

3. Causalidad e inferencia causal

Hemos analizado dos estadios de la investigación social: resumen de los pormenores históricos y extracción de inferencias descriptivas mediante la división del mundo en componentes sistemáticos y no sistemáticos (en los apartados 5 y 6, respectivamente, del capítulo anterior). Muchos estudiosos de los fenómenos sociales y políticos se detendrían en este punto, evitando hacer afirmaciones causales y pidiendo a unos hechos seleccionados y bien ordenados que «hablen por sí mismos».

Al igual que los historiadores, los científicos sociales necesitan resumir los pormenores históricos y hacer inferencias descriptivas. Sin embargo, cuando las ciencias sociales pretenden alcanzar ciertos objetivos, si no hay inferencia causal, el análisis es insuficiente. Esto significa que, al igual que una inferencia causal es imposible sin una buena inferencia descriptiva, ésta, por sí misma, suele resultar insatisfactoria e incompleta. No obstante, con esto no queremos decir que todos los científicos sociales tengan que intentar, en cualquiera de sus trabajos, dar explicaciones causales a los fenómenos que estudian. A veces la inferencia causal es demasiado difícil y en otras muchas situaciones la descriptiva constituye el fin último de la empresa investigadora. Es evidente que siempre tenemos que dejar claro si el objetivo de nuestro proyecto de investigación es describir o explicar. Muchos científicos sociales no se sienten cómodos con la inferencia causal. Les preocupa tanto la advertencia de que «correlación no es causalidad» que no plantean hipótesis o inferencias de tipo causal y señalan que su trabajo «estudia asociaciones y no causas». Otros caen con facilidad en afirmaciones aparentemente

estudio, como una persona, un país, un año o una organización política. Para mayor precisión y claridad, hemos elegido un ejemplo corriente de la investigación cuantitativa: el efecto causal que tiene sobre la proporción de votos que recibe un candidato demócrata a la Cámara de Representantes estadounidense el hecho de que ya esté en el poder (utilizar sólo un candidato demócrata simplifica el ejemplo). La variable dependiente será la proporción de votos que obtienen los demócratas en el sistema bipartidista de la Cámara. Por tanto, la variable explicativa clave es dicotómica: el demócrata está en el poder o no lo está (para lograr una mayor sencillez, en este apartado sólo tenemos en cuenta las circunscripciones en las que el candidato republicano fue derrotado en los últimos comicios).

El lenguaje causal puede ser confuso, y nuestra opción aquí no es muy original. A veces la «variable dependiente» se denomina «variable de resultados» y con frecuencia las «variables explicativas» son «variables independientes». Dividimos las explicativas entre «variables causales principales» (también denominada «causas» o «variable de tratamiento») y «variables de control». Para terminar, la variable causal principal siempre adopta dos o más valores, que suelen conocerse como «grupo de tratamiento» y «grupo de control».

Pensemos ahora únicamente en la cuarta circunscripción electoral de Nueva York para las elecciones al Congreso e imaginemos que en 1998 hubieran tenido lugar unos comicios en los que se hubieran enfrentado un congresista demócrata y un aspirante republicano (no en el poder). Supongamos que el candidato demócrata hubiera recibido en estas elecciones una fracción del voto expresada como y_4^d (el subíndice 4 indica la cuarta circunscripción de Nueva York y el superíndice d el hecho de que el demócrata sea ya congresista). Por tanto, y_4^d es un valor de la variable dependiente. Para definir el efecto causal (una cantidad *teórica*), imaginemos que damos marcha atrás en el tiempo hasta el comienzo de la campaña electoral y que todo sigue igual, a excepción de que el congresista demócrata decide no presentarse a la reelección y que su partido nombra a otro candidato (probablemente al ganador de las elecciones primarias). Expresamos la fracción de voto que el candidato demócrata (no congresista) recibió mediante y_4^d (Nótese que el candidato demócrata que no está en el poder).

Esta condición *contrafáctica* es la esencia de la definición de causalidad. En esta situación hipotética, entre voto real (y_4^r) y voto probable (y_4^d) es el efecto causal, concepto que definimos con mayor precisión y cuidado al definir mucho cuidado al definir contrafácticos por-que, aunque se oponen evidentemente a los hechos, tienen que ser razonables: debería ser posible que el acontecimiento contrafáctico hubiera ocurrido en unas determinadas circunstancias. Al definir la condición contrafáctica adecuada resulta esencial precisar qué se mantiene constante cuando alteramos el valor de la variable de tratamiento. En nuestro ejemplo, la variable causal principal (o de tratamiento) es estar en el poder, y

causales, calificando de «explicaciones» hipótesis o especulaciones no evaluadas, que se basan en diseños de investigación imprecisos. Creemos que ambas posturas esquivan el problema de la inferencia causal.

Evitar el lenguaje causal cuando la causalidad es el auténtico objeto de la investigación o bien hace irrelevante el estudio o bien le permite no respetar las reglas de la inferencia científica. Nuestra incertidumbre acerca de las inferencias causales nunca desaparecerá, pero esto no debe significar que evitemos extraerlas. Por el contrario, habría que hacerlas en cualquier lugar que puedan aportarlas, pero aportando al lector, al mismo tiempo, la mejor y más sencilla estimación de su incertidumbre. Hay que ser valiente al hacer inferencias causales, siempre que detallemos cuidadosamente su incertidumbre. Además, es importante que las hipótesis causales se disciplinen, acercándose lo más posible a las reglas de la inferencia causal. Entre los capítulos 4 y 6 casi toda nuestra atención se dedica a explicar las circunstancias en las que dicha inferencia es apropiada y a facilitar que las investigaciones de los cualitativistas tengan más posibilidades de aportar pruebas fiables de sus hipótesis causales.

En el apartado 1 damos una definición rigurosa de causalidad, que sirve igualmente para la investigación cualitativa que para la cuantitativa; después, en el apartado 2, clarificamos algunos conceptos de causalidad alternativos que figuran en la bibliografía sobre el tema y demostramos que no entran en conflicto con nuestra definición más fundamental. En el apartado 3 analizamos los presupuestos precisos sobre el mundo y las hipótesis que se necesitan para hacer inferencias causales fiables. Posteriormente, en el apartado 4, nos centramos en cómo aplicar a la inferencia causal los criterios que desarrollamos para juzgar la descripción. En el apartado 5 terminamos el capítulo con consejos más generales sobre cómo elaborar explicaciones, teorías e hipótesis de tipo causal.

1. Definición de causalidad

En este apartado, definimos la causalidad como un concepto *teórico* que es independiente de los datos utilizados para conocerlo. A continuación, entramos a considerar la *inferencia* causal a partir de nuestros datos (para un análisis de los problemas específicos de este tipo de inferencia, véanse los capítulos 4 al 6). En el apartado 1.1 planteamos con todo detalle nuestra definición de causalidad, junto a un simple ejemplo cuantitativo, mientras que en el 1.2 retomamos esta definición con un caso cuantitativo más elaborado.

1.1 Definición y un ejemplo cuantitativo

Nuestra definición teórica de causalidad se aplica de la forma más simple y clara a una sola unidad. Una unidad, tal como se definió en el apartado 4 del capítulo 2, es uno de los muchos elementos que hay que observar en un

siempre necesitan dividir el mundo en componentes sistemáticos y no sistemáticos, mientras que la definición de Holland no establece claramente esta distinción⁴. Para ver la importancia de esta división, pensemos en qué ocurriría se pudiéramos repetir la campaña electoral de 1998 en la cuarta circunscripción de Nueva York, con el congresista demócrata y el aspirante republicano al cargo. La suma total de votos tendría un resultado ligeramente diferente, a causa de ciertas características no sistemáticas de las campañas electorales: aspectos políticos que no se mantienen de unas campañas a otras, incluso si éstas parten de situaciones idénticas. Entre estas características no sistemáticas se pueden incluir las meteduras de pata verbales, un discurso o pronunciamiento sobre un problema que concita una atención sorprendente, una actuación inesperadamente desafortunada en un debate, mal tiempo durante un mitin del candidato o el día de las elecciones o también los resultados de alguna investigación periodística. Por lo tanto, podemos imaginar una variable que representara los valores del voto demócrata en diversas reproducciones hipotéticas de las mismas elecciones.

Como hemos señalado anteriormente (véase el apartado 6 del capítulo 2), esta variable se denomina «variable aleatoria», ya que tiene características no sistemáticas: le afectan las variables explicativas que no contienen nuestro análisis teórico o contiene una variabilidad fundamentalmente inexplicable⁵. Definimos la variable aleatoria que representa la proporción de votos recibida por el candidato demócrata que ya es congresista como Y_M^d (obsérvese la Y mayúscula), y la proporción de los que habría recibido un demócrata no congresista en reproducciones hipotéticas, como Y_M^a . Ahora definimos el *efecto causal aleatorio* en la circunscripción 4 como la diferencia entre estas dos variables aleatorias. Como queremos mantener cierto carácter general, de nuevo utilizamos una notación que se refiere a la unidad i en vez de a la circunscripción 4:

$$(3.2) \quad \text{Efecto causal aleatorio en la unidad } i = (Y_i^d - Y_i^a)$$

(Al igual que en la definición de variable aleatoria, un efecto causal del mismo tipo es el que cambia en cada reproducción hipotética del mismo experimento, pero también representa muchas interesantes características sistemáticas de las elecciones.) Si pudiéramos observar a la vez dos proporciones de voto diferentes en la circunscripción 4 —la primera referida a unas elecciones a las que se presentara un congresista demócrata y la segunda sobre otras sin él—, podríamos ver directamente el efecto causal realizado de la ecuación (3.1). Es evidente que, a causa del problema fundamental de inferencia causal, no podemos observar el efecto causal realzado, en consecuencia, en la ecuación 3.1, dicho efecto es una única aplicación *no observada* del efecto causal aleatorio de la ecuación 3.2. Dicho de otro modo, si en la circunscripción 4 se hacen diferentes reproducciones hipotéticas de las mismas elecciones, unas con un candidato congresista de-

cambia desde «ser congresista» a «no ser congresista». Durante esta transición, mientras que los elementos constantes (la fuerza relativa de los demócratas o los republicanos en esa circunscripción en elecciones anteriores, el carácter del proceso de nominación del candidato, las características de la circunscripción, el contexto económico y político del momento, etc.) hasta que el Partido Demócrata decide su candidato. No sometemos a ningún control las cualidades de los candidatos, como en qué medida se reconoce su nombre, su visibilidad o si se sabe cómo funciona el Congreso, ni tampoco cualquier otra consecuencia que tenga la nominación por parte del partido. Esto se debe a que, en parte, estos elementos son *consecuencia* de nuestra variable de tratamiento: estar en el poder, lo cual quiere decir que, entre las ventajas de esta situación, se incluye el reconocimiento del nombre, la visibilidad y todo lo demás. Si realmente mantuviéramos constantes estos elementos, estaríamos controlando algunos de los efectos más importantes que tiene estar en el poder Y , por tanto, haciendo caso omiso de ellos. En consecuencia, malinterpretaríamos su influencia global en el conjunto del voto. De hecho, controlar un número suficiente de las consecuencias que tiene estar en el poder podría llevarnos a la incorrecta conclusión de que dicha situación no tiene efecto alguno⁶.

De manera más formal, el efecto causal que tiene estar en la cuarta circunscripción de Nueva York —la proporción de votos recibida por el candidato demócrata que puede atribuirse al hecho de que ya sea congresista — sería la diferencia entre estas dos fracciones del voto: $(Y_i^d - Y_i^a)$. Por razones que pronto quedarán claras, esta diferencia la denominamos *efecto causal realizado*, y, para expresarla con una notación más general, utilizamos la unidad i en vez de sólo la circunscripción 4⁷:

$$(3.1) \quad \text{Efecto causal realizado en la unidad } i = Y_i^d - Y_i^a$$

Es obvio que este efecto se define únicamente en teoría, ya que, en cualquier elección real, sólo podríamos observar Y_i^d o Y_i^a o ninguno de los dos, pero nunca ambos a la vez. De este modo, esta sencilla definición de causalidad demuestra que nunca podemos conocer con certeza un efecto causal. Para Holland (1986), éste es el *problema fundamental de la inferencia causal*, y en verdad lo es porque, independientemente de lo perfecto que sea el diseño de la investigación, de la cantidad de datos que recojamos, de lo perspicaces que sean los observadores, de lo diligentes que sean los ayudantes y del grado de control experimental que tengamos, nunca conoceremos a ciencia cierta la inferencia causal. De hecho, gran parte de las cuestiones empíricas de los diseños de investigación que se analizan en este libro tienen que ver con este problema fundamental y casi todas las indicaciones que damos son intentos parciales de evitarlo.

Nuestra definición básica de causalidad se aparta de la de Holland, ya que en el apartado 6 del capítulo 2 hemos señalado que las ciencias sociales

mócrata y otras tantas con uno no congresista, el efecto causal realizado (no observado) se convertirá en un efecto causal aleatorio.

Describir la causalidad como una de las características sistemáticas de las variables aleatorias puede parecer excesivamente complicado, pero tiene dos ventajas. La primera es que nuestra definición de causalidad sea directamente análoga a las características sistemáticas de un fenómeno (como son la media o la varianza) que sirven de objeto a la inferencia descriptiva: las medias y las varianzas también son características sistemáticas de las variables aleatorias (como en el apartado 2 del capítulo anterior). La segunda es que permite la división de un problema inferencial causal en sus componentes sistemáticos y no sistemáticos. Aunque muchas de las características sistemáticas de una variable aleatoria puedan ser de interés, la más relevante para nuestro simple ejemplo es el *efecto causal medio* de la unidad *i*. Para explicar que significa esto retomamos nuestro ejemplo de las elecciones en Nueva York.

Recuérdese que la variable aleatoria representa la fracción del voto recibida por el candidato demócrata (ya sea congresista o no) en un número elevado de reproducciones hipotéticas de los mismos comicios. Definimos de la siguiente manera el valor esperado de esta variable aleatoria —la media de fracciones del voto en dichas reproducciones— para el no congresista:

$$E(Y^i) = \mu^i$$

y para el congresista como:

$$E(Y^j) = \mu^j$$

En consecuencia, el efecto causal medio de estar en el poder para la unidad *i* es una característica sistemática del efecto causal aleatorio, y se define como la diferencia entre estos dos valores esperados (utilizamos de nuevo, para generalizar, la unidad *i* en vez de la circunscpción 4):

Efecto causal

$$\text{medio en la unidad } i = \beta$$

$$= E(\text{Efecto causal aleatorio en la unidad } i)$$

$$= E(Y^i - Y^j)$$

$$= E(Y^i) - E(Y^j)$$

$$= \mu^i - \mu^j$$

de modo que en la primera línea de esta ecuación, β (beta) expresa este efecto causal medio. En la segunda indicamos que en la unidad *i* este indicador es sólo la media (valor esperado) del efecto causal aleatorio, mientras que en las líneas tercera y cuarta se muestra cómo calcular la media. La última línea es otra manera de expresar la diferencia entre las medias de los dos conjuntos de elecciones hipotéticas (la media de la diferencia de dos variables aleatorias es igual a la diferencia de sus medias). Para resumirlo con palabras: *el efecto causal es la diferencia que existe entre el componente sistemático de las observaciones que se hacen cuando la variable explicativa tiene un valor y el componente sistemático de observaciones compuestas cuando la variable explicativa tiene otro valor.*

La última línea de la ecuación 3.3 es similar a la ecuación 3.1, y, al ser así, el problema fundamental de la inferencia causal existe todavía en esta fórmula. De hecho, expresado de esta manera, el problema es aún más grande, porque, aunque pudiéramos esquivar el problema fundamental en un efecto causal realizado, todavía seguiríamos teniendo todos los hábitos les en la inferencia, incluido el de separar los componentes sistemáticos y no sistemáticos del efecto causal aleatorio. A partir de ahora, utilizaremos la expresión de Holland, «problema fundamental de la inferencia causal», para hacer referencia a la dificultad que el identificador *y también* a los problemas inferenciales habituales, que hemos añadido a la formulación de este autor. En el recuadro de la página 108 mostramos una notación más general de los efectos causales, que resultará útil en el resto del libro.

Hay otras muchas características sistemáticas de estos efectos causales aleatorios que podrían ser de interés en diversas circunstancias. Por ejemplo, quizá queramos saber la varianza de los posibles efectos causales (realizados) que tiene el hecho de estar en el poder sobre el voto demócrata en la unidad *i*, al igual que quisimos conocer la del propio voto mediante la ecuación 2.3 del apartado 6 del capítulo anterior. Para calcular este indicador del efecto causal, aplicamos la operación de la varianza:

$$(\text{varianza del efecto causal en la unidad } i) = V(Y^i - Y^j)$$

en la que evitamos introducir un nuevo símbolo que represente el resultado del cálculo de la varianza, $V(Y^i - Y^j)$. Con seguridad, los nuevos congresistas querían saber la varianza del efecto causal que tiene el hecho de estar en el poder para poder evaluar en qué medida su experiencia se parecerá a la de los congresistas anteriores y hasta qué punto pueden fiarse de la estimación del efecto causal medio que tuvo el hecho de estar en el poder en elecciones anteriores. Resulta especialmente importante comprender que esta varianza del efecto causal es una parte fundamental del mundo y que no representa una incertidumbre causada por la estimación.

En el apartado I hemos desarrollado nuestra precisa definición de causalidad y, al ser delicados y bastante complejos algunos de los conceptos que se barajaban, ilustramos la explicación con un ejemplo simple y corriente de la investigación cuantitativa, que nos ayudó a transmitir los conceptos que queremos recalcar sin tener que ocuparnos de los detalles contextuales y de la sensibilidad cultural que caracterizan la investigación cualitativa de calidad. En este apartado retomamos nuestra definición de causalidad, pero esta vez utilizamos un ejemplo cualitativo.

Los politólogos aprenderían mucho si pudieran volver a poner en marcha la historia manteniendo todos sus elementos constantes, a excepción de la variable explicativa controlada por el investigador. Por ejemplo, una de las cuestiones más importantes a las que se enfrentan los que participan de la política y el gobierno es la relacionada con las consecuencias de una determinada ley o normativa. El Congreso de los Estados Unidos aprueba una ley impositiva para que tenga ciertas consecuencias: que conduzca a unas inversiones concretas, incrementalmente los ingresos en cierta cantidad y cambie las pautas de consumo. ¿Acaso tiene tal efecto? Podemos observar lo que ocurre después de que se aprueba el impuesto, para comprobar si surgen las consecuencias deseadas, pero, aunque sea así, nunca estaremos seguros de que *provengan* de la ley. El cambio en la política de inversiones podría haberse producido en cualquier caso. Si fuera posible volver a poner en marcha la historia, con y sin la nueva normativa, tendríamos mucho más control a la hora de hacer una estimación del efecto causal de la ley. Evidentemente, no podemos hacerlo, pero la lógica nos ayudará a diseñar una investigación que de una respuesta aproximada a nuestra pregunta.

Pensemos ahora en un ejemplo general de la política comparada. Después del derrumbe del sistema soviético numerosos gobiernos de las repúblicas ex soviéticas de Europa del Este instauraron nuevas formas de gobierno. Ahora se hallan inmersos —y son conscientes de ello— en un importante experimento político: están aprobando nuevas constituciones con la intención de que estas creen sistemas democráticos estables. Uno de los dilemas constitucionales que se les presentan es elegir entre sistemas de gobierno parlamentarios o presidenciales. Determinar cuál de ellos tiene más posibilidades de conducir a una democracia estable ha suscitado un considerable debate entre los académicos del área (Linz, 1993; Horowitz, 1993; Lijphart, 1993). La polémica es compleja, principalmente porque hay muchos tipos de sistemas parlamentarios y presidenciales (como la naturaleza de las demás disposiciones constitucionales) que podrían acompañar a este dilema (que podrían acompañar a este dilema) que podrían acompañar a este dilema con el régimen electoral) que podrían acompañar a este dilema e interactuar con él. Nuestro objetivo no es analizar en profundidad estas opciones, sino dar una versión muy simplificada del dilema que plantean, con el fin de definir un efecto causal dentro de este ejemplo cualitativo. De este modo, su-

brayamos la diferencia que hay entre las características sistemáticas y no

sistemáticas del efecto causal.

El debate sobre sistemas presidenciales y parlamentarios tiene que ver con diversas características de ambos. Nos centraremos en dos de ellas: en qué medida representa cada ordenamiento los diversos intereses de la ciudadanía y si favorece la existencia de un liderazgo fuerte y con capacidad de decisión. Se suele decir que los sistemas parlamentarios permiten una mejor representación de todos los grupos e intereses sociales en el gobierno, ya que hay muchos escaños que llenar y los pueden ocupar representantes elegidos por diversos sectores. Por el contrario, el carácter drástico (todo o nada) de los sistemas presidenciales conlleva que algunos grupos se sientan apartados del gobierno, le sean desafectos y generen una mayor inestabilidad. Por otra parte, es posible que los regímenes parlamentarios —especialmente los que representan de manera adecuada a todo el arco de grupos e intereses sociales— se bloqueen y que no cuenten con un gobierno no decidido. También estas características pueden conducir a la desafección y la inestabilidad.

El objetivo principal de este apartado es formular una definición precisa de efecto causal. Para ello, imaginemos que podemos instaurar un sistema parlamentario y que, periódicamente, durante unos diez años, cambiamos su grado de estabilidad democrática (quizá la propia supervivencia o desaparición del sistema, así como los intentos de golpe de Estado que sufre u otros indicadores de inestabilidad); a la vez, en el mismo país, fundamos un sistema presidencial y también medimos su estabilidad en el mismo período y con los mismos elementos. El *efecto causal realizado* sería la diferencia entre los grados de estabilidad observados en el sistema presidencial y en el parlamentario. La imposibilidad de calibrar este efecto causal directamente es otro ejemplo del problema fundamental que plantea la inferencia causal. Como parte de esta definición, también necesitamos distinguir los efectos sistemáticos y no sistemáticos de la forma de gobierno. Para ello, imaginemos que llevamos a cabo este experimento hipotético muchas veces. Definimos el *efecto causal medio* como la media de los efectos causales realizados en las repeticiones de estos experimentos. Utilizar así la media hace que se anulen las características no sistemáticas del problema y permite que nuestro indicador incluya sólo las sistemáticas. Entre estas se encuentran la indecisión en un sistema parlamentario o la desafección de las minorías en uno presidencial, mientras que entre las no sistemáticas se puede incluir una repentina enfermedad del presidente que precipita el gobierno en el caos. Este último acontecimiento no sería un rasgo persistente en un sistema presidencial; aparecería en una de las repeticiones del experimento, pero no en otras.

Otro rasgo notable de este ejemplo es la varianza del efecto causal. A cualquier país que se esté planteando elegir entre estos sistemas políticos le interesará conocer el efecto causal medio que tiene uno u otro sobre la esta-

bilidad democrática; sin embargo, para cada país sólo hay una oportunidad sólo una reproducción de este experimento. En tal situación, puede que a los líderes políticos les interese algo más que el efecto causal medio. Quizá quieran comprender cuáles podrían ser los efectos causales máximo y mínimo o, al menos, la *variabilidad* de los efectos causales. Por ejemplo, quizá el sistema presidencial reduzca la estabilidad democrática media, pero la variabilidad de este efecto es enorme: a veces aumenta mucho la estabilidad y otras la disminuye de forma considerable. Esta variabilidad se convierte en un riesgo para el sistema político. En tales circunstancias, puede que los ciudadanos y los líderes políticos prefieran optar por un ordenamiento que genere una estabilidad media sólo un poco menor, pero con una variabilidad también inferior en su efecto causal, de manera que se minimicen las posibilidades de llegar a un resultado desastroso.

2. Clarificación de las definiciones alternativas de causalidad

En el apartado 1 hemos definido la causalidad en función de un efecto causal: el efecto causal medio es la diferencia que hay entre los componentes sistemáticos de una variable dependiente cuando la variable causal tiene dos valores diferentes. En este apartado, utilizamos nuestra definición de causalidad para clarificar diversas propuestas alternativas e ideas aparentemente complejas. Demostramos que las importantes afirmaciones — contempladas en este mismo capítulo — que han hecho otros autores acerca de los «mecanismos causales» (apartado 2.1), la causalidad «múltiple» (apartado 2.2) y la causalidad «simétrica» frente a la «asimétrica» (apartado 2.3) no entran en conflicto con nuestra más básica definición de causalidad.

2.1 «Mecanismos causales»

Algunos investigadores señalan que la idea principal de la causalidad es el conjunto de «mecanismos causales» que se dice que existen entre causa y efecto (véase Little, 1991, p. 15). De forma intuitiva, este punto de vista tiene sentido: cualquier explicación coherente que se dé a la causalidad tiene que precisar cómo se producen los efectos. Por ejemplo, supongamos que a un investigador le interesen las consecuencias que puede tener un nuevo acuerdo fiscal entre los Estados Unidos y Japón sobre el déficit de la balanza de pagos del primer país respecto al segundo. Según nuestra definición de causalidad, aquí el efecto causal se basaría en comparar la reducción del déficit de la balanza de pagos esperado, cuando se aplique el nuevo acuerdo fiscal, y la misma situación (en igual momento y con los mismos países) sin que haya entrado en vigor el tratado. A su vez, el meca-

mismo causal que opera en este proceso conllevaría — hasta que se produjera el efecto final sobre la balanza de pagos estadounidense respecto a Japón — la firma y ratificación del acuerdo, la información que dieran los periódicos, las reuniones de los actores afectados en las compañías múltiples nacionales y las acciones compensatorias que realizaran para reducir la carga fiscal de sus empresas (como el cambio en las reglas de fijación de precios de transferencia o el traslado de las fábricas de un país a otro), así como las acciones de otras compañías y trabajadores que quisieran aprovecharse de los movimientos de capital y de fuerza productiva entre los países, etc.

Desde el punto de vista de los procesos con los que opera la causalidad, hacer hincapié en los mecanismos causales tiene sentido intuitivamente: toda explicación coherente de la causalidad tiene que explicar de qué manera se producen sus efectos. Identificar mecanismos causales es una forma habitual de hacer análisis empíricos a la que se ha denominado con expresiones ligeramente diferentes: «rastreo de procesos» (que analizaremos en el apartado 3.3 del capítulo 6), «análisis histórico» y «estudios de caso de tallados». Muchos de los portadores de los estudios de caso bien hechos conllevan la identificación de mecanismos causales.

Sin embargo, para precisar cuáles son tales mecanismos es necesario utilizar la inferencia causal, empleando los métodos que se analizan posteriormente. Esto quiere decir que el investigador, para demostrar el carácter causal de cada posible vínculo en un mecanismo como ese, tendrá que definir el efecto causal subyacente y luego hacer una estimación del mismo. Si se quiere representar un mecanismo causal con coherencia interna es necesario utilizar, para cada uno de los eslabones de la cadena de acontecimientos, la definición de causalidad fundamental que hemos dado en el apartado 1 de este capítulo.

De ahí que nuestra definición de causalidad sea, lógicamente, previa a la identificación de los mecanismos causales. Además, en las ciencias sociales siempre existen infinitas de pasos causales entre cualquier par de eslabones de esa cadena de mecanismos. Si postulamos que una variable explicativa produce otra dependiente, para utilizar un enfoque a base de «mecanismos causales» tendríamos que listar los vínculos causales que hay entre ambas variables. En esta definición también necesitaríamos determinar una serie de enlaces causales, definir la causalidad en cada par de variables, así como la conexión entre parejas diferentes. Este enfoque nos conducirá rápidamente a un regreso al infinito y en ningún momento ofrece por sí solo una definición precisa de causalidad, relativa a una causa y un efecto.

En el ejemplo referido al efecto que tiene un sistema presidencial o parlamentario en la estabilidad democrática (apartado 1.2) los mecanismos causales de la hipótesis incluyen una mayor desafección de las minorías en

que el nivel de renta de una persona depende tanto de que culmine con éxito estudios superiores como del hecho de que sus padres también lo hicieron. Resulta insuficiente cumplir una condición y no la otra. En este caso, necesitamos comparar diversas categorías de nuestra variable causal: los encuestados que cumplen ambas condiciones, los dos grupos en los que sólo se da una u otra y aquel que no cuenta con ninguna. De este modo, el concepto de «causalidad múltiple» exige más de nuestros datos, ya que ahora nuestras variables causales se dividen en cuatro categorías, aunque esto no haga necesario un cambio en nuestra definición de causalidad. Para ésta, tendríamos que medir la renta esperada en la misma persona, al mismo tiempo y en cada una de las cuatro situaciones.

Sin embargo, ¿qué ocurre si diferentes explicaciones causales producen los mismos valores en la variable dependiente? Supongamos, por ejemplo, que en una población de trabajadores fabriles nuestra variable causal (dicotómica) fuera el hecho de tener o no tener una licenciatura universitaria. En esta situación sería bastante razonable que licenciados y no licenciados tuvieran la misma renta (nuestra variable dependiente). Una de las razones puede ser que esta variable explicativa (los estudios universitarios) no tiene un efecto causal en la renta de los trabajadores fabriles, quizá porque la educación superior no ayuda a un mejor rendimiento. Por el contrario, hay diversas razones que pueden explicar que ambos grupos tengan el mismo nivel de renta. Quizá el de los licenciados tenga que ver con su título y el de los que no fueron a la universidad con sus cuatro años de antigüedad en el puesto. En esta situación, ¿no podríamos llegar a la conclusión de que la «educación universitaria» no tiene un efecto causal en los niveles de renta de los que van a convertirse en trabajadores fabriles?

Por fortuna, nuestra definición de causalidad requiere una mayor precisión en la condición contrafáctica. En el presente ejemplo los valores de la variable causal principal que hay que cambiar son (1) tener un título universitario frente a (2) carecer de él pero llevar cuatro años en un puesto. La variable dependiente es la renta anual inicial. En consecuencia, nuestro efecto causal se define de la siguiente manera: registramos la renta de una persona que termina su carrera y va a trabajar a una fábrica. Después, retrocedemos cuatro años y ponemos al mismo individuo a trabajar en la misma fábrica, en vez de mandarle a la universidad, y al cabo de cuatro años medimos «de nuevo» sus ingresos. La diferencia esperada entre estos dos niveles de renta en el mismo individuo es nuestra definición de efecto causal medio. En la situación actual, hemos imaginado que este efecto causal es igual a cero, pero esto no significa que «los estudios universitarios no tengan un efecto sobre la renta», sólo que la diferencia media entre los grupos de tratamiento (1) y (2) es cero. De hecho, no hay una única definición lógica del «efecto causal de los estudios universitarios», ya que no se puede definir un efecto causal sin que haya, al menos, dos condiciones. No tienen por qué ser las que hemos mencionado, pero es preciso identificarlas claramente.

conlleva, pero no determinar éstos sin definir el efecto causal. Para nosotros, identificar los mecanismos que hacen que una causa tenga un efecto suele servir de apoyo a una teoría y es un procedimiento operativo muy útil. A veces, determinar mecanismos causales puede darnos más control sobre una teoría, al convertir las observaciones de otros niveles de análisis en consecuencias de dicha teoría. El concepto también puede generar nuevas hipótesis causales que investigar. Sin embargo, no debemos confundir la definición de causalidad con el procedimiento operativo de identificar mecanismos causales, que no es definitivo pero sí suele ser útil.

2.2 «Causalidad múltiple»

En un trabajo reciente, Charles Ragin (1987, pp. 34-52) aboga por una metodología que utilice muchas variables explicativas y pocas observaciones, con el fin de que se pueda tener en cuenta lo que el denomina «causalidad múltiple». Esto quiere decir que «El fenómeno que se investiga tiene determinantes alternativos; lo que Mill (1843) llamaba el problema de la "pluritudad de causas"». Para la teoría general de sistemas (George, 1982, p. 11), este problema se denomina «equifinalidad». En situaciones con causas múltiples, estos autores señalan que un mismo resultado pueden producirlo combinaciones de diferentes variables independientes.¹⁰

Para Ragin, cuando diversas variables explicativas pueden explicar el mismo resultado en una variable dependiente, hay métodos estadísticos que rechazan equivocadamente la hipótesis de que tales variables tengan un carácter causal. Ragin tiene razón al afirmar que algunos modelos estadísticos (o relevantes diseños de investigación cualitativa) quizá no alerten al investigador de la existencia de una «causalidad múltiple», pero sí los hay adecuados (algunos de los que Ragin analiza) para manejar con facilidad este tipo de situaciones.

Además, los rasgos fundamentales de la «causalidad múltiple» pueden compatibilizarse con nuestra definición de causalidad y son los mismos en la investigación cuantitativa y en la cualitativa. Esta idea no contiene rasgos o requisitos teóricos nuevos. Consideremos, por ejemplo, la hipótesis de

Como ejemplo de la afirmación de Lieberman, imaginemos que en las elecciones al Congreso estadounidense de 1998 no se hubiera presentado ningún congresista en la cuarta circunscripción de Nueva York y que el candidato demócrata recibiera el 55 por cien de los votos. Lieberman definió el efecto causal de estar en el poder como el aumento de votos que se produciría si el demócrata ganador en 1998 se presentara, siendo ya congresista, a las siguientes elecciones del año 2000. Este efecto sería «sintético» si la ausencia de un candidato congresista en las elecciones posteriores (las del año 2002) hiciera que los votos volvieran a ser del 55 por cien. El efecto podría ser «asintético» si, por ejemplo, el congresista demócrata reuniera dinero y mejorara la organización de la campaña de su partido; en consecuencia, si ningún congresista se presentara a las elecciones del 2002, el candidato demócrata podría recibir más del 55 por cien de los votos.

El argumento de Lieberman es inteligente y de gran importancia. Sin embargo, para nosotros no constituye una definición de *causalidad* sino que sólo es aplicable a algunas *inferencias* causales: al proceso de aprendizaje sobre un efecto causal a partir de las observaciones existentes. En el apartado 1 de este capítulo hemos definido la causalidad en una sola unidad. En el ejemplo actual, un efecto causal puede definirse desde el punto de vista técnico en función de los acontecimientos hipotéticos que han tenido lugar únicamente en las elecciones de 1998 en la cuarta circunscripción de Nueva York. Nuestra definición se basa en la diferencia que se da en el componente sistemático del voto en esa circunscripción cuando se presenta un congresista y cuando no se presenta, si las elecciones, la fecha y la circunscripción son las mismas.

Por el contrario, el ejemplo de Lieberman no incluye cantidades hipotéticas y, por tanto, no puede ser una definición causal. Sólo se ocupa de lo que ocurriría en dos elecciones reales si la variable explicativa pasara de ser no congresista a congresista y de lo que sucedería si se diera el proceso no congresista —de congresista— en otros dos comicios. Cualquiera análisis empírico de este ejemplo plantearía numerosos problemas inferenciales. Analizamos muchos de los problemas relativos a la inferencia causal entre los capítulos 4 y 6. En el presente ejemplo podríamos preguntarnos si el efecto estimado sólo nos parecía mayor por no haber explicado el hecho de que se hubieran registrado recientemente una gran cantidad de ciudadanos en la cuarta circunscripción, o si la oleada de apoyo para el congresista demócrata en las elecciones se nos antojaba más pequeña de lo que era porque, necesariamente, descartábamos aquellas circunscripciones en las que los demócratas habían perdido en las primeras elecciones.

De este modo, es importante considerar los conceptos de causalidad «simétrica» y «asimétrica» de Lieberman en el contexto de la inferencia causal. Sin embargo, no deben confundirse con la definición técnica de causalidad que ya hemos dado en el apartado 1 de este capítulo.

Otro posible par de condiciones causales es el que surge de comparar a una persona con título universitario con otra que carece de él pero que tiene la misma antigüedad en el puesto que la primera. En cierto sentido, esto es poco realista, ya que el no universitario habrá hecho algo en los cuatro años en los que no ha estudiado, pero quizá estemos dispuestos a imaginar que en ese tiempo ha tenido otro empleo de poca importancia. Dicho de otro modo, este contrafáctico alternativo se basa en comparar el efecto de tener estudios universitarios con el de no tenerlos, siempre que la antigüedad en la empresa sea la misma. Si este elemento no se mantiene constante en las dos condiciones causales, las estimaciones de cualquier diseño de investigación darán los resultados de nuestro primer contrafáctico, y no los de esta versión revisada. Si el objetivo es el segundo pero no introducimos controles, nuestro análisis empírico será defectuoso por el «sesgo de la variable omitida» (que abordamos en el apartado 2 del capítulo 5).

De este modo, las cuestiones que se agrupan bajo la etiqueta de «causalidad múltiple» no confunden nuestra definición de causalidad, aunque existan más de los análisis posteriores. El hecho de que algunas variables dependientes, y quizá todas las interesantes de este tipo que hay en las situaciones tan habituales es definir la condición contrafáctica mediante una precisa concepción de cada efecto causal. En el capítulo 5 demostramos que, para estimar el único efecto causal que tiene interés, no es necesario identificar «todos» los de una variable dependiente (aunque esto fuera posible). El investigador puede centrarse en el único efecto que le interesa, llegar a conclusiones consistentes y pasar después a otros efectos que puedan ser relevantes (véanse los apartados 2 y 3 del capítulo 5) ¹¹.

2.3 Causalidad «simétrica» y «asimétrica»

Stanley Lieberman (1985, pp. 63-64) distingue entre lo que clasifica como formas de causalidad «simétricas» y «asimétricas» y se centra en las diferencias que se producen en los efectos causales cuando se aumenta una variable explicativa o cuando se reduce. Según sus propias palabras:

Al examinar, por ejemplo, la influencia causal de X_1 [una variable explicativa] sobre Y [una dependiente], también hay que considerar si los cambios producidos desde cualquier dirección en un valor dado de X_1 tienen las mismas consecuencias para Y [...] Si la relación causal entre X_1 [la variable explicativa] e Y [la dependiente] es simétrica o realmente reversible, el efecto que tendrá sobre Y un aumento de X_1 desaparecerá si X_1 vuelve a su nivel anterior (siempre que el resto de las condiciones sea constante).

3. Supuestos necesarios para la estimación de efectos causales

Como evitar el problema fundamental de la inferencia causal y también el de separar los componentes sistemáticos de los no sistemáticos? Los capítulos que van del 4 al 6 se dedicarán integralmente a dar una respuesta completa a esta pregunta, pero aquí proporcionamos una visión general de los dos supuestos que pueden evitar este problema fundamental: la *homogeneidad de las unidades* (que analizaremos a continuación en el apartado 3.1) y la *independencia condicional* (apartado 3.2). Al igual que cualquier intento de esquivar el problema fundamental de la inferencia causal, tales supuestos siempre conllevarán otros que no son comprobables. Cada investigador tiene la responsabilidad de mostrar a los lectores con toda claridad las consecuencias esenciales de este punto débil de su diseño de investigación. Las inferencias causales no deben parecer mágicas y los supuestos pueden y deben justificarse mediante cualquier información colateral o investigación anterior que se pueda aportar, siempre que se reconozca explícitamente.

3.1 Homogeneidad de las unidades

Si no podemos volver a poner en marcha la historia en el mismo momento y lugar, dando diferentes valores a nuestra variable explicativa en cada ocasión —tal como requeriría una auténtica solución para el problema fundamental de la inferencia causal—, si se puede partir de un segundo supuesto: poner otra vez en marcha nuestro experimento en dos unidades diferentes que sean «homogéneas». Dos unidades son homogéneas si los valores esperados de las variables dependientes de cada una de ellas son los mismos cuando nuestra variable explicativa adopta un valor determinado (o sea, $\mu_1^i = \mu_2^i$ y $\mu_1^j = \mu_2^j$) Por ejemplo, si observamos que $X = 1$ (un congresista) en la circunscripción 1 y $X = 0$ (no congresista) en la 2, la premisa de homogeneidad de las unidades supone que podemos utilizar las proporciones observadas del voto en dos circunscripciones separadas para extraer una inferencia sobre el efecto causal β , que presuponemos es el mismo en ambas. En un conjunto de datos con n observaciones, la homogeneidad de las unidades consiste en presuponer que todas las que tengan igual valor en sus variables explicativas tendrán también el mismo valor esperado en la variable dependiente. Es evidente que esto es sólo un supuesto y que puede ser erróneo: quizá ambas circunscripciones sean diferentes por alguna razón desconocida que sesgue nuestra inferencia causal. De hecho, cualquier parte de circunscripciones reales *difieren* en ciertos sentidos, y para aplicar este supuesto ambas tendrían que ser iguales en cuanto a la media de muchas reproducciones hipotéticas de la campaña electoral. Por ejemplo, la plivvosi-

dad (que puede reducir la asistencia a las urnas en algunas áreas) no varía, como promedio, de una circunscripción a otra, a menos que haya diferencias climáticas sistemáticas entre las dos.

En la cita siguiente, Holland (1986, p. 947) da un ejemplo claro del supuesto de homogeneidad de las unidades (definiendo a partir de su idea de efecto causal realizado y no en función del efecto causal medio). Al no existir apenas aleatoriedad en el experimento del siguiente ejemplo, su definición y la nuestra se parecen (de hecho, como planteamos en el apartado 2 del capítulo 4, en un pequeño número de unidades, el supuesto de la homogeneidad de las unidades es más útil cuando el grado de aleatoriedad es bastante bajo).

Si [la unidad] es la habitación de una casa, t [de «tratamiento»] significa que pulso el interruptor de la luz en esa habitación, c [de control] significa que no lo hago y [la variable dependiente] indica si la luz está encendida o no durante un corto periodo después de que se aplica t o c , de manera que podría *creer* que se pueden conocer los valores de [la variable dependiente, tanto en t como en c] al pulsar el interruptor. Sin embargo, está claro que mi *creencia* sólo la puede compartir alguien por la plausibilidad de ciertos supuestos relativos a la situación. Si, por ejemplo, la luz se ha estado apagando y encendiéndose sin razón aparente mientras estoy pensando en realizar mi experimento, podría dudar de que se puedan conocer los valores de [la variable dependiente tanto en t como en c] después de pulsar el interruptor, tal menos hasta que fuera lo suficientemente ingenuo como para concebir un nuevo experimento!

En este ejemplo, el supuesto de homogeneidad de las unidades consiste en que si hubiéramos pulsado el interruptor (aplicado t , según la notación de Holland) en ambos periodos, el valor esperado (si la luz iba a estar encendida o no) habría sido el mismo. La homogeneidad de las unidades también presupone que si no hubiéramos apretado el interruptor (aplicado c) en los dos periodos, el valor esperado habría sido el mismo, aunque no necesariamente igual a cuando se aplicara t . Hay que señalar que, para asegurarnos de esto, tendríamos que haber puesto el interruptor en la posición de apagado después del primer experimento, pero también tendríamos que haber partido de un supuesto improbable: que pulsar el encendido del interruptor en el primer periodo no tiene como consecuencia los dos valores hipotéticos esperados en el siguiente (como sería el caso si se fundieran los plomos después de la primera pulsación). En general, el supuesto de homogeneidad de las unidades no puede comprobarse en una única unidad (aunque, en este caso, abriendo la pared e inspeccionando la instalación eléctrica podríamos generar algunas nuevas hipótesis en relación con el mecanismo causal).

El supuesto del *efecto constante* es una versión más débil, pero completamente aceptable, de la homogeneidad de las unidades. En lugar de presuponer que el valor esperado de la variable dependiente es el mismo en

den compararse con interruptores de la luz. Sin embargo, entender el grado de homogeneidad de nuestras unidades de análisis nos ayudará a hacer una estimación del nivel de incertidumbre o de otros sesgos similares que pueden atribuirse a nuestras inferencias.

3.2 Independencia condicional

Según el supuesto de *independencia condicional*, los valores que se adjudican a las variables explicativas son independientes de los de las dependientes (la misma expresión se utiliza a veces en estadística, pero su definición no es la misma que suele tener en la teoría de la probabilidad). Esto quiere decir que, después de tener en cuenta las variables explicativas (o de controlarlas), el proceso de asignar valores a la variable explicativa es independiente de las (en general, dos o más) variables dependientes, Y^i y Y^j . Utilizamos la expresión «asignar valores» a las variables explicativas para describir el proceso mediante el cual éstas consiguen unos valores determinados. En el trabajo experimental el investigador *asigna* realmente valores a estas variables: unos objetos se colocan en el grupo de tratamiento y otros en el de control, mientras que en el trabajo no experimental los pueden «asignar» la naturaleza o el medio. En estos casos, lo esencial es que los valores de las variables explicativas no sean consecuencia de las dependientes. En el apartado 4 del capítulo 5 se describe el problema de la «endogeneidad», que aparece cuando las variables explicativas son consecuencia, al menos parcialmente, de las dependientes.

Los análisis de un n grande en los que hay que utilizar un sistema aleatorio, tanto de selección como de asignación de valores, son la forma más fiable de garantizar la independencia condicional y no necesitan del supuesto de homogeneidad de las unidades. Este sistema aleatorio de selección y asignación nos ayuda a extraer inferencias causales porque cumple *automáticamente* tres supuestos que subyacen en el concepto de independencia condicional: (1) que el proceso de asignación de valores a las variables explicativas es independiente de las dependientes (o sea, que no existe problema de endogeneidad), (2) que no hay sesgo de selección —problema que analizamos en el apartado 3 del capítulo 4— (3) y tampoco sesgo de la variable omitida (apartado 2 del capítulo 5). De este modo, si podemos cumplir de alguna manera estas condiciones, ya sea mediante el sistema aleatorio de selección y asignación (tal como se analiza en el apartado 2 del capítulo 4) o utilizando algún otro procedimiento, podremos evitar el problema fundamental de la inferencia causal.

Por fortuna, el sistema aleatorio de selección y asignación *no* es necesario para que se cumpla el supuesto de independencia condicional y, si el proceso de «asignación» de valores a las variables explicativas no es independiente de las dependientes, aún podremos cumplir dicho supuesto como-

diferentes unidades si también lo es el de la variable explicativa, sólo tendremos que presuponer que el efecto causal es constante. Ésta es una versión más endébil del supuesto de homogeneidad de las unidades porque el efecto causal no es más que la diferencia entre los dos valores esperados. Si los de las unidades cuya variable explicativa vale lo mismo cambian de igual manera, se vulnerará el supuesto de homogeneidad de las unidades, pero seguirá siendo válido el del efecto constante. Por ejemplo, en unas elecciones al Congreso estadounidense dos circunscripciones pueden variar en cuanto a la proporción de voto esperada para candidatos demócratas no congresistas (digamos, un 45% frente a un 65%), pero puede que el hecho de tener ese cargo siga aumentando en un 10% el voto a los candidatos demócratas en cualquier circunscripción.

El concepto de homogeneidad de las unidades (o el supuesto de los efectos causales constantes, que es menos exigente) está en la raíz de la investigación científica. Por ejemplo, es la premisa que subyace en los estudios de caso comparados. Contrastamos diversas unidades cuyas variables explicativas tienen diferentes valores y observamos los de las dependientes. Creemos que los resultados que observamos en estos últimos valores proceden de las diferencias que hay entre los de las variables explicativas que se relacionan con las observaciones. Lo que hemos planteado aquí es que, en este caso, nuestra «creencia» se apoya necesariamente en el supuesto de homogeneidad de las unidades o de los efectos constantes.

Hay que señalar que puede que busquemos unidades homogéneas a lo largo del tiempo o del espacio. Podemos comparar los votos que recibe un candidato demócrata, cuando es congresista y cuando no lo es, en la misma circunscripción y en diferentes ocasiones, así como en varias circunscripciones al mismo tiempo (o combinar ambas posibilidades). Como un efecto causal sólo puede estimarse y no conocerse, no resulta sorprendente que, en general, el supuesto de homogeneidad de las unidades no sea compatible. En qué gama de unidades esperamos que tenga sentido nuestro supuesto probable; sin embargo, es importante que se haga explícita su naturaleza. ¿En qué gama de unidades esperamos un efecto uniforme?, ¿en cualquiera de las elecciones para el Congreso?, ¿para el Congreso sí pero no para el Senado?, ¿sólo en las del norte?, ¿sólo en las de los últimos veinte años?

Véase de qué manera el supuesto de homogeneidad de las unidades se relaciona con el análisis que hicimos en el apartado 1.3 del capítulo 1 de la «complejidad y de la «singularidad». Señalamos entonces que las generalizaciones de las ciencias sociales dependen de nuestra capacidad para simplificar la realidad coherentemente. En su forma más extrema, generalizar de ese modo con el fin de extraer inferencias causales precisa del respeto a las normas de homogeneidad de las unidades: para realizar el análisis, las observaciones que se examinan se hacen idénticas en aspectos importantes. Con frecuencia, resulta imposible lograr la homogeneidad de las unidades; las elecciones al Congreso, por no hablar de las revoluciones, apenas pue-

asunto en el capítulo 2 se aplique igualmente a los problemas de la inferencia causal que abordamos aquí. En este apartado, formalizaremos brevemente las diferencias relativamente escasas que hay entre estas dos situaciones.

En el apartado 7 del capítulo 2 el objetivo de nuestra inferencia era una media (el valor esperado de una variable aleatoria) que denominamos μ y que conceptualizamos como un número fijo pero desconocido. Se dice que un estimador de μ no está sesgado si es igual a la media de μ en muchas repeticiones hipotéticas del mismo experimento.

Al igual que entonces, seguimos conceptualizando el valor esperado de un efecto causal aleatorio, denominado β , como un número fijo pero desconocido. Por tanto, la falta de sesgo se define de manera análoga: un estimador de β carece de sesgo si es igual a la media de β en muchas repeticiones hipotéticas del mismo experimento. La eficiencia también se define de forma parecida a la variación que se registra de una reproducción hipotética a otra. Estos son conceptos muy importantes que, entre los capítulos 4 y 6, nos servirán de base para estudiar muchos de los problemas de la inferencia causal. En los dos recuadros siguientes se presentan definiciones formales.

5. Reglas para elaborar teorías causales

Muchos consejos sensatos para mejorar la investigación cualitativa son precisos, específicos y detallados, y se refieren a un aspecto manejable y, por tanto, delimitado de ese tipo de investigación. Sin embargo, aun cuando estamos entascados en la resolución de multitud de problemas concretos, hay que tener en mente una clara visión general: toda solución específica debe ayudarnos a solventar cualquiera que sea el problema general de la inferencia causal del que nos estemos ocupando. Hasta ahora, en este capítulo lo hemos proporcionado una precisa definición teórica de efecto causal y hemos analizado algunos de los problemas que conlleva la extracción de inferencias causales. Ahora vamos a dar un paso atrás para ofrecer una panorámica más amplia de ciertas normas relativas a la elaboración de teorías. Para nosotros (tal como se ha señalado en el apartado 2 del capítulo 1) la mejora de una teoría no acaba cuando comienza la recogida de datos.

Las teorías causales tienen como objetivo mostrar las causas de un fenómeno o conjunto de fenómenos. Cualquiera teoría, independientemente de que en principio haya sido deductiva o inductiva, conlleva un conjunto interrelacionado de hipótesis causales. Cada una de ellas postula la existencia de una relación entre variables que genera consecuencias observables: si unas determinadas variables explicativas tienen ciertos valores, se predice que las dependientes tendrán otros valores específicos. Para comprobar o evaluar cualquier hipótesis causal se necesitan inferencias del mismo tipo. La teoría general, que se compone de hipótesis, ha de tener coherencia interna porque si no se plantearán hipótesis contradictorias.

ciendo el proceso e incorporando una medida del mismo a nuestras variables de control. Por ejemplo, suponemos que nos interesa hacer una estimación de las consecuencias que tiene el nivel de segregación residencial en la intensidad del conflicto entre israelíes y palestinos en la Cisjordania ocupada por los primeros. Estaríamos transgrediendo considerablemente nuestro supuesto de independencia condicional si, para hallar el efecto causal, sólo nos ocupáramos de la relación entre estas dos variables. La razón es que puede que los israelíes y palestinos que decidan vivir en zonas segregadas lo hagan en virtud de una convicción ideológica que determina quién tiene realmente el derecho a estar en Cisjordania. Por lo tanto, el extremismo ideológico (en ambos bandos) puede llevar al conflicto. Una medida que creemos que representa la segregación residencial quizá sea, en realidad, un sustituto de la ideología. La diferencia entre ambas explicaciones puede ser bastante importante, ya que una nueva política de vivienda podría ayudar a remediar el conflicto si la segregación residencial fuera su auténtica causa, pero sería ineficaz, e incluso contraproducente, si su fuerza motriz es la ideología. Podríamos intentar solucionar este problema midiendo también de forma explícita la ideología de los residentes y controlando este factor. Por ejemplo, podríamos informarnos de qué apoyo tienen los partidos extremistas entre los israelíes y de la afiliación a la OLP entre los palestinos. Posteriormente, se podrían controlar las posibles consecuencias engañosas de la ideología comparando comunidades en las que haya un mismo nivel de extremismo ideológico pero un grado de segregación residencial distinto.

Cuando no es factible ni seleccionar ni asignar valores de manera aleatoria y tampoco se puede controlar el proceso de asignación y de selección, tenemos que recurrir a alguna versión del supuesto de homogeneidad de las unidades para hacer inferencias causales válidas. Como en las investigaciones sociales nunca se cumplirá perfectamente tal supuesto, habrá que poner un cuidado especial en especificar el grado de incertidumbre de las inferencias causales. Este supuesto se pondrá especialmente de manifiesto cuando analicemos, en el apartado 6 del capítulo 5, los procedimientos que se utilizan para analizar observaciones «equiparables».

4. Criterios para evaluar las inferencias causales

Recordemos que al definir la causalidad en función de variables aleatorias logramos establecer una estricta analogía entre ese fenómeno y otras características sistémicas, como son la media o la varianza, en las que nos centramos al extraer inferencias descriptivas. Esta analogía nos permite evaluar las inferencias causales precisamente con los mismos criterios que utilizamos en el apartado 7 del capítulo 2 para juzgar las descriptivas: la falta de sesgo y la eficiencia. De ahí que gran parte de lo que dijimos sobre este

De la misma manera, cuando un congresista demócrata se presenta en la circunscripción i , el valor esperado es μ_i^D

$$E(Y_i^D | X_i = 1) = \mu_i^D + (1)\beta$$

$$= \mu_i^D + \beta$$

$$= \mu_i^D + (1)\beta - \mu_i^D$$

$$= \mu_i^D$$

De este modo, la ecuación (3.5) proporciona un modelo útil de inferencia causal, siendo β —la diferencia entre las dos proporciones teóricas— nuestro efecto causal. Para terminar, simplificamos por última vez esta ecuación para poder remitirnos a ella en el futuro. Si suponemos que la media de Y_i^D es cero (o que se representa como una desviación de esta media que no limita en modo alguno la aplicabilidad del modelo), podremos prescindir del término constante en esta ecuación y escribirla de forma

más simple como:

$$E(Y_i^D) = X_i\beta \tag{3.6}$$

El parámetro β sigue siendo el valor teórico del efecto causal medio, una característica sistemática de las variables aleatorias y uno de nuestros objetivos en la inferencia causal. Este modelo constituye un caso especial del «análisis de regresión» que es habitual en la investigación cuantitativa, pero los coeficientes de regresión sólo coinciden a veces con las estimaciones de los efectos causales.

harías para que se pueda demostrar de la forma más rápida y fácil posible que están equivocadas. Es evidente que no tenemos que intentar equivocarnos, pero incluso una teoría incorrecta es mejor que una afirmación que no es ni cierta. Subrayar la necesidad de que las teorías sean falsables nos obliga a mantener una perspectiva adecuada sobre la incertidumbre y garantizar que vamos a considerar tales teorías como algo provisional, sin dejarlas convertirse en dogmas. Siempre hay que estar dispuesto a rechazar las teorías si existen suficientes pruebas científicas en su contra. Una de las preguntas que hay que plantearse en relación a una teoría (o a cualquier hipótesis que se derive de ella) es simplemente: ¿qué datos la falsarían? Esta

Notación para un modelo formal de efecto causal

Con el fin de hacer más comprensibles los apartados siguientes, a continuación generalizamos nuestra notación. Casi siempre tendremos n realizaciones de la variable Y_i . En nuestro ejemplo cuantitativo corriente, n es el número de circunscripciones en las elecciones al Congreso (435), mientras que la realización Y_i de la variable aleatoria Y_i es la proporción de votos demócratas observada (como 0,56%) en la circunscripción i , dentro del reparto entre ambos partidos. La proporción de votos esperada para el congresista en ese reparto (la media de todas las repeticiones hipotéticas) en la circunscripción i es μ_i^D . Definimos la variable explicativa como X_i y, en el presente ejemplo, la codificamos como cero cuando en la circunscripción i no se presenta un congresista demócrata y como uno cuando sí se presenta. De este modo, podemos expresar el efecto causal medio en la unidad i como:

$$\beta = E(Y_i^D | X_i = 1) - E(Y_i^D | X_i = 0) = \mu_i^D - \mu_i^R \tag{3.4}$$

e incorporarla al siguiente modelo formal simple:

$$E(Y_i^D) = \mu_i^D + X_i(\mu_i^D - \mu_i^R) \tag{3.5}$$

$$= \mu_i^D + X_i\beta$$

Así, cuando en la circunscripción i no se presenta un congresista e $X_i = 0$, el valor esperado se ve determinado por la sustitución de X_i por 0 en la ecuación, y la respuesta es la siguiente:

$$E(Y_i^D | X_i = 0) = \mu_i^D + (0)\beta$$

$$= \mu_i^D$$

5.1 Regla 1: elaborar teorías falsables

Las teorías e hipótesis que se ajustan a esta descripción abarcan una amplia gama. En este apartado presentamos cinco reglas que ayudan a elaborar buenas teorías y analizaremos cada una de ellas mediante ejemplos.

Con esta primera regla no sólo queremos decir que una «teoría» que no puede estar equivocada no es tal teoría, sino que las teorías hay que dise-

Un análisis formal de la falta de sesgo en las estimaciones causales

En este recuadro demostramos la falta de sesgo del estimador que representa el parámetro del efecto causal que aparece en el apartado 1 de este capítulo. La notación y la lógica de estas ideas son bastante paralelas a las de la definición formal de falta de sesgo que se dio en el apartado 7 del capítulo 2, dentro del contexto de la inferencia descriptiva. El simple modelo lineal que tiene una variable explicativa y otra dependiente es como sigue:

$$E(Y_i) = \beta X_i$$

Nuestra estimación de β es simplemente la de regresión por mínimos cuadrados:

$$b = \frac{\sum_{i=1}^n Y_i X_i}{\sum_{i=1}^n X_i^2} \tag{3.7}$$

Para determinar si b es un estimador no sesgado de β es preciso tomar el valor esperado, haciendo su media en las reproducciones hipotéticas:

$$E(b) = E \left(\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} \right) \tag{3.8}$$

lo cual demuestra que b es un estimador no sesgado de β .

pregunta debe aplicarse a cualquier teoría o hipótesis, pero, sobre todo, tiene que hacerlo el investigador que primero formula una teoría. Karl Popper es el autor que más se relaciona con la idea de falsabilidad (Popper, 1968), y para él existe una asimetría fundamental entre confirmar

Un análisis formal de la eficiencia

A continuación, calibramos la eficiencia del estimador estándar que representa el parámetro del efecto causal β que aparece en el apartado 1 de este capítulo. En la ecuación (3.8) hemos demostrado que este estimador no está sesgado, y ahora calculamos su varianza:

$$V(b) = V \left(\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} \right) \tag{3.9}$$

$$= \frac{1}{\sum_{i=1}^n X_i^2} V(\lambda)$$

$$= \frac{V(\lambda)}{\sum_{i=1}^n X_i^2}$$

$$= \frac{\sum_{i=1}^n X_i^2}{\sigma^2}$$

En consecuencia, la varianza de este estimador es una función de dos componentes. En primer lugar, cuanto más aleatoria sea cada unidad en nuestros datos (la mayor es σ^2), más variable será nuestro estimador b , y esto no debería sorprendernos. Además, cuanto mayor sea la varianza observada en la variable explicativa $\left(\sum_{i=1}^n X_i^2 \right)$, menos variable será nuestra estimación de b . En el caso extremo de que no haya variabilidad en X , nada podrá ayudarnos a estimar el efecto que tienen los cambios producidos en la variable explicativa sobre la dependiente, y, en este ejemplo, la fórmula pronosticará una varianza infinita (incertidumbre completa). De manera más general, este componente indica que la eficiencia es mayor cuando, en la variable explicativa, tenemos datos que proceden de una gama de valores más amplia. Por tanto, lo más habitual es que sea mejor evaluar nuestras hipótesis causales en tantas situaciones diferentes como sea posible. Se puede pensar en este último asunto como si consistiera en trazar una línea utilizando una regla, dos puntos sobre una página y poco pulso. Si los dos puntos están muy juntos (X tiene una pequeña varianza), los errores que se registren al colocar la regla serán mucho mayores que si estuvieran más separados (situación en la que X tiene una varianza grande).

- Para los científicos (y especialmente para los sociales) que evalúan teorías formuladas con propiedad, la asimetría fundamental de Popper resulta prácticamente irrelevante. O'Hear (1989, p. 43) hizo una afirmación similar acerca de la aplicación de las ideas de Popper a las ciencias físicas:

Popper siempre suele hablar teniendo en mente *aplicaciones* de teorías universales. Pero hay que insistir una vez más en que proponer y comprobar teorías universales es sólo uno de los objetivos de la ciencia. Puede que, debido a condiciones que han variado considerablemente a lo largo del tiempo y el espacio, no haya teorías universales verdaderas, y esta es una posibilidad que no podemos subestimar. Sin embargo, aunque así fuera, la ciencia aún podría cumplir muchos de sus objetivos, al darnos conocimientos y hacer verdaderas predicciones acerca de las condiciones que constituyen y rodean nuestro restringido ámbito espacio-temporal.

Seguramente, esta misma idea puede aplicarse aun con más razón a las ciencias sociales.

Además, la evaluación que hace Popper de las teorías no establece una distinción fundamental entre las que acaban de formularse y las que ya han resistido numerosas pruebas empíricas. Cuando estamos comprobando la distinción determinista entre verdad o ficción en una teoría universal (de la cual no hay ejemplos interesantes), la idea de Popper es apropiada, pero no resulta útil para lo que nosotros pretendemos, que es buscar los límites de aplicación de una teoría. Como hemos señalado muchas veces en este libro, propugnamos que todas las inferencias relativas a hipótesis específicas se hagan planeando el mejor cálculo (una estimación) y midiendo su incertidumbre. La conclusión a la que llegamos respecto a si la inferencia concuerda o no concuerda con nuestra teoría, tendrá el mismo efecto sobre la información sobre la verdad de la teoría y tienen que influir en la certidumbre de nuestras creencias.¹³

Pensemos en la hipótesis de que las campañas electorales de los demócratas y de los republicanos durante las elecciones presidenciales estadounidenses tienen poco impacto global sobre su resultado. Esta hipótesis implica otras muchas de carácter más específico, como las que propugnan que los anuncios televisivos y radiofónicos, así como los debates, tampoco tienen mucha influencia en los votantes. En realidad, cualquier comprobación de la teoría debe contrastar una de estas hipótesis. Hay una comprobación de esta teoría que ha demostrado que se pueden hacer pronósticos más precisos del resultado utilizando las variables de que se dispone únicamente al hacerse las convenciones y, por lo tanto, antes de las campañas (Gelman y King, 1993). Esta comprobación concuerda con la teoría (si podemos pretender las elecciones antes de la campaña, no puede decirse que ésta tenga mucha importancia), pero no la verifica completamente. Puede que haya algún aspecto de la campaña que tenga alguna pequeña influencia que expli-

una teoría (verificación) y desmentirla (falsación). Lo primero es práctica-mente irrelevante, mientras que lo segundo es la clave de la ciencia. Popper cree que una teoría, una vez que se enumera, pasa a convertirse de inmediata to en parte integrante del cuerpo de conocimientos científicos aceptados. Al ser generales las teorías y específicas las hipótesis, las primeras implican, técnicamente, la existencia de un número infinito de las segundas. Sin embargo, sólo es posible someter a pruebas empíricas a un número limitado de hipótesis. En este sentido, «las teorías no son verificables» porque nunca podemos comprobar todas sus consecuencias observables (Popper, 1968, p. 252). Al comprobar cada hipótesis se puede demostrar que es coherente con la teoría, pero un determinado número de resultados empíricos coherentes no cambiarían nuestra opinión, ya que la teoría sigue constituyendo un conocimiento científico aceptado. Por otra parte, si se demuestra que una sola hipótesis es incorrecta y, por tanto, que no se ajusta a la teoría, ésta será falsada y dejará de pertenecer al bagaje del conocimiento humano. «Por lo tanto, la sucesión de comprobaciones no altera en absoluto la posición de ninguna hipótesis, pero el fracaso de una sola de esas pruebas sí puede tener muchas consecuencias» (Miller, 1988, p. 22). Popper no pretendía que la falsación fuera un concepto determinista y reconocía que toda inferencia empírica es, en cierto modo, incierta (Popper, 1982). En su análisis del desmentido escribió que «aunque se admita la asimetría [entre falsación y verificación], sigue siendo imposible, por varias razones, que cualquier sistema teórico sea falsado de manera concluyente» (Popper, 1969, p. 42).

Las ideas de Popper nos parecen fundamentales para formular teorías. Siempre hay que diseñarlas de manera que puedan ser falsadas y también debemos aprender de la insistencia de Popper en el carácter provisional de toda ellas. Sin embargo, para evaluar las teorías existentes en las ciencias sociales no resulta tan significativa la asimetría entre verificación y falsación. Ninguna de ellas contribuye a nuestro conocimiento científico. En cierto sentido general, la cuestión no es tanto si una teoría es o no falsa —casi todas las teorías sociales interesantes cuentan al menos con una consecuencia observable que parece equivocada—, sino en qué medida *nos ayuda dicha teoría a explicar el mundo*. Si siguiéramos la regla de Popper, las teorías que se basan en el supuesto de la elección racional hubieran sido rechazadas hace tiempo, ya que han sido falsadas en muchos ejemplos específicos. Sin embargo, los científicos sociales optan con frecuencia por mantener este supuesto, convenientemente modificado, porque les concede un poder considerable en muchas clases de problemas de investigación (véase Cook y Levi, 1990). Lo mismo puede decirse de casi cualquier otra teoría social de interés. De hecho, en las ciencias sociales, intentar falsar teorías es buscar los límites de su aplicabilidad. Aprendemos cuando alguna consecuencia observable indica que la teoría no se cumple y también cuando funciona correctamente.

En el análisis anterior hemos planteado una forma importante de abordar las teorías y también hicimos una advertencia: nuestro enfoque subraya el carácter contingente de teorías e hipótesis. Más adelante, propugnamos que se intente dar a ambas una aplicación amplia. Esta es una estrategia útil para investigar, pero siempre debemos recordar que no es probable que las teorías sociales tengan una aplicación universal. Las teorías que se dicen aplicables a cualquier circunstancia, en cualquier lugar—algunas variantes del marxismo y de la teoría de la elección racional son ejemplos de esta

Para nosotros, no es así como se hacen las ciencias sociales o como deberian hacerse. Después de hacer mi comprobación a favor o en contra, no prescindiríamos de la teoría que afirma que las campañas no tienen impacto, sino que la modificaríamos para decir, quizá, que las campañas normales carecen de impacto, a no ser que haya pruebas concluyentes de que uno de los candidatos se ha comportado de forma inmoral. Sin embargo, como esta modificación haría más restrictiva nuestra teoría, necesitaríamos evaluarla con un nuevo conjunto de datos antes de confiar en su validez. La teoría seguida siendo convincente, y de alguna manera, con cada nueva evaluación empírica sabríamos más sobre sus límites. Cada prueba que pasa una teoría influye tanto en la estimación de su validez como en la incertidumbre de ese cálculo, y el impacto puede alcanzar nuestro grado de interés en que la teoría funcione.

Sin embargo, si se demostrara que un único acontecimiento de la campaña—como una importante acusación de comportamiento inmoral—tiene alguna influencia en los votantes, la teoría quedaría falsada. Según Popper, aunque esta teoría no se falsara de forma concluyente (lo cual recordo que era imposible), aprenderíamos más de esta manera que con mi comprobación que concordaran con ella.

Podríamos llevar a cabo otras muchas comprobaciones, introduciendo en el modelo de pronóstico variables que midieran aspectos de la campaña, como la cantidad relativa de espacios televisivos o radiofónicos, la elocuencia de los candidatos y la evaluación de los resultados de los debates. Si todas estas hipótesis no tuvieran ningún efecto, Popper diría que nuestra opción no cambia de forma reseñable, de manera que la teoría de que las campañas presidenciales no tienen impacto se mantendría. De hecho, si hicieramos mi comprobación similar y todas concordaran con la teoría, esta todavía podría estar equivocada, ya que no hemos puesto a prueba todas las infinitas variables posibles que miden una campaña. De manera que, aunque dispusiéramos de muchos resultados coherentes con la teoría, aún *podría* ser cierto que las campañas de elecciones presidenciales influyen en el comportamiento del votante.

Además, la predicción puede haber sido fruto de la suerte o quizá, en los años en que se han recogido los datos, las campañas no hayan incorporado ninguna táctica innovadora (y, por tanto, impredecible).

Si nuestro estudio desmiente algún aspecto de la teoría podemos optar por mantenerla, pero incorporando una excepción. Este procedimiento resulta aceptable siempre y cuando se reconozca el hecho de que estamos reduciendo las pretensiones de la teoría. De este modo, ésta será menos valiosa porque explica menos, según nuestra terminología, tendremos menos *control* sobre el problema que queremos comprender¹⁴. Además, este enfoque puede producir una «teoría» que no sea más que una mezcla de diversas excepciones y conclusiones. En algún momento hay que estar dispuesto a prescindir por completo de ciertas teorías e hipótesis. Si hay demasiadas excepciones, la teoría debe rechazarse. De manera que, en sí misma, la *construcción*, que es la preferencia normativa por teorías que tienen pocos componentes, no siempre es aplicable. Lo único que necesitamos es nuestro concepto general de maximización del control, del que se deriva

ble, añadiendo numerosas excepciones y casos especiales.

En consecuencia, hay que tener mucho cuidado al adaptar teorías para que concuerden con nuevos datos y no hay que forzarlas más allá de lo plausible, añadiendo numerosas excepciones y casos especiales.

Si los llevamos demasiado lejos haremos que nuestras teorías sean invulnerables al desmentido. En tales ajustes, pero éstos pueden ser peligrosos. Si los llevamos demasiado lejos haremos que nuestras teorías sean invulnerables al desmentido. En ajustes, seguir aceptando esa elaboración teórica. La ciencia avanza mediante hipótesis aunque no se sostenga en un caso determinado y, en virtud de ese o ajustar lo que consideramos que es el ámbito de aplicación de una teoría o llegar a la conclusión de que quizá los datos sólo son pobres por casualidad. Podemos hacer varias cosas: ramos la teoría inmediatamente por la borda. Podemos hacer varias cosas: una prueba empírica no concuerda con nuestras expectativas teóricas, no funciones empíricas no las confirman ni desmienten definitivamente. Cuando proceso de evaluación de teorías e hipótesis es flexible: ciertas comprobaciones empíricas no las confirman ni desmienten definitivamente. Cuando

Sin embargo, hay que añadir una advertencia. Hemos señalado que el proceso de evaluación de teorías e hipótesis es flexible: ciertas comprobaciones empíricas no las confirman ni desmienten definitivamente. Cuando una prueba empírica no concuerda con nuestras expectativas teóricas, no tenemos la teoría inmediatamente por la borda. Podemos hacer varias cosas: llegar a la conclusión de que quizá los datos sólo son pobres por casualidad o ajustar lo que consideramos que es el ámbito de aplicación de una teoría o prescindir de ella.

Si los llevamos demasiado lejos haremos que nuestras teorías sean invulnerables al desmentido. En tales ajustes, pero éstos pueden ser peligrosos. Si los llevamos demasiado lejos haremos que nuestras teorías sean invulnerables al desmentido. En ajustes, seguir aceptando esa elaboración teórica. La ciencia avanza mediante hipótesis aunque no se sostenga en un caso determinado y, en virtud de ese o ajustar lo que consideramos que es el ámbito de aplicación de una teoría o llegar a la conclusión de que quizá los datos sólo son pobres por casualidad. Podemos hacer varias cosas: ramos la teoría inmediatamente por la borda. Podemos hacer varias cosas: una prueba empírica no concuerda con nuestras expectativas teóricas, no funciones empíricas no las confirman ni desmienten definitivamente. Cuando proceso de evaluación de teorías e hipótesis es flexible: ciertas comprobaciones empíricas no las confirman ni desmienten definitivamente. Cuando

14. Véase, por ejemplo, Popper (1972), p. 10.

empírico. Solo desde un punto de vista abstracto son más precisas que las teorías sociales no formales: sus predicciones acerca del mundo real no son más específicas porque sus condiciones no se corresponden, ni siquiera de forma aproximada, con las reales.

Simplificar es esencial para los modelos formales, al igual que lo es en toda investigación, pero hay que tener cuidado con las inferencias que se extraen de la realidad a partir de esos modelos. Por ejemplo, presuponer que todas las variables omitidas no tienen consecuencias sobre los resultados puede ser muy útil al construir modelos, y, precisamente, esto es lo que hacemos en muchos de los modelos formales cualitativos de este libro. Esos supuestos no suelen justificarse por ser rasgos del mundo, sólo se dice que son instrumentos útiles para construir modelos a partir de él. En consecuencia, sus resultados se refieren concretamente a situaciones en la que las variables omitidas son irrelevantes y pueden parecerse o no a los del mundo real. No tenemos que comprobar el supuesto para elaborar el modelo y sus consecuencias, pero sí es *esencial* que lo pongamos a prueba durante la evaluación empírica. Para que el modelo formal sea útil, el supuesto no tiene que ser correcto, pero no podemos emplear en la elaboración de diseños de investigación empírica supuestos teóricos no comprobados o justificados. Por el contrario, para que las teorías formales sean útiles en los estudios empíricos, suele ser preciso complementarias con otros elementos.

Un buen modelo formal debe ser abstracto, de manera que los rasgos clave del problema se pongan de manifiesto y pueda aplicarse fácilmente el razonamiento matemático. Así, consideremos un modelo formal sobre las consecuencias de la representación proporcional en los sistemas de partidos que implique que la primera fragmenta los seguros. La variable causal clave es el ordenamiento electoral, o sea, si es un sistema de representación proporcional en el que los partidos obtienen escaños en función del porcentaje de votos que logran o si es un sistema mayoritario en el que sólo hay un ganador en cada circunscripción. La variable dependiente es el número de partidos políticos, que suele denominarse «grado de fragmentación del sistema de partidos». La hipótesis principal es que los ordenamientos electorales que se basan en la representación proporcional generan más partidos políticos que los de tipo mayoritario. Para una mayor sencillez, este modelo podría incluir únicamente variables que midieran algunos rasgos esenciales del sistema electoral y también el grado de fragmentación del sistema. El modelo sólo generaría una *hipótesis*, no una conclusión, acerca de la relación que existe entre la representación proporcional y la fragmentación del sistema de partidos en el mundo real, y esa hipótesis habría que comprobarla utilizando métodos cualitativos o cuantitativos empíricos.

Sin embargo, aunque una de las consecuencias de este modelo sea que la representación proporcional fragmenta los partidos políticos y aunque no se utilizaran otras variables en el modelo, en un análisis empírico sería absurdo

completamente la idea de concisión cuando es útil. Pensar que, en gran medida, la ciencia consiste en explicar muchos fenómenos mediante unos pocos pone de manifiesto que las teorías de pocos componentes no son mejores ni peores. Para maximizar el control hay que intentar formular teorías que expliquen lo más posible con el menor número de elementos. A veces esta formulación se logra mediante la concisión, pero otras no. Podemos pensar en ejemplos en los que una teoría ligeramente más complicada explica una parte del mundo mucho más amplia. En tal situación, seguramente utilizáramos una teoría poco concisa, ya que esta maximiza mejor el control que la más concisa.¹⁵

5.2 Regla 2: construir teorías que tengan coherencia interna

Una teoría sin coherencia interna no sólo es falsa: es falsa. De hecho, ésta es la única situación en la que se conoce la veracidad de una teoría sin necesidad de datos empíricos: si tiene dos o más componentes que generan hipótesis contradictorias, no hay datos del mundo empírico que puedan sostenla. No tendría que ser polémico garantizar la coherencia interna de las teorías, pero esa cualidad suele ser difícil de alcanzar. Se puede lograr mediante modelos formales, matemáticos. Es en la economía donde más desatrollada está la *creación de modelos formales*, aunque es cada vez más habitual en sociología, psicología, ciencia política, antropología y en otras disciplinas (véase Ordeshook, 1986). Los politólogos han desarrollado numerosas teorías esenciales a partir de modelos matemáticos en áreas como la elección racional, la social, los modelos electorales espaciales, la economía pública o la teoría de juegos. Con estas investigaciones se han logrado importantes resultados y una gran cantidad de hipótesis plausibles. Una de las aportaciones más valiosas de los modelos formales es su capacidad para poner de manifiesto la incoherencia de teorías planteadas mediante la expresión verbal.

Sin embargo, al igual que ocurre con otras hipótesis, los modelos formales no son explicaciones verificadas si no evalúan empíricamente sus predicciones. En la investigación social, la formalización nos ayuda a razonar con más claridad y garantiza realmente que nuestras ideas tengan coherencia interna, pero no resuelve los problemas de evaluación empírica de las teorías. En las ciencias sociales, los supuestos de los modelos formales suelen ser sencillas expresiones matemáticas o maneras de garantizar que se puede llegar a un equilibrio. Hay pocos autores que crean que el mundo político es matemático en la misma medida en que lo creen algunos físicos del mundo que ellos analizan. Por lo tanto, estos modelos formales no son más que modelos: abstracciones que hay que diferenciar del mundo que estudiamos. De hecho, algunas teorías formales hacen predicciones a partir de supuestos enormemente simplificados y a veces no tienen gran valor

riable se mantenga constante, puede resultar que lo sea. Sin embargo, esto no supondrá un problema a menos que hayamos predeterminado tal situación mediante la selección de criterios. Por ejemplo, suponemos que seleccionan observaciones en dos categorías de una variable explicativa y que resulta que la dependiente es constante en los dos grupos. Este no será más que un caso en el que el efecto causal estimado es cero.

Para terminar, tenemos que elegir una variable dependiente que represente la *variancia que queremos explicar*. Aunque esta indicación parezca evidente, el asunto es bastante delicado, como pone de manifiesto Stanley Lieberman (1985, p. 100):

Un simple experimento gravitatorio, presente en el Ontario Science Centre de Toronto, inspira un ejemplo heurístico. En este montaje se deja caer una moneda y una pluma dentro de un tubo en el que se ha hecho el vacío, y ambas llegan al fondo casi al mismo tiempo. Como el vacío no es total, es de suponer que la moneda llegará al fondo un poquito antes que la pluma. En cualquier caso, imaginémosnos un estudio en el que se describa varios objetos, sin que se tenga el control tan férreo que otorga el vacío (que es lo que ocurriría en una investigación social no experimental). Si los científicos sociales perciben que varía el tiempo que tarda cada objeto en llegar al suelo, lo normal es que quieran saber qué características determinan esa diferencia. Es probable que rasgos como la densidad y la forma de los objetos influyan en la velocidad de la caída si no existe el vacío. Si el investigador social tiene suerte, la conjunción de esos factores explicará completamente las diferencias de velocidad que presenta la caída de cada uno de los objetos y, si es así, se pondrá muy contento porque se habrán explicado todas las variaciones que hay entre un objeto y otro. Mediante el razonamiento habitual en las ciencias sociales, el investigador llegará a la conclusión de que se entiende completamente el fenómeno porque se han explicado todas las diferencias que existen entre los objetos que se estudian. Seguramente, debe de haber algo que falle en nuestros procedimientos si podemos abordar este problema sin siquiera referirnos a la gravedad.

En este ejemplo, los procedimientos del investigador solo serían erróneos si la variable más interesante fuera la gravedad. Si ésta fuera la variable explicativa que nos preocupara, el experimento no la cambiaría (ya que sólo ocurre en un lugar) y, por tanto, no nos daría información sobre ella. Sin embargo, el experimento que describe Lieberman tendría un mayor interés si quisieramos comprender las variaciones que se producen en el tiempo que tardan diferentes objetos en tocar el suelo, cuando se dejan caer desde una misma altura a diferentes niveles de presión del aire. De hecho, aunque lo supieramos todo de la gravedad, este experimento nos seguiría dando una información valiosa, pero sí, como presupone el autor, lo que nos interesa realmente es hacer una inferencia sobre el efecto causal de la gravedad, necesitamos una variable dependiente que cambiara nuestras observaciones cada vez que variara el grado de atracción de la gravedad. Del mismo modo, en las ciencias sociales hay que asegurarse de que realmente nos in-

utilizar solamente dos. Un estudio que indique que los países que utilizan la representación proporcional tienen sistemas de partidos más fragmentados estará dejando de lado el problema de la endogeneidad (apartado 4 del capítulo 5), puesto que quizá los países que instauran sistemas electorales de tipo proporcional lo hagan precisamente porque ya tienen un sistema de partidos fragmentado. Del mismo modo, el sesgo de la variable omitida sería un problema porque también es probable que los países en los que existen profundas divisiones raciales, étnicas o religiosas tengan un sistema de partidos fragmentado y que su sistema electoral sea de tipo proporcional.

De este modo, parecen cumplirse las dos condiciones para que exista sesgo de la variable omitida (apartado 2 del capítulo 5): dicha variable se correlaciona tanto con la dependiente, y, por tanto, todo análisis que prescinda de la variable relativa a la división social producirá inferencias sesgadas.

Tendría que quedar claro que los modelos formales son extremadamente útiles para aclarar nuestras ideas y para desarrollar teorías con coherencia interna. Quizá sólo un modelo formal pueda poner de manifiesto y corregir las incoherencias internas de muchas teorías especialmente complejas que se expresan con palabras. Al mismo tiempo, no es probable que esas elaboraciones teóricas proporcionen modelos correctos para la comprobación empírica, y podemos asegurar que no sirven para evitar ninguno de los problemas empíricos de la inferencia científica.

5.3 Regla 3: seleccionar cuidadosamente las variables dependientes

Es evidente que al investigar todo debe hacerse con cuidado, pero elegir las variables, principalmente las dependientes, es una decisión especialmente importante. A continuación damos tres indicaciones (basadas en errores que se cometen con demasiada frecuencia en escritos cuantitativos y cualitativos):

En primer lugar, *las variables dependientes deben ser realmente dependientes*. Elegir una variable dependiente que en realidad produce cambios en las explicativas es un error muy habitual. Analizamos las consecuencias específicas de la endogeneidad y ciertas maneras de esquivarla en el apartado 4 del capítulo 5, pero hay que subrayar aquí este problema porque la forma más fácil de evitarlo es elegir variables explicativas claramente exógenas y variables dependientes manifestamente endógenas.

En segundo lugar, *no hay que seleccionar observaciones que estén basadas en la variable dependiente para que esta se mantenga constante*. Esto también puede parecer bastante obvio, pero hay investigadores que suelen elegir observaciones en las que la variable dependiente no cambia en absoluto (como en el ejemplo analizado en el apartado 3.1 del capítulo 4). Aunque el diseño de investigación no se haga con el propósito de que esta va-

La cuarta regla, que surge de nuestra insistencia en la falsabilidad, la coherencia y la variación de la variable dependiente, es maximizar lo concreto. Siempre que sea posible, hay que elegir conceptos observables en vez de no observables. En las teorías de las ciencias sociales se utilizan con frecuencia conceptos abstractos, no observados, como los de utilidad, cultura, intenciones, motivaciones, identificación, inteligencia o interés nacional. Pueden ser muy útiles para formular teorías, pero también una carga para la evaluación empírica de teorías e hipótesis, a menos que se puedan definir de manera que sea posible observarlos y medirlos o, al menos, hacer ambas cosas con sus consecuencias. Plantear explicaciones mediante ideas como cultura, interés nacional o utilidad resultará sospechoso, a menos que tales conceptos se puedan calibrar al margen de la variable dependiente que estamos explicando. Cuando se utilizan tales términos en las explicaciones, es fácil caer en la tautología o no presentar consecuencias observables y difíciles. La acción de un individuo o de una nación puede explicarse a partir de un deseo de maximizar la utilidad, materializar unas intenciones o tabajar en beneficio del interés nacional; pero la prueba de que esa acción ha logrado tales fines reside en el hecho de que el actor o la nación hayan participado en ella. El investigador es responsable de formular una teoría que especifique con claridad y precisión qué consecuencias observables van a indicar su veracidad y a distinguirla de otras alternativas lógicas.

5.4 Regla 4: maximizar lo concreto

Interesa más nuestra variable dependiente que los factores contextuales que el diseño de investigación mantiene constantes. En consecuencia, para obtener una estimación no sesgada del impacto de las variables explicativas es preciso que todo el abanico de posibles variaciones de la variable dependiente pueda ser resultado del experimento. Poner límites artificiales al abanico de posibilidades de la variable dependiente o a sus valores genera lo que definiremos (en el apartado 3 del capítulo 4) como sesgo de selección. Por ejemplo, si nos interesan las condiciones que propician el estallido de conflictos armados, no podemos elegir sólo observaciones en las que ese sea precisamente el resultado. Tal estudio podría darnos mucha información sobre cómo varían las observaciones de conflictos armados (al igual que el experimento de la gravedad nos informaba de las que se registran en la velocidad con que caen diversos objetos), pero no nos permitiría examinar las causas de esos conflictos. Si esto es lo que queremos entender, será mejor utilizar un diseño en el que selección-mos observaciones a partir de nuestras variables explicativas, dejándole a la dependencia la posibilidad de abarcar todas las situaciones posibles: en las que hay poco o ningún conflicto, en las que amenaza con producirse o en las que realmente estalla.

Aquí tenemos que dejar claro lo que queremos decir. El desfase entre concepto e indicador es inevitable en gran parte de los estudios sociales, pero hay buenas razones para utilizar conceptos generales en vez de específicos: nos permiten ampliar nuestro marco de referencia y la aplicabilidad de las teorías. De este modo, podemos referirnos al poder legislativo en vez de utilizar categorías más estrechas, como parlamento, o instituciones concretas, como el Bundestag alemán. Podemos hablar de «órganos decisivos» en lugar de legislativo cuando queremos que nuestra teoría sea aplicable a una gama de instituciones aún más amplia (de hecho, en el siguiente apar-

dicados organismos. Aquí tenemos que dejar claro lo que queremos decir. El desfase entre concepto e indicador es inevitable en gran parte de los estudios sociales, pero hay buenas razones para utilizar conceptos generales en vez de específicos: nos permiten ampliar nuestro marco de referencia y la aplicabilidad de las teorías. De este modo, podemos referirnos al poder legislativo en vez de utilizar categorías más estrechas, como parlamento, o instituciones concretas, como el Bundestag alemán. Podemos hablar de «órganos decisivos» en lugar de legislativo cuando queremos que nuestra teoría sea aplicable a una gama de instituciones aún más amplia (de hecho, en el siguiente apar-

Con esta regla no queremos decir en modo alguno que conceptos como los de intencionalidad o motivación carezcan de importancia. Sólo pretendemos señalar que, en una ciencia empírica como la nuestra, la principal norma explicativa tiene que ser la verificación o falsación empírica. Siempre será más difícil y menos provechoso buscar pruebas empíricas para conceptos abstractos, incommensurables e inobservables que para muchas ideas específicas y concretas concebidas de manera imperfecta. Cuanto más abstractos sean nuestros conceptos, más difusas serán sus consecuencias observables y menos posibilidades habrá de falsar la teoría.

Los investigadores suelen utilizar la siguiente estrategia. Parten de algún concepto abstracto como los mencionados anteriormente, aceptan que no pueden medirlo de forma directa y, en consecuencia, señalan indicadores específicos de dicho concepto que sí pueden medirse y los utilizan en sus explicaciones. La elección de ese indicador concreto se basa en que es observable. A veces, es lo único observable (por ejemplo, sólo hay datos disponibles de este fenómeno o es el único tipo de acontecimiento histórico del que se guarda información). Este aspecto de la investigación empírica es perfectamente respetable y, de hecho, suele ser necesario.

Sin embargo, tiene una parte mala, pues con frecuencia el indicador específico está lejos del concepto original y sólo se relaciona con él de forma indirecta e incierta. Quizá no sea en absoluto un indicador válido del concepto abstracto, pero el investigador, después de pedir perdón rápidamente por el desfase que hay entre uno y otro, adjudica al primero el nombre del segundo y sigue adelante como si estuviera midiendo el concepto directamente. Por desgracia, esta objetualización es frecuente en las ciencias sociales, quizá más en la investigación cuantitativa que en la cualitativa, pero demasiado habitual en ambas. Por ejemplo, el investigador utiliza las cifras que tiene sobre correos, comercio, turismo e intercambio de estudiantes para elaborar un índice de «integración social» en Europa; realiza una encuesta para averiguar si a los entrevistados les preocupa más el medio ambiente que ganar dinero, clasificándolos después como «materialistas» o «postmaterialistas», u observa que las instituciones federales estadounidenses se diferencian entre sí por el período medio de permanencia de los trabajadores en sus empleos y con este dato mide la «institucionalización» de

Cuanto más impreciso sea nuestro lenguaje, menos posibilidades habrá de que nos equivocemos, pero también de que nuestro trabajo sea útil. Es mejor equivocarse que ser impreciso.

En nuestra opinión, al presentar la lógica de un proyecto de investigación, defender su importancia y describir los acontecimientos en detalle, hay que cultivar una redacción elegante (y disfrutarla), práctica que constituye un bien escaso en las ciencias sociales. El aburrimiento nunca ha hecho avanzar ninguna ciencia. Sin embargo, cuando el objeto de estudio se transforma en una inferencia causal o descriptiva, en la que nos interesan observaciones y generalizaciones que esperamos se mantengan, es preciso utilizar una terminología y una forma de pensar concretas y específicas¹⁶.

5.5 Regla 5: formular teorías de forma tan incluyente como sea posible

Aceptando los condicionantes que implica asegurarse de que una teoría es falsable y de que maximizamos la concreción, hay que formular tal teoría de manera que explique una parte del mundo lo más grande posible. Nos damos cuenta de que hay cierta tensión entre esta quinta regla y el hecho de que antes insistíamos a la concreción. Sólo podemos decir que ambos objetivos son importantes, aunque en muchos casos pueden contraponerse y hay que tenerlos presentes para llegar a un equilibrio.

Por ejemplo, no debemos presentar nuestra teoría como si sólo fuera aplicable al Bundestag alemán, cuando hay razones para creer que pudiera serlo en todo poder legislativo independiente. Para formularla no es preciso dar pruebas de todas sus consecuencias, siempre que se haga una estimación razonable de la incertidumbre que conlleva. Puede que hayamos dado pruebas consistentes de que la teoría es aplicable al Bundestag alemán. Aunque no las tengamos de que funciona en otros ámbitos, tampoco las tenemos de que no sea así. Un marco de referencia más amplio resulta útil si no perdemos de vista el hecho de que es necesario evaluar su aplicabilidad. De hecho, expresarla en relación a un marco de referencia hipotéticamente más amplio podría obligarnos a pensar en qué rasgos estructurales de la teoría la harían aplicable o no a otros poderes legislativos independientes. Sería adecuada, por ejemplo, para el Senado estadounidense, cuyos integrantes no se eligen de una sola vez sino de manera escalonada, para la asamblea de New Hampshire, que es mucho más numerosa en relación al número de electores, o para la Cámara de los Comunes británica, en la que es mucho más fuerte el voto de partido? Presentar qué rasgos sistemáticos de la teoría creemos que la hacen aplicable en diferentes áreas es un importante ejercicio. Quizá así nos demos cuenta de que estamos equivocados, pero esto es bastante mejor que no formular la teoría con suficiente precisión desde el principio.

La ciencia se basa en estas clasificaciones abstractas, si no estariamos volviendo al resumen de los portmenores históricos. Sin embargo, nuestros términos abstractos y generales deben relacionarse en algún momento con conceptos específicos y mensurables que hagan posible la comprobación empírica. Esta relación —y la distancia que se recorre para establecerla— siempre hay que tenerla en mente y hacerla explícita. Además, el problema teórico que se plantea debe justificarse el determinado nivel de abstracción elegido, que, a su vez, tiene que ayudar a relacionar una investigación concreta —en la que un determinado indicador es el actor clave— con un problema general. Para nosotros será una responsabilidad ver que se lleven a cabo nuevas investigaciones, que utilicen otros indicadores específicos, con el fin de reforzar el supuesto de que los nuevos están realmente relacionados con un concepto más amplio. Los términos abstractos anteriores —«integración social», «postmaterialismo» e «institucionalización»— pueden medirse de forma razonable con los indicadores específicos indicados. No negamos que haya que dar un salto desde el indicador específico hasta el concepto abstracto general —en las ciencias sociales esto es preciso para continuar la investigación—, pero hay que darlo con cuidado, justificándolo y «recordando» siempre de donde se parte. Por lo tanto, no nos oponemos a las abstracciones, pero pedimos que el lenguaje de la investigación social sea lo más concreto y preciso posible. Si, como suele ocurrir en las ciencias sociales, no hay más remedio que utilizar elaboraciones técnicas invisibles, al menos tendremos que *elegir ideas con consecuencias observables*. Por ejemplo, la «inteligencia» nunca se ha observado directamente, pero no deja de ser por ello un concepto muy útil. Existen numerosos tests y otras formas de evaluar sus implicaciones. Por otra parte, si podemos elegir entre «institucionalización de la presidencia» y «número de empleados de la Casa Blanca», suele ser mejor quedarse con lo segundo. Podemos señalar que este concepto se relaciona con el de institucionalización de la presidencia —que es más general—, pero no hay que concretar el concepto limitado hasta el punto de considerar que es idéntico al más amplio. Además, si número de empleados significara institución, tendríamos que poder encontrar otras medidas de este segundo concepto que respondieran a las mismas variables explicativas que el primero. Más adelante analizaremos cómo se «maximiza el control» amplian- do las variables dependientes.

Nuestra defensa de la concreción se extiende en general a las palabras que utilizamos para describir nuestra teoría. Si el lector tiene que pasar mucho tiempo averiguando los significados precisos de una teoría, ésta tendrá menos utilidad. Al describir teorías, hay que reducir al mínimo la polémica sobre nuestras intenciones. Para lograr esta mayor especificidad, aunque no estemos llevando a cabo investigación empírica, hay que dedicar cierto tiempo a considerar, explícitamente, las consecuencias observables de la teoría e incluso los posibles proyectos de investigación que podríamos realizar.

Puede que parezca que esta regla entra en conflicto con la preferencia que mostró Robert Merton ([1949], 1968) por las «teorías de alcance medio», pero incluso una lectura rápida de este autor nos indicará que no es así. Merton estaba reaccionando contra una tradición sociológica en la que había «teorías», como la de «la acción» de Parsons, que se formulaban de manera tan amplia que no podían falsarse. En la ciencia política, la «teoría de sistemas» de Easton (1965) se halla en la misma tradición (véase Ecks-tein, 1975, p. 90). Este es un ejemplo del tipo de críticas que a Merton ([1949], 1968, p. 43) le gustaba hacer: «Por lo que se ve, la teoría de la acción de los roles no concuerda con orientaciones teóricas tan generales como el marxismo, el análisis funcionalista, el conductismo social, la sociología integral de Sorokin o la teoría de la acción de Parsons». Merton no critica la teoría de la fijación de los roles, que consideraba de alcance medio, sino que ataca las «amplias orientaciones teóricas» con las que concuerda casi cualquier teoría más específica u observación empírica. Merton está a favor de las teorías de «alcance medio», pero creemos que estaría de acuerdo en que hay que formular teorías de la manera más general posible, siempre que sigan siendo falsables y concretas. Para retomar un concepto que hemos mencionado anteriormente, enunciar teorías de este modo es maximizar el control. Si la teoría puede comprobarse —y es evidente que el peligro de las más generales es que quizá estén expresadas de manera que no se pueda—, cuanto más general sea, mejor. O sea, a mayor amplitud, más control.

4. Precisar qué va a observarse

Hasta ahora hemos señalado cuáles son para nosotros las normas de la investigación científica, según se aplican tanto a la investigación cualitativa como a la cuantitativa (capítulo 1), hemos definido la inferencia descriptiva (capítulo 2) y aclarado lo que entendemos por causalidad e inferencia causal (capítulo 3). A continuación, vamos a plantear problemas específicos del diseño de investigaciones. A partir de ahora, para ilustrar nuestros argumentos utilizaremos muchos ejemplos tomados de la bibliografía existente o elaborados de forma hipotética. El presente capítulo se ocupa de cómo tenemos que seleccionar los casos u observaciones para nuestro análisis. Es- tas decisiones tienen gran importancia, ya que una mala selección de casos puede impedir que, en estadios posteriores, la más imaginativa de las inferencias logre extraer inferencias causales válidas. En el capítulo 5 señalamos algunas de las principales causas de sesgo e ineficiencia que habrá que evitar o, al menos, entender para poder aquilatar nuestras estimaciones. Posteriormente, en el capítulo 6, proponemos algunas ideas para incrementar las observaciones disponibles, las cuales suelen encontrarse ya en los datos que hemos recogido. A continuación, retomamos un asunto que se introdujo en el capítulo 1: la necesidad de extraer la mayor cantidad posible de consecuencias observables de nuestras teorías y, en la medida de lo posible, de comprobar cuantas más mejor.

En el apartado 3.2 del capítulo 3 analizamos la «independencia condicional»: supuesto que señala que se eligen las observaciones y se asignan los valores de las variables explicativas al margen de los de las dependien-

tos. Por ejemplo, esta independencia se vulnera si las variables explicativas se eligen en función de reglas que se correlacionan con las variables dependientes o si éstas producen las explicativas. Es habitual que algunos tratadistas que trabajan con gran cantidad de observaciones utilicen la aleatoriedad en la selección de unidades y en la adjudicación de valores a las variables explicativas, para asegurarse de que se cumple el supuesto de independencia condicional; después utilizan métodos estadísticos para mitigar el problema fundamental de la inferencia causal. Por desgracia, seleccionar observaciones y adjudicar valores de forma aleatoria tiene graves implicaciones en investigaciones con un n pequeño. Si estas estrategias no resultan apropiadas, podemos lograr la homogeneidad de las unidades seleccionando intencionalmente las observaciones (como se analizó en el apartado 3.1 del capítulo 3). En cierto sentido, este último procedimiento es nuestra «última línea defensiva» si queremos crear las condiciones para la extracción de inferencias causales válidas.

Recordemos la base del supuesto de homogeneidad de las unidades: si dos unidades tienen el mismo valor de la variable explicativa principal, el valor esperado de la dependiente será el mismo. La versión más estricta de este supuesto implica, por ejemplo, que si al pulsar un interruptor se enciende una bombilla de 60 vatios, lo mismo ocurrirá si se acciona un segundo interruptor. En este ejemplo, la posición del interruptor es la variable explicativa clave, y la situación de la luz (encendida o apagada) es la dependiente. Según el supuesto de homogeneidad de las unidades, la situación esperada de cada luz será la misma siempre que los interruptores se encuentren en la misma posición. La versión menos estricta de este supuesto —que suele ser más plausible pero igualmente aceptable— es la del *efecto constante*, en la que una variación similar de los valores de la variable explicativa en las dos observaciones conduce al mismo efecto causal en unidades diferentes, aunque puede que los niveles de dichas variables no sean los mismos. Supongamos, por ejemplo, que en nuestros interruptores haya tres posiciones y que midamos la variable dependiente según el voltaje generado. Si un interruptor pasa de estar «apagado» a «bajo» y el otro de «bajo» a «alto», se cumplirá el supuesto del efecto constante si el aumento de voltaje es igual en las dos habitaciones, aunque en una observación vaya de 0 a 60 y en la otra de 60 a 120.

Cuando no se cumple ni el supuesto de independencia condicional ni el de homogeneidad de las unidades, habrá graves problemas en la inferencia causal. Sin embargo, éstos serán aún más acusados —de hecho, no podremos extraer inferencias causales en absoluto— cuando nuestro diseño de investigación sea impreciso. Un diseño de investigación preciso es el requisito imprescindible de la inferencia causal; de ahí que en el apartado 1 de este capítulo comencemos por analizar diseños de investigación imprecisos. Después, abordamos el problema del sesgo de selección que surge del no cumplimiento de los supuestos de independencia condicional y de homoge-

Ahora merece una especial atención el tema principal de este capítulo: los problemas que conlleva la selección de casos o de observaciones para el análisis. Como las denominaciones pueden ser confusas, es importante revisar desde el principio algunas cuestiones terminológicas. Con frecuencia, cuando se analizan diseños de investigación cualitativa, se habla de «casos» —como al referirse a los estudios de caso o al «método de los casos». Sin embargo, la palabra «caso» suele utilizarse de manera ambigua, ya que puede significar una sola observación. Tal como se explicó en el apartado 4 del capítulo 2, una «observación» se define como la medida que se toma de una variable dependiente en una unidad e incluye información sobre los valores de las variables explicativas. Sin embargo, un caso también puede referirse a una única unidad en la que se midan muchas variables, o incluso a un amplio ámbito de análisis.

Por ejemplo, hay investigadores que pueden escribir sobre un «estudio de caso de la India» o de la Segunda Guerra Mundial. Para ciertos propósitos, estos dos objetos de estudio pueden constituir observaciones únicas; por ejemplo, en una investigación sobre la distribución de la población por países o sobre el número de muertos en combates en las guerras contemporáneas. Sin embargo, por lo que respecta a muchos asuntos de interés para los científicos/sociales, tanto la India como la Segunda Guerra Mundial contienen por separado muchas observaciones que conlleven sus propias unidades y variables. Un investigador podrá comparar los resultados electorales de cada partido en los estados indios o los de las batallas durante la Segunda Guerra Mundial. En un diseño de ese tipo, podría resultar engañoso referirse a la India o a la Segunda Guerra Mundial como estudios de caso, ya que uno y otro definen simplemente los límites dentro de los cuales se hace un gran número de observaciones.

A la hora de elegir qué se va a observar, lo que realmente nos preocupa son las *observaciones* utilizadas para extraer inferencias, en cualquier caso que sea el nivel de análisis que nos interese; de ahí que recomendemos a los científicos sociales que se basen en las observaciones que van a poder hacer y no en la imprecisa terminología de los casos. Sin embargo, los cualitativistas suelen comenzar por elegir lo que consideran «casos» —observa-

ciones concebidas en un nivel de análisis muy agregado— y después se dan cuenta de que para obtener observaciones suficientes deben desagregarse.

Supongamos, por ejemplo, que un investigador quiera comprender cómo influyen las pautas de crecimiento económico de los países democráticos pobres en sus instituciones políticas. En principio, quizá piense en utilizar como único caso la India entre 1950 y 1990, teniendo en mente observaciones de esa sola unidad (la India) en relación a dos variables: la tasa de crecimiento económico y un índice de cambio o de estabilidad de las instituciones políticas. Sin embargo, quizá sólo pueda encontrar un número escaso de democracias pobres y, en este nivel de análisis, no disponer de observaciones suficientes para extraer inferencias causales válidas. Una vez que ha reconocido este problema, quizá de forma tardía, podrá optar por utilizar cada uno de los estados indios como unidad de análisis; quizá también dividiendo su período de estudio en cuatro o cinco subperíodos. Si estas observaciones desagregadas fueran consecuencia de la misma teoría que se dispone a comprobar, el procedimiento le reportaría muchas observaciones dentro de su «estudio de casos» de la India. De este modo, la investigación resultante podría aportar información suficiente para sostener inferencias causales válidas sobre la política india y sería muy diferente de un estudio de caso convencional cuya estrecha concepción parte de observaciones hechas en una sola unidad sobre diversas variables.

Como la «observación» se ha definido con mayor precisión que el «caso», en este capítulo nos referiremos normalmente a la «selección de observaciones». Sin embargo, dado que los investigadores suelen comenzar por elegir ámbitos de estudio que contienen múltiples observaciones potenciales y que la terminología al uso los denomina «casos», con frecuencia hablaremos de selección de casos en lugar de observaciones cuando nos referimos a la práctica de los cualitativistas.

1. Diseños de investigación imprecisos

Un diseño de investigación es un plan que muestra, mediante el análisis de un modelo y de unos datos, de qué manera esperamos utilizar nuestro material para extraer inferencias. En la investigación cualitativa los diseños de investigación no siempre están explícitos, pero sí lo están, al menos, en cada investigación concreta. Sin embargo, algunos diseños son imprecisos, o sea, que no puede saberse nada de sus hipótesis causales.

Por desgracia, esta práctica está muy extendida tanto en la investigación cuantitativa como en la cualitativa. No obstante, existe una diferencia en la imprecisión de una y otra. Cuando la segunda es imprecisa, el problema suele ser evidente porque el programa de ordenador no dará estimaciones. Sin embargo, los programas informáticos no siempre funcionan como de-

ben, y hay muchos ejemplos de investigaciones cuantitativas con modelos estadísticos imprecisos que llegan a conclusiones incoherentes. Por desgracia, en los estudios cualitativos no disponemos de nada tan automático como un programa de ordenador para descubrir diseños de investigación imprecisos, pero ser consciente del problema hace que sea más fácil identificar este tipo de diseños e idear soluciones. Además, los cualitativistas suelen tener una ventaja sobre los cuantitativistas: con frecuencia, disponen de suficiente información para hacer que sus diseños sean precisos.

Supongamos que al recoger datos nos proponemos contrastar la validez de una hipótesis. Habría que diseñar la investigación para que nos diera el máximo control al distinguir entre diversos resultados posibles que fueran relevantes para la hipótesis. Sin embargo, hay dos situaciones en las que un diseño de investigación es impreciso y, por tanto, no nos concede ese control:

1. Cuando el número de inferencias que tenemos que hacer es mayor que el de consecuencias observadas.
2. Cuando en nuestros datos hay dos o más variables explicativas que se correlacionan perfectamente entre sí —desde el punto de vista estadístico, este problema se denomina «multicolinealidad» (las variables podrían incluso ser diferentes, pero, si podemos predecir una a partir de la otra sin error en los casos que tenemos, el diseño es impreciso).

Hay que señalar que estas situaciones, así como el concepto general de diseño impreciso, sólo tienen que ver con el objetivo de hacer inferencias causales. Un diseño de investigación que pretenda resumir portadores históricos no puede ser impreciso, a menos que directamente se retengan observaciones no relevantes. Toda recogida de datos que se haya concebido con el fin de revelar preguntas interesantes (véase el apartado 1.1 del capítulo 2) no puede ser imprecisa si tenemos, al menos, cierta información. Es evidente que, en cualquier caso, después puede haber imprecisión cuando volvamos sobre nuestros datos (o recogamos otros nuevos) para contrastar una hipótesis causal.

1.1 Más inferencias que observaciones

Consideremos la primera situación, aquella en la que hay más inferencias que consecuencias observadas. La inferencia es el proceso mediante el cual utilizamos hechos conocidos para saber algo sobre otros desconocidos. Lo que podemos descubrir a partir de una información restringida no es limitado, y resulta que la regla precisa es que un hecho (o consecuencia observada) sólo puede ofrecer información *independiente* acerca de otro hecho. De manera más general, cada observación puede ayudarnos a extraer, como

máximo, una inferencia, un número de observaciones n nos ayudará a hacer una cantidad de inferencias menor que n si las observaciones no son independientes. En la práctica, se suele necesitar mucho más que una observación para hacer una inferencia causal que sea razonablemente cierta.

Tener más inferencias que consecuencias observadas es una dificultad habitual en los estudios de caso cualitativos. Sin embargo, no es un problema inherente a este tipo de investigación, sino al estudio que se conceptualiza u organiza inadecuadamente a partir de muchas consecuencias observables de una teoría. En primer lugar, describiremos este problema y, posteriormente, analizaremos las soluciones.

Supongamos, por ejemplo, que tenemos tres estudios de caso y que cada uno describe el esfuerzo conjunto que realizan dos países para desarticular una estructura armamentística con tecnología muy avanzada. En cada estudio se incluyen interesantes descripciones de dichas estructuras, de las negociaciones que hay entre los países y del producto final. Durante nuestra investigación elaboramos una lista de siete razones importantes que hacen que los países tengan éxito en su proyecto conjunto de construcción de sistemas defensivos estratégicos, las cuales podrían ser variables explicativas muy plausibles. También podríamos haber entrevistado a altos cargos de los diferentes países y así habérmolos dado cuenta de que ellos también creen que éstas son las variables importantes. Esta forma de abordar el problema nos daría siete hipótesis plausibles y también observaciones sobre ocho variables: las siete explicativas y la dependiente. Sin embargo, en estas circunstancias, aunque recogeríamos los datos de la manera más cuidadosa, no nos libraríamos de un problema fundamental. Aunque este enfoque —que, fundamentalmente, se basa en una comparación estructurada y bien centrada— es valioso, su impreciso diseño de investigación no nos proporciona una metodología para la inferencia causal. Este diseño, que utiliza siete variables causales y sólo tres observaciones, no puede determinar cual de las hipótesis es correcta (si es que hay alguna que lo sea).

Al intentarnos a explicaciones imprecisas, a veces buscamos otras causas posibles para el acontecimiento que intentamos explicar. Esto es exactamente lo contrario de lo que la lógica de la explicación tendría que sugerirnos. Una descripción mejor o más completa de cada caso de estudio no es la solución, ya que, al haber más parámetros que observaciones, casi cualquier explicación del impacto de cada una de las siete variables estaría tan en consonancia con los datos como las demás. Si el diseño de la investigación es impreciso, no habrá descripción, por profusa y detallada que sea, ni método, aunque sea inteligente; ni investigador, independientemente de su pericia, que pueda sacar mucho de cualquiera de las hipótesis causales. Intentar incluir todas las variables explicativas puede llevarnos a cruzar la línea que nos separa de un diseño de investigación impreciso.

Un número mayor de estudios de caso puede ayudarnos a solucionar el problema que plantea el diseño de investigación del párrafo anterior, pero

incluirlos quizá cueste un tiempo y unos recursos de los que no disponemos o quizá sólo haya tres ejemplos de los fenómenos que estudiamos. El problema de la imprecisión se podría solucionar reorientando la investigación hacia el estudio del efecto de unas determinadas variables explicativas sobre un abanico de acciones de los estados, más que de las causas de ciertos conjuntos de consecuencias, como es el éxito de proyectos conjuntos. Una solución alternativa que no cambia de manera tan drástica el centro de atención del estudio consistiría en añadir un nuevo conjunto de observaciones que hubieran sido medidas en diferentes niveles de análisis. Además de usar la estructura armamentística, se podrían identificar todas las decisiones importantes que se producen al desartillar cada una de esas estructuras. Este procedimiento puede ser de bastante ayuda si en estas decisiones hay más información relevante para la inferencia causal, y, siempre que nuestra teoría tenga alguna consecuencia sobre el carácter de tales decisiones si se atiende a las partes que lo componen. Al añadir nuevas observaciones procedentes de diferentes niveles de análisis podemos aportar muchas pruebas de tales consecuencias. Este método es una de las formas más útiles de rediseñar la investigación cualitativa y de evitar (hasta cierto punto) tanto la imprecisión como el sesgo de la variable omitida, que analizaremos en el apartado 2 de este capítulo. De hecho, el aumento del número de observaciones mediante el diseño de investigación es el tema principal del capítulo 6 (especialmente de su apartado 3).

1.2 Multicolinealidad

Supongamos que, centrándonos en los efectos de ciertas causas preseleccionadas en lugar de en las de los efectos observados, conseguimos resolver el problema de las pocas observaciones añadiendo otras procedentes de diferentes niveles de análisis o mediante algún otro cambio en el diseño de la investigación. Seguirá siendo necesario ocuparse del otro problema que conduce a la imprecisión de los diseños de investigación: «la multicolinealidad». Hemos tomado este término de la investigación estadística, especialmente del análisis de regresión, pero vamos a utilizarlo de forma mucho más general. El uso que hacemos de él se refiere particularmente a cualquier situación en la que podamos predecir perfectamente una variable explicativa en función de una o más de las restantes. No partimos de ningún supuesto de linealidad, como si se hace en el significado habitual del término estadístico.

Un análisis formal del problema que se plantea cuando hay más inferencias que observaciones

La mejor manera de comprender este problema es mediante un caso sencillo. Con el fin de maximizar la intuición, en la prueba siguiente evitamos generalizar; es decir, presentar la prueba más global, lo cual no obsta para que el carácter intuitivo del ejemplo lo haga mucho más aplicable.

Supongamos que nos interesa hacer inferencias sobre dos parámetros en un modelo causal que tiene dos variables explicativas y una única variable independiente

$$E(Y) = X_1\beta_1 + X_2\beta_2 \quad (4.1)$$

pero no tenemos más que una observación para hacer la estimación (o sea, $n = 1$). Supongamos además que, para una mayor claridad, nuestra observación consiste en $X_1 = 3$, $X_2 = 5$ e $Y = 35$. Para terminar, supongamos que en este ejemplo sucede que Y es igual a su valor esperado (lo cual ocurriría por azar o si no hubiera variabilidad aleatoria en Y). De este modo, $E(Y) = 35$. En la práctica, nunca responderemos de esta última información (por la aleatoriedad inherente a Y), de manera que si, en este caso, tenemos problemas al hacer una estimación de β_1 y de β_2 , seguramente fracasaremos en el caso general cuando no dispongamos de dicha información sobre el valor esperado.

Por tanto, el objetivo, será hacer una estimación de los valores de parámetro en la siguiente ecuación:

$$E(Y) = X_1\beta_1 + X_2\beta_2 \quad (4.2)$$

$$35 = 3\beta_1 + 5\beta_2$$

El problema es que esta ecuación no tiene una única solución. Por ejemplo, los valores ($\beta_1 = 10$, $\beta_2 = 1$) la satisfacen, pero también lo hacen ($\beta_1 = 5$, $\beta_2 = 4$) y ($\beta_1 = -10$, $\beta_2 = 13$). Esto resulta bastante problemático, ya que los diferentes valores de los parámetros pueden señalar consecuencias esenciales muy diferentes acerca de los efectos causales de estas dos variables; en el último caso, cambia incluso un signo. En realidad, tanto estas como un número infinito de soluciones satisfacen igualmente la ecuación. De este modo, no hay nada en el problema que nos ayude a distinguir entre las soluciones, porque todas ellas concuerdan con nuestra única observación.

Supongamos, por ejemplo, que en el estudio sobre cooperación armamentística antes mencionado dos de las hipótesis sean: (1) es más probable que progrese la colaboración entre los países cuando sus tamaños son diferentes que cuando son similares y (2) su éxito es más probable cuando los países son vecinos que cuando sí lo son. Las variables explicativas que están detrás de estas hipótesis se centran en el impacto negativo que tiene la rivalidad sobre la colaboración; ambas son bastante razonables y puede que incluso las respalden entrevistas en profundidad o la bibliografía sobre política industrial. Sin embargo, supongamos que sólo logramos identificar un pequeño conjunto de datos en el que la unidad de análisis sea un par de países. Supongamos, además, que únicamente recogemos dos tipos de observaciones: (1) países vecinos de tamaño diferente y (2) países vecinos de tamaño similar. Si sucediera que (por diseño o por azar) todas nuestras observaciones pertenecieran a dichas categorías, sería imposible utilizar los datos para encontrar prueba alguna que constatará o negará cualquier hipótesis. La razón es que las dos variables explicativas se correlacionan perfectamente: toda observación en la que los posibles asociados tengan un tamaño similar preocupará a los vecinos, y al revés. El tamaño y la proximidad geográfica son variables conceptualmente muy diferentes, pero, al menos en este conjunto de datos, no se distinguen la una de la otra. En este punto, lo mejor que podría hacerse es recoger más observaciones en las que estados de tamaño similar fueran vecinos. Si esto fuera imposible, la única solución sería buscar consecuencias observables en algún otro nivel de análisis.

Aunque se solucione el problema de la imprecisión del diseño de investigación, puede que nuestras inferencias causales sigan siendo muy inciertas, por cuestiones como la de un número insuficiente de observaciones o la colinealidad entre las variables causales. Para tener más confianza en las estimaciones siempre habrá que intentar *maximizar el control* sobre nuestro problema. En consecuencia, siempre tendremos que observar tantas consecuencias de nuestra teoría como sea posible. Es evidente que, en la práctica, el tiempo y los recursos que podemos dedicar a la recogida de datos siempre están condicionados, pero la necesidad de tener más observaciones que inferencias debe alertarnos frente a situaciones en las que haya que dejar de recabar información detallada sobre un caso determinado para comenzar a recogerla sobre otros parecidos. La preocupación por la imprecisión también tiene que influir en cómo definimos nuestra unidad de análisis: tendremos problemas al hacer inferencias causales válidas si en nuestro estudio sólo disponemos de una unidad de análisis en la que los acontecimientos son prácticamente únicos, ya que resultará muy difícil encontrar muchos ejemplos. Aunque lo que nos interesa sea el comunismo, la Revolución Francesa o las causas de la democracia, también compensará dividir el problema en unidades manejables y más numerosas.

También recomendamos que se maximice el control limitando el número de variables explicativas con las que se quiere hacer inferencias causales.

Un análisis formal de la multicolinealidad

Para una mejor comprensión, utilizaremos la misma estrategia que en el último análisis formal, es decir, dar sólo una prueba de un único caso concreto. También aquí la intuición puede llevarnos más allá de este simple ejemplo, que es muy similar al anterior. De nuevo utilizamos un ejemplo muy parecido al anterior. Retomamos el modelo de la ecuación (4.1), pero esta vez tenemos un número muy elevado de observaciones y cada una de nuestras dos variables explicativas es una combinación lineal perfecta de la otra. De hecho, para que el problema resulte aún más transparente, suponemos que las dos son la misma, de manera que $X_1 = X_2$. Podríamos haber definido X_1 y X_2 en función de dos variables esencialmente diferentes (como género y embarazo), pero en una muestra de datos ambas podrían resultar la misma (si se diera el caso de que todas las mujeres encuestadas estuvieran embarazadas). ¿Podemos distinguir los efectos causales de estas diferentes variables?

Hay que señalar que la ecuación (4.1) se puede escribir de la siguiente manera:

$$E(Y) = X_1\beta_1 + X_2\beta_2 \tag{4.3}$$

$$= X_1(\beta_1 + \beta_2)$$

La segunda línea de esta ecuación debería dejar claro que, independientemente de lo que sean $E(Y)$ y X_1 , numerosos valores de β_1 y de β_2 pueden satisfacerla. (Por ejemplo, si $\beta_1 = 5$ y $\beta_2 = -20$ satisfacen la ecuación (4.3), también ocurre así cuando $\beta_1 = -20$ y $\beta_2 = 5$.) De este modo, aunque ahora disponemos de muchas más observaciones que parámetros, la multicolinealidad nos deja el mismo problema que cuando teníamos más parámetros que unidades: no hay ninguna forma de estimación que nos pueda ofrecer un solo cálculo aproximado de los parámetros.

dades. En la práctica, los que experimentan no seleccionan aleatoriamente, sino que eligen dentro de una población adecuada —como los estudiantes de segundo curso de una facultad—, pero aquí nos centramos en una situación ideal. Analizamos la selección, posponiendo al final del capítulo 5 el estudio de la adjudicación de valores de las variables explicativas.

En la investigación cualitativa, y de hecho en gran parte de la cuantitativa, la selección aleatoria puede no ser factible, porque el universo de casos no está claramente especificado. Por ejemplo, si quisiéramos elaborar una muestra aleatoria de las élites que tienen que ver con la política exterior de los Estados Unidos, no las encontraríamos en una lista comparable a la de las

Al limitar estas variables, hay que procurar no incurrir en el sesgo de la variable omitida (apartado 2 del capítulo 5). Las reglas que aparecen en el apartado 3 del siguiente capítulo tendrían que ayudarnos en este punto. Un proyecto que logra sus fines es el que explica mucho utilizando pocos elementos, y lo ideal es que se utilice una sola variable explicativa para aclarar numerosas observaciones en las dependientes.

Un diseño de investigación que explique mucho utilizando un gran número de elementos no será muy informativo, pero un diseño impreciso no nos permitirá separar efectos causales en absoluto. Para evitar este problema hay que seleccionar observaciones de las mismas variables u otras que sean consecuencia de nuestra teoría. Después de formalizar la multicolinealidad (véase el recuadro), nos ocuparemos con más detalle del análisis de los métodos de selección de observaciones y del problema del sesgo de selección.

2. Límites de la selección aleatoria

En los estudios en los que n es grande evitamos el sesgo de selección mediante una selección aleatoria de las observaciones, porque un criterio aleatorio no se correlaciona con todas las variables explicativas o dependientes posibles. La aleatoriedad es un enfoque muy valioso porque ofrece un procedimiento de selección que *automáticamente* no se correlaciona con todas las variables. Esto quiere decir que, cuando n es grande, las posibilidades de que un criterio de selección se correlacione con cualquier variable observada son extremadamente pequeñas. En consecuencia, la selección aleatoria de observaciones elimina automáticamente el sesgo de selección en este tipo de estudios. En un mundo en el que hay muchas variables potenciales para confundirnos —algunas de ellas desconocidas— la aleatoriedad tiene muchas virtudes para los científicos sociales. Si, como ocurre a menudo en la investigación política, hemos de abandonar la aleatoriedad, habrá que hacerlo con cuidado.

En las ciencias sociales los experimentos controlados sólo se realizan ocasionalmente. Sin embargo, son un modelo útil para comprender ciertos aspectos del diseño de la investigación no experimental. Los mejores experimentos suelen combinar la selección de observaciones y la adjudicación de valores a las variables explicativas según criterios aleatorios con un gran número de observaciones (o pruebas experimentales). Aunque ningún experimento consiga resolver el problema fundamental de la inferencia causal, quienes los realizan suelen poder seleccionar sus observaciones (en vez de recibir las de los procesos sociales) y dar tratamiento a las variables explicativas) a las unidades; de ahí que merezca la pena centrarse en estas dos ventajas de los experimentos: control sobre la selección de observaciones y adjudicación de valores de las variables explicativas a las unidades.

Los cualitativistas apenas recurren abiertamente a la aleatoriedad como criterio de selección, pero han de asegurarse de que los criterios que utilizan no tengan consecuencias similares. Supongamos, por ejemplo, que a un investigador le interesen los países de Europa del Este con cultura católica que fueron dominados por la Unión Soviética después de la Segunda Guerra Mundial: Checoslovaquia, Hungría y Polonia. Se dará cuenta de que en los años setenta y ochenta hubo una considerable variación en las políticas de cada uno de estos países: en Polonia, surgió un movimiento antigubernamental bien organizado (Solidaridad), en Checoslovaquia actuaba un grupo mucho más reducido de intelectuales (Carta 77), mientras que en Hungría no se desarrolló ningún movimiento nacional tan amplio. El problema consiste en explicar estas discrepancias.

Para examinar la naturaleza de los movimientos antigubernamentales es preciso estudiar en profundidad la prensa, los documentos de los partidos comunistas recientemente desclasificados y también hacer muchas entrevistas a los participantes en dichos movimientos —para lo cual hay que conocer su idioma. Además, la dificultad de investigar la Europa Oriental contemporánea estriba en que se necesitará un año para estudiar cada uno de los países. Por lo tanto, para realizar este trabajo parece factible centrarse sólo en dos de ellos. Por fortuna, por razones no relacionadas con este proyecto, la investigadora conoce ya el checo y el polaco, de manera que decide estudiar Carta 77 en Checoslovaquia y Solidaridad en Polonia. Resulta evidente que este proceder dista mucho de ser una adjudicación aleatoria, pero, al menos, probablemente la razón para elegir estos países no tenga relación con la variable dependiente. Sin embargo, en este ejemplo resulta que el criterio de selección (conocer ciertos idiomas) sí se correlaciona con la variable dependiente y que, por lo tanto, habrá un sesgo de selección. En este caso, si no fuera por los condicionantes lingüísticos, habría sido mejor hacer una selección fundamentada, no aleatoria.

Esta investigadora podría evitar el sesgo de selección olvidando que sabe checo y poniéndose a aprender húngar, ¡pero esta solución no resulta en absoluto atractiva! En esta observación la alternativa más realista sería aprovecharse de que se conoce el sesgo para evaluar en qué dirección actúa, corregirlo al menos parcialmente y matizar las conclusiones en consecuencia. Desde el principio, la investigadora sabe que ha reducido sistemáticamente el grado de variación de su variable dependiente, lo cual tendería a hacer que subestimara sus estimaciones causales, al menos como promedio (aunque esto pueden cambiarlo otros problemas de la misma investigación). Además, para que cualquier variable explicativa sea plausible, tendrá que realizarse, al menos, la suficiente investigación secundaria en Hungría como para darse cuenta de si la dirección del sesgo va a favor o en contra de su hipótesis. Por ejemplo, esta podría señalar, partiendo de los casos checo y polaco, que los movimientos antigubernamentales masivos surgen cuando los regímenes comunistas tienen un carácter indulgente, relativa-

Podríamos juntar listas pro-circunscripciones electorales para el Congreso. Podríamos siempre correlacionar el sesgo de que en ellas hubiera ciertos sesgos. Por ejemplo, el universo de selección podría basarse en listas estatales de ciudadanos a los que se hubiera consultado acerca de cuestiones de política exterior. Ciertamente, se podría considerar que estas personas pertenecen a una élite de la política exterior, pero si el problema de la investigación tuviera que ver con la relación que existe entre procedencia social y preferencia por una política, nuestra lista podría estar sesgada a favor de individuos de clase alta que suelen apoyar la política del gobierno. Además, quizá no podríamos estudiar una muestra de élites tomada aleatoriamente de una lista, porque los gastos de desplazamiento podrían ser muy altos. Entonces, habría que seleccionar sólo a los que vivirían en nuestra región, lo cual introduciría probablemente otros sesgos.

La selección aleatoria no es necesariamente la mejor técnica, ni siquiera cuando es factible. Es frecuente que los cualitativistas se resistan a utilizar (con razón) el concepto de selección aleatoria, negándose a correr el riesgo de no contar con casos importantes que este método podría descartar (¿para qué estudiar las revoluciones si no incluimos la Revolución francesa?). En realidad, si sólo tenemos un número reducido de observaciones, la selección aleatoria puede no resolver el problema de selección y puede ser incluso peor que otros métodos de selección. Creemos que, intuitivamente, muchos cualitativistas se dan cuenta de esto cuando se quejan de los que consideran descaminados sermones sobre las virtudes de la aleatoriedad que les dirigen algunos investigadores cuantitativos. De hecho, utilizando un modelo formal de investigación cualitativa muy sencillo, demostraremos a continuación que las observaciones suele producir graves sesgos.

Supongamos que tenemos tres unidades en las que hay observaciones sobre una variable dependiente que puede expresarse como alta, media o baja, pero que sólo dos de ellas se seleccionan para el análisis ($n = 2$). Ahora necesitamos un criterio de selección. Si hacemos que I se refiera a una unidad seleccionada para el análisis y que 0 indique una omisión, sólo serán posibles tres criterios de selección: $(1,1,0)$, lo cual significa que todas las opciones alta y media pero no la baja $(0,1,1)$ y $(1,0,1)$. El problema es que sólo el último criterio de selección, en el que se omite la segunda unidad, no se correlaciona con la variable dependiente. En este ejemplo, en el que n tiene un valor reducido, al ser equivalentes la selección aleatoria de observaciones y la elección aleatoria de uno de estos tres criterios de selección posibles, ¡existirá una probabilidad de dos contra tres de que la selección de observaciones produzca un sesgo de selección! Sería mucho más probable que una selección más cuidadosa de las observaciones, que utilizara el conocimiento previo de los posibles valores de la variable dependiente, optara por el tercer criterio de selección, evitando así el sesgo.

plio más evidente se produce cuando, sabiendo el resultado que queremos obtener de la investigación (confirmar la hipótesis que preferimos), seleccionamos con más o menos sutileza observaciones que respondan a combinaciones de las variables independientes y dependientes que constatan la conclusión deseada. Supongamos que creemos que la inversión estadounidense es una causa primordial de violencia interna en los países del Tercer Mundo; en consecuencia, seleccionamos, por un lado, un conjunto de naciones que ha recibido grandes sumas de los Estados Unidos y en las que se ha registrado mucha violencia y, por otro, un grupo en el que no ha habido ni inversión ni violencia. Existen otras observaciones que ponen de manifiesto las otras combinaciones (mucha inversión y ninguna violencia o bastante inversión y mucha violencia), pero se dejan «comodamente» de lado. El sesgo de selección casi nunca es tan flagrante como en este ejemplo, pero como en la investigación cualitativa los criterios de selección suelen estar implícitos y muchas veces se elige sin intentar evaluar conscientemente los sesgos potenciales, existen muchas posibilidades de que el sesgo se filtre en nuestros procedimientos de selección.

3.1 Seleccionar en función de la variable dependiente

En estudios con un n grande la selección aleatoria nos permite hacer caso omiso de la relación que existe en nuestro análisis entre los criterios de selección y otras variables. Al prescindir de la selección aleatoria, hay que considerar de qué manera se relacionan los criterios utilizados con cada variable. Esto nos conduce a una regla fundamental y evidente: *la selección tiene que permitir que se produzca, al menos, alguna variación en la variable dependiente*. Este requisito resulta tan obvio que podría pensarse que no es necesario mencionarlo. ¿Cómo podríamos explicar los cambios que se producen en una variable dependiente si ésta no cambia? Por desgracia, la bibliografía está llena de obras en las que se comete precisamente el error de no dejar que cambie: por ejemplo, investigaciones que intentan explicar los estallidos bélicos estudiando sólo guerras, el comienzo de las revoluciones investigando únicamente estos fenómenos o las pautas de asistencia a las urnas mediante entrevistas a los que no votan.⁶

En el capítulo 1 señalamos que los buenos científicos sociales suelen crecerse cuando existen anomalías que hay que explicar. Uno de los efectos de esta tendencia es que pueden seleccionar, especialmente los cualitativos, observaciones que tienen un resultado sorprendente, como las revoluciones sociales que tuvieron lugar en Francia durante el siglo XVIII y las que ocurrieron en ese mismo país y en China durante el XX (Skocpol, 1979). Esta es una selección de observaciones que se realiza en función de la variable dependiente y, por tanto, corre el riesgo de introducir el sesgo de selección que analizamos en este apartado. Cuando las observaciones se eli-

mente poco represivo, y no cuando son fuertes y represivos. La investigación tendría que saber que, aunque Hungría tenía el régimen comunista más benevolente de Europa del Este, no hubo en este país un movimiento anti-gubernamental masivo. En consecuencia, si es posible, la investigadora tendrá que aumentar el número de observaciones con el fin de evitar el sesgo de selección, y aunque no todas puedan estudiarse exhaustivamente, el hecho de conocerlas parcialmente podrá, al menos, mitigar el problema. Resultaría muy productivo acompañar este estudio detallado de dos casos con el de otros muchos menos portemonstrados que se basaran en fuentes secundarias y, quizá, añadir también un análisis bastante más agregado (y necesariamente superficial) de un gran número de casos. Si los estudios de caso detallados producen una clara hipótesis causal, puede que sea mucho más fácil recabar sólo información de las pocas variables que se consideren importantes en un número mucho mayor de observaciones en varios países (véase el apartado 3 del capítulo 4 para un análisis similar y un tratamiento más formal). Otra solución podría ser reorganizar la enorme información recogida en los dos estudios de caso para que se convirtiera en un gran número de consecuencias observables de la teoría. Por ejemplo, si fuera correcta la teoría de que la represión del régimen impide que aumenten los movimientos antigubernamentales, éstos no habrían podido desarrollarse mucho en las ciudades o regiones en las que la policía secreta era entusiasta y eficiente, mientras que habrían actuado con mayor libertad allí donde el comportamiento de esas fuerzas era más relajado —siempre teniendo en cuenta las circunstancias de cada país.

3. Sesgo de selección

¿Cómo tenemos que seleccionar las observaciones para incluirlas en un estudio? Si queremos saber las opiniones de los cargos municipales, ¿a cuáles tenemos que entrevistar? Si estamos haciendo una comparación entre estudios de caso de grandes guerras, ¿cuáles hay que seleccionar? Si nos interesan los vetos presidenciales en los Estados Unidos, ¿hay que elegir todos?, ¿sólo los que se han producido desde la Segunda Guerra Mundial?, ¿hacer una muestra aleatoria?, ¿o acaso contar únicamente los invalidados por el Congreso? Cuando se comienza a diseñar un proyecto de investigación no hay ninguna pregunta que esté tan presente como ésta: ¿qué casos (o, más exactamente, qué observaciones) tenemos que incluir en nuestro estudio? En la investigación cualitativa la elección de las observaciones resulta crucial para el resultado del estudio y condiciona en qué medida éste puede generar resultados precisos y fiables.

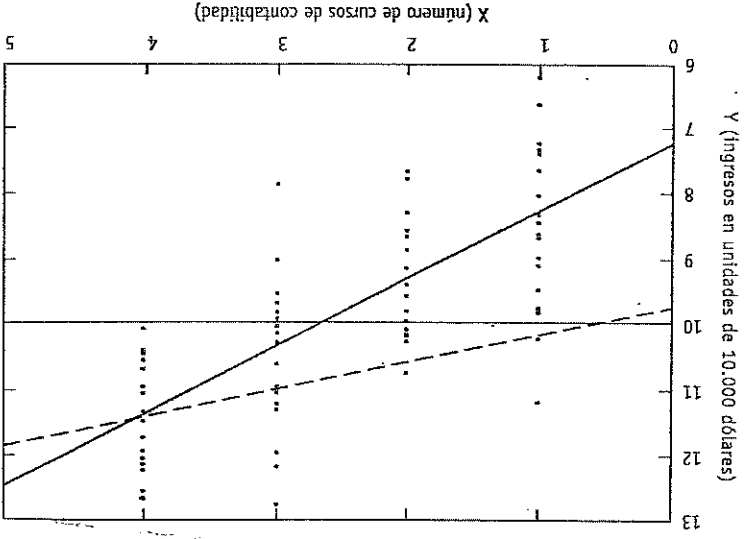
Como hemos visto en el apartado anterior, la selección aleatoria no suele ser apropiada cuando abordamos investigaciones con un n reducido, pero prescindir de este método abre la puerta a muchos tipos de sesgo. El ejem-

gen a partir de un determinado valor de la variable dependiente, no se puede conocer ninguna de las causas de ésta sin tener en cuenta otros ejemplos en los que adopte valores diferentes. Por ejemplo, en su investigación, Theodor Skocpol (1979) soluciona en parte este problema incluyendo unos pocos datos sobre «momentos de crisis revolucionaria» (Skocpol, 1984, p. 380) en la Inglaterra del siglo XVII, en la Prusia/Alemania del XIX y en el Japón del mismo siglo. Considera estas observaciones «casos de control», aunque los analiza con mucho menos detalle que los principales. El hecho de que exista un sesgo producido por seleccionar en función de la variable dependiente no implica que nunca haya que tener en cuenta los valores de dicha variable al diseñar investigaciones. Lo que supone es que, como se indica más adelante y en el capítulo 6, tenemos que ser conscientes de los sesgos que introduce esa selección en la variable dependiente y buscar, en la medida de lo posible, maneras de corregirlos.

Existe también una versión más leve y más habitual del problema que conlleva seleccionar en función de la variable dependiente. En algunos casos, el diseño de una investigación contempla el cambio de esta variable pero la trunca: es decir, nuestras observaciones se limitan a una gama de cambios menor de la que puede alcanzar dicha variable en el mundo real. En estas situaciones, se puede decir algo sobre las causas de la variable dependiente, pero es probable que las inferencias estén sesgadas, ya que, como el criterio de selección no tiene en cuenta las variables explicativas, *cuálquier criterio de selección que se correlacione con la variable dependiente aventura la media de estimaciones de los efectos causales* (véanse Achen, 1986; King, 1989, capítulo 9). En la investigación cuantitativa, esto supone que las estimaciones numéricas de los efectos causales estarán más cerca de cero que en la realidad. En la cualitativa el sesgo de selección implicará que el auténtico efecto causal es mayor de lo que cree el investigador (por supuesto, a menos que éste sea consciente de lo que aquí se expone y corrija sus estimaciones en consecuencia). Si sabemos que existe un sesgo de selección y no hay forma de esquivarlo mediante una muestra mejor, esto supone que la media de nuestra estimación mostrará, al menos, que el verdadero efecto causal tiene un límite más bajo. El grado de subestimación de este efecto causal depende de la gravedad del sesgo de selección (hasta qué punto se correlaciona el criterio de selección con la variable dependiente), y de ahí habrá que tener pruebas detalladas o, por lo menos, cierta idea.

Los casos de sesgo de selección externo — en los que el diseño no permite que cambie la variable dependiente — son fáciles de manejar: ¡evítelos! Si los utilizamos no conoceremos efectos causales. La forma modificada de este tipo de sesgo, en la que las observaciones se seleccionan de forma que están relacionadas con la variable dependiente, puede ser más difícil de eliminar, ya que quizá no tengamos acceso a todas las observaciones que queremos. Sin embargo, por fortuna, los efectos de este sesgo no son tan devastadores, porque algo si se puede aprender, quizá nuestras inferencias

Gráfico 4.1 Sesgo de selección



estén sesgadas, pero de una forma previsible que *podemos* compensar. Los siguientes ejemplos ilustran esta situación.

Dado que con frecuencia nos veremos obligados a elegir observaciones que se correlacionan con la variable dependiente y que, por tanto, contienen sesgo de selección, merece la pena comprobar si, de todos modos, podemos sacarles información valiosa. El gráfico 4.1, que presenta un simple modelo de sesgo de selección, demuestra que si es posible. Cada punto representa un observación (una persona, por ejemplo); el eje horizontal es la variable explicativa (como puede ser el número de cursos de contabilidad que se reciben en una escuela de negocios), y el vertical, la dependencia (por ejemplo, el salario inicial en el primer trabajo con dedicación exclusiva, citado en unidades de 10.000 dólares estadounidenses). La línea de regresión que muestra la relación que hay entre estas dos variables es la continua que cruza el diagrama de puntos. Cada curso de contabilidad tiene un valor medio de 10.000 dólares en concepto de salario inicial. El diagrama de puntos que rodea esta línea continua de regresión indica que, como es habitual, ésta no se ajusta perfectamente a la situación de todos los estudiantes. En gráficos como éste la desviación *vertical* que hay entre los puntos y la línea representa los errores de predicción (dados unos determinados valores para las variables explicativas) que, por tanto, se minimizan a la hora de adecuar dicha línea a los puntos.

Ahora supongamos que un alumno que acaba de matricularse en una escuela de negocios quiera estudiar cómo va a aumentar su renta cuando ter-

El problema que acabamos de describir es habitual en la investigación cuantitativa (véase Geddes, 1990) y puede surgir de un procedimiento tan aparentemente inocuo como seleccionar casos a partir de los datos disponibles, cuando la disponibilidad de esos datos está relacionada con la variable dependiente. Supongamos, por ejemplo, que nos interesen los factores que han determinado la participación del presidente de los Estados Unidos en decisiones de política internacional importantes durante los últimos años y

3.1.1 Ejemplos de sesgo de selección producido por el investigador

Este es un ejemplo concreto de cómo podemos subestimar un efecto causal cuando seleccionamos en función de la variable dependiente. Por fortuna, si hay algo que nuestro estudiante puede hacer para solucionar este problema. Supongamos que, después de terminar sus estudios en la escuela de negocios, se aburre de ganar dinero y se matricula en una facultad de ciencias sociales en la que aprende lo que es el sesgo de selección. Esta muy ocupado preparando sus exámenes globales, así que no tiene tiempo de retomar adecuadamente su investigación. No obstante, sabe que su salario inicial se habría incrementado en una cantidad considerablemente superior a la cifra de 5.000 dólares en que él había valorado cada curso de contabilidad. Ahora llega a la conclusión de que, como su criterio de selección era bastante estricto (de hecho, determinista), habría ganado más dinero en los negocios si se hubiera apuntado a más cursos de contabilidad —pero, como ha decidido no maximizar sus ingresos (¿quién podría entrar en una facultad con este propósito?), se alegra de no haber conocido el sesgo de selección antes de que cambiáran sus valores.

que nos proponemos estudiar aquellas en cuyas reuniones consta que participó el presidente. El problema de este diseño de investigación es que el criterio de selección (la disponibilidad de información) probablemente se correlacione con unos niveles relativamente bajos de participación presidencial (la variable dependiente), ya que es probable que el presidente haya tenido una mayor presencia en las reuniones más secretas —a las que no se podrá acceder— que en aquellas cuyas deliberaciones se hayan hecho públicas. Por lo tanto, en el conjunto de las observaciones que se refieren a la información disponible, habrá una excesiva representación de los acontecimientos en los que se haya registrado un participación presidencial menor, lo cual sesgará las inferencias que hagamos acerca de los factores determinantes de dicha actividad.

El razonamiento utilizado en nuestro ejemplo de la escuela de negocios puede ayudarnos a comprender las consecuencias que tendrá un inevitable sesgo de selección en la investigación cuantitativa. Supongamos que en el estudio anterior nos interesara saber si los presidentes se implican más cuando los acontecimientos conlleven una amenaza de utilización de la fuerza que cuando ésta no se plantea. Supongamos también que, según las pruebas disponibles, quizá basadas en dos docenas de observaciones, sí existe esta relación entre los hechos pero su magnitud es sorprendentemente baja. Para calibrar el grado de sesgo de selección de este estudio, en primer lugar haríamos una lista de situaciones relativas a la política exterior en las que el presidente hubiera actuado o hecho declaraciones públicas, independientemente de que se disponga de alguna información sobre los procesos de toma de decisiones. Esta lista evitaría una de las causas de sesgo de selección que hemos identificado: el mayor secretismo en las decisiones que conlleven una amenaza de utilización de la fuerza. Esta nueva lista no incluiría todos los problemas en los que se implicó el presidente, ya que no registraría las operaciones encubiertas ni aquellas en las que no se tomó ninguna medida, pero sí sería una lista más larga que la primera porque en esta necesitaríamos información sobre la toma de decisiones. Podríamos después comparar ambas listas para determinar si (tal como sospechábamos) existía un sesgo que favorecía a aquellos casos sobre los que había información relativa a la toma de decisiones e iba en contra de aquellos en los que se utilizaba la fuerza o se amenazaba con utilizarla. Si fuera así, sería razonable inferir que, probablemente, la verdadera relación entre los hechos y el sesgo de selección suele aparecer en la política comparada cuando los investigadores tienen que viajar a determinados lugares para abordar su objeto de estudio. Con frecuencia, su capacidad de elegir unidades es limitada, ya que algunos gobiernos no dan información a los investigadores extranjeros. Por desgracia, esta negativa puede correlacionarse con la variable dependiente que interesa al investigador. Alguien que quiera explicar la liberalización de los regímenes autoritarios basándose en

las tácticas utilizadas por los grupos de disidentes podría obtener resultados te, que es la liberalización. Es evidente que no vamos a recomendar que se realice trabajo clandestino en lugares inhóspitos, pero sí aconsejamos que se tenga en cuenta este problema y que se utilice la imaginación para buscar fuentes de datos alternativas cuando no se dispone de las que están en el lugar de los hechos. Reconocer estas dificultades también podría llevarnos a revisar el diseño de nuestra investigación, con el fin de enfrentarnos a la realidad del acceso a las fuentes para fines académicos en el mundo. Si no podemos solucionar el problema de los datos, al menos podemos utilizar nuestras indagaciones sobre el sesgo de selección para saber hacia dónde se sesgarán nuestros resultados y quizá así podamos corregir parcialmente ese inevitable sesgo de selección que conlleva un estudio de este tipo. Es decir, que si esta clase de sesgo es inevitable, hay que analizarlo, determinar su dirección y, si es posible, su magnitud, y utilizar también esta información para modificar nuestras primeras estimaciones en la dirección correcta.

El sesgo de selección es un problema tan endémico que puede que sea útil abordar algunos ejemplos más de él. Pensemos en un trabajo reciente de Michael Porter (1990), autor que quería conocer el origen de lo que denominaba «ventaja competitiva» en los actuales sectores industriales y empresas. Para estudiar este tema, diseñó un proyecto a gran escala con diez equipos nacionales. Al elegir los diez países que iba a analizar optó, según sus palabras, por «los que ya habían demostrado su capacidad de competir en diversos sectores o, como era el caso de Corea y Singapur, los que presentaban síntomas de mejora en este sentido» (Porter, 1990, p. 22). En su impaciencia por ahondar en el problema que le interesaba, Porter fue consciente de que seleccionaba en función de la variable dependiente, haciendo que esta variable observable fuera casi constante. En consecuencia, cualquier intento que haga Porter, o todo aquel que utilice sus datos en este nivel de análisis, para explicar el grado de éxito de cada uno de los diez países producirá efectos causales considerablemente sesgados.

Sin embargo, lo que hizo Porter —intentar determinar las circunstancias y políticas que se relacionan con el éxito competitivo— tiene que ver, en cierto sentido, con el método del compromiso de Mill. Este no es un mal sistema para abordar el problema, ya que le permitió a Porter desarrollar alguna hipótesis sobre las causas de la ventaja competitiva, por medio de la observación de las características que compartían las naciones seleccionadas; sin embargo, su diseño de investigación le impidió evaluar efectos causales concretos.

Más grave resulta el error de lógica del método: sin un grupo de control de naciones (o sea, con la variable explicativa adscrita a otros valores) el

3.1.2 Ejemplos de sesgo de selección producido por el mundo

¿Acaso nos impide evitar el sesgo de selección el optar por registros de observaciones en vez de por una muestra? Se podría pensar que sí, ya que parece que en ese caso no hay en absoluto selección, pero no siempre ocurre de ese modo. Por ejemplo, suponíamos que queremos hacer una inferencia descriptiva mediante una estimación del grado de apoyo que tiene el Partido Liberal en el estado de Nueva York. Nuestra variable dependiente es el porcentaje de votos que ha recibido el candidato (o candidatos) que avala el partido en las circunscripciones en que se divide el estado para elegir su asamblea. El problema es que en muchas circunscripciones electorales el partido opta por no avalar a nadie y que, si no lo hace en aquellos lugares en los que está seguro de que va a perder (y este parece ser el caso), se produce un sesgo de selección cuando eliminamos circunscripciones en las que

reciente en los estudios de evaluación de políticas» (Achen y Sntidal, 1989, p. 162).

es frecuente en los estudios de evaluación de políticas» como el que cuando se utilizan mal los casos para hacer una estimación del índice de suasión función (en estadíos preliminares del proceso). «En consecuencia, de forma sistemática, se han excluido del análisis aquellas en las que la disuasión funciona mal los casos para hacer una estimación del índice de éxito de la disuasión, el diseño genera un "sesgo de selección" como el que este sesgado conjunto de observaciones presenta graves problemas porque, embargó, para fundamentar inferencias (y sin una corrección adecuada), gar formular hipótesis sobre las causas de los resultados observados. Sin dor puede describir los episodios de interés más significativos sin quizá lo- que ocurre cuando Porter hace incapie en el éxito competitivo, el observa- este enfoque, hay que caracterizarlo bien, al menos inicialmente: al igual la acción, pasando por el de emisión de ciertas señales. Para describir mejor producido en un primer estado del proceso que lleva la disuasión no se ha frecuencia «crisis agudas», o sea, aquellas en las que la disuasión no se ha (Achen y Sntidal, 1989, p. 151). Los que investigan este asunto analizan con para inducir a los oponentes a que se comporten de la forma deseada» encuentra un ejemplo notable de sesgo de selección: «el uso de amenazas En los escritos sobre política exterior que se ocupan de la disuasión se o si está en un punto intermedio».

si haya avalado a algún candidato. En este ejemplo, la selección utilizada para nuestra investigación pueden ser exactamente las mismas que si hubiéramos producido el problema nosotros.

El sesgo que se produce cuando la selección de casos se correlaciona con la variable dependiente es una de las dificultades más habituales a las que se enfrentan los investigadores que buscan pruebas en fuentes históricas, y entre ellos nos incluíamos casi todos. La razón es que los procesos «históricos» indican lo que ha de observarse en función de un conjunto de criterios selectivos que no siempre se desprende claramente del registro. Sin embargo, es esencial descubrir cómo se han producido esos datos. Tomemos un ejemplo de otro campo: unas culturas crean esculturas de piedra y otras de madera. Con el paso del tiempo, las primeras sobreviven, mientras que las segundas se pudren. En Europa, esta tendencia llevó a algunos historiadores del arte a subestimar la calidad y complejidad del arte africano antiguo, que solía utilizarse para hacer máscaras, porque la «historia» había ido eliminando de forma selectiva unas esculturas, a la vez que mantenía otras. El investigador cuidadoso debe siempre evaluar los sesgos de selección que pudiera haber en las pruebas de que dispone: ¿qué clase de acontecimientos se han podido registrar? ¿cuáles es posible que se hayan dejado de lado?

Pensemos en otro ejemplo. Los científicos sociales suelen partir de un punto final que quieren «explicar»: por ejemplo, el peculiar sistema organizativo de los estados modernos. El investigador observa que hace bastante tiempo (digamos, en 1500) había una gran variedad de unidades organizativas en Europa, pero que, posteriormente (por ejemplo, en 1900), todas las importantes, o casi todas, eran estados-nación. Lo que el investigador debe hacer es partir de las unidades existentes en 1500 y explicar las formas organizativas posteriores mediante un número de variables limitado. Entre los dos períodos muchas de las unidades de análisis habrán desaparecido, porque perdieron guerra o fueron incorporadas a entidades mayores, mientras que otras habrán sobrevivido. Por lo tanto, mediante una categorización cuidadosa, podremos obtener una variable dependiente que indicará si una entidad que se había convertido en estado-nación existía todavía en 1900 y, si no era así, cuándo había desaparecido.

Sin embargo, sin darse cuenta, lo que muchos historiadores hacen es bastante diferente. Como ha señalado Charles Tilly (1975, p. 15), comienzan por hacer investigación retrospectiva: seleccionan «para su comparación un pequeño número de estados de Europa occidental que todavía existían en el siglo XIX o que existen en el XX». Por desgracia para esos investigadores, «Inglaterra, Francia, e incluso España, son supervivientes de una lucha despiadada en la que la mayoría de los contendientes perdió». En la Europa de 1500 había unas cuantas unidades políticas, más o menos independientes, mientras que en la de 1900 existían alrededor de veinticinco. No había estado alemán en 1500, ni siquiera en 1800. Comparar, res-

do que, en realidad, fue bastante excepcional.

Por lo tanto, con este procedimiento, la selección se realiza en función de un valor de la variable dependiente: haber sobrevivido hasta 1900, lo cual sesga los resultados de la investigación reduciendo la media de efectos que se atribuyen a las variables explicativas que distinguen los estados que han sobrevivido de los que han durado menos tiempo. Tilly y sus colegas (1975), al darse cuenta del problema que planteaba el sesgo de selección, abandonaron el enfoque retrospectivo para optar por una formulación prospectiva del objeto de su investigación. Sin embargo, suponíamos que no hubiera sido posible realizar este enorme esfuerzo o que los autores hubieran deseado recabar los mejores datos disponibles para prepararse para un estudio más amplio. Podrían haber vuelto a analizar los estudios retrospectivos de que disponían, intentando que sus estimaciones de los efectos causales estaban, en la mayoría de las observaciones, sesgadas a la baja. Tendrían que recordar que, aunque los criterios antes descritos fueran completamente aplicables, cualquier nueva utilización que se hiciera de ellos podría exagerar o subestimar el efecto causal. Sin embargo, para el verdadero efecto causal, la mejor alternativa que podía salir de los defectuosos estudios retrospectivos era la de subestimar, al menos, la media de los efectos causales, si se presupone que las reglas mencionadas eran aplicables y que los criterios de selección se correlacionaban con la variable dependiente.

3.2 Selección en función de una variable explicativa

Seleccionar observaciones para un estudio mediante las categorías de la variable explicativa principal no produce problemas inferenciales porque nuestro procedimiento de selección no predetermina el resultado del estudio, ya que no hemos restringido el grado de variación posible de la variable dependiente. Al limitar el alcance de la variable causal clave podemos reducir el carácter general de nuestra conclusión o la legítima certeza con la que se puede mantener, pero no introducimos ningún sesgo. Seleccionar casos a partir de los valores de esta variable nos permite controlarla dentro de ese proceso. No se introducirá sesgo, aunque la variable causal se correlacione con la dependiente, porque ya hemos controlado esa variable explicativa. De este modo, aunque la selección se realice en función de una variable que se correlaciona con la dependiente, es posible evitar el sesgo si se controla la primera en el análisis.

Si nos remitimos al gráfico 4.1 podemos ver fácilmente que seleccionar en función de una variable explicativa no causa sesgo. Si restringimos esta

guro, no sería difícil hacer que esta variable fuera un control en nuestro análisis de las consecuencias que genera tener líderes militares o civiles. Se incluiría este control si estudiaríamos con qué frecuencia triunfan los golpes de Estado dirigidos por uno u otro tipo de líderes, en países en los que hay o no una jerarquía burocrática. La presencia de este control nos ayudaría a evitar el sesgo de selección, y su efecto causal proporcionaría información quizá importante sobre el proceso que se había utilizado realmente para seleccionar las observaciones.

3.3 Otros tipos de sesgo de selección

En todos los ejemplos anteriores, el sesgo de selección aparecía al elegir unidades según algún criterio que se correlacionaba con la variable dependiente desde el principio o después de haber tenido en cuenta las explicativas. Cuando la selección recibe este tipo de influencia, los efectos causales estimados siempre tienen valores que están por debajo del nivel real. Este es, con mucho, el tipo de sesgo de selección más frecuente, tanto en la investigación cualitativa como en la cuantitativa. Sin embargo, hay que señalar otro sesgo de este tipo, ya que sus consecuencias pueden ser precisamente las contrarias y producir una *sobreestimación* del efecto causal.

Supongamos que el efecto causal de alguna variable se transforma de centrado en ella. En el apartado 1 del capítulo 3 definimos el efecto causal en una sola unidad y dejamos que varíara en cada una de ellas. Supongamos, por ejemplo, que nos interesara el efecto causal que tiene la pobreza sobre la violencia política en los países de América Latina. Esta relación podría ser más acusada en algunos de ellos, como allí donde la historia reciente ofrece ejemplos de este tipo de violencia. En esta situación, en la que los efectos causales varían de una unidad a otra, un criterio de selección que se correlacione con el tamaño del efecto causal conducirá a un sesgo en las estimaciones de los efectos causales *medios*. Por lo tanto, si sólo llevamos a cabo nuestro estudio en países en los que ha habido violencia política en la historia reciente, pero pretendemos generalizar nuestras conclusiones a toda Latinoamérica, es posible que exageremos el efecto causal que estamos estudiando. Si seleccionamos unidades con efectos causales acusados y hacemos una media de ellos durante la estimación, el efecto causal medio resultante será una *sobreestimación*. Del mismo modo, si seleccionamos unidades con efectos débiles, el cálculo aproximado del efecto causal medio tendrá un valor menor que el real.

cita para excluir todas aquellas observaciones en las que la variable explicativa es igual a uno, la lógica de esta cantidad no se verá afectada y no cambiará la línea que atraviesa los puntos. En cierto modo, al tener ahora menos observaciones y menos información disponible para el problema inferencial, será un poco menos segura, pero, como promedio, no habrá sesgo.

En consecuencia, se puede evitar el sesgo seleccionando casos en función de la variable causal clave, pero también podemos lograr el mismo objetivo utilizando las categorías de una variable de control (siempre que, en términos causales, esta preceda a la variable causal clave, como ha de ser siempre en las variables de control). Los experimentos casi siempre seleccionan en función de las variables explicativas. Las unidades se crean cuando se manipulan dichas variables (administrando un fármaco, por ejemplo) y después se observa lo que le ocurre a la variable dependiente (si mejora la salud del paciente). En este caso, sería difícil seleccionar a partir de esta última variable, ya que su valor no se conoce hasta después del experimento. Sin embargo, la mayoría de los experimentos no son en absoluto perfectos, y podemos cometer el error de seleccionar en función de la variable dependiente, aplicando tratamientos a los pacientes que, sin que nos demos cuenta, se basan en la respuesta esperada.

Para dar otro ejemplo, pensemos que, si estamos investigando las consecuencias de la discriminación racial en las notas escolares de los niños negros, será bastante razonable seleccionar varias escuelas en las que haya poca discriminación y otras en las que haya mucha. Aunque nuestro criterio de selección se correlacione con la variable dependiente (los negros obtienen peores notas en las escuelas en las que hay segregación), esto no ocurre *después* de que tengamos en cuenta el efecto de las variables explicativas, ya que los valores de una de ellas determinan el criterio de selección. También podemos evitar el sesgo haciendo la selección a partir de una variable irrelevante para nuestro estudio (que no influye en nuestra variable dependiente). Por ejemplo, para estudiar el efecto de la discriminación en las notas, supongamos que alguien elige todas las escuelas cuyo nombre comienza por la letra «A». Evidentemente, aunque este criterio no es recomendable, si la variable irrelevante no representa a alguna otra que se correlaciona con la dependiente, no producirá sesgo.

Cuando resulta muy útil seleccionar en función de una variable irrelevante hay que realizar un análisis secundario de los datos que existen. Por ejemplo, supongamos que nos interesan los factores que hacen que triunfe un golpe de Estado. Nuestra hipótesis principal es que estas intenciones suelen tener más éxito con un jefe militar que con uno civil. Supongamos que encontramos un estudio de golpes de Estado que hubiera seleccionado los casos en función de en qué medida exista en cada país una jerarquía burocrática antes del golpe. Podríamos utilizar tales datos aunque este factor fuera irrelevante para nuestra investigación. Sin embargo, para ir sobre se-

4. Selección intencionada de observaciones

En la investigación política, lo normal es que no tengamos control sobre los valores de nuestras variables explicativas; los adjudica la «naturalaleza» o la «historia», pero no nosotros. En esta situación tan común la principal influencia que podemos recibir al diseñar la investigación afecta a la selección de los casos y de las observaciones. Como hemos visto en el apartado 2 de este capítulo, cuando logramos centrarnos sólo en un reducido número de observaciones, no suele ser necesario recurrir a criterios aleatorios. En general, la selección debe hacerse de una forma *intencionada*, que esté en consonancia con los objetivos y estrategia de nuestra investigación.

Seleccionar de manera intencionada las observaciones implica que, como mínimo, conozcamos de antemano los valores de ciertas variables relevantes y descartamos la selección aleatoria. Tendremos menos posibilidades de equivocarnos si los casos se seleccionan a partir de categorías de las variables explicativas. De este modo, la propia investigación consistirá en encontrar los valores de la variable dependiente. Sin embargo, en la práctica, no solemos tener más que datos fragmentarios sobre los valores de muchas de nuestras variables, incluso antes de seleccionar las observaciones. Esto puede ser peligroso, ya que sin darnos cuenta podemos introducir un sesgo de selección que quizá refuerce nuestra hipótesis previa. A continuación analizamos los diversos métodos de selección intencionada de observaciones.

4.1 Seleccionar observaciones en función de la variable explicativa

Tal como acabamos de señalar, los mejores diseños «intencionados» eligen sus observaciones de manera que la variable explicativa (y cualquiera de las de control) pueda cambiar, al margen de cuáles sean los valores de las dependientes. Estos sólo se descubren durante la investigación; después se extrae la primera inferencia causal, examinando de qué manera varía la distribución de resultados en la variable dependiente si se dan ciertos valores en las explicativas.

Supongamos, por ejemplo, que nos interesarán las consecuencias que tuvieron durante la Guerra Fría los tratados formales de control de armamentos en las decisiones que tomaron los Estados Unidos y la Unión Soviética para conseguir armas. En consecuencia, nuestra variable causal clave sería la existencia de un tratado formal de control de armamento que afectara al sistema defensivo de un país determinado. Podríamos elegir un conjunto de tipos de armas —unas contempladas por las limitaciones de los tratados y otras no— que cambiara en relación con nuestra variable explicativa. La dependiente, que no hemos utilizado para seleccionar, podría ser la frecuencia con la que se reemplaza el armamento. Si los dos conjuntos de observaciones están bien equiparados en las variables de control y se resuel-

ven satisfactoriamente problemas como el de la endogeneidad, este diseño nos permitirá extraer inferencias válidas sobre las consecuencias de los acuerdos de control de armamentos.

A veces sólo nos interesará una de las diversas variables explicativas, que parece influir de manera sustancial en la dependiente. Cuando se dé tal situación, habrá que controlar la variable que no es primordial (o la que no lo es en ese momento). Un ejemplo de este procedimiento nos lo proporciona Jack Snyder (1991), quien hizo una selección de los países que consideraba «principales competidores por el poder» en la época contemporánea con el fin de estudiar su grado de «expansión excesiva» (la variable dependiente). Una variable muy importante para explicar este proceso expansivo es el poderío militar, pero esta causa es tan evidente y se halla tan bien documentada que Snyder no quiso invertir más recursos en hacer una nueva estimación de sus efectos. De hecho, controló este factor eligiendo únicamente naciones en las que su valor era muy alto. Al mantener prácticamente constante esta transcendental variable de control, Snyder no pudo hacer inferencias sobre las consecuencias que tenía el poderío militar sobre el exceso de expansión, pero sí logró centrarse en las variables explicativas que le interesaban sin padecer los efectos del sesgo de la variable omitida. Aparte de estos aspectos relativos al diseño, la investigación de Snyder tuvo un carácter exploratorio, ya que ni siquiera identificó todas sus variables explicativas antes de comenzarla (Snyder, 1991, pp. 61-65). Probablemente, este carácter abierto hizo que considerara ideas que, de no haber sido así, no se hubiera planteado, pero también tuvo como consecuencia que, finalmente, las preguntas resueltas tuvieran una respuesta que no era tan evidente como hubiera podido ser. En concreto, el abanico de variaciones posibles de las variables explicativas que le interesaban probablemente no fue tan amplio como lo era en potencia. Además, no se evaluó la teoría más que en el conjunto de datos con el que se había formulado.

Como venimos señalando en este libro, las recomendaciones «puristas» —seleccionar siempre en función de las variables explicativas y nunca a partir de las dependientes— suelen ser poco realistas para la investigación cualitativa. No todo está perdido cuando hay que tener en cuenta los valores de la variable dependiente al recoger los datos o cuando aquellos de los que se dispone ya tienen en cuenta dichos valores. Aún puede obtenerse información sobre los efectos causales, pero si no tenemos mucho cuidado es probable que aparezca algún sesgo.

4.2 Seleccionar un abanico de valores de la variable dependiente

Como alternativa a la elección de observaciones en función de la variable explicativa, se podría partir para seleccionarlas de un abanico de valores de la dependiente. Las investigaciones suelen comenzar del siguiente modo:

encontramos algunas fascinantes variaciones en un comportamiento que queremos explicar. En este diseño de tipo retrospectivo (que en epidemiología se denomina estudio de «control de casos»), tomamos aquellas observaciones que adoptan valores especialmente altos o bajos en la variable dependiente. Como hemos subrayado, aunque este proceso de selección pueda ayudarnos a hacer inferencias causales, carece de utilidad para hacer las *descriptivas* que se refieren a la variable dependiente. Además, la falta de datos descriptivos sistemáticos y la mayor posibilidad de que aparezcan otros problemas debidos a posibles no linealidades o efectos causales variables hacen que las inferencias causales que genera este procedimiento no suelen ser válidas.

Un diseño de investigación retrospectivo puede ayudarnos a recabar información valiosa sobre la plausibilidad empírica de una inferencia *causal*, ya que bien podríamos darnos cuenta de que los valores altos y bajos de la variable dependiente están asociados, respectivamente, con los valores de uno u otro signo que se observan en las posibles variables explicativas. Sin embargo, para que este diseño nos conduzca a inferencias causales que tengan sentido —aunque sean necesariamente limitadas—, es esencial que seleccionemos las observaciones sin tener en cuenta los valores de las variables explicativas. No hay que buscar observaciones que se ajusten (o no se ajusten) a nuestra teoría inicial. Las observaciones han de ser tan representativas como sea posible en la población sobre la que queremos generalizar. Si percibimos que los valores altos y bajos de las variables explicativas potenciales se corresponden, respectivamente, con uno u otro tipo de valores en la variable dependiente, quizá tengamos que diseñar un estudio en el que las observaciones sólo se seleccionen a partir de la variable o variables explicativas, con el fin de calibrar si nuestra hipótesis es correcta. Lo mínimo es que, al principio, los resultados sean inciertos o, de lo contrario, no aprenderemos nada. Para que la incertidumbre afecte a las inferencias causales hay que dejar que los valores de nuestra variable explicativa o dependiente se vean determinados por la situación del estudio.

Por ejemplo, podríamos observar variaciones sorprendentes en los conflictos violentos que tienen lugar entre los estados y especular con la posibilidad de que los produzcan diferentes formas de gobierno. Podría valer la pena comenzar, de forma exploratoria, con un examen cuidadoso de algunas relaciones bilaterales en las que la guerra es frecuente y otras que se caractericen por un carácter pacífico excepcional. Supongamos que viéramos que las observaciones de guerra están asociadas a relaciones en las que participa, al menos, una autocracia en período de modernización y que las de paz afectan a relaciones entre dos democracias estables. Esta explicación exploratoria generaría una hipótesis más precisa que la inicial. No podríamos decir que hemos confirmado nuestra hipótesis, ya que seguiríamos sin tener una visión clara de las pautas generales (al haber seleccionado las observaciones a partir de la variable dependiente), pero sí podríamos vernos

4.3 Seleccionar observaciones utilizando tanto las variables explicativas como las dependientes

Resulta peligroso seleccionar observaciones utilizando intencionadamente tanto las variables explicativas como las dependientes, porque, al hacerlo, error más grande consiste en seleccionar observaciones en las que las variables explicativas y la dependiente cambian al unísono de una manera que se sabe se ajusta a la hipótesis que la investigación trata de contrastar. Por ejemplo, quizá queramos comprobar si es cierto que los sistemas autoritarios (que acaban con las organizaciones sindicales y con las reivindicaciones laborales) conducen a mayores índices de crecimiento económico. Podríamos seleccionar observaciones que cambien en ambos tipos de variables, pero seleccionándolas a propósito para que todas las relaciones autoritarias arrojen altos índices de crecimiento y para que las referidas a sistemas no autoritarios registren valores bajos en el mismo indicador. Un diseño así no describirá ni explicará nada, ya que, sin examinar un conjunto de observaciones representativo no se puede determinar si el crecimiento económico es tanto o más probable en aquellas observaciones en las que hay un régimen democrático que permite la actividad sindical.

A pesar del riesgo que conlleva hacer una selección a partir de las variables explicativas y de la dependiente, puede que haya unos pocos ejemplos de estudios con un número limitado de observaciones en los que tenga cierto sentido utilizar un método que baraje información sobre los valores de uno y otro tipo de variables, aunque esta técnica sea peligrosa y haya que tener mucho cuidado al utilizarla. Por ejemplo, supongamos que la distribución de los valores de nuestra variable dependiente estuviera seriamente sesgada, de modo que la mayoría de las observaciones tomaran uno de sus valores. Si seleccionáramos los valores a partir del cambio de la variable explicativa y dejaríamos que los de la dependiente «cayeran donde pudiesen», quizá no obtuviéramos variación en esta última. Nada en este resultado descarta el análisis de los datos. De hecho, cuando los valores de la variable dependiente son los mismos, al margen de cuales sean los de las explicativas, nos encontraremos ante un caso claro de efecto causal cero. Esta única situación sólo sería preocupante si creyéramos que el auténtico efecto causal es muy pequeño, pero no cero. En estudios con un *n* pequeño,

proporcionado de observaciones de la categoría en la que sí lo hubo, con el fin de incluir ejemplos de ambas en el conjunto de observaciones final. Este procedimiento habrá que llevarlo a cabo de forma que fuera independiente de nuestro conocimiento de las observaciones en lo relativo a la variable explicativa. Por ejemplo, podríamos tomar observaciones de forma aleatoria del grupo en el que no se había registrado conflicto y seleccionar todas las del grupo en el que sí lo había habido. Posteriormente, si en el conjunto de observaciones final se observara una estrecha relación entre las pautas de pertenencia a organizaciones y el conflicto de tipo militar, podríamos hacer inferencias causales.

El estudio que realizó Ahi Kohli (1987) sobre el papel del Estado indio en las políticas contra la pobreza pone de manifiesto los condicionantes que sufre la selección de observaciones en investigaciones con un n pequeño, las consecuencias que tienen para la extracción de inferencias causales válidas y algunas formas de superarlos. A Kohli le interesaba la influencia de las estructuras de autoridad gubernamentales y de los tipos de regímenes en la persistencia de políticas contra la pobreza en los países en desarrollo. Su argumento, dicho en pocas palabras, es que los regímenes cuya ideología les hace comprometerse claramente con la ayuda a los pobres, que poseen tabas a la participación en el poder de las clases altas y que disponen de gran capacidad organizativa crearán políticas eficientes para alcanzar su objetivo. Los regímenes que carecen de ese compromiso ideológico, que tienen apoyos interclasistas y que no disponen de una organización bien trabada no lograrán desarrollar de manera eficiente este tipo de políticas, aun-

que se comprometan abiertamente a ello.

Kohli se centró en la India, que es donde están sus intereses a la hora de investigar y para los que cuenta con conocimientos lingüísticos. Sus primeras observaciones fueron los estados indios. Como señalaba: «El carácter federal del sistema político indio permite realizar dentro del país un análisis desagregado y comparativo. Por debajo del gobierno federal, los de los estados (o provincias) de la India tienen un papel considerable en la formulación y ejecución de las políticas agrarias. Las variaciones de tipo político que se producen en el nivel estatal pueden hacer que cambie el grado de eficiencia de los programas que luchan contra la pobreza» (Kohli, 1987, pp. 3-4). El autor utilizó una versión poco estricta (pero apropiada) de la homogeneidad de las unidades, la del «efecto constante», según la cual el efecto causal era idéntico en los estados en los que se daban diferentes niveles de sus factores explicativos clave (o sea, los componentes ideológico, de clase y organizativo que, según la hipótesis, conducían a políticas de lucha contra la pobreza). Sólo pudo evaluar su hipótesis causal comparando la situación de la variable dependiente en diferentes estados, partiendo, a la vez, de «efecto constante» en cada uno de ellos.

Para Kohli, una muestra de los estados indios resultaba útil porque éstos eran relativamente similares. Por lo menos «se acercan más que la mayoría

no es probable que podamos distinguir con mucha certeza entre un efecto estimado con valor cero y otro efecto pequeño que no tenga este valor. En esta situación, la solución más directa será incrementar el número de observaciones. También se pueden seleccionar observaciones que se basen en valores muy extremos de las variables explicativas, de manera que sea más fácil reconocer cualquier pequeño efecto causal. Si estas medidas no son suficientes, la selección a partir de las variables explicativas y de la dependiente (pero sin utilizarlas simultáneamente) puede mejorar el diseño de investigación lo suficiente como para que se encuentre el efecto que buscamos (véase el apartado 3 del capítulo 6 si se desea encontrar más sugerencias).

Por lo tanto, quizá tenga sentido utilizar técnicas de muestreo para elegir observaciones partiendo primero de la variación de la variable explicativa, pero haciéndolo de tal manera que se incluyan también las observaciones que tengan el valor atípico de la variable dependiente. Sin embargo, al hacer lo así, es importante no predefinir el valor de la variable explicativa con la que se asocia la dependencia. Además, si se utiliza este procedimiento, hay que ser consciente de las posibilidades de sesgo que introduce y, por lo tanto, del valor limitado de nuestras inferencias. Dicho de otro modo, en estos casos atípicos, podemos seleccionar en función de los valores de las variables explicativas y de los de la dependiente, pero no utilizándolos a la vez.¹⁰

Por ejemplo, supongamos que planteáramos la hipótesis de que, en ciertas organizaciones internacionales, existe una determinada pauta de pertenencia que inhibe de forma significativa el estallido de conflictos violentos entre cualquier pareja de estados miembros. Siguiendo el método que predefinimos, el de seleccionar sólo en función de la variable explicativa, nuestras observaciones se compondrían de parejas de estados cuya pertenencia a organizaciones internacionales cambiaría de un periodo específico a otro. Supongamos también que fuera difícil determinar si existe la pauta de pertenencia indicada, de modo que sólo podríamos examinar un número de observaciones relativamente pequeño: sólo algunas parejas de países y no cientos o miles de ellas. La dificultad para nuestro método surgiría si los conflictos fueran escasos: por ejemplo, si sólo se hubiera registrado uno entre dos países durante el periodo de estudio. En esta situación, podríamos seleccionar parejas de países que cambiaran en relación a la variable explicativa (pertenencia a una organización), pero nos daríamos cuenta de que ninguna de ellas había sufrido conflictos violentos.

En tales condiciones, quizá fuera aconsejable algún procedimiento de selección mixto. Podríamos recoger observaciones a partir de algún cambio en la variable explicativa (certain parejas de naciones con pautas específicas de pertenencia a organizaciones y otras que no las tuvieran) y seleccionar un número de observaciones mayor del que teníamos intención de estudiar. Después podríamos dividir estas potenciales observaciones en dos categorías, en función de si se había registrado o no un conflicto armado entre las naciones en un periodo concreto, y elegir después un número des-

4.4 Selección observaciones de manera que la variable causal clave sea constante

A veces los científicos sociales diseñan investigaciones de manera que la variable explicativa que constituye la base de la selección se mantiene constante. Este enfoque es claramente deficiente, porque el efecto causal de una variable explicativa que no cambia no puede juzgarse. Por lo tanto, no es probable que sea productivo —al menos en sí mismo— un diseño que pretenda mostrar el efecto que tiene un rasgo constante del medio. Sin embargo, la mayoría de los trabajos forman parte de un cuerpo bibliográfico o tradición investigadora (véase el apartado 2.1 del capítulo 1), y, por consiguiente, es probable que de antemano se disponga de cierta información útil. Por ejemplo, el abanico usual de valores de la variable dependiente puede ser muy conocido cuando la explicativa adopta un determinado valor. Quizá el investigador que realiza un estudio para averiguar cuál es ese abanico de valores cuando la variable explicativa tiene otro valor sea el primero en hacer una estimación del efecto causal.

Pensemos en el siguiente ejemplo, en el que una investigación realizada manteniendo constante el valor de la variable explicativa condujo a una hipótesis razonable, pero provisional, sobre el efecto causal, la cual, a su vez, fue refutada en un estudio posterior en el que dicha variable tenía otro valor. En un antiguo estudio sobre el impacto de la industrialización, Inkeltz y Rossi (1956) habían comparado el prestigio que tenían ciertas ocupaciones en diversos países industrializados. Se habían dado cuenta de que este indicador era muy parecido en países que, aparte de ser industrializados, eran muy diferentes. Llegaron a la conclusión de que la industrialización era la variable causal que determinaba la específica jerarquía del prestigio que habían observado. A falta de cambios en su variable explicativa (todas las naciones estudiadas eran industrializadas), no hubieron resultado apropiado hacer una firme inferencia causal, pero sí era razonable llegar a una conclusión provisional que hacía la hipótesis más plausible. Sin embargo, otros investigadores reprodujeron este estudio en Filipinas e Indonesia (países no industrializados) —cambiando así el valor de la variable explicativa— y encontraron una jerarquía del prestigio similar, lo cual ponía en cuestión el efecto causal acauchado a la industrialización (véase Zelditch, 1971).

El ejemplo anterior muestra de qué manera una serie de proyectos de investigación puede superar los problemas que plantea hacer inferencias válidas cuando el estudio original carece de cambios en la variable explicativa. David Latin (1986) proporciona un instructivo ejemplo de cómo un único investigador puede solucionar este tipo de problema en una sucesión de estudios. En el que realizó sobre el impacto del cambio religioso en el medio político de los yoruba nigerianos, analiza por qué no había podido superar este problema en su investigación anterior sobre Somalia. Según el mismo señala, la religión, que es su variable explicativa, es una constante en toda

de los países independientes [...] al presupuesto que permite comparar si se mantienen ciertas constantes» (Kohli, 1987, p. 4). ¿Pero, qué estados había que escoger? Las investigaciones en profundidad que quería realizar (a partir de dos viajes bien planificados al país) excluían la posibilidad de acudir de todos ellos. Al tener estos condicionantes, sólo pudo elegir tres estados. Haber seleccionado estas tres unidades de manera aleatoria no hubiera sido inteligente, ya que este procedimiento sólo ofrece garantías cuando se utiliza en un estudio con un n grande. La mayoría de los estados indios presentaban las características que impiden el desarrollo de políticas contra la pobreza y, por lo tanto, no solían tenerlas. De hecho, sólo en el gobierno de Bengala Occidental se observaban rasgos que pudieran impulsarlas. Como señala Kohli, Bengala debía estar presente en la muestra. Después añadió dos estados más: Uttar Pradesh, donde había pocos programas de lucha contra la pobreza, y Karnataka, que representaba un punto intermedio entre los anteriores. Los tres estados fueron seleccionados independientemente a partir de la variable dependiente «... porque representan una gradación, desde el valor máximo al mínimo, en los esfuerzos que hacen los gobiernos para mitigar la pobreza rural» (Kohli, 1987, p. 7).

El problema de este estudio era que los valores de las variables explicativas también se conocían, de hecho, la selección se realizó a partir de ellas y las también de la variable dependiente. En estas circunstancias, el diseño era impreciso y no daba información sobre la hipótesis causal. Es decir, esta no podía evaluarse con observaciones que, desde el principio, se sabía que concordaban con ella.

Entonces, ¿tenía algún valor el estudio? No mucho si Kohli sólo estaba contrastando su hipótesis en estos tres estados. Por fortuna, hizo bastantes más cosas. Su estudio se conceptualizó en función de tres observaciones, pero, como muchos otros que a primera vista parecen tener un n pequeño, se barajaban muchas más. De hecho, era un estudio con un n grande. Kohli fue más allá del simple descubrimiento de que en el caso de los tres estados analizados las variables explicativas y la dependiente concordaban con sus hipótesis, y lo hizo ocupándose de las numerosas consecuencias observables de su hipótesis que aparecían tanto en los estados indios que estudiaba como en otros países. Como esta forma de abordar estudios en los que n es *aparentemente* pequeño constituye el tema central del siguiente capítulo, en el apartado 3.1 del capítulo 6 describiremos la estrategia que siguió Kohli para enfrentarse a ello.

Sin embargo, en un nivel de análisis agregado, Kohli podría haber hecho más para mejorar sus inferencias causales. Por ejemplo, probablemente sabía, o podría haber calibrado, los valores de sus variables explicativas y dependiente en casi todos los estados indios. Su libro habría mejorado en gran medida si hubiera incluido un capítulo corto para analizarlos brevemente. Así habría dado una idea clara de la veracidad general de su hipótesis causal y habría podido seleccionar sus tres estudios de caso según reglas más sistemáticas.

Somalia y, además, tiene una relación multicolineal (véase el apartado 1 de este mismo capítulo) con otras variables, lo cual hace imposible que se aislemos su efecto causal. «La investigación de campo en Somalia me hizo preguntarme sobre el impacto independiente que tenía el cambio religioso en la política, pero llevar a cabo más investigación de este tipo en el mismo país no me habría permitido afrontar esa pregunta de manera sistemática. ¿Cómo puede calibrarse el impacto del Islam en una sociedad en la que todos los musulmanes? Además, toda la población habla somalí y casi todos comparten una herencia nómada. Prácticamente todos los somalíes han recibido la misma tradición poética. Todos los rasgos comunes que les llevan a actuar pueden atribuirse a tradiciones somalíes poéticas, nómadas o lingüísticas, más que a la religión» (1986, p. 186). Latin supera este problema centrándose en los yoruba de Nigeria, que están divididos entre musulmanes y cristianos. En el capítulo 5 veremos cómo lo hace.

4.5 Seleccionar observaciones de manera que la variable dependiente sea constante

Tampoco podemos aprender nada sobre el efecto causal de un estudio que seleccione sus observaciones de modo que la variable dependiente no varíe, pero la bibliografía del área puede aportar la suficiente información como para permitir que se hagan inferencias causales válidas.

De ahí que haya que cambiar, si es posible, un estudio acerca de por qué nunca se llegó a un determinado resultado, para hacer que varíen la variable dependiente y las explicativas. Por ejemplo, si la pregunta de la investigación es por qué los terratenientes no usaban las cantidades óptimas de fertilizantes para mantener la productividad de las plantaciones de Carolina del Sur antes de la guerra civil estadounidense, poco podremos aprender del comportamiento de ese estado si sólo investigamos en él y si todos sus terratenientes se comportaban del mismo modo. En tal caso, no habría varianza en la variable dependiente y esta falta de cambios sería por completo imputable al investigador y, por tanto, no aportaría ninguna información adicional. Si al observar ambos estados para explicar la variación respecto al uso del producto, al menos hay una diferencia entre los estados que constituirían nuestra variable causal que puede explicar el uso de fertilizantes. Por otra parte, si todos los estudios anteriores se hubieran realizado en estados que no utilizaban este producto, ocuparse de uno en el que los terratenientes sí recurrían a él podría constituir una considerable aportación a lo escrito sobre el tema. Por lo mismo, este enfoque posibilitaría una estimación del efecto causal.

Podemos mencionar otro ejemplo: a pesar del miedo de toda una generación y de los sombríos pronósticos de muchos políticos, las armas nucleares no se han utilizado como arma de guerra desde 1945. Sin embargo, aun-

que no se haya producido un conflicto nuclear, parece interesante intentar comprender en qué condiciones podría tener lugar. Está claro que éste es un caso extremo de selección a partir de la variable dependiente en el que ésta se mantiene constante. Sin embargo, como muchos autores del área señalan fervientemente, puede que las armas nucleares no se hayan utilizado porque el valor de la variable explicativa (un mundo en el que habla, al menos, dos grandes potencias nucleares) se ha mantenido constante durante todo el periodo. A menos que reformulemos el problema, no servirá de nada hacer una estimación de la inferencia causal por medio de «variables» explicativas o dependientes constantes. En el apartado 6.3 del capítulo 6 mostraremos cómo solucionar este problema en el presente ejemplo.

Los investigadores sociales a veces utilizan un enfoque retrospectivo similar al de los Centros de control de enfermedades (CCE), que basan su selección en valores extremos, pero constantes, de la variable dependiente. Un CCE identifica un «conglomerado cancerígeno» (un grupo de personas que padece el mismo tipo de cáncer en la misma área geográfica), después busca en el medio algún factor químico o de otro tipo (la variable explicativa principal) que pueda haber producido todos los cánceres (la variable dependiente). Estos estudios, en los que las observaciones se seleccionan a partir de valores extremos de la variable dependiente, son razonablemente válidos porque existen bastantes datos sobre los niveles normales de las variables explicativas. Aunque casi todos los estudios de las CCE sean negativos o poco concluyentes, a veces sí encuentran algún componente químico sospechoso. Cuando no hay pruebas anteriores de que ese elemento produzca cáncer, el CCE suele encargar un estudio en el que las observaciones se seleccionan, si es posible, a partir de la variable explicativa (variación en la presencia o ausencia del componente), con el fin de que la inferencia causal sea más fiable.

Hay veces en que los investigadores sociales utilizan este enfoque cuando detectan la existencia de un determinado «conglomerado político» (una comunidad o región en la que hay una larga tradición de radicalismo o violencia de signo político, o cualquier otro rasgo) e intentan encontrar lo que tiene de «especial» esa área. Al igual que en el estudio de los CCE, si en la investigación aparecen correlaciones sugeridas, no hay que considerar que confirmen nuestra hipótesis; únicamente que indiquen que merece la pena diseñar un estudio que seleccione a partir de la supuesta variable explicativa, dejando a la vez que varíe la dependiente (el radicalismo o la violencia de signo político).

5. Comentarios finales

En este capítulo hemos analizado cómo se pueden seleccionar observaciones para conseguir un diseño que minimice el sesgo relativo al proceso de selección. Como no se pueden conseguir diseños perfectos, hemos criticado

los procesos de selección sugiriendo, a la vez, estrategias imperfectas pero útiles que pueden darnos cierto control sobre el problema que plantea el estudio. Al fin y al cabo, nuestro objetivo es lograr un diseño de investigación que seleccione en función de las variables explicativas presentes en nuestra teoría y que permita el cambio de la variable dependiente. Sin embargo, al encaminarnos hacia esa meta, puede que sea útil utilizar diseños que tengan en cuenta los valores observados de la variable dependiente; pero al investigar que lo haga hay que aconsejarle suma cautela. Nuestro objetivo primordial es obtener más información relevante para contrastar nuestra teoría, sin por ello introducir un sesgo que ponga en peligro la calidad de las inferencias.

El diseño de la investigación científica

5. Entender qué debe evitarse

En el capítulo 4 analizamos cómo había que elaborar un estudio mediante un diseño de investigación preciso en el que los procedimientos de selección de observaciones hicieran posibles inferencias causales válidas. Sin embargo, realizar bien esta tarea es necesario pero no suficiente para extraer tales inferencias, ya que hay posteriores errores de análisis que pueden destruir el trabajo que antes se ha hecho bien. En este capítulo vamos a estudiar de qué manera, una vez seleccionadas las observaciones para el análisis, es posible comprender las causas de la ineficiencia y del sesgo y reducir a proporciones manejables. Posteriormente, nos ocuparemos de cómo podemos controlar la investigación con el fin de enfrentarnos adecuadamente a estos problemas.

Para analizar la falta de eficiencia y el sesgo vamos a recordar los criterios de evaluación de inferencias que presentamos en el apartado 7 del capítulo 2 y en el 4 del capítulo 3. Si tenemos un diseño de investigación preciso, tenemos que centrarnos en los dos problemas clave que se van a estudiar en este capítulo: *sesgo e ineficiencia*. Para entender estos conceptos resulta útil imaginarse las inferencias como estimaciones de puntos concretos flanqueados por los extremos de un intervalo. Por ejemplo, podría mos suponer que una persona tiene cuarenta años, dos arriba, dos abajo. Esta cifra es nuestro mejor cálculo (la estimación), que el intervalo que hay entre treinta y ocho y cuarenta y dos sitúa en su centro, con una estimación de nuestra incertidumbre (la amplitud del intervalo). Queremos elegir el intervalo de manera que la verdadera edad esté dentro de él durante mucho