uno de estos métodos consiste en evaluar la validez de un indicador en base a su asociación con un indicador que supuestamente mide perfectamente el concepto en consideración. Otro de los métodos evalúa la validez de un indicador en base al grado en que sus asociaciones con otros indicadores de otros conceptos se adecua a las predicciones sobre dichas asociaciones. BOLLEN, finalmente sugiere utilizar como criterio la asociación entre el concepto y su indicador, obtenida de modo empírico utilizando LISREL. Dado que ninguna de las estrategias empíricas es completamente acertada, la estrategia teórica es cuanto menos una condición necesaria a la hora de evaluar la validez de un indicador. La utilización de métodos empíricos debería completarla, siempre con plena conciencia de sus limitaciones.

La fiabilidad de un indicador es la consistencia con que mide un concepto. No es igual a su validez y, de este modo, podemos tener medidas que son a la vez fiables pero inválidas. Por ejemplo, el peso proporcionado por una balanza bien calibrada es un indicador fiable del peso real de un objeto

Validez de contenido

validez por  
Exposición  
3º voto

determinado. Pero este mismo peso es un indicador fiable pero no válido de la temperatura de un cuarto.

La fiabilidad de un indicador sólo se puede medir de forma empírica y existen diversos métodos. BOLLEN cita unos cuantos de estos métodos: El primero de ellos es la técnica consistente en repetir una misma medición dos veces y en calcular el coeficiente de correlación entre las dos mediciones. El segundo, consiste en utilizar dos indicadores de un mismo concepto en dos ocasiones distintas y en calcular la correlación entre los valores obtenidos en las dos mediciones. Los coeficientes de correlación obtenidos con los dos métodos constituyen medidas de fiabilidad. El tercer método enumerado por BOLLEN es el consistente en tomar varios indicadores de un mismo concepto y dividirlos de forma arbitraria en dos grupos. Dentro de cada grupo se combinan los valores obtenidos para cada uno de los indicadores y, finalmente, se calcula el coeficiente de correlación entre la variable compuesta obtenida para cada grupo. Este coeficiente de correlación señala la fiabilidad de los indicadores incluidos en los dos grupos. BOLLEN, por su parte, sugiere tomar como medida de fiabilidad el coeficiente de correlación múltiple entre un indicador y el concepto o conceptos que éste mide supuestamente. En un capítulo ulterior exploraremos de modo más detallado las técnicas sugeridas por BOLLEN para determinar la validez y fiabilidad de un indicador determinado.

Es, por lo tanto, importante reconocer que las medidas que utilizemos para cada concepto utilizado en nuestra investigación contienen un determinado grado de error. Algunas veces este error es debido al azar; en otras ocasiones se trata de un error sistemático. Nuestra función como investigadores consiste en proporcionar definiciones claras de los conceptos que utilizemos, asegurarnos de que nuestros indicadores miden cada una de las dimensiones de un concepto determinado o, si no es así, ser explícitos respecto a las limitaciones de nuestros indicadores y, finalmente, utilizar varias medidas alternativas del mismo concepto, que, en determinados casos puedan ser introducidas dentro de un mismo modelo estadístico que tenga en cuenta la existencia de error de medida.

## Causalidad y diseños de investigación no experimentales

Cualquier diseño de investigación interesado en la explicación de relaciones causales entre variables intentará demostrar que:

1. El indicador  $x$  sirve para medir el concepto  $X$  y lo hace con fiabilidad.
2. El indicador  $y$  sirve para medir el concepto  $Y$  y lo hace con fiabilidad.
3. La relación entre  $X$  e  $Y$  no sólo existe sino que además es causal, es decir, no se debe al efecto causal simultáneo de terceras variables  $Z$ .

*Ejemplo de relación causal entre X e Y:*



Por ejemplo, el nivel de estudios de una persona (X) determina sus ingresos (Y), o el nivel de estudios de una persona determina sus ingresos debido a los conocimientos que proporciona (T). Tanto uno como otro modelo implican la existencia de una relación causal entre el nivel de estudios e ingresos.

*Ejemplo de relación no causal entre X e Y:*



Por ejemplo, alguien podría argumentar que la relación entre nivel de estudios e ingresos es espúrea, es decir que aunque existe una asociación entre las dos variables, ésta no se debe al efecto causal del nivel de estudios sobre los ingresos, sino a que tanto el nivel de estudios como los ingresos vienen determinados por el origen social de las personas.

Una vez descartada la posibilidad de que la relación entre las dos variables sea espúrea, el investigador diseña su investigación de manera que pueda explicar dichas relaciones causales mediante la intervención de terceras variables. Es decir que, por ejemplo, el investigador no se contenta con demostrar que existe una relación causal entre el nivel de estudios de una persona y sus ingresos, sino que también intenta explicar esta relación causal por la mediación de terceras variables (por ejemplo, el nivel de conocimientos proporcionado por un determinado nivel de estudios).

Pasamos de un modelo como el siguiente:

$$X \rightarrow Y$$

a otro modelo más complejo como el siguiente:



Además, el investigador se interesa muchas veces en analizar la relación causal entre dos variables, dados varios contextos diferentes. Por ejemplo, intenta averiguar si la relación causal entre el nivel de estudios y los ingre-

Z  
Causa Suete
 sos de las personas es igual en España que en Estados Unidos. En este ejemplo, el país considerado constituye la variable  $T$  utilizada para analizar la relación causal entre  $X$  e  $Y$  en distintos contextos.

Eliminar  
Variables
 Sólo la distribución aleatoria de los objetos de análisis (ya sean personas, agregados de personas, o lo que sea) en un grupo experimental y otro de control permite alcanzar estos objetivos. Por ejemplo la mejor manera de determinar si el nivel de estudios de las personas determina sus ingresos consistiría en asignar aleatoriamente, es decir mediante cualquier procedimiento que se base en la suerte o azar, quién va a la Universidad, quién abandona sus estudios al final de la secundaria, para después de algún tiempo medir los ingresos obtenidos por estas personas después de haber iniciado su ejercicio profesional. Por supuesto, tanto este experimento como muchos otros chocan con obstáculos éticos que hacen que la mayoría de las veces nos tengamos que contentar con aproximaciones al diseño experimental. En estas situaciones, la asignación aleatoria de los objetos de estudio al grupo experimental y al grupo de control no es posible y ello hace que nunca estemos seguros de que la asociación entre dos variables es causal y no espúrea. Lo más que podemos hacer es ir eliminando hipótesis alternativas.

97m  
Factor
↘
 Un último objetivo ligado a muchos diseños de investigación es el comparar el efecto causal de una variable con el de otras. Por ejemplo, podemos intentar determinar si el nivel de estudios es más importante a la hora de determinar los ingresos de las personas que los ingresos de los padres de dichas personas. En última instancia tratamos de hallar un número de variables determinado que nos permita predecir de la manera más precisa posible la variación de la variable dependiente  $Y$ . Por ejemplo, intentamos encontrar un número determinado de variables que nos permita predecir con máxima certeza los ingresos de las personas.

En resumen, el investigador diseña su trabajo de modo que pueda determinar si dos o más variables están relacionadas, de manera que pueda determinar si dicha relación es causal, de manera que determine qué variables median o preceden a esa relación causal, de manera que pueda determinar la variabilidad de la relación causal en varios contextos distintos, de manera que determine la existencia de otras variables explicativas y su efecto causal relativo sobre la variable dependiente, y de manera que maximice nuestra capacidad de predicción de la variable dependiente  $Y$ .

La mayoría de las veces el investigador analiza relaciones causales dentro de diseños no experimentales, es decir en situaciones en las cuales los objetos de estudio no han sido asignados aleatoriamente a los grupos experimental y de control. Por ejemplo, no podemos decidir de forma aleatoria sobre el nivel de estudios alcanzado por una persona. Independientemente de cuestiones éticas, es obvio que la sociedad se opondría a que un investigador decidiese aleatoriamente cuándo deben acabar sus estudios los individuos escogidos para un estudio de las relaciones entre nivel de estudios

x 2  
x 3-  
stand  
a  
unit of

e ingresos. Los individuos escogidos para nuestro estudio vienen ya con un nivel de estudios dado que no podemos determinar libremente. En dichas situaciones, la única manera de establecer una relación causal entre nivel de estudios ( $X$ ) e ingresos ( $Y$ ) consiste en eliminar cuantas variables explicativas potenciales sea posible. Otras vías posibles para reforzar empíricamente la hipótesis sobre la existencia de una relación causal entre las dos variables consistiría en verificar el cumplimiento de predicciones respecto a las variables ( $T$ ) que median aquella relación o en verificar el cumplimiento de predicciones respecto al tipo de relación existente entre las dos variables cuando variamos el contexto donde esta relación se desarrolla. Estas dos últimas alternativas en ningún modo tienen el mismo valor analítico derivado de descartar el máximo número de variables explicativas alternativas.

En resumen, cuando nos hallamos en condiciones no experimentales no nos queda más remedio que introducir variables de control ( $T$ ) en el análisis, para asegurarnos que la relación entre  $X$  e  $Y$  no es espúrea. Las variables  $T$  incluidas dentro de modelos estadísticos cumplen varias funciones. En primer lugar pueden cumplir una función interpretativa. Es decir que introducimos  $T$  en el análisis para mostrar de qué modo se produce la relación entre  $X$  e  $Y$  o, en otras ocasiones, para mostrar qué variables antecedentes originan la relación entre  $X$  e  $Y$ :

a1.



a2.



En otras ocasiones introducimos variables  $T$  porque pensamos que la relación entre  $X$  e  $Y$  es espúrea:

a3.



En otras ocasiones introducimos variables  $T$  porque queremos determinar la capacidad explicativa relativa de distintas variables:

a4.



Somos nosotros los que a partir de argumentaciones teóricas decidimos el rol que juega la variable  $T$  dentro de nuestro análisis. Los datos nunca nos dirán cuál de los tres modelos presentados en las líneas precedentes es más adecuado. Lo único que nos dirá el análisis estadístico es la validez de nuestro modelo así como el valor de los distintos componentes de nuestro modelo. Hay ocasiones en las cuales nos es fácil decidir cuál de los cuatro modelos precedentes es más adecuado. Estas ocasiones se producen cuando existe clara precedencia temporal de unas variables sobre otras. Sin embargo, el diseño de investigación utilizado en algunas ocasiones facilita y en otras dificulta la determinación de un orden de precedencia. Los diseños longitudinales (por ejemplo, dos tandas de entrevistas al mismo grupo de personas) la facilitan, mientras que los diseños no longitudinales (por ejemplo, una sola entrevista a un grupo determinado de personas) la dificultan.

Imaginemos, por ejemplo, que queremos analizar la relación causal entre la ideología política y los ingresos. Se podría argumentar que la gente de izquierdas, al valorar menos el bienestar material, tenderá a comportarse de manera a obtener menos ingresos, mientras que la gente de derechas, al valorar más dicho bienestar material, tenderá a comportarse de manera a obtener más ingresos. Del mismo modo se podría argumentar que la gente que tiene menos ingresos tenderá a desarrollar una ideología más de izquierdas, mientras que la gente que tiene más ingresos tenderá a desarrollar una ideología más de derechas. Un diseño no longitudinal difícilmente podrá demostrarnos el orden de precedencia causal de una variable sobre la otra. Lo más que podremos observar en la mayoría de las ocasiones es el grado de asociación entre las dos variables. Los datos nunca nos dirán el orden de precedencia apropiado. Sin embargo, si tenemos datos para las mismas personas recogidos en dos momentos distintos, podemos al menos establecer la relación causal de una variable sobre la otra en un período de tiempo determinado.

En resumen, los diseños de investigación no experimentales, mayoritariamente utilizados por científicos sociales, impiden el que podamos estar seguros sobre la existencia de una relación causal entre dos variables. Nuestro esfuerzo va encaminado la mayoría de las veces a introducir variables  $T$  en nuestros modelos que nos permitan determinar la importancia relativa de nuestra variable  $X$  a la hora de explicar  $Y$ , así como la posibilidad de que la relación entre  $X$  e  $Y$  sea espúrea. Una vez descartadas una serie de variables explicativas alternativas para la asociación existente entre  $X$  e  $Y$ , podemos empezar a introducir otras variables  $T$  en nuestros modelos que nos ayuden a interpretar la relación entre  $X$  e  $Y$ . Ahora bien, los datos nunca nos dirán el rol que juega  $T$  dentro de estos modelos, del mismo modo que no nos dicen si  $X$  causa  $Y$  o viceversa. Solamente nuestra teoría puede indicarnos tanto lo uno como lo otro.