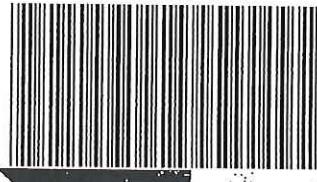


136 copias  
(Doble faz)

CS 013

Alianza  
Editorial  
manuales  
antropología  
arte  
biología  
ciencia política  
crítica literaria  
economía  
educación  
filosofía  
física  
geografía  
historia  
lingüística  
matemáticas  
física  
ecología  
física  
biología

051201089



El libro universitario  
Alianza Editorial

Gary King, Robert O. Keohane  
y Sidney Verba son profesores de Ciencia  
de la Universidad de Harvard.

Dra. Cte. Gustavo de Armas  
"Metodología"  
La Papelería  
PAPELERIA EN GENERAL  
FOTOCOPIAS  
GUAYABO 1607  
TEL. 4026228

En un momento en el que el desacuerdo acerca de la idoneidad de los métodos cuantitativos y cualitativos amenaza con socavar la coherencia de las ciencias sociales, Gary King, Robert O. Keohane y Sidney Verba desarrollan un enfoque único tanto para la inferencia descriptiva como para la causal. Sostienen que es posible elaborar descripciones válidas de los fenómenos sociales (dentro de la tradición cualitativa) y buenas explicaciones causales de los mismos (dentro de la tradición cuantitativa). Por otro lado, abordan las cuestiones que plantea *El diseño de la investigación social*: ¿Qué preguntas se deben hacer? ¿Cómo se pueden evitar los sesgos? ¿Cuántos casos son necesarios y cómo se deben seleccionar? ¿Cuáles son las formas adecuadas de evaluar la incertidumbre de las conclusiones? ¿Cómo se puede identificar el error?



Gary King, Robert O. Keohane y Sidney Verba  
*El diseño de la investigación social*

89 M 8901089  
136 copias CB CP SOC TS  
② 3 4

**Gary King  
Robert O. Keohane  
Sidney Verba**

# El diseño de la investigación social

## La inferencia científica en los estudios cualitativos

Ciencias Sociales  
Alianza Editorial

## **El diseño de la investigación social**

**La inferencia científica  
en los estudios cualitativos**



LIBROS · MUSICA · OBJETOS  
*Soriani y Michelin / 902 6889*

**Manuales / Ciencias Sociales**

**Gary King, Robert O. Keohane  
y Sidney Verba**

# **El diseño de la investigación social**

**La inferencia científica  
en los estudios cualitativos**

**Versión de Jesús Cuéllar Menezo**

**El libro universitario**

**Alianza Editorial**

Reservados todos los derechos. El contenido de esta obra está protegido por la Ley, que establece penas de prisión y/o multas, además de las correspondientes indemnizaciones por daños y perjuicios, para quienes reprodujeren, plagiaren, distribuyeren o comunicaren públicamente, en todo o en parte, una obra literaria, artística o científica, o su transformación, interpretación o ejecución artística fijada en cualquier tipo de soporte o comunicada a través de cualquier medio, sin la preceptiva autorización.

© de la traducción: Jesús Cuéllar Menezo, 2000  
© 1994 by Princeton University Press  
© Ed. cast.: Alianza Editorial, S.A., Madrid, 2000  
Calle Juan Ignacio Luca de Tena, 15; 28027 Madrid; teléf. 91 393 88 88  
ISBN: 84-206-8697-2  
Depósito legal: M. 1.102-2000  
Fotocomposición: EPCA, S. A.  
Polígono «Las Monjas»; 28850 Torrejón de Ardoz (Madrid)  
Printed in Spain

## Índice

Prefacio .....	9
1. La ciencia en las ciencias sociales .....	13
1. Introducción .....	13
1.1 Dos formas de investigar, una sola lógica inferencial .....	13
1.2 Definición de investigación científica en las ciencias sociales .....	17
1.3 Ciencia y complejidad .....	20
2. Principales componentes del diseño de investigación .....	23
2.1 Mejorar las preguntas de la investigación .....	24
2.2 Mejorar la teoría .....	29
2.3 Mejorar la calidad de los datos .....	33
2.4 Utilizar mejor los datos existentes .....	37
3. Temas de estudio de este volumen .....	39
3.1 La utilización de consecuencias observables para vincular la teoría y los datos .....	39
3.2 La maximización del control .....	39
3.3 Admitir la incertidumbre .....	42
3.4 Pensar como un científico social: escepticismo e hipótesis contrapuestas .....	42
2. La inferencia descriptiva .....	45
1. El conocimiento general y los hechos particulares .....	46
1.1 «Interpretación» e inferencia .....	48
1.2 «Singularidad», complejidad y simplificación .....	53
1.3 Estudios de caso comparados .....	55

2. La inferencia: el fin científico de la recogida de datos .....	57
3. Modelos formales de investigación cualitativa .....	60
4. Un modelo formal sobre la recogida de datos.....	62
5. El resumen de los pormenores históricos.....	64
6. La inferencia descriptiva .....	66
7. Criterios para juzgar las inferencias descriptivas .....	74
7.1 Inferencias no sesgadas.....	74
7.2 Eficiencia .....	76
3. Causalidad e inferencia causal .....	87
1. Definición de causalidad.....	88
1.1 Definición y un ejemplo cuantitativo.....	88
1.2 Un ejemplo cualitativo .....	88
2. Clarificación de las definiciones alternativas de causalidad.....	94
2.1 «Mecanismos causales» .....	96
2.2 «Causalidad múltiple» .....	96
2.3 Causalidad «simétrica» y «asimétrica» .....	98
3. Supuestos necesarios para la estimación de efectos causales .....	100
3.1 Homogeneidad de las unidades .....	102
3.2 Independencia condicional .....	102
4. Criterios para evaluar las inferencias causales .....	105
5. Reglas para elaborar teorías causales .....	106
5.1 Regla 1: elaborar teorías falsables .....	107
5.2 Regla 2: construir teorías que tengan coherencia interna .....	108
5.3 Regla 3: seleccionar cuidadosamente las variables dependientes .....	116
5.4 Regla 4: maximizar lo concreto .....	118
5.5 Regla 5: formular teorías de forma tan incluyente como sea posible .....	120
	123
4. Precisar qué va a observarse .....	125
1. Diseños de investigación imprecisos .....	128
1.1 Más inferencias que observaciones .....	129
1.2 Multicolinealidad .....	131
2. Límites de la selección aleatoria .....	134
3. Sesgo de selección .....	138
3.1 Seleccionar en función de la variable dependiente .....	139
3.1.1 Ejemplos de sesgo de selección producido por el investigador .....	142
3.1.2 Ejemplos de sesgo de selección producido por el mundo .....	145
3.2 Seleccionar en función de una variable explicativa .....	147
3.3 Otros tipos de sesgo de selección .....	149
4. Selección intencionada de observaciones .....	150
4.1 Seleccionar observaciones en función de la variable explicativa .....	150
4.2 Seleccionar un abanico de valores de la variable dependiente .....	151
4.3 Seleccionar observaciones utilizando tanto las variables explicativas como las dependientes .....	153
4.4 Seleccionar observaciones de manera que la variable causal clave sea constante .....	157

4.5 Seleccionar observaciones de manera que la variable dependiente sea constante .....	158
5. Comentarios finales .....	159
5. Entender qué debe evitarse .....	161
1. Error de medida .....	162
1.1 Error de medida sistemático .....	167
1.2 Error de medida no sistemático .....	169
1.2.1 Error de medida no sistemático en la variable dependiente .....	170
1.2.2 Error de medida no sistemático en una variable explicativa .....	173
2. Exclusión de variables relevantes: el sesgo .....	178
2.1 Cómo calibrar el sesgo de las variables omitidas .....	178
2.2 Ejemplos de sesgo de la variable omitida .....	186
3. Inclusión de variables irrelevantes: la ineficiencia .....	193
4. Endogeneidad .....	197
4.1 Corregir inferencias sesgadas .....	199
4.2 Precisar cuáles son los componentes de la variable dependiente .....	200
4.3 Convertir la endogeneidad en un problema de variable omitida .....	201
4.4 Seleccionar observaciones para evitar la endogeneidad .....	202
4.5 Precisar cuáles son los componentes de la variable explicativa .....	205
5. Asignación de valores a la variable explicativa .....	207
6. Controlar la situación que se investiga .....	211
7. Observaciones finales .....	218
6. Aumentar el número de observaciones .....	221
1. Inferencias causales con diseños de una sola observación .....	222
1.1 Estudios de caso «cruciales» .....	222
1.2 Razonamiento analógico .....	225
2. ¿Cuántas observaciones son suficientes? .....	226
3. Aumentar el número de observaciones a partir de unas pocas .....	230
3.1 Las mismas medidas con otras unidades .....	232
3.2 Nuevas medidas con las mismas unidades .....	236
3.3 Nuevas medidas con nuevas unidades .....	237
4. Observaciones finales .....	241
Notas .....	243
Bibliografía .....	255
Índice analítico .....	265

## Prefacio

En este libro se propone un mismo enfoque para la extracción de inferencias válidas, tanto causales como descriptivas, en la investigación cualitativa, donde las medidas numéricas resultan imposibles o no recomendables. Lo que queremos señalar es qué la lógica de un buen proyecto de investigación no difiere fundamentalmente por que éste sea cuantitativo o cualitativo. Nuestro enfoque se aplica por igual a estas dos formas de estudio aparentemente distintas.

Con este libro pretendemos animar a los investigadores cualitativos a que se tomen en serio la inferencia científica y la incorporen a su trabajo. Confiamos en que nuestra lógica inferencial única y el intento de demostrar que puede ser útil para los cualitativistas ayuden a mejorar los trabajos en nuestra disciplina y, quizás también, los de otros estudios sociales. Por lo tanto, esperamos que este trabajo lo lean y examinen de forma crítica politólogos y científicos sociales de todas las tendencias y niveles profesionales: desde los investigadores cualitativos hasta los estadísticos, desde los estudiantes de los últimos cursos de licenciatura y los que comienzan el doctorado hasta académicos ya asentados. Utilizamos algunas fórmulas matemáticas porque resultan especialmente útiles para clarificar ciertos conceptos de los métodos cualitativos; sin embargo, no presuponemos en el lector conocimientos matemáticos o estadísticos, y es posible saltarse la mayoría de las fórmulas sin perder el hilo argumental.

Las autoridades universitarias suelen decir que la docencia y la investigación son complementarias. De hecho, ambas actividades son prácticamente inseparables.

mente coincidentes porque las dos entrañan la adquisición de nuevos conocimientos y su transmisión a otras personas, aunque de una manera ligeramente diferente en uno y otro caso. Este libro atestigua el carácter sincrónico de ambas actividades. Desde 1989, hemos estado trabajando en este libro y dirigiendo juntos en el departamento de gobierno de la Universidad de Harvard un seminario de tercer ciclo titulado «Métodos cualitativos en las ciencias sociales». El seminario ha sido muy animado y con frecuencia ha continuado en los pasillos y en los largos mensajes que hemos intercambiado entre nosotros y con nuestros estudiantes. Las batallas intelectuales han sido siempre amistosas, pero las reglas del combate dictaban que «aceptar el desacuerdo» y hacer concesiones eran graves delitos. Si uno de nosotros no estaba verdaderamente convencido de algo, creíamos que nuestra obligación era continuar el debate. Al final, los autores aprendimos bastante unos de otros y de los estudiantes acerca de los métodos de investigación cualitativos y cuantitativos, y cambiamos muchos de nuestros puntos de vista iniciales. Al margen de sus objetivos principales, este libro constituye una declaración del enfoque unánime de la inferencia científica en la investigación cualitativa al que tanto nos costó llegar.

La primera versión del libro se terminó en 1991, y en los años posteriores la hemos revisado considerablemente. Gary King fue el primero en sugerir que lo escribiéramos, esbozó las primeras versiones de la mayoría de los capítulos y llevó la voz cantante en el largo proceso de revisión. Sin embargo, la redacción del libro ha sido alterada en tan gran medida por Robert Keohane y Sidney Verba, así como por el propio Gary King, que nos resultaría imposible determinar de manera fiable la autoría de muchos fragmentos y apartados.

Durante este largo proceso pasamos borradores a compañeros de todos los Estados Unidos, a los que estamos muy agradecidos por la extraordinaria generosidad de sus comentarios. Damos también las gracias a los estudiantes de doctorado, tanto de Harvard como de otras universidades, que han ido conociendo el texto y cuya reacción ha sido importante a la hora de revisarlo. Es muy aventurado pretender mencionar a todos aquellos que han colaborado en un proyecto como éste (calculamos que existe una posibilidad de 0,92 de que sin darnos cuenta omitámos a alguien que nos haya hecho comentarios importantes). Sin embargo, quisieramos mencionar a las siguientes personas: Christopher H. Achen, John Aldrich, Hayward Alker, Robert H. Bates, James Battista, Nathaniel Beck, Nancy Burns, Michael Cobb, David Collier, Gary Cox, Michael C. Desch, David Dessler, Jorge Domínguez, George Downs, Mitchell Duneier, Matthew Evangelista, John Ferejohn, Andrew Gelman, Alexander George, Joshua Goldstein, Andrew Green, David Green, Robin Hanna, Michael Hiscox, James E. Jones, padre, Miles Kahler, Elizabeth King, Alexander Kozheviakin, Stephen D. Krasner, Herbert Kritzer, James Kuklinski, Nathan Lane, Peter Lange, Tony Lavelle, Judy Layzer, Jack S. Levy, Daniel Little, Sean Lynn-Jones, Lisa L.

Martin, Helen Milner, Gerardo L. Munck, Timothy P. Nokken, Joseph S. Nye, Charles Ragin, Swarna Rajagopalan, Shamara Shantu Riley, David Rocke, David Rohde, Frances Rosenbluth, David Schwieder, Collins G. Shackelford, padre, Kenneth Shepsle, Daniel Walsh, Carolyn Warner, Steve Aviv Yetiv, Mary Zerbino y Michael Zürn. Vaya nuestro agradecimiento a Steve Voss por confeccionar el índice analítico y al personal de Princeton University Press: Walter Lippincott, Malcolm DeBevoise, Peter Dougherty y Alessandra Bocco. También damos las gracias a la National Science Foundation de los Estados Unidos por la ayuda a la investigación SBR-9223637 que concedió a Gary King. Robert O. Keohane agradece a la John Simon Guggenheim Memorial Foundation el período de beca que le permitió concluir el trabajo en este libro.

Los tres autores (ordenados y combinados de diversas maneras) tuvimos la inmensa fortuna de poder presentar versiones anteriores de este libro en seminarios y grupos de trabajo de las reuniones de la Midwest Political Science Association (Chicago, 2-6 de abril de 1990); del Grupo sobre Metodología Política (Universidad de Duke, 18-20 de julio de 1990); de la American Political Science Association (Washington, D.C., del 29 de agosto al 1 de septiembre de 1991); del Seminario sobre metodología y filosofía de las ciencias sociales (Universidad de Harvard, Center for International Affairs, 25 de septiembre de 1992); de la serie de coloquios del Interdisciplinary Consortium for Statistical Applications (Universidad de Indiana, 4 de diciembre de 1991); del ciclo de seminarios del Institute for Global Cooperation and Change (Universidad de California, Berkeley, 15 de enero de 1993); así como en la Universidad de Illinois, Urbana-Champaign (18 de marzo de 1993).

GARY KING  
ROBERT O. KEOHANE  
SIDNEY VERBA

Cambridge, Massachusetts

# 1. La ciencia en las ciencias sociales

## 1. Introducción

Este libro trata de la investigación en las ciencias sociales y tiene un objetivo práctico: diseñar investigaciones que generen inferencias válidas sobre la vida social y política. Nos centramos en la ciencia política, pero nuestro argumento es aplicable a disciplinas como la sociología, la antropología, la historia, la economía y la psicología, así como a áreas que no se consideran una disciplina, como las técnicas periciales, la investigación en temas de educación y el razonamiento clínico.

Nuestra obra no se ocupa de la filosofía de las ciencias sociales, ni tampoco es un manual para tareas específicas de la investigación como el diseño de encuestas, la realización del trabajo de campo o el análisis de datos estadísticos. En realidad, trata del diseño de la investigación: de cómo plantear preguntas y moldear los estudios académicos para extraer inferencias descriptivas y causales válidas. Por lo tanto, se sitúa en un zona intermedia entre las abstractas polémicas filosóficas y las técnicas prácticas para centrarse en la lógica intrínseca que subyace en toda investigación social científica.

### 1.1 Dos formas de investigar, una sola lógica inferencial

Nuestro principal objetivo es relacionar las tradiciones de lo que se suele denominar investigación «cuantitativa» y «qualitativa» mediante una misma lógica inferencial. Ambas tendencias parecen bastante diferentes y, de he-

cho, a veces se consideran enfrentadas. Para nosotros, tales diferencias sólo son una cuestión de estilo y de técnicas específicas. Los dos tipos de investigación participan de una misma lógica subyacente, que suele explicarse y formalizarse claramente al analizar métodos cuantitativos. Sin embargo, esta lógica inferencial es igual a la que respalda las buenas investigaciones cualitativas, y tanto los cualitativistas como los cuantitativistas se beneficiarían si se le concediera una atención más explícita al diseñar investigaciones.

Los estilos de la investigación cuantitativa y cualitativa son muy diferentes. La primera se sirve de números y métodos estadísticos. Sigue basarse en medidas numéricas de ciertos aspectos de los fenómenos; parte de casos concretos para llegar a una descripción general o para comprobar hipótesis causales y busca medidas y análisis que otros investigadores puedan reproducir fácilmente.

Por el contrario, la investigación cualitativa abarca una amplia gama de enfoques, pero, por definición, ninguno de ellos se basa en medidas numéricas. Este tipo de trabajo se centra generalmente en un caso o en un reducido número de ellos; se sirve de entrevistas en profundidad o de análisis detallados de materiales históricos; utiliza un método discursivo e intenta estudiar de forma global o exhaustiva un acontecimiento o unidad. Los cualitativistas, aunque sólo tengan un pequeño número de casos, suelen sacar a la luz en sus estudios una gran cantidad de información. A veces, en las ciencias sociales este tipo de trabajo tiene que ver con estudios de área o de caso en los que el centro de atención es un hecho determinado o una decisión, institución, lugar, problema o ley. Al igual que en la investigación cuantitativa, el objeto de estudio puede ser importante en sí mismo: es un cambio crucial en una nación, unas elecciones, una decisión trascendental o una crisis mundial. ¿Por qué se derrumbó tan súbitamente el régimen germano oriental en 1989? De forma más general, ¿por qué casi todos los regímenes comunistas de Europa del Este se vinieron abajo en 1989? A veces, aunque desde luego no siempre, se puede elegir un acontecimiento como ejemplo de un cierto tipo de hechos, como sería el caso de una revolución política o de una determinada comunidad que decide oponerse a la instalación de un depósito de residuos. Este tipo de trabajo suele estar relacionado con estudios de área en los que se aborda la historia y la cultura de una determinada parte del mundo. El lugar o acontecimiento en cuestión se analiza de cerca y con todo detalle.

Durante varias décadas, los polítólogos han venido contrastando las ventajas de los estudios de caso con las de los estadísticos, las de los estudios de área con las de los trabajos comparativos, así como las de las investigaciones políticas «científicas», que usan métodos cuantitativos, con las de las «históricas», que se basan en una profunda comprensión de las fuentes escritas o los contextos. Algunos cuantitativistas creen que en las ciencias sociales el único camino que lleva a la verdad es el análisis estadístico sistemático. Los partidarios de la investigación cualitativa se oponen violenta-

mente a esta idea. Esta diferencia de opiniones genera una viva polémica, pero, por desgracia, también divide la disciplina en dos ramas, una cuantitativa-sistémica-generalizadora y otra cualitativa-humanista-discursiva. A medida que la primera aumenta la complejidad de sus análisis estadísticos (y su trabajo se hace menos comprensible para los que no han estudiado las técnicas), la segunda se convence cada vez más de que tales análisis resultan irrelevantes para acontecimientos que, como aquellos que interesan a sus partidarios, no parecen poder reproducirse o generalizarse.

Uno de los principales propósitos de este libro es demostrar que las diferencias entre la tradición cuantitativa y la cualitativa son sólo de tipo estilístico y que su importancia en términos metodológicos y de contenido es mínima. Puede considerarse —con razón— que toda buena investigación procede de la misma lógica inferencial subyacente. Tanto los estudios cuantitativos como los cualitativos pueden ser sistemáticos y científicos. La investigación histórica puede ser analítica si lo que pretende es evaluar diversas explicaciones mediante un proceso inferencial causal válido. La historia, o la sociología histórica, no es incompatible con las ciencias sociales (Skocpol, 1984, pp. 374-86).

Para romper estas barreras es preciso que comencemos por cuestionar el concepto mismo de investigación «cuantitativa». Hemos utilizado el término en el título de este libro para hacer referencia a nuestro objeto de estudio, pero sin querer decir con ello que la investigación «cuantitativa» se difiere fundamentalmente de la «cuantitativa», excepto en el estilo.

La mayoría de las investigaciones no corresponde claramente a una u otra categoría, y las mejores suelen combinar características de los dos tipos. En el mismo proyecto de investigación, algunos de los datos recogidos pueden ser objeto de análisis estadístico, pero no así otros elementos igualmente relevantes. Las pautas y tendencias del comportamiento social, político o económico son más susceptibles de análisis cuantitativo que el flujo de ideas entre personas o la influencia determinante que tiene un líder extraordinario. Para entender un mundo social en cambio constante tenemos que barajar información que no es fácilmente cuantificable y también la que sí lo es. Además, todas las ciencias sociales necesitan comparar, lo cual supone que se evalúe qué fenómenos se parecen «más» o «menos» en cuestión de grado (o sea, diferencias cuantitativas) o de tipo (diferencias cualitativas).

Dos magníficos estudios recientes son ejemplo de este punto de vista. En *Coercive Cooperation* (1992), Lisa L. Martin quiso explicar el grado de cooperación internacional en materia de sanciones económicas, analizando cuantitativamente 99 casos en los que se intentó aplicar tales medidas después de la Segunda Guerra Mundial. Aunque este análisis cuantitativo ofrecía mucha información valiosa, algunas de las inferencias causales que indicaban los datos eran ambiguas; de ahí que Martin decidiera estudiar en detalle seis casos de sanciones con la intención de recabar más pruebas de relevancia para su inferencia causal. Para escribir *Making Democracy Work*

### El diseño de la investigación social

(1993), Robert D. Putnam y sus colegas entrevistaron a 112 parlamentarios regionales italianos en 1970, a 194 en 1976 y a 234 entre 1981 y 1982, así como a 115 líderes comunitarios en 1976 y a 118 entre 1981 y 1982. También enviaron por correo un cuestionario a más de 500 líderes comunitarios de todo el país en 1983. Además, ex profesor para este trabajo, se llevaron a cabo cuatro encuestas de alcance nacional. No obstante, entre 1976 y 1989 los autores del libro realizaron estudios de caso pormenorizados sobre la vida política de seis regiones. Los investigadores, con la intención de someterse a la «prueba del trauma interocular», «llegaron a conocer con detalle las maniobras y personalidades políticas que habían animado la política regional en las dos últimas décadas» (Putnam, 1993, p. 190).

Las lecciones de estos análisis deberían ser claras: cualquiera que sea el tema de estudio, ni la investigación cuantitativa será mejor que la cualitativa ni al contrario. Dado que muchos asuntos de interés para las ciencias sociales, si han de tener sentido, no pueden formularse de manera que sea posible la comprobación estadística de hipótesis mediante datos cuantitativos, no queremos animar a nadie a que utilice únicamente técnicas cuantitativas. No pretendemos sacar de la biblioteca a todos los científicos sociales para llevarlos a centros informáticos ni sustituir conversaciones irrepetibles por entrevistas estructuradas. En realidad, lo que señalamos es que los estudios no estadísticos tendrían resultados más fiables si los investigadores prestaran más atención a las reglas de la inferencia científica, que a veces se formulan mejor con el estilo de la investigación cuantitativa. Los precisos métodos estadísticos que sustentan esta tendencia conforman modelos formales abstractos que son aplicables a todo tipo de investigación, incluso a aquellas en las que las variables no pueden medirse cuantitativamente. El carácter extremadamente abstracto, e incluso poco realista, de los modelos estadísticos es lo que hace que las reglas inferenciales resalten con tanta claridad.

Las reglas de la inferencia que analizamos no son relevantes para todos los problemas que los científicos sociales consideran importantes. Muchos asuntos cruciales de la vida política —relacionados con conceptos como el de acción (*agency*), obligación, legitimidad, ciudadanía o soberanía y con la correcta relación entre las sociedades nacionales y la política internacional— tienen un carácter filosófico más que empírico. Sin embargo, estas reglas sí que tienen importancia para toda investigación en la que el objetivo sea conocer lo que ocurre en el mundo real.<sup>7</sup> De hecho, lo que distingue a las ciencias sociales de la observación ocasional es que la primera pretende hacer inferencias válidas mediante el uso sistemático de procedimientos de investigación contrastados.<sup>8</sup> El hecho de que aquí nos centremos en la indagación empírica supone dejar de lado muchos problemas de la filosofía de las ciencias sociales, así como las polémicas que se ocupan del papel del postmodernismo, del carácter y existencia de la verdad, del relativismo y de asuntos similares. Partimos de la base de que es posible lograr cierto conocimiento del mundo exterior, pero tal conocimiento siempre será incierto.

### 1. La ciencia en las ciencias sociales

Además, nada en nuestro conjunto de reglas implica que tengamos que llevar a cabo el experimento perfecto (si es que existe tal cosa) o reunir todos los datos relevantes antes de poder extraer inferencias científicas válidas. Merece la pena estudiar un asunto importante aunque se disponga de poca información. Si aplicamos cualquier diseño de investigación a esta situación las conclusiones serán relativamente inciertas, pero si se admite honestamente tal incertidumbre esta clase de estudio será muy útil. Con frecuencia, la investigación social se caracteriza por disponer de una información limitada. Como el mundo social cambia rápidamente, los análisis que nos ayudan a comprender las transformaciones han de describirlos e intentar interpretarlos en su contexto, aunque la incertidumbre de nuestras conclusiones sea alta. La necesidad de resolver un problema puede ser tan grande que quizás haya datos —recogidos con los métodos científicos más útiles— que se queden obsoletos antes de organizarlos. Si una persona enloquecida corre hacia nosotros blandiendo un hacha, puede que entregarle un cuestionario de cinco páginas sobre psicopatía no sea la mejor estrategia. Joseph Schumpeter citó una vez a Albert Einstein, que había dicho que «siempre que nuestras proposiciones sean ciertas no dirán nada acerca de la realidad, y siempre que no digan nada acerca de la realidad, no serán ciertas» (Schumpeter [1936], 1991, pp. 298-299). Sin embargo, aunque la certeza no pueda alcanzarse, la seguridad en nuestras conclusiones, así como su fiabilidad, validez y sinceridad, podrán incrementarse si prestamos atención a las reglas de la inferencia científica. Las ciencias sociales que propugnamos pretenden extraer del mundo inferencias descriptivas y causales. Los que no comparten la premisa de que es posible un conocimiento parcial e imperfecto ni aspiran a una comprensión descriptiva y causal tendrán que buscar en otra parte inspiración o batallas sobre paradigmas en las que participar.

En resumen, aquí no se dan recetas para la investigación científica empírica. Ofrecemos ciertos preceptos y reglas cuya intención es disciplinar el pensamiento, no sofocarlo. Tanto en la investigación cuantitativa como en la cualitativa aplicamos de manera imperfecta reglas inferenciales teóricas a diseños de investigación y a datos empíricos inherentemente imperfectos. Toda reglamentación que tenga sentido admitirá excepciones, pero hay que pedir que éstas se justifiquen de manera explícita, que se evalúen sus consecuencias para la fiabilidad de la investigación y que se plantea la incertidumbre de las conclusiones. No buscamos un dogma, sino un pensamiento disciplinado.

#### 1.2 Definición de investigación científica en las ciencias sociales

Nuestra definición de «investigación científica» es un ideal al que toda investigación cuantitativa y cualitativa, incluso la más cuidadosa, sólo puede aproximarse. Sin embargo, necesitamos definir lo que es una buena investi-

## El diseño de la investigación social

gación, y para ello utilizamos la palabra «científica» como calificativo<sup>1</sup>. Esta palabra tiene muchas connotaciones injustificables, inapropiadas o, para muchos investigadores cualitativos, completamente incendiarias. De ahí que nosotros le demos una definición explícita. Tiene que quedar claro que no consideramos más científica la investigación cuantitativa que la cualitativa. El estilo de una buena investigación —o sea, la que es científica— puede ser cuantitativo o cualitativo. Sin embargo, en lo tocante al diseño, la investigación tiene estas cuatro características:

### 1. El objetivo es la inferencia

El diseño de la investigación científica tiene como objetivo la extracción de inferencias descriptivas o explicativas a partir de la información empírica que se tenga del mundo. Para hacer estudios científicos suele ser indispensable describir con cuidado ciertos fenómenos, pero la acumulación de hechos no es, en sí misma, suficiente. Éstos pueden ser reunidos (por investigadores cualitativos o cuantitativos) de forma más o menos sistemática, y, evidentemente, la primera es mejor que la segunda; sin embargo, nuestra definición de ciencia conlleva un paso adicional, que es el de utilizar los datos inmediatos para hacer inferencias que conduzcan a algo más amplio que no se observa directamente. Ese algo puede entrañar una inferencia descriptiva —utilizar observaciones del mundo para revelar otros hechos que no se han observado— o una inferencia causal —conocer efectos causales a partir de los datos observados. El alcance de la inferencia puede delimitarse en el espacio y el tiempo—el comportamiento electoral en las elecciones estadounidenses desde 1960, los movimientos sociales en Europa del Este desde 1989— o puede ser más general—el comportamiento humano desde la aparición de la agricultura. En cualquiera de los casos, lo que distingue la investigación científica es el objetivo de hacer inferencias que vayan más allá de las observaciones específicas que se han recogido.

### 2. Los procedimientos son públicos

Para generar y analizar datos la investigación científica utiliza métodos explicativos, codificados y públicos que, por lo tanto, pueden evaluarse. Gran parte de la investigación social cualitativa sigue procedimientos de búsqueda o inferenciales menos precisos. Como afirmó Robert K. Merton ([1949], 1968, pp. 71-72): «El análisis sociológico de datos cualitativos suele depender de un mundo privado de ideas penetrantes pero insonables y de interpretaciones ineptas [...] [Sin embargo.] la ciencia [...] es pública, no privada». La afirmación de Merton no es aplicable a todos los cualitativistas (y, por desgracia, si a algunos cuantitativistas), pero muchos se comportan como si carecieran de método y, a veces, como si la utilización de medios explícitos menoscabara su creatividad. No obstante, no pueden dejar de utilizar alguno. De alguna forma observan los fenómenos, plantean preguntas, extraen información sobre el mundo a partir de esas ob-

### 1. La ciencia en las ciencias sociales

*T. Klum*

servaciones y hacen inferencias sobre causas y consecuencias. Si el método y la lógica de las observaciones e inferencias de un investigador no se hacen explícitos, la comunidad académica no tiene forma de juzgar la validez de lo que se ha hecho. No podrá evaluar los principios de selección utilizados para registrar las observaciones, ni cómo se han procesado éstas, ni la lógica de las conclusiones. No podremos aprender de sus métodos ni reproducir sus resultados. Esta investigación no será un hecho *público*. Aunque sea una buena lectura, no será una aportación a las ciencias sociales.

Todos los métodos —sean o no explícitos— tienen sus limitaciones. La ventaja de lo explícito es que pueden captarse tales limitaciones y, si es posible, remediarlas. Además, los métodos se pueden enseñar y compartir. Este proceso permite que los resultados de la investigación se comparan con los de otros investigadores, que se reproduzcan los estudios de proyectos ajenos y que los académicos aprendan.

### 3. Las conclusiones son inciertas

La inferencia es, por definición, un proceso imperfecto. Su objetivo es utilizar datos cuantitativos y cualitativos para conocer el mundo que los ha producido. Evidentemente, es imposible llegar a conclusiones perfectamente ciertas utilizando datos inciertos. De hecho, la incertidumbre es un aspecto crucial de toda investigación o conocimiento del mundo. Si este factor no se calcula de manera razonable, no se podrá interpretar ninguna descripción del mundo real ni ninguna inferencia causal que le afecte. Un investigador que no contempla abiertamente la incertidumbre estará afirmando que lo sabe todo a la perfección o que no tiene ni idea de la certeza o incertidumbre de sus resultados. En cualquier caso, las inferencias que carecen de un cálculo de incertidumbre no son ciencia tal como aquí se define.

### 4. El contenido es el método

Para terminar, la investigación científica propugna un conjunto de normas inferenciales de las que depende su validez. Explicar las más importantes es una de las tareas primordiales de este libro<sup>2</sup>. El principal contenido de la «ciencia» son sus métodos y reglas, no su objeto de estudio, ya que podemos utilizar tales métodos para estudiar prácticamente todo. Esta idea fue aceptada hace unos cien años cuando Karl Pearson (1892, p. 16) explicó que «el campo de la ciencia es ilimitado; su materia es infinita; cada conjunto de fenómenos naturales, cada fase de la vida social, cada estadio del desarrollo pasado o presente es materia científica. La unidad de todas las ciencias se basa únicamente en su método, no en su materia».

A su vez, estas cuatro características tienen otra consecuencia: cuando la ciencia es buena, es una *empresa social*. El trabajo de todo investigador o equipo de investigadores está sometido a limitaciones de conocimiento e ideas y los errores son inevitables, pero es probable que otros los señalen.

### El diseño de la investigación social

Comprender el carácter social de la ciencia puede resultar liberador, ya que implica que nuestro trabajo no tiene que estar por encima de la crítica para hacer una aportación importante —ya sea en cuanto a la descripción de un problema o a su conceptualización, a la teoría o a la evaluación de una en concreto. Una aportación será posible siempre que nuestro trabajo contempla explícitamente (o pretenda reinterpretar) las preocupaciones de la comunidad académica y utilice medios públicos para hacer inferencias que respeten las reglas científicas y la información de que disponemos. Incluso la aportación de un artículo menor será siempre más grande que la de un «trabajo importante» que permanezca para siempre en un cajón o confinado en un ordenador.

### **1.3 Ciencia y complejidad**

Las ciencias sociales pretenden explicar situaciones del mundo social que consideramos más o menos complejas. Sin embargo, hay que reconocer que lo que percibimos como complejidad no es algo inherente a los fenómenos: el mundo no está dividido de forma natural en conjuntos de acontecimientos simples o complejos. Por el contrario, la complejidad que se percibe en una situación depende en parte de en qué medida podemos simplificar la realidad adecuadamente, y nuestra capacidad de simplificar depende de si podemos extraer con coherencia resultados y variables explicativas. Tener más observaciones puede ayudarnos en este proceso, pero suele resultar insuficiente. Por lo tanto, «la complejidad» depende, en parte, del estado de nuestra teoría.

Los métodos científicos pueden ser tan valiosos para acontecimientos intrínsecamente complejos como para otros más simples. Es probable que la complejidad haga más inciertas nuestras inferencias, pero no tiene por qué hacerlas menos científicas. La incertidumbre y unos datos limitados no tienen que llevarnos a abandonar una investigación científica. Por el contrario: la compensación más grande a la hora de utilizar las reglas de la inferencia científica se logra precisamente cuando los datos son limitados, los instrumentos de observación tienen defectos, las medidas no están claras y las relaciones son inciertas. Si hay relaciones claras y datos no ambiguos, quizás el método no sea tan importante, ya que incluso reglas o inferencias defectuosas podrían producir respuestas medianamente aceptables.

Consideremos algunos acontecimientos complejos y, en cierto sentido, únicos que tuvieron enormes ramificaciones. La caída del Imperio romano, la Revolución francesa, la Guerra Civil de los Estados Unidos, la Primera Guerra Mundial, el holocausto y la reunificación alemana de 1990 son ejemplos de este tipo. El que estos acontecimientos tuvieran lugar parece ser el resultado de una compleja interacción de muchas fuerzas uya conjunción se antoja crucial para que así fueran. Esto quiere decir que series de

### 1. La ciencia en las ciencias sociales

acontecimientos y de fuerzas, cuyas causas son independientes entre sí, convergieron en un lugar y momento determinados, de manera que su interacción parece que ocasionó los acontecimientos que se observan (Hirschman, 1970). Además, suele ser difícil creer que tales acontecimientos fueran consecuencias inevitables de fuerzas históricas a gran escala: algunos parecen haber dependido, en parte, de la idiosincrasia de ciertas personalidades, instituciones o movimientos sociales. De hecho, desde el punto de vista de nuestras teorías, con frecuencia el azar parece haber tenido su papel: factores que se encuentran fuera del alcance de la teoría vincularon de forma determinante la sucesión de los acontecimientos.

Una de las formas de comprender estos hechos es mediante la búsqueda de generalizaciones, que conceptualizan cada caso dentro de un *tipo de acontecimientos* del que se puede decir que tiene una serie de características generales. Este método suele funcionar bien para guerras o revoluciones ordinarias, pero algunos de estos procesos, al ser mucho más extremos que otros, constituyen «valores atípicos» en la distribución estadística. Además, hay guerras o revoluciones precursoras de gran importancia que pueden tener una influencia tan considerable en fenómenos posteriores del mismo tipo —pensamos de nuevo en la Revolución francesa— que hay que tener cuidado al compararlas con sus secuelas, ya que, en cierto sentido, éstas pueden ser fruto de la imitación. Expandir el tipo de acontecimiento puede ser útil, pero no siempre resulta apropiado.

También se puede proceder de manera científica con acontecimientos peculiares de gran magnitud utilizando un análisis *contrafáctico*, que es «esta construcción mental de una serie de acontecimientos que se altera mediante la modificación de una o más "condiciones"» (Weber [1905], 1949, p. 173). La aplicación de esta idea de forma sistemática y científica se pone de manifiesto en un ejemplo especialmente extremo de un peculiar acontecimiento de la geología y la biología evolutiva, ciencias naturales con orientación histórica. Stephen J. Gould ha indicado que una de las formas de distinguir entre características evolutivas sistemáticas y acontecimientos estocásticos, aleatorios, puede ser imaginar cómo sería el mundo si se fijaran todas las condiciones, hasta un determinado punto, y se volviera a dar cuerda al resto de la historia. Señala que si fuera posible «volver a poner en marcha la cinta de la vida» y dejar que la evolución tuviera de nuevo lugar desde el principio, los organismos del mundo actual serían completamente diferentes (Gould, 1989a).

Un acontecimiento único en el que se han centrado últimamente los que estudian la evolución es la súbita desaparición de los dinosaurios hace 65 millones de años. Gould (1989a, p. 318) afirma que «debemos aceptar que nuestra conciencia no se habría desarrollado en el planeta si una catástrofe cósmica no se hubiera cobrado la vida de los dinosaurios». Si tal afirmación fuera cierta, la extinción de estos animales habría sido para los seres humanos tan importante como cualquier otro hecho histórico. Sin embargo,

su desaparición no se ajusta con propiedad al tipo de acontecimientos que se podrían estudiar directamente de forma sistemática y comparativa mediante la aplicación de leyes generales.

No obstante, la extinción de los dinosaurios puede abordarse científicamente, ya que se pueden desarrollar hipótesis alternativas y contrastarlas con sus consecuencias observables. Una de las hipótesis que explican este fenómeno, que desarrollaron Luis Álvarez y sus colaboradores de Berkeley a finales de los años setenta (W. Álvarez y Asaro, 1990), postula que hubo una colisión cósmica: un meteorito chocó con la Tierra a unos 72.000 kilómetros por hora, lo cual generó una explosión mayor que la de una guerra nuclear total. Si esta hipótesis fuera correcta, su consecuencia observable sería que el iridio (un elemento habitual en los meteoritos, pero escaso en la Tierra) se encontraría en una determinada capa de la corteza terrestre que corresponde a los sedimentos depositados hace 65 millones de años; de hecho, el descubrimiento de este metal en las capas que se había pronosticado se ha considerado una prueba que confirma parcialmente esta teoría. Aunque este acontecimiento es sin duda único, hay otras muchas consecuencias observables. Para dar sólo un ejemplo, tendría que ser posible encontrar el cráter del meteorito en algún lugar de la Tierra (y ya se han encontrado algunos candidatos)<sup>1</sup>.

El problema que plantea la causa (o causas) de la extinción de los dinosaurios sigue sin resolverse, aunque la polémica ha generado investigaciones muy valiosas. Por lo que a nosotros respecta, la importancia de este ejemplo radica en que las generalizaciones científicas son útiles incluso cuando se estudian acontecimientos muy poco habituales que no pertenecen a un tipo importante. La hipótesis de Álvarez no puede contrastarse con un conjunto de acontecimientos normales, pero si tiene consecuencias observables en otros fenómenos evaluables. Sin embargo, hay que señalar que una hipótesis no se considera una explicación razonablemente cierta hasta que no se evalúa empíricamente y se somete a ciertas pruebas estrictas. Como mínimo, sus consecuencias deben ser compatibles con nuestro conocimiento del mundo exterior, y cuando la hipótesis es realmente buena debe predecir lo que Imre Lakatos (1970) denomina «hechos nuevos», o sea, aquellos que no se habían observado hasta entonces.

La cuestión es que incluso acontecimientos aparentemente únicos, como la extinción de los dinosaurios, pueden estudiarse de forma científica si nos ocupamos de mejorar las teorías, los datos y la forma de utilizarlos. Matizar nuestra teoría mediante la clarificación conceptual y la especificación de variables puede generar más consecuencias observables e incluso poner a prueba teorías causales referidas a acontecimientos únicos como la extinción de los dinosaurios. Perfeccionar nuestros datos nos permitirá contemplar un mayor número de consecuencias observables, mientras que utilizarlos mejor hará posible que extraigamos más consecuencias de ellos. El hecho de estudiar acontecimientos muy complejos no hace irrelevante la

## 1. La ciencia en las ciencias sociales

realización de un cuidadoso diseño de investigación. Si recogemos datos sobre la mayor cantidad posible de consecuencias observables de nuestra teoría, el estudio mejorará, independientemente de que estemos barajando muchos o pocos fenómenos —o incluso uno.

## 2. Principales componentes del diseño de investigación

Cuando la investigación social es de calidad, constituye un proceso creativo en el que la intuición y el descubrimiento surgen dentro de una consolidada estructura de estudio científico. Para el científico social de primera categoría el diseño de una investigación no es un programa que haya que seguir mecánicamente para recoger y evaluar datos. Por el contrario, el investigador debe ser lo suficientemente flexible como para rechazar las viejas formas de mirar el mundo y plantear nuevas preguntas, revisar los diseños de investigación apropiadamente y recoger después datos diferentes de los que había pensado en un principio. Sin embargo, para que sus conclusiones sean válidas y aceptables para los expertos del área, todas estas revisiones y nuevas consideraciones han de producirse siguiendo procedimientos explícitos que respeten las normas de la inferencia. Un proceso de investigación dinámico tiene lugar dentro de una normativa estable.

Con frecuencia, los científicos sociales comienzan sus investigaciones con un diseño ponderado, recogen algunos datos y extraen conclusiones. Sin embargo, el proceso no suele carecer de obstáculos, y este orden no siempre es el mejor, ya que las conclusiones pocas veces se desprenden fácilmente del diseño de investigación y de los datos que se recogen de acuerdo con él. Una vez que el investigador ha reunido esos datos, tal como indicaba su proyecto, a menudo se encontrará con que el enganche entre las principales preguntas de la investigación, la teoría y los datos disponibles es imperfecto. En este momento, suele cundir el desánimo y se cree, equivocadamente, que otros científicos sociales hallan un vínculo estrecho e inmediato entre datos e investigación. Esta percepción se debe a que con frecuencia los investigadores retiran los andamios después de levantar sus edificios intelectuales, sin dejar apenas rastro de las penalidades e incertidumbres de la construcción. De ahí que el proceso investigador parezca más mecánico y previsible de lo que realmente es.

Algunos de nuestros consejos van dirigidos a los investigadores que intentan relacionar la teoría y los datos. A veces pueden concebir procedimientos de recogida de datos más apropiados para la mejor evaluación de una teoría; en otras ocasiones pueden utilizar los datos de que disponen y reformular una pregunta teórica (o incluso plantear otra completamente diferente que no se contempló en un principio) con el fin de generar un proyecto de investigación más importante. Si el estudio sigue las normas inferenciales, seguirá siendo científico y producirá inferencias fiables acerca del mundo.

### El diseño de la investigación social

Siempre que sea posible, los investigadores deben también perfeccionar el diseño de su investigación antes de llevar a cabo cualquier trabajo de campo. Sin embargo, los datos disciplinan el pensamiento a su manera. Es muy habitual que un diseño excelente haga aguas cuando se recogen las primeras observaciones: no es que la teoría esté equivocada, sino que los datos no son apropiados para responder a las preguntas planteadas en un principio. Entender desde el comienzo lo que se puede y no se puede hacer en las etapas finales puede ayudar al investigador a prever, al menos, algunos problemas cuando haga el primer diseño de investigación.

Para un mejor análisis, dividimos todos estos diseños en cuatro partes: la pregunta de la investigación, la teoría, los datos y la utilización de los datos. Estos componentes no suelen desarrollarse por separado y los estudiantes no se ocupan de ellos con un orden preestablecido. De hecho, para los cualitativistas que comienzan el trabajo de campo antes de elegir una pregunta precisa para su investigación, primero van los datos y después lo demás. Sin embargo, esta peculiar ruptura, que explicamos entre los apartados 2.1 y 2.4 de este capítulo, es especialmente útil para comprender el carácter de los diseños de investigación. Con el fin de determinar con precisión lo que *podría* hacerse si se reorientaran los recursos, nuestros consejos en el resto de este apartado presuponen que el investigador dispone de tiempo y medios ilimitados. Evidentemente, en cualquier investigación real siempre hay que hacer concesiones. Creemos que si el investigador entiende los consejos que damos para los cuatro componentes, esto le ayudará a hacer concesiones que mejoren al máximo el proyecto de su investigación, aunque ésta sea objeto de condicionantes externos.

#### 2.1 Mejorar las preguntas de la investigación

A lo largo de este libro se plantea qué hacer una vez que identificamos el objeto de estudio. Cuando hay una pregunta para la investigación, ¿cómo podemos llevar ésta a cabo para dar explicaciones válidas a los fenómenos sociales y políticos? Nuestro análisis comienza con una pregunta para la investigación y continúa después con las fases de diseño y realización. Pero, ¿dónde se originan las preguntas de un estudio? ¿Cómo elige un investigador el tema de análisis? No hay respuesta sencilla a estas preguntas. Al igual que otros autores, Karl Popper (1968, p. 32) ha señalado que «no existe algo que pueda llamarse método lógico para tener nuevas ideas [...] El descubrimiento contiene un "elemento irracional" o una "intuición creativa"». En los primeros estadios del proceso de investigación, las normas de elección están menos formalizadas que las de otras actividades de ese mismo proceso. En el área de la elección social existen textos que tratan sobre cómo diseñar experimentos de laboratorio; en estadística hay criterios para seleccionar la muestra de una encuesta que trate sobre las actitudes respec-

to a una política, y también existen manuales para llevar a cabo observación participante en una oficina. Sin embargo, no hay reglas que indiquen cómo elegir el proyecto de investigación que se va a realizar y, si decidimos llevar a cabo trabajo de campo, tampoco las hay para indicarnos dónde tenemos que hacerlo.

Podemos proponer maneras de hacer una muestra de comunidades con el fin de estudiar el impacto de diferentes políticas educativas o formas de conceptualizar los conflictos étnicos que nos lleven a la formulación y comprobación de hipótesis relativas a su incidencia. Sin embargo, no existen normas que nos indiquen si hay que estudiar la política educativa o el conflicto étnico. Desde el punto de vista de la metodología de las ciencias sociales, hay formas mejores y peores de estudiar la caída del régimen de Alemania del Este en 1989, al igual que las hay para abordar la relación entre las opiniones de un candidato en materia fiscal y sus probabilidades de éxito electoral. Sin embargo, no hay ninguna manera de determinar si es mejor estudiar la caída del régimen germano oriental o el papel de los imuestos en las elecciones estadounidenses.

El tema concreto que estudia un científico social puede tener un origen personal e idiosincrásico. No es casual que, habitualmente, sean los miembros de un grupo los que comiencen a estudiarlo: las mujeres han sido llevadas la batuta en la historia de su propio género, los negros en la de su etnia y los inmigrantes en la de la emigración. Los temas también pueden verse influidos por inclinaciones o valores personales. Es probable que quien estudia el Tercer Mundo tenga más ganas de viajar y tolere mejor condiciones de vida difíciles que el que se ocupa de la elaboración de políticas en el Congreso estadounidense, del mismo modo que quien analice la cooperación internacional puede tener una especial aversión a los conflictos violentos.

Estas experiencias y valores personales suelen proporcionar la motivación para convertirse en un científico social y, posteriormente, para elegir una determinada pregunta de investigación. En consecuencia, pueden ser las «auténticas» razones para embarcarse en un determinado proyecto, y es justo que sea así. Sin embargo, independientemente de lo personales o peculiares que sean las razones para elegir un tema, los métodos científicos y normas inferenciales que se analizan en este libro ayudarán a los académicos a hacer proyectos de investigación más sólidos. Desde el punto de vista de la posible aportación a las ciencias sociales, las razones personales no son ni necesarias ni suficientes para justificar la elección de un tema. En la mayoría de los casos, no tendrían que aparecer en los escritos académicos. Para decirlo de manera más directa pero menos delicada, a nadie le importa nuestra opinión: a la comunidad académica sólo le interesa lo que podemos demostrar.

Aunque no existan normas precisas para elegir un tema, sí hay maneras de determinar —aparte de las preferencias personales— la posible trans-

cendencia que tiene para la comunidad científica una empresa investigadora. Lo ideal sería que todos los proyectos de investigación de las ciencias sociales cumplieran dos condiciones. En primer lugar, un proyecto de investigación tiene que plantear una pregunta «importante» para el mundo real. El tema ha de ser relevante para la vida política, social o económica, para entender algo que afecte de manera significativa a la vida de muchas personas o para comprender o predecir acontecimientos que pudieran resultar dañinos o beneficiosos (véase Shively, 1990, p. 15). En segundo lugar, un proyecto de investigación tiene que hacer una aportación concreta a lo escrito en un área académica identificable, aumentando la capacidad colectiva de dar explicaciones científicas verificables a algún aspecto del mundo. Esta segunda condición no supone que todas las investigaciones que aporten algo a nuestra reserva de explicaciones para las ciencias sociales pretendan en realidad hacer inferencias causales. En ocasiones, la situación del conocimiento en un determinado campo hace que se precisen muchos datos y descripciones antes de afrontar el reto de dar una explicación. A veces, la aportación de un proyecto sólo es la inferencia descriptiva, mientras que en otras el objetivo ni siquiera es hacer tal inferencia, sino únicamente observar de cerca ciertos acontecimientos o hacer un resumen de hechos históricos. Sin embargo, en este caso se cumple nuestra segunda condición, porque tales acontecimientos son un requisito imprescindible para la explicación.

La primera condición dirige nuestro interés al mundo real de los fenómenos políticos y sociales, y también a cómo se registran los acontecimientos y problemas actuales e históricos que configuran la vida de las personas. Determinar si una pregunta de investigación cumple dicho criterio es algo principalmente social. La segunda condición nos sitúa en la bibliografía académica de las ciencias sociales, es decir, en las perplejidades intelectuales que no se han planteado todavía, en las que están por resolver, así como en las teorías y métodos científicos de que se dispone para resolverlas.

A los polítólogos no les resulta difícil encontrar un objeto de estudio que cumpla la primera condición. A lo largo de los últimos cuatrocientos años diez guerras importantes se han cobrado la vida de casi treinta millones de personas (Levy, 1985, p. 372); algunas «guerras limitadas», como las que libraron los Estados Unidos y Vietnam del Norte, o Irán e Irak, han producido cada una un número de víctimas cercano al millón; del mismo modo, si hubiera una guerra nuclear, ésta podría eliminar a miles de millones de seres humanos. La mala gestión de los políticos, tanto en el ámbito nacional como en el internacional, ha generado —como en los años treinta— privaciones económicas de carácter global y crisis regionales y locales, tal como lo demuestran las trágicas experiencias de gran parte de África y América Latina durante los años ochenta. En general, las variaciones entre países en lo tocante a instituciones políticas se relacionan con las diversas condiciones de la vida ordinaria, las cuales se reflejan en la diferente esperanza de vida o mortalidad infantil que tienen países con un grado de desa-

rrollo económico similar (Russett, 1978, pp. 913-28). Dentro de los Estados Unidos, la eficacia de los programas que combaten la pobreza o la desarticulación social parece haber variado de forma considerable. No cabe duda de que será importante toda investigación que contribuya, aunque sea de forma marginal, al conocimiento de estos problemas.

Anunque los científicos sociales tienen muchas cuestiones significativas que investigar, las herramientas para entenderlas son escasas y poco elaboradas. Hay muchos escritos sobre la guerra o el sufrimiento en las sociedades que poco contribuyen a la comprensión de tales problemas porque no los describen de forma sistemática o no hacen inferencias causales o descriptivas válidas. Las ideas brillantes pueden mejorar la comprensión al aportar nuevas e interesantes hipótesis, pero la brillantez no es un método de investigación empírica. Todas las hipótesis han de ser contrastadas empíricamente antes de que hagan una aportación al conocimiento. Este libro no da consejos sobre cómo ser brillante. Sin embargo, lo que sí puede hacer es recalcar la importancia de que la investigación se lleve a cabo de manera que constituya una aportación al conocimiento.

Nuestro segundo criterio para elegir una pregunta de investigación, «hacer una aportación», significa explícitamente que hay que situar el diseño del estudio dentro del marco de la bibliografía científica existente. Esto garantiza que el investigador comprende el «estado de la cuestión» y reduce al mínimo la posibilidad de repetir lo que ya se ha hecho. También garantiza que el trabajo sea importante para otros, aumentando así el éxito del conjunto de la comunidad académica. Se puede hacer una aportación explícita a lo escrito sobre el tema de varias maneras. A continuación se enumeran algunas de ellas:

1. Eligiendo una hipótesis que los estudiosos consideren importante en la bibliografía pero de la que no se haya realizado un estudio sistemático. Si encontramos pruebas a favor o en contra de la hipótesis, estaremos haciendo una aportación.
2. Eligiendo una hipótesis aceptada en la bibliografía que sospechemos es falsa (o creemos que no se ha demostrado adecuadamente) e investigando si realmente lo es o si otra teoría es correcta.
3. Intentando resolver o presentando más pruebas a favor de uno de los bandos en una polémica que esté presente en lo escrito hasta el momento; demostrando quizás que toda esa polémica carecía de fundamento desde el principio.
4. Diseñando investigaciones que arrojen luz o evalúen premisas no cuestionadas por la bibliografía.
5. Señalando que en la bibliografía no se ha concedido atención a un asunto importante y proceder a continuación a aportar a ese campo un estudio sistemático.
6. Señalando qué las teorías o pruebas relativas a cierto objetivo en un campo podrían aplicarse a otro para solucionar un problema existente, pero aparentemente alejado.

### **El diseño de la investigación social**

Si nos preocupamos demasiado por hacer una aportación a la bibliografía académica sin prestar cierta atención a asuntos importantes para el mundo real, correremos el riesgo de formular preguntas insignificantes desde el punto de vista político. Por el contrario, centrarse en el contexto político actual sin preocuparse de hasta qué punto un problema social es susceptible de investigarse sistemáticamente dentro del marco de un cuerpo de conocimiento conduce a trabajos descuidados que poco añaden a una comprensión profunda.

Nuestros dos criterios para elegir preguntas de investigación no son necesariamente opuestos. A largo plazo, la comprensión de los fenómenos del mundo real aumenta con la elaboración y evaluación de hipótesis explicativas que propugna el método científico. Sin embargo, de forma más inmediata, puede haber contradicción entre lo que es útil desde el punto de vista práctico y lo que finalmente tiene valor. Por ejemplo, Mankiw (1990) señala que la teoría macroeconómica y la macroeconomía aplicada se apartaron considerablemente la una de la otra en los años setenta y ochenta: modelos cuya incoherencia teórica se había demostrado siguieron utilizándose para hacer pronósticos sobre la economía estadounidense, mientras que los nuevos modelos teóricos, concebidos para corregir esos fallos, continuaron teniendo un carácter especulativo y no se desarrollaron lo suficiente como para realizar predicciones ajustadas con ellos.

Cuando un investigador elige un tema, puede parecer que los criterios de aplicabilidad al mundo real y aportación al progreso científico son opuestos. Algunos investigadores comenzarán con un problema del mundo real que sea muy significativo socialmente, como la amenaza de guerra nuclear, la disparidad de ingresos entre hombres y mujeres o la transición a la democracia en Europa del Este. Otros quizás partan de un problema intelectual planteado en la bibliografía de las ciencias sociales: la contradicción entre diversos estudios experimentales de tomas de decisión en condiciones de incertidumbre o la existente entre diversas teorías de voto referidas a las elecciones para el Congreso estadounidense y los últimos resultados en las urnas. Evidentemente, no hay una distinción estricta entre los criterios. Algunas preguntas de investigación responden a los dos desde el principio, pero, a la hora de diseñar el estudio, los investigadores suelen comenzar estando más cerca de uno que de otro<sup>4</sup>.

Independientemente del punto de partida, el proceso de diseñar una investigación para dar respuesta a una pregunta debe cumplir nuestras dos condiciones, y es evidente que la dirección de este movimiento dependerá de dónde comience. Si lo que nos motiva es una perplejidad científica, tendremos que preguntarnos cómo hacer que el objeto de la investigación sea más relevante para asuntos importantes del mundo real: por ejemplo, cómo podría un experimento de laboratorio dilucidar mejor las elecciones estratégicas de los decisores del mundo real o las consecuencias que, para el comportamiento, puede tener esta teoría. Si partimos de un problema real, ten-

### **1. La ciencia en las ciencias sociales**

dremos que preguntarnos cómo puede estudiarse con métodos científicos modernos para hacer una aportación al fondo de explicaciones de las ciencias sociales. Bien pudiera ser que apartarse demasiado de uno u otro criterio no fuera la mejor opción. Los que hacen experimentos de laboratorio pueden señalar que es prematuro buscar referencias externas y que se avanzará más si se perfecciona la teoría y el método en un medio controlado; quizás, en un estudio de larga duración, esto sea cierto. Por el contrario, quien esté preocupado por un problema del mundo real puede indicar que antes de explicar es necesario describir con precisión, y puede que también tenga razón, ya que la descripción exacta es un paso importante de los estudios explicativos.

En cualquier caso, si es posible, todo proyecto de investigación tiene que intentar cumplir nuestras dos condiciones: ocuparse de un asunto relevante del mundo real y estar concebido para realizar una aportación, directa o indirectamente, a una determinada área bibliográfica académica. En este libro, al ser nuestra principal preocupación que la investigación cualitativa sea más científica, nos ocuparemos más del investigador que parte del «mundo real», pero nuestro análisis es aplicable a ambas tendencias.

Cuando el punto de partida es un problema significativo del mundo real y no un área bibliográfica ya establecida, resulta esencial que elaboremos un plan de estudio factible. Si se propone un asunto que no puede desarrollarse hasta convertirlo en un proyecto de investigación concreto, que permita la extracción de inferencias descriptivas y causales válidas, hay que modificarlo o abandonarlo. Del mismo modo, también hay que cambiar toda propuesta que no haga alguna aportación a los escritos académicos. Al elegir provisionalmente un tema, estamos dialogando con su bibliografía. ¿Qué preguntas de interés para nosotros se han respondido ya? ¿Cómo se puede plantear y pulir nuestra pregunta para que parezca posible contestarla con las herramientas disponibles? Podemos partir de un asunto candente, pero tendremos que hacer frente tanto a lo escrito en las ciencias sociales como a problemas inferenciales.

### **2.2 Mejorar la teoría**

En las ciencias sociales, una teoría es una especulación razonada y precisa sobre la respuesta que cabe dar a la pregunta de una investigación, e incluye una declaración de por qué tal respuesta es correcta. Las teorías suelen conllevar hipótesis descriptivas o causales más específicas. Una teoría debe estar en consonancia con los datos disponibles anteriormente sobre una pregunta de investigación. «Una teoría que no tiene en cuenta los datos existentes es una incoherencia. Si dispusiéramos de algo equivalente a la legislación sobre "veracidad en la publicidad", esta incoherencia no se llamaría teoría» (Lieberson, 1992, p. 4; véase también Woods y Walton, 1982).

El desarrollo de la teoría suele presentarse como el primer paso de la investigación. A menudo, en la práctica, es lo primero, pero no tiene por qué serlo. De hecho, no podemos elaborar una teoría sin conocer antes los trabajos realizados sobre el tema y sin haber reunido algunos datos, ya que en este estadio ni siquiera se sabe cuál es la pregunta de la investigación. No obstante, independientemente de la cantidad de datos que se hayan reunido ya, hay algunas formas generales de evaluar y aumentar la utilidad de una teoría. Ahora sólo vamos a esbozarlas, pero las analizaremos en detalle en capítulos posteriores.

*falta bibliografía*

En primer lugar, hay que elegir teorías que puedan estar equivocadas. En realidad, se aprende mucho más de las que *están* equivocadas que de aquellas que se enuncian de forma tan general que puede que ni siquiera lo estén en principio<sup>5</sup>. Es necesario poder responder directamente a la pregunta: ¿qué datos nos convencerían de que estamos equivocados?<sup>6</sup>. Si no hay respuesta para esta pregunta, tampoco hay teoría.

En segundo lugar, para asegurarse de que una teoría es falsable, hay que elegir una que pueda generar tantas *consecuencias observables* como sea posible. Tal elección posibilitará que la teoría se someta a más pruebas, utilizando más datos y más variados, que esté más veces en peligro de ser falsada y que se puedan recoger más datos para darle una mayor consistencia.

En tercer lugar, al diseñar teorías es preciso ser lo más concreto posible. Las teorías e hipótesis que se formulan con vaguedad no sirven más que para ofuscar. Se puede demostrar más fácilmente que una teoría está equivocada cuando se ha enunciado con precisión y hace predicciones específicas; por consiguiente, esta teoría será mejor.

Algunos investigadores recomiendan que se siga el principio de «concisión». Por desgracia, este término se ha utilizado tantas veces en conversaciones informales y en escritos académicos que el principio ha quedado eclipsado (véase Sober [1988] para un análisis completo). Jeffreys (1961, p. 47) dio la definición más clara de concisión: «Las teorías más simples tienen en principio más probabilidades»<sup>7</sup>. Por lo tanto, la concisión es un juicio, e incluso un supuesto, sobre la naturaleza del mundo: se presupone que es simple. Elegir teorías cuya premisa es la existencia de un mundo simple es una regla que se aplica con claridad en situaciones en las que se está muy seguro de que el mundo es así realmente. Los físicos parecen cómodos con la concisión, pero los biólogos suelen creer que es absurda. En las ciencias sociales, los hay que la defienden con vehemencia para sus subáreas (por ejemplo, Zellner, 1984), pero nosotros creemos que sólo es útil en ciertas ocasiones. Dada la precisa definición de concisión como supuesto acerca del mundo, nunca deberíamos insistir en que fuera un principio general a la hora de diseñar teorías, aunque sea útil en aquellas situaciones en las que haya indicios de que el mundo que estudiamos es simple.

Lo que queremos decir es que no recomendamos al investigador que busque la concisión como un bien en sí mismo, porque, a menos que ya se-

pamos mucho de un asunto, no parece que haya muchas razones para practicarla. Ni siquiera necesitamos este principio para evitar teorías excesivamente complicadas, porque está implícito en la máxima de que la teoría debe ser tan complicada como indican los datos. Las situaciones en las que no hay información suficiente sobre la complejidad de la teoría que se investiga pueden conducir a lo que se denomina «diseños de investigación imprecisos» (véase el apartado 1 del capítulo 4), pero éstos son problemas del diseño de la investigación y no supuestos acerca del mundo.

Todos los consejos que hemos dado hasta ahora pueden aplicarse si aún no hemos reunido nuestros datos ni hemos comenzado a analizarlos. Sin embargo, si están ya recogidos, podemos utilizar esas reglas para modificar nuestra teoría y conseguir más datos, generando de este modo más consecuencias observables sobre la nueva teoría. Por supuesto, este proceso es caro, cuesta mucho tiempo y quizás vayamos a desperdiciar los datos que ya se han recogido. ¿Qué ocurre entonces cuando está claro que hay que mejorar nuestra teoría pero no podemos permitirnos recoger más datos? Esta situación —en la que con frecuencia se encuentra el investigador— ha de tratarse con cautela y dominio de uno mismo. Cualquier investigador inteligente puede elaborar una teoría «plausible» con el conjunto de datos que surge de un hecho; sin embargo, esto no diría nada sobre la veracidad de la teoría, que puede ajustarse muy bien a los datos y ser completamente falsa —y, de hecho, su falsedad ser demostrable con otros muchos datos. A los seres humanos se les da bien reconocer pautas, pero no tan bien detectar la falta de las mismas (¡la mayoría vemos pautas en manchas de tinta salteadas!). A una teoría hay que racionarle los ajustes realizados sobre la marcha que no sean coherentes con los datos disponibles, y hay que hacerlos con bastante disciplina<sup>8</sup>.

Todavía nos queda el problema de qué hacer cuando hayamos terminado tanto la recogida de datos como nuestro análisis y queramos mejorar nuestra teoría. En esta situación, recomendamos que se respeten dos normas: en primer lugar, si nuestra predicción se halla condicionada por diversas variables y queremos librarnos de una de las condiciones, podemos hacerlo. Por ejemplo, si al principio nuestra hipótesis era que los países democráticos con sistemas avanzados de protección social no luchan entre sí, sería aceptable extender tal hipótesis a todas las democracias contemporáneas y de este modo contrastar nuestra teoría en un mayor número de casos y aumentar las posibilidades de que sea falsada. El propósito general es que, después de ver los datos, podamos modificar la teoría para que se pueda aplicar a una gama más amplia de fenómenos. Esta alteración de nuestras tesis, al exponerlas aún más a la falsación, no debería llevarnos a explicaciones para salir del paso que sólo pretendan «salvar» una teoría inadecuada restringiéndola a fenómenos con los que ya se ha comprobado que concuerda.

Sin embargo, la práctica opuesta no suele ser apropiada. Despues de observar los datos, no deberíamos únicamente añadir una condición restrictiva

y continuar como si esa matización hubiera demostrado que nuestra teoría es correcta. Si la teoría original fuera que las democracias contemporáneas no luchan entre sí porque tienen sistemas constitucionales, sería menos aceptable que, al encontrar excepciones a nuestra «regla», limitáramos la proposición a las democracias que disponen de un sistema avanzado de protección social una vez que se ha constatado mediante la inspección de los datos que esta matización parece convertir nuestra proposición en correcta. [O supongamos que nuestra teoría inicial fuera que las revoluciones sólo tienen lugar en situaciones de grave crisis económica, pero nos damos cuenta de que no es verdad en uno de nuestros estudios de caso. En esta situación no sería razonable añadir únicamente condiciones generales como: las revoluciones nunca tienen lugar durante períodos de prosperidad excepto allí donde el ejército es débil, los líderes políticos son represivos, la economía se basa en un reducido número de productos y el clima es calido. Tal formulación es una forma caprichosa (y engañosa) de afirmar que «mi teoría es correcta, excepto en el país X». Ya que hemos descubierto que nuestra teoría no se aplica en dicho país, no es de mucha ayuda convertir esta falsación en una generalización espuria. Si no nos esforzamos por recabar nuevos datos, no dispondremos de pruebas aceptables que sostengan la nueva versión de la teoría.]

Por lo tanto, nuestra regla básica en cuanto a la alteración de una teoría una vez que se han observado los datos es: podemos hacer que la teoría sea menos restrictiva (de forma que contemple una gama más amplia de fenómenos y se vea expuesta a más oportunidades de falsación), pero esto no ha de hacerse sin haber recogido más datos con los que contrastar la nueva versión de la teoría. Si no podemos recabar más datos, estaremos en punto muerto y no podemos proponer ningún método mágico para dejar de estarlo. En ciertos momentos, aceptar que estamos equivocados es lo mejor y, de hecho, algunos resultados negativos pueden ser bastante útiles para un área académica. ¿Quién no habría de preferir una sólida conclusión negativa antes que varias de carácter insustancial que se basen en teorías elaboradas para salir del paso?]

Además, si estamos equivocados, no tenemos por qué dejar de escribir una vez que hayamos admitido nuestra derrota. Podemos añadir un apartado a nuestro artículo o un capítulo en el libro para ocuparnos de futuras investigaciones empíricas y de la presente especulación teórica. En este contexto, disponemos de bastante más libertad y podemos señalar nuevas condiciones que sería plausible introducir en nuestra teoría —si creemos que pueden solucionar el problema—; plantear la modificación de otra teoría existente o apuntar hacia una gama de teorías completamente diferente. En tal situación, no podemos llegar a ninguna conclusión con mucha certeza (excepto, quizás, al afirmar que la teoría que planteamos al principio está equivocada), pero sí podemos permitirnos el lujo de inventar nuevos diseños de investigación o proyectos de recogida de datos que podrían utilizarse

para determinar si nuestras especulaciones son correctas. Este ejercicio puede ser bastante útil, especialmente a la hora de señalar áreas de las que los futuros investigadores podrían ocuparse.

Como hemos señalado anteriormente, ya se sabe que las ciencias sociales no siguen normas estrictas: ¡la necesidad de creatividad a veces exige que se prescinda del manual! Y los datos pueden disciplinar el pensamiento. De ahí que a los investigadores, una vez que se enfrentan a los datos, se les ocurra cómo deberían haber construido la teoría en un principio. Aunque sea de forma restrictiva, tal modificación quizá valga la pena si podemos convencernos a nosotros mismos y a los demás de que reformar la teoría tal como lo proponemos es algo que habrían hecho antes de recoger los datos si se nos hubiera ocurrido. Sin embargo, hasta que la teoría no se contraste con nuevos datos, su posición seguirá siendo incierta, y así habrá que considerarla.

Una de las consecuencias importantes de estas normas es que los proyectos piloto suelen ser muy útiles, especialmente en investigaciones en las que los datos han de recogerse mediante entrevistas u otros medios particularmente costosos. La recogida de datos preliminar puede llevarnos a alterar las preguntas de la investigación o a modificar la teoría. Posteriormente, se pueden recabar otros datos para contrastar la nueva teoría, evitándose así el problema de utilizar la misma información para elaborar y comprobar una teoría.

### 2.3 Mejorar la calidad de los datos

Los «datos» son informaciones sobre el mundo recogidas de forma sistemática y pueden ser de tipo cualitativo o cuantitativo. A veces se reúnen datos para evaluar una teoría muy concreta, pero no es inusual que los académicos los recojan antes de saber exactamente qué es lo que les interesa encontrar. Además, aunque los datos se reúnan para contrastar una hipótesis determinada, puede que al final a los investigadores les interesen preguntas que no se les había ocurrido anteriormente.

En cualquier caso, ya sea que los datos se reúnan para un determinado propósito o que se utilicen para otro que no está claro cuando se realiza la recogida, existen ciertas normas que mejoran la calidad de tales datos. En principio, podemos considerar esas normas al margen de las que se han mencionado en el apartado 2.2 para mejorar la teoría. En la práctica, para toda recogida de datos se necesita cierto grado de teoría del mismo modo que para formular cualquier teoría se precisan algunos datos (véase Combs, 1964).

Nuestra primera y más importante directriz para mejorar la calidad de los datos es:  registrar y detallar el proceso con el que se generan los datos. Sin esta información no podemos determinar si la utilización de procedi-

mientos normalizados de análisis producirá inferencias sesgadas. [Sólo si conocemos el proceso de obtención de los datos podremos hacer inferencias descriptivas y causales válidas.] En una encuesta de opinión de tipo cuantitativo, para registrar el proceso de obtención de datos es preciso conocer exactamente de qué manera se ha realizado la muestra y qué preguntas se han hecho. En un estudio de caso comparado de tipo cualitativo es crucial enunciar las normas precisas que hemos utilizado para elegir el reducido número de casos que se van a analizar. En el capítulo 6 damos más directrices para la selección de casos en investigación cualitativa, pero aún más importante que elegir un buen método es registrar y detallar con cuidado cualquiera que se utilice, así como la información necesaria para que otros lo apliquen<sup>9</sup>.

En el apartado 2.2 defendimos teorías de las que pudieran extraerse muchas consecuencias observables. [Nuestra segunda directriz para mejorar la calidad de los datos es: *para evaluar mejor una teoría hay que recoger datos acerca de la mayor cantidad posible de consecuencias observables.*] Esto supone recoger tantos datos como sea posible en los contextos más diversos. Cada consecuencia adicional de nuestra teoría que observemos aportará un nuevo contexto en el que poner a prueba su veracidad. Cuantas más consecuencias observables encontremos que sean coherentes con la teoría, más penetrante será la explicación y más ciertos sus resultados.

[Al añadir datos sobre nuevas consecuencias observables de una teoría, podemos (a) recoger más observaciones sobre la misma variable dependiente o (b) registrar más variables dependientes. Podemos, por ejemplo, desagregar los datos en períodos de tiempo más cortos o áreas geográficas más pequeñas.] También se puede recabar información sobre variables dependientes de un interés menos directo; si los resultados son como los que prevé la teoría, tendremos más confianza en ella.

Pensemos, por ejemplo, en la teoría de la disuasión racional: los potenciales iniciadores de una guerra calculan los costes y beneficios de atacar a otros estados, y en estos cálculos pueden influir amenazas de represalia creíbles. La comprobación más directa de esta teoría sería evaluar si, dadas ciertas amenazas de guerra, la decisión de atacar se asocia con factores como el equilibrio de fuerzas militares entre los posibles contendientes o con los intereses que están en juego para el país atacado (Huth, 1988). Sin embargo, aunque utilizar únicamente casos en los que se lanzan amenazas conforme un conjunto de consecuencias observables de la teoría, éstas son sólo una parte de las que se podrían recoger (y si sólo se contemplan éstas se podría producir un sesgo de selección), ya que aquellas situaciones en las que se disuade incluso de amenazar serían excluidas de los datos. Por lo tanto, también podría merecer la pena recoger información sobre otra variable dependiente (o sea, sobre otro conjunto de consecuencias observables) que calibrara si los estados amenazan cuando tienen incentivos para hacerlo.

[Siempre que se carezca de datos suficientes sobre la disuasión en política internacional, también puede ser útil poner a prueba otra teoría que parte de presupuestos similares respecto a la motivación y que se aplique a otra variable dependiente en diferentes condiciones, sin dejar de ser una consecuencia observable de la misma teoría. Por ejemplo, podríamos concebir un experimento de laboratorio para ver si, en circunstancias simuladas, la capacidad militar y unas negociaciones firmes, en vez de acentuar las «amenazas», hacen que no se produzcan. También podríamos examinar si, en situaciones análogas, actores como los oligopolios que compiten por su cuota de mercado o las familias del crimen organizado que luchan por el territorio utilizan estrategias de disuasión y hasta qué punto tienen éxito en diversas circunstancias. De hecho, los economistas que trabajan en el área de la organización industrial han utilizado la teoría de juegos no cooperativa, en la que también se basa la de la disuasión, para estudiar problemas como el de la entrada en ciertos mercados o las estrategias de precios (Fudenberg y Tirole, 1989). Como las teorías se parecen tanto, los datos empíricos en los que se basan las predicciones de la teoría de juegos acerca del comportamiento de las empresas aumentarían la plausibilidad de otras hipótesis relacionadas con la acción del Estado en política internacional. Seguiría habiendo incertidumbre al aplicar las conclusiones de un campo al otro, pero el problema es lo suficientemente importante como para justificar que se intenta sacar ideas y pruebas de donde sea posible.]

Es evidente que una recogida de datos que siempre prescinda del análisis impedirá la finalización de una investigación útil en vez de facilitarla. En la práctica, las limitaciones de tiempo y de recursos siempre condicionarán los esfuerzos por recoger datos. Aunque más información, más casos, más entrevistas, una nueva variable y otras formas relevantes de recopilación de datos siempre vayan a dar más certeza a nuestras inferencias en algún sentido, puede que haya investigadores prometedores que se vean perjudicados en igual medida por el exceso de información que por su escasez. Insistir en leer un libro más o en hacerse con otra matriz de datos brutos sin escribir una línea es recetar improductividad.

[La tercera directriz es: *maximizar la validez de nuestras mediciones*. La validez tiene que ver con calibrar lo que creemos que estamos calibrando.] El índice de paro puede ser un buen indicador de la situación económica, pero uno y otra no son sinónimos. En general, lo más fácil es maximizar la validez ajustándose a los datos sin permitir que se nos interpongan conceptos no observados o no mensurables. Si un entrevistado responde a nuestra pregunta aduciendo ignorancia, lo que sabemos es que *dijo* que no lo sabía, y para eso si tenemos una medida válida. Sin embargo, lo que *quiso decir* en realidad constituye un concepto completamente diferente que no puede medirse con un alto grado de confianza. Por ejemplo, en países con gobiernos represivos, aducir ignorancia puede ser, para algunas personas, una forma de criticar el sistema político; para otros es como decir «No sé».

### El diseño de la investigación social

*[En la parte superior izquierda de este cuadro se encuentra una nota manuscrita que dice: "Nuestra cuarta directriz es: asegurarse de que los métodos de recogida de datos son fiables. Fidabilidad significa que si se aplica el mismo procedimiento de la misma manera siempre se obtendrá la misma medida. Cuando se utiliza un procedimiento fiable en diferentes ocasiones, sin que nada haya ocurrido entretanto que transforme la «verdadera» situación del objeto que estamos calibrando, se observará el mismo resultado<sup>10</sup>. Las medidas fiables también ofrecen los mismos resultados cuando las realizan diferentes investigadores, y ello depende, evidentemente, de que se puedan seguir procedimientos explícitos<sup>11</sup>.]*

*[Nuestra última directriz es que todos los datos y análisis deben ser, en la medida de lo posible, reproducibles.]* La capacidad de reproducción no solo se aplica a los datos, con el fin de que podamos comprobar si nuestras medidas son fiables, sino que también tiene que ver con el proceso de razonamiento que se utiliza para alcanzar conclusiones. Partiendo del informe de nuestra investigación, otro investigador debería poder copiar los datos y rastrear la lógica de nuestras conclusiones. La capacidad de reproducción es importante aunque nadie repita nuestro estudio. Los procedimientos y métodos utilizados sólo se podrán evaluar si el estudio se describe con el suficiente detalle como para que pueda reproducirse.

En algunos tipos de investigación reproducir los datos puede resultar difícil o imposible: quizás los entrevistados mueran o desaparezcan y no sea posible reproducir las observaciones directas de acontecimientos del mundo real que han hecho testigos o participantes en ellos. La capacidad de reproducción también significa cosas diferentes en cada tradición investigadora. En los estudios cuantitativos se centra en repetir el análisis a partir de los mismos datos. Como bien sabe cualquiera que haya intentado alguna vez reproducir los resultados cuantitativos de obras publicadas, la tarea suele ser bastante más difícil de lo que debería —aunque las obras sean de relieve— y siempre resulta más útil de lo que parece al principio (acerca de la reproducción en investigación cuantitativa, véase Dewald y otros, 1986).

En la investigación cualitativa tradicional la analogía la proporcionan las notas a pie de página y las revisiones bibliográficas. Utilizando estas herramientas, los sucesivos investigadores tendrían que poder localizar las fuentes utilizadas en una obra publicada y hacer sus propias evaluaciones sobre las inferencias que han surgido de esa información. Si los estudios se basan en la observación directa, la reproducción es más difícil. Un investigador puede tomar prestadas las notas de campo de otro o la grabación de sus entrevistas para comprobar si este material fundamenta las conclusiones extraídas en primera instancia. Este nuevo análisis de resultados, utilizando los mismos datos, no suele hacerse, ya que gran parte de la información del trabajo de campo se basa en conversaciones, impresiones y otros actos de participación no registrados. Sin embargo, se podría avanzar bastante si hubiera más investigadores que intentaran hacer este tipo de reproducción, y

### 1. La ciencia en las ciencias sociales

así se animaría también a otros a tomar notas de campo más detalladas. Alguna vez se ha reproducido un proyecto de investigación completo, incluyendo la recogida de datos. La reproducción no puede ser perfecta porque es imposible retrotraerse en el tiempo; no obstante, el intento puede resultar bastante fructífero. Quizás la reproducción más exhaustiva que se haya llevado a cabo de un proyecto cualitativo sea la del estudio sociológico de la ciudad estadounidense de Middletown, en Indiana, que comenzaron Robert y Helen Lynd. Su primer estudio de «Middletown» se publicó en 1929 y se reprodujo en un libro que apareció en 1937. Unos cincuenta años después del primer proyecto se está publicando una larga serie de libros y artículos que reproduce los primeros estudios (véase Caplow y otros, 1983a, 1983b y las citas que incluyen). La investigación cualitativa no tiene por qué ser tan exhaustiva, pero este gran proyecto de investigación debería servir como modelo de lo que se puede hacer.

Toda investigación tendría que aspirar a ser lo más reproducible posible: los investigadores habrían de registrar siempre los métodos, normas y procedimientos exactos que han utilizado en la recogida de información y extracción de inferencias para que otros pudieran hacer lo mismo y llegar (esperamos) a la misma conclusión. La capacidad de reproducción también supone que los académicos que utilizan fuentes inéditas o privadas tendrían que comprometerse a que, en el futuro, otros pudieran acceder a este material en condiciones similares; aprovecharse de un acceso privilegiado sin buscar que otros lo tengan imposibilita la reproducción y pone en cuestión la calidad científica del trabajo. Lo normal es que las investigaciones no sean reproducidas, pero nuestra responsabilidad es comportarnos como si alguien pudiera querer hacerlo. Aunque la obra no sea reproducida, aportar los materiales para que si lo sea hará posible que los lectores comprendan y evalúen nuestro trabajo.

#### 2.4 Utilizar mejor los datos existentes

Solucionar los problemas que plantean los datos mediante la recopilación de otros nuevos y mejores es casi siempre preferible a utilizar de otro modo los datos defectuosos de que disponemos: sin embargo, no siempre es posible recoger otros. Los científicos sociales suelen enfrentarse a datos problemáticos y a pocas posibilidades de adquirir otros mejores, de manera que tienen que sacar el máximo partido a los que tienen.

Cómo utilizar mejor los datos ya recogidos es el tema principal de las clases de métodos estadísticos y, de hecho, es la mayor aportación de la estadística inferencial a las ciencias sociales. Los preceptos de este asunto, qué tan claros están en el estudio de la estadística inferencial, también se aplican a la investigación cualitativa. En el resto del libro nos ocuparemos de ellos con más detalle, y aquí sólo vamos esbozar brevemente las

directrices que hay que seguir para utilizar mejor datos recogidos con anterioridad.

En primer lugar, siempre que sea posible, para hacer inferencias hay que utilizar datos que no estén «sesgados», es decir, que como promedio sean correctos. Para entender esta idea tan concreta de la investigación estadística, imaginemos que se aplica la misma metodología (ya sea en la investigación cuantitativa o en la cualitativa) para analizar y extraer conclusiones de datos procedentes de varias matrices de datos brutos. Es probable que, a causa de pequeños errores en los datos o en la aplicación del procedimiento, utilizar sólo una vez esta metodología nunca dará resultados del todo correctos. Un procedimiento «no sesgado» será correcto si se considera como una media de muchas aplicaciones; aunque, por separado, éstas no sean correctas. El procedimiento no inclinará sistemáticamente el resultado en una u otra dirección.

Es evidente que la extracción de inferencias no sesgadas depende tanto de la primera recogida de datos como de su uso posterior y que, como hemos señalado anteriormente, siempre es mejor prever los problemas antes de comenzar a reunir los datos. Sin embargo, aquí mencionaremos brevemente estas dificultades porque al utilizar los datos es preciso poner un especial cuidado en el análisis de posibles fuentes de sesgo que hayan pasado desapercibidas cuando se recogían dichos datos. Una de esas fuentes, que puede falsear las inferencias, es la del sesgo de selección, que consiste en elegir observaciones de tal manera que se distorsiona sistemáticamente la población de la que proceden. Aunque un ejemplo obvio es el de escoger a propósito sólo aquellos casos que apoyan nuestra teoría, el sesgo de selección puede ser mucho más sutil. Hay otro sesgo que puede surgir de las variables omitidas y que consiste en excluir alguna variable de control que pudiera influir en una aparente conexión causal entre las variables explicativas y la que queremos aclarar. Entre los capítulos 2 y 6 analizaremos estas y otras posibles dificultades que se presentan en la extracción de inferencias no sesgadas.

La segunda directriz se basa en el concepto estadístico de «eficiencia»: para que los datos se utilicen de manera eficiente hay que maximizar la información que se ha tenido en cuenta para hacer las inferencias descriptivas o causales. Para maximizar la eficiencia no sólo es preciso utilizar todos nuestros datos, sino usar también toda la información relevante que hay en ellos para mejorar las inferencias. Por ejemplo, si los datos están desagregados en pequeñas unidades geográficas, tenemos que utilizarlos de esa forma, no sólo como un agregado nacional. Los agregados más pequeños tendrán un mayor grado de incertidumbre, pero si son, al menos en parte, consecuencias observables de la teoría, contendrán información que puede traerse a colación en el problema inferencial.

### 3. Temas de estudio de este volumen

Finalizamos este capítulo general señalando los cuatro puntos importantes en la realización de un diseño de investigación que hemos planteado aquí y que analizaremos en detalle a lo largo del libro.

#### 3.1 La utilización de consecuencias observables para vincular la teoría y los datos

En este capítulo hemos subrayado que para que una teoría merezca la pena debe contemplar consecuencias relacionadas con las observaciones que esperamos encontrar si es correcta. Estas *consecuencias observables* de la teoría han de guiar nuestra recogida de datos y ayudarnos a distinguir entre hechos relevantes e irrelevantes. En el apartado 6 del capítulo 2 analizamos cómo afecta la teoría a la recogida de datos y de qué manera éstos disciplinan la imaginación teórica. Aquí lo que queremos recalcar es que la teoría y la investigación empírica deben estar estrechamente relacionadas. Cualquier teoría que de verdad trabaje para nosotros tendrá consecuencias en la investigación empírica, y ésta no podrá llegar a buen puerto sin una teoría que la guíe en la elección de sus preguntas. Tanto la teoría como la recogida de datos son aspectos esenciales del proceso que nos lleva a decidir si una teoría puede considerarse de forma provisional verdadera o falsa, aunque esté sujeta en ambos casos a la incertidumbre que caracteriza a todas las inferencias.

A toda teoría hay que preguntarle cuáles son sus consecuencias observables y en cualquier investigación empírica hay que saber si las observaciones son relevantes para las consecuencias de la teoría y, si es así, qué nos permiten inferir sobre la corrección de la misma. En cualquier estudio social científico las consecuencias de la teoría y de la observación de los hechos han de ser elementos imbricados: las conclusiones de las ciencias sociales no pueden considerarse fiables si no se basan en teorías y datos estrechamente relacionados, que se hayan forjado mediante la formulación y análisis de las consecuencias observables de una teoría.

#### 3.2 La maximización del control

El investigador que busca consecuencias adicionales para una hipótesis persigue uno de los objetivos más importantes de las ciencias sociales: *explicar tanto como sea posible de la forma más escuetu posible*. La ciencia social de calidad pretende aumentar la trascendencia de lo que se explica mediante la información utilizada en la explicación. Si podemos exponer con precisión, a través de una o varias variables causales, lo que en principio

### El diseño de la investigación social

pio parece un complicado efecto, el *control* que tendremos sobre un problema será muy elevado. Al revés, si podemos explicar muchos efectos a partir de una variable o de unas pocas, también dispondremos de un control considerable. En general, el control es escaso en las ciencias sociales y lo es aún más en determinadas áreas temáticas. Esto puede deberse a que los investigadores no saben aún cómo aumentarlo, a que la naturaleza no está organizada adecuadamente o a ambas razones. Las áreas que se estudian habitualmente de forma cualitativa suelen ser aquellas en las que hay poco control. Para explicar cualquier cosa se suele necesitar un conjunto de variables explicativas; utilizamos muchos elementos para explicar poco. En estos casos tendríamos que proponernos diseñar investigaciones con más control.

Hay varias maneras de aumentar el control sobre el problema de nuestra investigación. La principal consiste en incrementar el número de consecuencias observables de nuestra hipótesis y buscar el modo de confirmarlas. Como se ha descrito anteriormente, esta tarea puede suponer, (1) una mejoría de la teoría para que tenga más consecuencias observables, (2) una mejora de los datos para que realmente se detecten más consecuencias y se utilicen para evaluar la teoría y (3) una utilización más provechosa de esos datos, con el fin de extraer más consecuencias de aquellos que estén disponibles. Ninguna de estas posibilidades, ni tampoco el concepto general de maximización del control, son iguales a la idea de concisión, la cual, como explicamos en el apartado 2.2, es un presupuesto relacionado con la naturaleza del mundo más que una norma para diseñar investigaciones.

[La maximización del control es tan importante y tan general que recomendamos fervientemente que los investigadores hagan listas constantes de aquellas posibles consecuencias observables de sus hipótesis que pudieran detectarse en sus datos o en otros. Quizá se puedan comprobar algunas de esas nuevas consecuencias en la matriz de datos brutos original, siempre que la consecuencia no «surja de» los datos y sea una hipótesis que, de manera independiente, indique la teoría u otra matriz de datos. Sin embargo, utilizar otros datos es aún mejor. En este sentido, tendríamos que considerar también consecuencias que pudieran aparecer en nuevos datos —relativos a otras unidades o a otros aspectos de los que estudiamos, datos sobre diferentes grados de agregación y de otros períodos, como las predicciones sobre el futuro próximo— y evaluar las hipótesis en esos ámbitos. Cuantas más pruebas encontremos en diversos contextos, más convincente será nuestra explicación y más confianza podremos tener nosotros y los demás en las conclusiones.]

En principio, algunos investigadores podrían oponerse a la idea de recoger consecuencias observables en cualquier sitio o en niveles de agregación diferentes de aquel para el que se ha diseñado la teoría. Por ejemplo, Lieberson (1985) aplica a la investigación cualitativa la idea estadística de «eficiencia ecológica» —utilizar de forma incorrecta datos agregados para hacer

### 1. La ciencia en las ciencias sociales

inferencias sobre individuos —con el fin de prevenir contra las inferencias que utilizan diferentes niveles de agregación<sup>12</sup>. Estamos de acuerdo en que utilizar datos agregados puede llevarnos a hacer inferencias incorrectas sobre los individuos: cuando son éstos los que nos interesan, suele ser mejor estudiarlos directamente si podemos hacernos con los datos adecuados. Sin embargo, si lo que queremos es que nuestra inferencia plantea una hipótesis con bastante validez, quizás sea mejor que nuestra teoría tenga consecuencias en muchos niveles de análisis diferentes y, a menudo, podemos utilizar datos de todos para dar alguna información sobre ella. De este modo, aunque lo que más nos interese esté en un nivel de análisis agregado, con frecuencia podremos lograr un mayor control sobre la veracidad de nuestra teoría si utilizamos datos de los otros.

[Por ejemplo, si elaboramos una teoría para explicar las revoluciones, tendremos que buscar consecuencias observables de ella no sólo en los resultados más generales sino en fenómenos como las respuestas de los revolucionarios a entrevistas en profundidad, en las reacciones de los que residen en pequeñas comunidades de zonas apartadas del país o en declaraciones oficiales de los cargos de los partidos. Tendríamos que estar dispuestos a incorporar cualquier información que pudiéramos lograr, siempre que nos ayudara a conocer la veracidad de nuestra teoría. Si podemos contrastarla examinando los resultados de las revoluciones, perfecto. Pero, en la mayoría de los casos, existe poca información a ese nivel, quizás una o dos observaciones cuyo valor suele ser ambiguo o estar mal medido. La existencia de una revolución es compatible con muchas teorías diferentes. Sólo al profundizar en el caso que nos ocupa o aportando información relevante sobre otros será posible diferenciar las teorías que ya existían con anterioridad.]

El único problema, a la hora de utilizar información procedente de otros niveles y fuentes para estudiar una teoría que está diseñada con un determinado grado de agregación, reside en dilucidar si tales observaciones contienen *alguna* información relevante para evaluar las consecuencias de nuestra teoría. Si esas nuevas observaciones ayudan a comprobar la teoría, habrá que utilizarlas aunque no sean consecuencias de gran interés. Por ejemplo, quizás no nos importen en absoluto las opiniones de los revolucionarios, pero si sus respuestas a nuestras preguntas son compatibles con nuestra teoría de la revolución, será más posible que ésta sea correcta y la recogida de información adicional no habrá sido inútil. De hecho, una observación en el nivel de análisis más agregado —que tenga lugar una revolución predicha, por ejemplo— no es más que una consecuencia observada de la teoría y, dada la poca información que contiene, no habría que concederle más importancia que a otras consecuencias observables. Tenemos que recabar información sobre tantas consecuencias observables de nuestra teoría como sea posible.

### 3.3 Admitir la incertidumbre

Tanto en la investigación cuantitativa como en la cualitativa el conocimiento y la inferencia son siempre inciertos. La medición en ambos tipos de estudio es propensa al error, pero sus causas pueden ser diferentes. El entrevistador cualitativo, que hace una larga entrevista en profundidad a una persona cuyo contexto ha estudiado, tiene menos posibilidades de calibrar mal la auténtica ideología política de ese sujeto que otro que entreviste de forma estructurada a una persona elegida de forma aleatoria y de la que no sepa nada (aunque lo contrario también es posible si, por ejemplo, el entrevistador se fija excesivamente de un entrevistado que no es fiable). Sin embargo, el encuestador tiene menos que el que investiga en profundidad a extrapolar de forma inapropiada a una población amplia, lo que sólo es aplicable a casos particulares. Ninguno de ellos es inmune a las incertidumbres de la medición o al inherente carácter probabilístico del mundo social.)

Todos los buenos científicos sociales —ya sean cuantitativos o cualitativos— contemplan un margen aproximado de incertidumbre en sus inferencias. En la ciencia política, quizás el problema más importante de los estudios cualitativos sea la constante falta de estimaciones de incertidumbre razonables que se observa en las inferencias de los investigadores (véase King, 1990). (Siguiendo las normas de este libro se puede hacer una inferencia válida casi en cualquier situación, independientemente de lo limitados que sean los datos, pero tendríamos que evitar sacar conclusiones muy ambiciosas a partir de datos poco consistentes.) El problema no es que sea imposible extraer inferencias fiables en la investigación cualitativa, sino que siempre habría que calcular de forma razonable el grado de incertidumbre que existe en cada una de ellas. Neustadt y May (1986, p. 274), al ocuparse de campos en los que es difícil hacer estimaciones cuantitativas precisas, proponen un método útil para animar a los decisores políticos (que a menudo se enfrentan a la necesidad de llegar a conclusiones sobre qué política seguir, utilizando datos inadecuados) a juzgar la incertidumbre de sus conclusiones. Preguntan: «¿Cuánto dinero propio pondrías en ello?». Esta pregunta tiene sentido siempre que también nos planteemos: «¿Con qué posibilidades?».

### 3.4 Pensar como un científico social: escepticismo e hipótesis contrapuestas

La incertidumbre de las inferencias causales tiene como consecuencia que los buenos investigadores sociales no las acepten fácilmente. Cuando se dice que A produce B, alguien que «piense como un científico social» se preguntará si esa relación es auténticamente causal. Es fácil hacer esas preguntas respecto a las investigaciones ajenas, pero resulta más importante

plantearlas en relación a los propios proyectos.) Existen muchas razones para tomarse una explicación causal con escepticismo, por plausible que pueda parecer a primera vista. Leemos en el periódico que los japoneses comen menos carnes rojas y que sufren menos ataques cardíacos que los estadounidenses. Esta observación es interesante por sí misma y, además, la explicación resulta creíble: comer demasiados filetes conduce a la alta tasa de enfermedades cardíacas de los Estados Unidos. El científico social esceptico pregunta por la precisión de los datos (¿cómo conocemos los hábitos alimentarios?, ¿qué muestra se ha utilizado?, ¿se clasifican los ataques al corazón de forma parecida en Japón y en los Estados Unidos, de manera que podamos comparar fenómenos similares?). Suponiendo que los datos sean precisos, ¿qué otros factores podrían explicar los efectos?: ¿hay más variables (otras diferencias dietéticas, rasgos genéticos o de forma de vida) que puedan explicar este resultado? ¿Acaso hemos intercambiado la causa y el efecto sin darnos cuenta? Es difícil imaginarse que el hecho de no sufrir un ataque al corazón pueda hacer que comamos menos carne roja, pero es posible. Quizás la gente pierda, con los años, las ganas de comer hamburguesas y filetes. Si así fuera, los que (por la razón que sea) no han sufrido un ataque al corazón vivirían más y comerían menos carne. Este hecho produciría la misma relación que llevó a los expertos a la conclusión de que la carne es la culpable de los ataques al corazón.

No pretendemos cuestionar esos estudios médicos; simplemente queremos poner un ejemplo de cómo los científicos sociales se enfrentan al problema de la inferencia causal: con escepticismo y teniendo en cuenta las explicaciones alternativas que podrían haberse pasado por alto. De este modo, la inferencia causal se convierte en un *proceso* en el que cada conclusión es una oportunidad de investigar aún más, con el fin de profundizar en esa inferencia y ponerla a prueba. Intentaremos, mediante aproximaciones sucesivas, acercarnos cada vez más a una inferencia causal precisa.

## 2. La inferencia descriptiva

[En las ciencias sociales, la investigación, ya sea cuantitativa o cualitativa, tiene dos objetivos: describir y explicar. Unos investigadores se proponen describir el mundo y otros explicarlo, y ambos propósitos son esenciales. No podemos dar explicaciones causales con sentido sin una buena descripción, y ésta, a su vez, pierde gran parte de su interés si no está vinculada a alguna relación causal. La descripción suele venir primero porque es difícil proponer explicaciones antes de saber algo acerca del mundo y qué ha de explicarse en función de qué características. Sin embargo, la relación entre descripción y explicación es interactiva. A veces nuestras explicaciones nos inducen a buscar descripciones de diferentes partes del mundo y, al contrario, éstas pueden llevarnos a nuevas explicaciones causales.]

Tanto la descripción como la explicación dependen de las reglas de la inferencia científica. En este capítulo nos centramos en la descripción y en la inferencia descriptiva. Describir no es en absoluto algo mecánico o carente de problemas, ya que para hacerlo hay que elegir entre un número infinito de hechos que podrían registrarse. La descripción científica tiene varios aspectos fundamentales. Uno de ellos es que conlleva un proceso inferencial; describir consiste, en parte, en inferir información sobre hechos no observados a partir de aquellos que sí se han contemplado. Otro aspecto es su capacidad para distinguir lo que tienen de sistemático y de no sistemático los hechos observados.

Debe quedar claro que no estamos de acuerdo con los que denigran la «simple» descripción. Aunque explicar —relacionar causas y efectos— sea

el objetivo último, describir resulta determinante en toda explicación y es una actividad fundamental en sí misma. Lo que distingue los estudios científicos de otros tipos de investigación no es el enfrentamiento entre descripción y explicación, sino el hecho de que se hagan inferencias sistemáticas según procedimientos válidos. La inferencia, ya sea descriptiva o causal, cuantitativa o cualitativa, es el objetivo último de toda ciencia social de calidad. La recogida sistemática de hechos es una empresa muy importante sin la que la ciencia no sería posible, pero no es ciencia en sí misma. El trabajo de archivo o el resumen de hechos históricos pueden conformar, si son correctos, una buena historia descriptiva, pero ninguna de estas actividades es suficiente para constituir ciencias sociales.

En este capítulo, distinguimos entre descripción —recogida de datos— e inferencia descriptiva. En el apartado 1 analizamos la relación entre los objetivos aparentemente contradictorios de la investigación académica: descubrir conocimientos generales y conocer hechos particulares. De este modo, en el apartado 2, podemos explicar con más detalle el concepto de inferencia. En el resto del libro nuestra intención es presentar ideas a través de la expresión verbal y también mediante modelos de investigación algebraicos muy simples. En el apartado 3 entramos a considerar la naturaleza de esos modelos. Posteriormente, se analizan los relativos a la recogida de datos, al resumen de datos históricos y a la extracción de inferencias descriptivas en los apartados 4, 5 y 6, respectivamente. Para terminar, en el apartado 7, proporcionamos algunos criterios concretos para evaluar inferencias descriptivas.

## 1. El conocimiento general y los hechos particulares

*unir las reglas*

El mundo que estudian los científicos sociales se compone de elementos particulares: votantes concretos o entidades gubernamentales específicas, así como ciudades, tribus, grupos, estados, provincias y naciones determinadas. La ciencia social de calidad pretende ir más allá de estos pormenores para alcanzar un conocimiento más general. Sin embargo, la generalización no quita importancia al detalle. De hecho, el único objetivo que persigue ir de lo particular a lo general es aumentar nuestro conocimiento de ambas cosas. Los integrantes específicos del mundo social —o, más concretamente, los hechos concretos de esos integrantes— son la base en la que deben descansar las generalizaciones. Además, casi siempre aprendemos más de un caso específico cuando estudiamos conclusiones más generales. Si quisieramos saber por qué ha dimитido el ministro de Asuntos Exteriores de Brasil, sería útil saber por qué lo han hecho otros ministros brasileños, por qué lo hacen los mismos cargos en otros países o por qué las personas, en general, dejan sus puestos en los gobiernos o incluso en otros ámbitos laborales. Cada una de estas razones nos ayudará a comprender diferentes

tipos de hechos generales y de principios de comportamiento humano, y no dejarán de ser importantes aunque nuestro único objetivo sea comprender por qué ha dimítido el último ministro de Asuntos Exteriores brasileño. Por ejemplo, estudiando a otros ministros podríamos saber que todo el gobierno brasileño ha dimítido para protestar por las acciones del presidente, algo de lo que no nos habríamos dado cuenta si sólo hubiéramos examinado lo que hacía el de Exteriores.

En las ciencias sociales algunas investigaciones pretenden informar sobre un tipo de acontecimientos específico sin decir nada en particular sobre un hecho o unidad concreta. Los estudios de comportamiento electoral que utilizan encuestas masivas explican las decisiones del conjunto de los votantes, no el sufragio de un individuo concreto. Los estudios de las finanzas del Congreso estadounidense explican la influencia del dinero en los resultados electorales de todas las circunscripciones. La mayoría no mencionará la séptima circunscripción de Pensilvania ni ninguna otra a no ser de pasada o como una excepción a la regla general. Estos estudios siguen la pauta de Przeworski y Teune (1982): eliminar los nombres propios. Sin embargo, aunque no pretendan comprender el funcionamiento de ningún distrito en particular, no tendrían que pasar por alto —como, por desgracia, se hace de vez en cuando en esta tradición— que los hechos relativos a los diversos distritos que se incluyen en el análisis tienen que ser precisos.

*en la otra parte*

Hay otras investigaciones que intentan decirnos algo acerca de un ejemplo concreto. Se centran en la Revolución francesa o en otro acontecimiento «importante» con la intención de explicar cómo o por qué sucedió. Los estudios de esta tradición serían impensables —y realmente carecerían de interés para la mayoría de sus lectores habituales— sin nombres propios. Un político puede tratar con acierto las pautas de relación en el conjunto de las campañas para el Congreso estadounidense sin ocuparse de circunscripciones o candidatos concretos, pero imaginémonos el análisis que hizo Robert Caro (1983) de las elecciones tejanas de 1948 al Senado sin Lyndon Johnson y Coke Stevenson<sup>1</sup>. Acontecimientos concretos como la Revolución francesa o las primarias demócratas de 1948 para el Senado en Texas, pueden tener un interés intrínseco: suscitan nuestra curiosidad y, si fueran condiciones previas para que tuvieran lugar acontecimientos posteriores (como las guerras napoleónicas para la presidencia de Johnson), quizás fuera preciso conocerlas para comprender dichos acontecimientos. Además, tener un conocimiento general sobre las revoluciones, rebeliones o guerras civiles nos proporcionará información de valor incalculable para centrar mejor nuestro estudio sobre las causas concretas de la Revolución francesa.

Vamos a abordar estos asuntos analizando la «interpretación», que se presenta como alternativa a la inferencia científica (apartado 1.1), los conceptos de singularidad y complejidad del objeto de estudio (apartado 1.2) y el área general de los estudios de caso comparados (apartado 1.3).

## 1.1 «Interpretación» e inferencia

*+ interpretación  
+ contexto  
+ significado  
de los  
acontecimientos*

En las ciencias humanas hay historiadores y antropólogos que afirman que lo único que buscan son conocimientos concretos a través de lo que denominan «interpretación». Los interpretacionistas tratan de encontrar síntesis precisas de los pormenores históricos y situar los acontecimientos que describen en un contexto inteligible dentro del que pueda explicarse el significado de las acciones. Como ha escrito Ferejohn (en Goldstein y Keohane, 1993, p. 228): «Queremos que las teorías de las ciencias sociales proporcionen explicaciones causales de los acontecimientos [...] y que expliquen las razones de la acción social o su significado. No sólo queremos saber lo que llevó al agente a realizar un acto determinado, sino sus razones para hacerlo». Geertz (1973, p. 17) también escribe que «... no nos interesa quitarle al comportamiento humano, antes de comenzar a examinarlo, precisamente las propiedades que nos atraen».

*Walter*

*ej.: Enca  
partid. del  
migrante del  
lugar*

Los investigadores que hacen hincapié en la «interpretación» pretenden arrojar luz sobre los aspectos deliberados del comportamiento humano utilizando el concepto de *Verstehen* («empatía: comprender el significado de las acciones e interacciones a partir del punto de vista de los que participan en ellas») [Eckstein, 1975, p. 81]). Los interpretacionistas quieren explicar las razones de las acciones intencionadas relacionándolas con todos los conceptos y prácticas en los que están inmersas. También emplean normas de evaluación: «Las normas más evidentes son la coherencia y el campo de aplicación; una explicación interpretativa tiene que dar la máxima coherencia o inteligibilidad a un conjunto de prácticas sociales y, si lo que explica es un conjunto de prácticas determinado, tiene que ser compatible con otras prácticas o tradiciones de la sociedad» (Moon, 1975, p. 173).)

Para los interpretacionistas, quizá la recomendación operativa más importante sea que los investigadores aprendan mucho sobre una cultura antes de elaborar preguntas de investigación, ya que sólo mediante una profunda inmersión cultural y un intenso conocimiento de la materia se pueden plantear preguntas correctas y formular hipótesis útiles. Duneier (1993), por ejemplo, estudió el comportamiento colectivo de los hombres de clase trabajadora negros y blancos en un bar autoservicio no segregado de Chicago. Al sumergirse en esta cultura local durante cuatro años, le sorprendieron ciertas cosas que antes no se le habían ocurrido. Por ejemplo, observó que, aunque estos hombres estaban muy en contra del Partido Republicano, su actitud ante muchos problemas era socialmente conservadora.

Algunos investigadores llevan aún más lejos el papel de la interpretación, llegando a indicar que, para las ciencias sociales, constituye un paradigma de investigación completamente diferente, «... no una ciencia experimental en busca de leyes sino otra de tipo interpretativo en busca de significado» (Geertz, 1973, p. 5). Sin embargo, para nosotros, la ciencia (tal como la hemos definido en el apartado 1.2 del capítulo 1) y la interpre-

## 2. La inferencia descriptiva

*interventivas*

tación no son empresas fundamentalmente diferentes que persigan objetivos distintos. Ambas dependen de la elaboración de cuidadosas descripciones, de una comprensión profunda del mundo, de que se planteen buenas preguntas, de la formulación de hipótesis falsables a partir de teorías más generales y de la recogida de las pruebas necesarias para evaluar tales hipótesis. La aportación primordial de la ciencia es un conjunto de procedimientos con los que se pueden dar respuestas a preguntas descriptivas y causales elaboradas apropiadamente.

*preguntas  
observación  
y trabajo*

Al insistir en la metodología inferencial no pretendemos menoscabar la importancia del proceso de formulación de preguntas fructíferas. Por el contrario, estamos de acuerdo con los interpretacionistas en que éste resulta crucial para entender en profundidad una cultura, antes de formular hipótesis o de diseñar un proyecto de investigación sistemático que encuentre una respuesta. Sólo queremos añadir que no se puede evaluar la veracidad de afirmaciones basadas en métodos como el de la observación participante si no se utiliza la lógica de la inferencia científica que nosotros describimos. Encontrar respuestas adecuadas para preguntas erróneas es un ejercicio pueril. La interpretación que parte del concepto de *Verstehen* suele ser fuente de penetrantes hipótesis. Por ejemplo, las detalladas observaciones que hizo Richard Fenno del Congreso estadounidense (Fenno, 1978) a través de lo que él denomina «empaparse y hurgar» han hecho grandes aportaciones al estudio de esta institución, principalmente porque han ayudado a plantear mejores preguntas de investigación. En su estudio de las regiones italianas, Putnam (1993, p. 12) afirma que para practicar el método de «empaparse y hurgar» «... es necesario que el investigador se sumerja en los pequeños detalles de una institución, con el fin de participar, como los que viven en ella a diario, en sus costumbres y prácticas, en sus éxitos y fracasos. Esta inmersión agudiza nuestras intuiciones y ofrece innumerables pistas sobre cómo se articula la institución y de qué manera se adapta a su medio». Cualquier definición de ciencia que no deje lugar a las ideas relativas a la elaboración de hipótesis será tan frívola como una interpretación a la que no le preocupe descubrir la verdad.

*Ejemplos  
de los  
que no  
observan  
a lo que  
nosotros  
nos  
relacionan  
con el  
observar  
a los que  
nos  
observan*

Sin embargo, una vez que las hipótesis se han formulado, para demostrar que son correctas (dejando un margen a la incertidumbre) se precisan inferencias científicas válidas. Además, los procedimientos inferenciales de los científicos sociales interpretacionistas deben respetar las mismas reglas que los de otros investigadores cualitativos y cuantitativos. Esto significa que, aunque estamos de acuerdo en que las ciencias sociales de calidad necesitan interpretaciones penetrantes u otros métodos para generar buenas hipótesis, también hay que subrayar que la ciencia es esencial para que la interpretación sea precisa. Si pudieramos entender el comportamiento humano únicamente a través del concepto de *Verstehen*, nunca podríamos falsificar nuestras hipótesis descriptivas o aportar pruebas de ellas que no formaran parte de nuestra propia experiencia. En consecuencia, las conclusiones nun-

ca dejarían de ser hipótesis no comprobadas y las interpretaciones tendrían un carácter personal y no científico.

Uno de los mejores y más famosos ejemplos de la tradición interpretativa es el análisis que hizo Clifford Geertz del estudio de Gilbert Ryle sobre las diferencias entre un tic nervioso y un guiño. Geertz (1973, p. 6) escribió:

Pensemos en [...] dos chicos que contraen rápidamente los párpados de sus respectivos ojos derechos. En uno de ellos es un tic involuntario; en el otro, una señal de tipo conspirativo para un amigo. Los dos movimientos son, como tales movimientos, idénticos. Si se realizara una observación «fenomenística» de estos movimientos aislados, como la haría una cámara fotográfica, no se podría distinguir entre el tic y el guiño o si, en realidad, ambos han sido una cosa u otra. Sin embargo, la diferencia que existe entre un tic nervioso y un guiño es enorme, aunque no sea fotografiable, como sabe cualquiera que haya tenido la desgracia de que confundieran la primera con la segunda. El que guiña se está comunicando de una forma precisa y determinada: (1) deliberadamente, (2) con alguien en concreto, (3) para transmitir un mensaje concreto, (4) según un código social establecido y (5) sin que lo note el resto de los presentes. Como señala Ryle, el que guinya hace dos cosas: contraer el párpado y guñar, mientras que el que tiene un tic sólo hace una: contraer el párpado. Cuando existe un código público por el que contraer los párpados se considera una señal de tipo conspirativo, hacer este movimiento es guñar.

Geertz está señalando un importante problema conceptual. Sin el concepto de «guiño», al que da sentido una teoría de la comunicación, el estudio más preciso de «las contracciones de los párpados en los seres humanos» carecería de sentido para los que estudian las relaciones sociales. En este ejemplo, la teoría, que surgió de meses de «empaparse y hurgar» y de minuciosos estudios culturales, resulta esencial para la pregunta básica de si las contracciones de los párpados pueden ser siquiera «tics» o «guiños». Este ejemplo indica claramente la gran importancia que tiene la interpretación, ya que proporciona nuevas formas de mirar el mundo (nuevos conceptos que considerar e hipótesis que evaluar). Sin una profunda inmersión en una determinada circunstancia quizá ni siquiera pudieramos pensar en qué teorías hay que sopesar. En el ejemplo anterior, si no pensáramos en la diferencia entre tics y guiños, todo estaría perdido. Si la interpretación —o cualquier otra cosa— nos ayuda a abordar nuevos conceptos o hipótesis, entonces su utilidad resulta incuestionable, lo cual se ha confirmado una y otra vez no sólo en relación a este método sino a otras formas similares de pormenorizada comprensión cultural.<sup>1</sup>

El investigador, una vez que ha hecho una distinción teórica relevante, como la que afecta al guiño y al tic, necesita *evaluar* la hipótesis que indica que están teniendo lugar guiños. Es en esta evaluación donde la lógica de la inferencia científica no tiene rival. En otras palabras, la mejor manera de determinar el significado de las contracciones de los párpados es la que se basa en los métodos sistemáticos descritos en este libro. Si lo crucial fuera

distinguir entre un tic nervioso y un guiño, resultaría fácil diseñar un procedimiento de investigación para hacerlo. Si, por ejemplo, creemos que ciertas contracciones de los párpados son guiños dotados de un significado político, será posible observar entonces otras circunstancias similares, ya que un elaborado mecanismo de señalización como éste (un «código público») probablemente se utilizará de nuevo una vez desarrollado. A la vista de esta posibilidad, podríamos registrar las ocasiones en las que se contrae el párpado de este actor, observar si el otro actor principal le mira en el momento adecuado y si el primero le responde. Podríamos incluso diseñar una serie de experimentos para ver si los individuos de esta cultura están acostumbrados a comunicarse de esta manera. Comprender la cultura, describir cuidadosamente el acontecimiento y estar muy familiarizados con situaciones similares nos ayudará a plantear las preguntas correctas y nos dará incluso más confianza en nuestras conclusiones. Pero sólo con los métodos de la inferencia científica podremos evaluar la hipótesis y determinar si es o no correcta.

La interpretación que da Geertz al guiño se expresa mejor como una hipótesis causal (que definimos precisamente en el apartado 1 del capítulo 3): el hipotético efecto causal del guiño sobre el otro actor político es la respuesta de éste, dada la contracción del párpado menos su reacción a falta de otro movimiento (y de otros cambios). Si la contracción del párpado fuera un guiño, el efecto causal sería positivo; si sólo hubiera un tic nervioso, el efecto sería igual a cero. Si decidieramos hacer una estimación de este efecto causal (y, por tanto, averiguar si se ha producido un guiño o un tic), para extraer la mejor inferencia respecto a la interpretación del comportamiento observado, habría que entender todos los problemas inferenciales que se analizan en profundidad en el resto del libro.

Si lo que interpretamos como guiños fueran en realidad tics nerviosos involuntarios, nuestros intentos de extraer inferencias causales de las contracciones de los párpados, utilizando una teoría de la interacción social voluntaria, fracasarian sistemáticamente: no podríamos generalizar, y lo sabríamos<sup>2</sup>.

No es probable que un diseño de investigación que se proponga distinguir entre tics y guiños sea importante para la mayoría de las investigaciones de las ciencias políticas, pero el mismo problema metodológico aparece en gran parte de las áreas de trabajo de los polítólogos. Con frecuencia se nos pide que interpretemos el significado de una acción. En política exterior, los que toman las decisiones se envían mensajes unos a otros. Ante un determinado mensaje, cabe preguntarse si es una amenaza, un punto de negociación o una afirmación destinada a atraerse al público del propio país. Conocer las normas culturales, las convenciones de las comunicaciones internacionales y la historia de cada actor nos ayudará a hacer tal interpretación; del mismo modo que nos será útil observar de cerca otros rasgos secundarios de la comunicación. Pensemos también en el siguiente problema

### El diseño de la investigación social

cuantitativo: en los Estados Unidos los electores parecen estar enviando un mensaje al no acudir a las urnas, pero ¿qué significa esta escasa participación? ¿Acaso refleja una alienación respecto al sistema político?, ¿representa un cálculo de los costes y beneficios de votar en el que los primeros son mayores que los segundos?, ¿muestra una decepción respecto a los últimos candidatos o campañas?, ¿podría ser consecuencia de un cambio en la edad mínima para votar?, ¿o una señal de que nada inquieta lo suficiente a los votantes como para llevarlos a las urnas? El hecho de que un ciudadano decida no votar, al igual que un guiño o un mensaje diplomático, puede significar muchas cosas. Un investigador inteligente siempre tendrá que esforzarse por plantear las preguntas adecuadas y después para diseñar cuidadosamente un estudio científico con el que averiguar qué significaba en realidad una determinada acción ambigua.

Interpretación  
descripción  
explicación  
También nos gustaría ocuparnos brevemente de las afirmaciones extremas que hacen unos pocos partidarios de la interpretación cuando señalan que los objetivos de algunas investigaciones deberían ser sentimientos y significados sin consecuencias observables. En justicia, esta caracterización no afecta más que a una reducida minoría de los investigadores de esta tendencia, pero sus afirmaciones son lo suficientemente efusivas como para que valga la pena darles una contestación explícita. Al igual que las opiniones excesivamente entusiastas de los primeros positivistas, que mantenían la insostenible postura de que no había sitio en la investigación científica para los conceptos no observables, estos argumentos resultan inapropiados para la investigación empírica. Psathas (1968, p. 510), por ejemplo, señala que:

cualquier comportamiento que sólo se centre en lo evidente y en lo que se manifiesta en actos concretos, directamente observables, es, como mínimo, ingenuo. Por lo tanto, el desafío para el científico que pretende comprender la realidad social es captar el significado que, para el actor, tiene su propia acción.

Puede que Psathas tenga razón al señalar que los científicos sociales que sólo se centran en comportamientos evidentes u observables se están perdiendo mucho, pero ¿cómo habriamos de saber esto si no podemos ver más? Por ejemplo, si dos teorías sobre el concepto que uno tiene de sí mismo tienen idénticas manifestaciones observables, ningún observador tendrá información suficiente para distinguirlas. Esto es así para cualquier observador, independientemente de lo inteligente o sensible a factores culturales que sea, de sus habilidades para la interpretación, de lo bien que «clasifique» sus propias suposiciones o del empeño que ponga en su empresa. La interpretación, el presentimiento, la descripción minuciosa, la observación participante y la que no lo es, la entrevista en profundidad, la empatía, la cuantificación y el análisis estadístico, al igual que los demás procedimientos y métodos, no sirven para distinguir entre dos teorías que no tengan conse-

### 2. La inferencia descriptiva

cuencias observables diferentes. Por otra parte, si las hay, los métodos que describimos en este libro proporcionan formas de distinguirlas.

En la práctica, los etnógrafos (y todos los buenos científicos sociales) si que buscan comportamientos observables con el fin de establecer diferencias entre las teorías. Puede que se sumerjan en la cultura, pero todos se basan en diversas formas de observación, y cualquier nueva apreciación del contexto cultural procede directamente de ellas o de otros medios comparables. No siempre resulta fácil identificar observaciones relevantes. Por el contrario, encontrar las apropiadas quizá sea la parte más difícil de un proyecto de investigación, especialmente (y necesariamente) en aquellas áreas de estudio tradicionalmente dominadas por la investigación cualitativa.

#### 1.2 «Singularidad», complejidad y simplificación

Algunos investigadores cualitativos rechazarían la idea de que un conocimiento general sea necesario o útil (ni siquiera posible) como punto de partida para comprender un determinado acontecimiento. Lo que sostienen es que los hechos o unidades que estudian son «únicos», y, en cierto sentido, tienen razón. Sólo hubo una Revolución francesa, sólo hay una Tailandia y nadie que haya leído las biografías pertinentes o que viviera los años sesenta puede questionar el hecho de que sólo hubo un Lyndon B. Johnson. Pero aún van más lejos. La explicación, según su punto de vista, se limita a ese único acontecimiento o unidad: no al porqué de las revoluciones sino al de la Revolución francesa; no a por qué en ocasiones la democratización parece languidecer, sino a por qué ocurre así en Tailandia; no a por qué ganan los candidatos sino a por qué ganó Johnson en 1948 y 1964. Los investigadores de esta tradición creen que perderían su capacidad de explicar lo específico si intentaran ocuparse de lo general: de revoluciones, democratización o primarias del Senado estadounidense.

Sin embargo, el término «singularidad» es engañoso. La Revolución francesa, Tailandia y Johnson son realmente únicos. En cierto sentido, lo son todos los fenómenos, todos los acontecimientos; pero también lo fueron las elecciones al Congreso en la séptima circunscripción de Pensilvania en 1988, así como la opción que tomó cada uno de los millones de electores que votaron en los comicios presidenciales de ese mismo año. Considerado de forma global, cada aspecto de la realidad social es infinitamente complejo y se relaciona de alguna manera con acontecimientos naturales y sociológicos anteriores. (Por lo tanto, la singularidad de las cosas forma parte de la condición humana: no distingue entre las situaciones que son susceptibles de generalización científica y aquellas en las que no es posible generalizar. De hecho, como señalamos al analizar las teorías sobre la extinción de los dinosaurios en el capítulo 1, incluso los acontecimientos únicos pueden estudiarse de forma científica si se presta atención a las con-

secuencias observables de las teorías que se han desarrollado para explicarlos.

*(simplificar)* «Lo que realmente plantea la singularidad es el problema de la complejidad. La cuestión no es si los acontecimientos son de por sí únicos, sino el hecho de que sea posible o no extraer de un amasijo de acontecimientos las características principales de la realidad social que queremos comprender.

Wolfe  
(simplificar)  
también  
de acuerdo

Una de las primeras, y más difíciles, tareas de la investigación social es la de simplificar, la cual nos expone a ser criticados por practicarla en exceso y por omitir aspectos significativos de una situación. No obstante, ningún investigador puede evitar la simplificación, que ha sido parte integrante de todas las obras académicas conocidas —cuantitativas y cualitativas, antropológicas y económicas, de las ciencias sociales y de las naturales o físicas— y que probablemente siempre lo será. Ni siquiera la descripción más exhaustiva que hagan los mejores intérpretes de la cultura desde el conocimiento profundo de un determinado contexto podrá dejar de simplificar, objetualizar y reducir de forma drástica la realidad observada. De hecho, la diferencia que hay entre el grado de complejidad del mundo y el de la descripción más densa sigue siendo mucho más grande que la que existe entre dicha descripción y el análisis cuantitativo o formal más abstracto. Ninguna descripción —cualquiera que sea su densidad— y ninguna explicación —independientemente del número de factores explicativos que contempla— pueden llegar a captar, siquiera levemente, toda la «florecente y agitada» realidad del mundo. No queda más remedio que simplificar. La simplificación sistemática es un paso crucial hacia el conocimiento útil. Como ha señalado un historiador económico, si la insistencia en la singularidad «se lleva hasta el extremo de hacer caso omiso de todas las regularidades, se negará a las ciencias sociales la posibilidad misma de que existan y los historiadores no serán más que baladistas sin objetivos» (Jones, 1981, p. 160).

El analista tiene que simplificar sus descripciones, siempre que sea posible, sólo después de comprender la riqueza histórica y cultural. Puede que los científicos sociales, para establecer diferencias, utilicen únicamente unos pocos elementos de la historia de ciertos grupos de acontecimientos. No obstante, un conocimiento profundo y no estructurado del contexto histórico y cultural de los fenómenos que quieren tratar de forma sintética y científica suele ser indispensable para evitar simplificaciones claramente equivocadas. Poco se podría confiar en las generalizaciones de un científico social que estudiara las revoluciones o las elecciones al Senado estadounidense si no supiera mucho sobre la Revolución francesa o las elecciones de 1948 en Texas y no le importara su ignorancia.

En resumen, creemos que, siempre que sea posible, la investigación social tiene que ser general y específica al mismo tiempo: debe informarnos sobre los tipos de acontecimientos y también sobre hechos específicos en lugares concretos. Queremos aislarnos del tiempo sin perder el vínculo con él, y la insistencia en uno de estos objetivos puede variar de una investiga-

ción a otra, aunque es probable que ambos estén presentes. Además, más que opuestos son complementarios. De hecho, puede que la mejor manera de entender un determinado acontecimiento sea utilizar también los métodos de la inferencia científica para estudiar pautas sistemáticas en acontecimientos paralelos similares.

→ igual de lo de la literatura básica,  
el resto del resto sobre el resto.

### 1.3 Estudios de caso comparados

(U) En gran medida, lo que hacen los polítólogos es describir de forma sistemática acontecimientos que son importantes desde el punto de vista político. A la gente le interesa la caída de la Unión Soviética, la reacción popular en los países árabes ante la guerra autorizada por la ONU para expulsar a Irak de Kuwait y los resultados de las últimas elecciones al Congreso de los Estados Unidos. Confía en que los polítólogos describan con un conocimiento más profundo que el de las informaciones periódísticas la relación que existe entre estos y otros acontecimientos relevantes, ya sean contemporáneos o históricos. Al describir los acontecimientos tenemos que ser lo más precisos y sistemáticos que sea posible. Esto supone que si logramos encontrar medidas cuantitativas válidas para lo que queremos saber, tenemos que utilizarlas: ¿qué proporción de los periódicos soviéticos critica las políticas gubernamentales?, ¿qué ponen de manifiesto las encuestas de opinión llevadas a cabo en Jordania y Egipto sobre la actitud de estos países hacia la Guerra del Golfo?, ¿qué porcentaje de congresistas estadounidenses fue reelegido?

(U) Aunque quantificar produzca precisión, no favorece necesariamente la exactitud, ya que inventar índices cuantitativos que no se relacionen directamente con los conceptos o acontecimientos que pretendemos medir puede conducir a graves errores de medida y a problemas en las inferencias causales (véase el apartado I del capítulo 5). Del mismo modo, existen métodos más y menos precisos para describir acontecimientos no cuantificables. Hay disciplinados cualitativistas que intentan analizar con cuidado las constituciones y leyes, en vez de informar únicamente de lo que los observadores dicen sobre ellas. Los investigadores, al realizar estudios de caso sobre políticas gubernamentales, plantean a los entrevistados preguntas agudas y bien definidas, cuyas respuestas serán relativamente claras, y sondan de forma sistemática todo comentario inesperado que pueda sugerir hipótesis relevantes. Los estudios de caso son esenciales para la descripción y, por tanto, para las ciencias sociales. No tiene sentido intentar explicar lo que antes no se ha descrito con un razonable grado de precisión.

Describir con perspicacia acontecimientos complejos no es algo trivial. En áreas como la política comparada o las relaciones internacionales la labor de descripción es especialmente importante porque todavía hay muchas cosas que necesitamos saber; nuestra capacidad explicativa es débil y una

## El diseño de la investigación social

buenas descripciones dependen en parte de una buena explicación. Algunas de las razones que explican esta necesidad de conocer y la debilidad de nuestra capacidad explicativa son las mismas: en las relaciones internacionales, por ejemplo, las pautas de poder, de alineamiento y de interdependencia han cambiado mucho últimamente, por lo que, al alterarse el contexto sistémico en el que tienen lugar las relaciones que se observan entre los estados, cada vez resulta más necesario describir correctamente las nuevas situaciones. La causalidad suele ser difícil de determinar porque tanto los estados como otros actores pretenden prever y contrarrestar las acciones de los demás y, a la hora de explicar el comportamiento de dichos estados, las expectativas pueden tener un papel tan importante como las acciones que se observan. Pretender explicar algún aspecto de las relaciones internacionales presuponiendo que no hay interacciones estratégicas ni previsión de reacciones será mucho menos útil que describir con cuidado acontecimientos que tenemos razones para creer que son importantes y que están interconectados. Una buena descripción es mejor que una mala explicación.

Una de las ventajas del estudio de caso en profundidad, que suele pasar por alto, es que desarrollar una buena hipótesis causal complementa una buena descripción, no rivaliza con ella. Encuadrar un estudio de este tipo en una pregunta explicativa puede producir una descripción más centrada y relevante, aunque al final el estudio no logre extraer una sola inferencia causal válida.

Nosotros afirmamos que los estudios de caso comparados pueden generar inferencias causales válidas cuando se utilizan los procedimientos descritos en el resto del libro, aunque, tal como se practican en la actualidad, no suelen cumplir las normas de una inferencia válida (que explicaremos en el capítulo 3). De hecho, gran parte de lo que los científicos sociales con enfoque histórico o interpretativo denominan trabajo «explicativo» sigue siendo fundamentalmente descriptivo porque no cumple esas normas de aplicación universal. Desde esta perspectiva, resulta fundamental lo que señalan muchos investigadores cuando aconsejan que, al describir o explicar, los estudios de caso comparados sean más sistemáticos.

Alexander George, por ejemplo, recomienda un método de «comparación estructurada y centrada» que hace hincapié en la disciplina cuando se trata de recoger datos (George y McKeown, 1985; véase también Verba, 1967). George y sus colaboradores subrayan la necesidad de que haya una recogida sistemática de la misma información —de las mismas variables— en unidades seleccionadas cuidadosamente. También insisten en que, para que finalmente sea posible la inferencia causal, son necesarias directrices teóricas —plantearse preguntas explicativas cuidadosamente elaboradas— con el fin de describir de forma sistemática<sup>1</sup>.

El método de la comparación estructurada y centrada es una forma sistemática de emplear lo que George y McKeown denominan el «procedimiento de la coherencia». A través de este sistema, el investigador «define y es-

## 2. La inferencia descriptiva

tandariza las necesidades de los datos de los estudios de caso [...] formulando preguntas relevantes desde el punto de vista teórico con el fin de dar pautas para el análisis de cada caso» (George y McKeown, 1985, p. 41). El problema que señalan George y McKeown está bien traído: «Para hacer una comparación controlada mediante un tamaño muestral pequeño hay que seguir un procedimiento sistemático de recogida de datos». Esa «comparación estructurada y centrada» precisa datos de las mismas variables en diferentes unidades. Por lo tanto, no es un método diferente al que subrayamos aquí, ya que constituye una forma de sistematizar la información en estudios de caso descriptivos de manera que sea factible utilizarla para hacer inferencias descriptivas o causales. Muchos de los valiosos consejos que, como éste, se dan para hacer estudios de caso comparados son elementales, pero se suelen pasar por alto.

### 2. La inferencia: el fin científico de la recogida de datos

La inferencia es un proceso en el que se utilizan hechos que conocemos para aprender sobre los que desconocemos. Los hechos conocidos se someten a las preguntas, teorías e hipótesis de nuestra investigación y conforman nuestros datos y observaciones de tipo cuantitativo o cualitativo.

Cuando buscamos un conocimiento general, en sí mismo o para comprender mejor determinados hechos, es preciso evitar de alguna manera que nos apabulle el enorme mare magnum de observaciones potenciales y reales que hay sobre el mundo. Por fortuna, la solución a este problema reside precisamente en la búsqueda del conocimiento general. Esto quiere decir que la mejor manera científica de organizar los hechos es hacer que sean consecuencias observables de alguna teoría o hipótesis. La simplificación científica entraña la provechosa elección de una teoría (o hipótesis) con el fin de evaluarla; después, la teoría nos lleva a seleccionar los hechos que son consecuencia de ella. Organizarlos en función de las consecuencias observables de una determinada teoría tiene importantes y beneficiosos resultados para el diseño y la realización de la investigación. En primer lugar, con este criterio de selección de los hechos podemos darnos cuenta rápidamente de que si existen más observaciones de las consecuencias de una teoría no podemos más que beneficiarnos de ello al evaluar la teoría en cuestión. Como más información de este tipo no puede ser perjudicial, esos datos nunca se desechan, y así mejora el proceso investigador.

En segundo lugar, no hay por qué poseer una teoría completa antes de recoger los datos, ni tampoco nuestra teoría tiene por qué mantenerse inalterable durante el proceso. La teoría y los datos interactúan. Al igual que ocurre con la gallina y el huevo, siempre es necesaria alguna teoría antes de recoger los datos y también se precisan algunos datos antes de teorizar. Los manuales de investigación nos dicen que utilicemos los datos para compro-

bar nuestras teorías, pero aprender de ellos puede ser un objetivo tan importante como evaluar teorías e hipótesis previas. Para que se produzca este aprendizaje hay que reorganizar los datos como consecuencias observables de la nueva teoría. Esta reorganización es muy habitual al principio de muchas investigaciones, generalmente después de haber recogido algunos datos preliminares; después, prosigue la recogida de datos para evaluar la nueva teoría. Siempre se debe intentar seguir recogiendo datos, incluso después de la reorganización, para contrastar la nueva teoría y evitar así utilizar el mismo material para evaluar ésta y aquella de la que partió<sup>4</sup>.

En tercer lugar, insistir en que se recojan datos que sean consecuencias observables de una hipótesis deja mucho más claro el territorio común que comparten los estilos de investigación cuantitativo y cualitativo. De hecho, una vez que dejamos de pensar en los casos, unidades o registros de la forma generalmente estrecha, e incluso ingenua, que es habitual, nos damos cuenta de que la mayoría de los estudios cualitativos podrían aportar una gran cantidad de consecuencias observables a las teorías que estamos evaluando; sin embargo, muchas de ellas pueden pasarse desapercibidas al investigador. En consecuencia, organizar los datos mediante una lista de las observaciones específicas de una teoría ayuda a poner de manifiesto el fin científico esencial de gran parte de la investigación cualitativa. En cierto sentido, estamos pidiéndole al investigador que estudie un acontecimiento concreto —quizá una decisión gubernamental específica— que se pregunta: «Si mi explicación de por qué la decisión resultó de una determinada manera es correcta, ¿que más podría observar en el mundo real?». Se pueden encontrar consecuencias observables adicionales en otras decisiones, pero también en otros aspectos de la que se estudia, por ejemplo: cuándo se tomó, cómo, o de qué manera se justificó. La máxima crucial que ha de guiar tanto la creación de una teoría como la recogida de datos es la búsqueda de más consecuencias observables de esa teoría.

Cada vez que desarrollamos una nueva teoría o hipótesis resulta provechoso hacer una lista de todas las consecuencias de la misma que, en principio, pudieran observarse. La lista, que podría después limitarse a los elementos sobre los que se han recogido datos —o pudiera fácilmente hacerse—, constituye a continuación la guía básica operativa para el proyecto de estudio. Si recoger un dato más nos ayuda a considerar otra manera de evaluar la teoría, merece la pena hacerlo (siempre sujetos a los condicionantes de tiempo, dinero y esfuerzo). Cuando una entrevista u otro tipo de observación puede ser de interés, pero no es una consecuencia observable potencial de esta teoría (o de otra que sea relevante), entonces tiene que ser evidente que no nos ayudará a evaluarla.

Dentro del proceso de simplificación que se desarrolla al organizar nuestros datos como consecuencias observables de la teoría, necesitamos sistematizar dichos datos. Podemos pensar en convertir los fenómenos brutos del mundo real en «clases» que estén compuestas por «unidades» o en «casos»

que, a su vez, se compongan de «atributos», «variables» o «parámetros». La clase podría ser de «votantes», las unidades, una muestra de esos «votantes» en diversas circunscripciones para las elecciones al Congreso estadounidense; y los atributos o variables podrían ser la renta, la identificación partidista o cualquier otra cosa que represente una consecuencia observable de la teoría que se evalúa. La clase también podría ser un determinado colectivo, como las comunidades o países; las unidades podrían ser una selección de éstos; y los atributos o variables podrían ser el tamaño de los mismos, el tipo de gobierno, su situación económica, composición étnica o cualquier otro elemento cuantificable y de interés para el investigador. Estos conceptos, así como elaboraciones teóricas como las tipologías, estructuras y todo tipo de clasificaciones, son útiles mecanismos temporales cuando estamos recogiendo datos, pero no plantean una hipótesis clara que evaluar. En general, no animamos a los investigadores a que organicen sus datos de esta manera, ya que sólo necesitamos el concepto organizador inherente a nuestra teoría. Es decir, si las observaciones no son consecuencia de nuestra teoría, resultan irrelevantes. Si son irrelevantes o no observables, hay que prescindir de ellas, pero si son relevantes, tenemos que utilizarlas. Nuestros datos no tienen por qué estar en el mismo nivel de análisis. Hay datos desagregados u observaciones sobre un período de tiempo diferente, o incluso sobre otra parte del mundo, que pueden proporcionar más consecuencias observables de una teoría. Quizá no nos interesen en absoluto estas consecuencias secundarias, pero si concuerdan con la teoría, tal como se predice, nos ayudarán a otorgarle más confianza a sus capacidades y aplicabilidad. Nuestros datos tampoco necesitan ser «simétricos»: siempre que todos sean consecuencias observables de nuestra teoría, podremos utilizar un detallado estudio de una provincia, una comparación de dos países, entrevistas personales con miembros de la administración pertenecientes a un único sector de políticas e, incluso, algún elemento cuantitativo. En este proceso vamos de lo particular a lo general, ya que caracterizar ciertas unidades a partir de rasgos comunes es un proceso generalizador. La consecuencia es que aprendemos mucho más acerca de las teorías generales y de los hechos concretos.

En general, lo que queremos es concentrar en nuestra hipótesis tanta información como sea posible, lo cual puede suponer la realización de más estudios de caso; pero esto suele ser demasiado difícil, requiere mucho tiempo o es caro. Es evidente que no debemos considerar información irrelevante. Por ejemplo, tratar el número de escaños conservadores en la Cámara de los Comunes británica como una variable mensual en vez de como algo que cambia en las elecciones nacionales aumentaría considerablemente el número de observaciones, pero no tendría sentido porque no aportaría mucha más información. Por otra parte, desagregar los resultados de las elecciones presidenciales estadounidenses por estados, e incluso por condados, aumenta el número de casos y también la cantidad de información que se aporta al problema.

Estos datos desagregados pueden parecer irrelevantes, ya que el objetivo es conocer las causas de la victoria de un determinado candidato en la carrera presidencial, que es una cuestión fundamentalmente agregada. Sin embargo, la mayoría de las explicaciones que se dan al resultado de las elecciones presidenciales tienen diferentes consecuencias observables en cada una de las unidades desagregadas. Si, por ejemplo, pronosticamos el resultado de esta votación en función de variables económicas como el índice de paro, el uso de este indicador en cada estado aportará muchas más observaciones sobre las consecuencias de nuestra teoría que el índice agregado del conjunto del país. Verificar que la teoría tiene sentido en otras situaciones —aunque éstas no tengan un interés directo— aumenta la confianza en que la teoría es correcta y en que explica adecuadamente la única consecuencia observable que nos interesa.

### 3. Modelos formales de investigación cualitativa

(Un modelo es una simplificación de algún aspecto del mundo y una aproximación al mismo. Los modelos nunca son, netamente, «verdaderos» o «falsos», aunque los buenos sólo recopilan las características «correctas» de la realidad que representan.)

Pensemos por ejemplo en un juguete de plástico y pegamento que sea un modelo de quince centímetros de un avión. Ésta es una pequeña reproducción del auténtico aparato, que no puede volar ni contiene nada y cuyos componentes tampoco se mueven. Nadie la confundiría con el avión real; preguntar si cada aspecto del modelo es verdadero es como plantearse si la modelo que posó para la *Gioconda* de Leonardo da Vinci tenía realmente una sonrisa tan seductora. Aunque la tuviera, no esperamos que el cuadro de Leonardo sea una representación exacta de nadie, ya sea una modelo real o la Virgen María, del mismo modo que no esperamos que el modelo de un avión refleje todas sus características reales. Sin embargo, nos gustaría saber si esa reproducción recopila las características correctas de un avión para un problema concreto. Si queremos enseñárselas a un niño cómo es realmente un avión, este modelo podría ser adecuado. Si está hecho a escala, también podría ser útil para que los diseñadores aeronáuticos hicieran pruebas de túneles aerodinámicos. La característica principal de un avión real que se aprecia en este modelo es la forma, y, para ciertos propósitos, éste es realmente uno de los rasgos apropiados. Evidentemente, este modelo no refleja multitud de detalles de un avión, entre ellos el tamaño, el color, la sensación de estar en una aeronave, la solidez de sus diversos componentes, el número de plazas a bordo, la potencia de sus motores, el tejido de los cojines de los asientos, ni tampoco otros muchos sistemas esenciales como el eléctrico, el de aire o el de conducción de agua. Si queremos entender estos aspectos del avión, necesitaremos un conjunto de modelos completamente diferente.

¿Podemos evaluar un modelo sin conocer qué características del objeto queremos evaluar? Está claro que no. Por ejemplo, podríamos pensar que un modelo que se ocupara de la cantidad de desperdicios de un avión no sería de gran utilidad. Sin embargo, si que sería muy importante para fines educativos o para realizar pruebas de túneles aerodinámicos. Como incluso el polvo de las moquetas puede hacer que un avión pese más y que, por tanto, haya que pagar más combustible, los modelos de este tipo son importantes para la industria aeronáutica y se han realizado (ahorrándose así millones de dólares).

Hay versiones más o menos restrictivas de todos los modelos. Los más restrictivos son más claros, concisos y abstractos, pero también menos realistas (a menos que el mundo sea realmente conciso). Los poco restrictivos son detallados, contextuales y más realistas, pero también menos claros y más difíciles de calibrar con precisión (véase King, 1989: apartado 5 del capítulo 2). En qué punto de esta gradación decidimos construir un modelo depende de nuestro objetivo y de la complejidad del problema que estudiemos.

[Existen modelos físicos, gráficos, verbales o algebraicos. Por ejemplo, la descripción cualitativa de los sistemas judiciales europeos en un libro que se ocupe de ese tema constituye un modelo del mismo. Independientemente de lo profunda que sea la descripción o del talento del autor, el contenido del libro siempre será una abstracción o una simplificación de los sistemas jurídicos reales. Como para comprender se necesita resumir, lo que indica que un libro es bueno es tanto lo que se deja fuera de él como lo que se incluye.]

[Mientras que los investigadores cualitativos a menudo emplean modelos verbales, nosotros utilizaremos los algebraicos en el análisis que sigue, con el fin de estudiar los de tipo verbal y mejorarlos. Al igual que ocurre con los juguetes aeronáuticos y con los libros que se dedican en su totalidad a estudiar la Revolución francesa, nuestros modelos algebraicos cualitativos, no tienen que confundirse con la propia investigación cualitativa. Su único propósito es señalar con especial claridad los problemas que hay que evitar y las oportunidades que pueden aprovecharse. Además, con frecuencia ocurre que nos ayudan a descubrir ideas que, de otro modo, no se nos habrían ocurrido.]

Partimos de la base de que los lectores no han tenido experiencia previa con modelos algebraicos, aunque a los que conoczan los estadísticos algunos de los siguientes les resultarán familiares. Sin embargo, la lógica inferencial de estos modelos se aplica tanto a la investigación cuantitativa como a la cualitativa. El hecho de que, probablemente, los cuantitativistas estén más familiarizados con nuestra terminología no significa que se les vaya a dar mejor aplicar la lógica de la inferencia científica. Además, estos modelos *no* se adaptan más directamente a la investigación cuantitativa que a la cualitativa: en ambos casos representan abstracciones útiles de la

investigación a la que se aplican. Para facilitar su presentación, todos los modelos algebraicos van precedidos de descripciones verbales a las que siguen recuadros con una notación matemática normalizada. Aunque no lo recomendamos, los recuadros pueden saltarse sin que se pierda el hilo argumental.

#### 4. Un modelo formal sobre la recogida de datos

Antes de formalizar nuestra presentación de las inferencias descriptiva y causal —los dos objetivos principales de la investigación social—, desarrollaremos un modelo para la recogida y resumen de datos que es bastante sencillo, pero que resulta eficaz para analizar los problemas inferenciales. Nuestro modelo algebraico no es tan formal como los estadísticos, aunque si clarifica nuestras ideas y facilita su transmisión. Con la expresión recogida de datos hacemos alusión a una gran variedad de métodos, entre los que se incluye la observación general y la participante, las entrevistas en profundidad, las encuestas a gran escala, la historia basada en fuentes secundarias, los experimentos aleatorios, la etnografía, el análisis de contenido y cualquier otro método que sirva para reunir datos fiables. La regla más importante para toda recogida de datos es dejar claro cómo se han creado y de qué manera hemos accedido a ellos. Toda información que recojamos tiene que concretar las consecuencias observables de nuestra teoría. Desarrollar una nueva pregunta de investigación puede ayudarnos, pero no servirá para responder a la primera si no es una consecuencia observable de ella.

*En que nos fijamos*  
*en una muestra*  
*de personas*  
*a una unidad*  
*de observación*  
*de datos*  
*en un dato*

*Muy bien*

[Hacemos modelos con los datos por medio de variables, unidades y observaciones. Un ejemplo sencillo es la renta anual de cuatro personas diferentes. Los datos pueden representarse simplemente mediante cuatro cantidades: 9.000, 22.000, 21.000 y 54.292 dólares. En el caso más general, podríamos etiquetar la renta de las cuatro personas (numeradas del 1 al 4) como  $y_1$ ,  $y_2$ ,  $y_3$  e  $y_4$ . Una variable codificada para dos entrevistas no estructuradas podría utilizar los valores «participativo», «cooperador» o «intransigente», y se etiquetaría como  $y_1$  e  $y_2$ . En estos ejemplos la variable es  $y$ , las unidades son los individuos, y las observaciones, los valores de las variables en cada unidad (renta en dólares o grado de cooperación). El símbolo  $y$  se denomina variable porque su valor cambia en cada unidad y, en general, puede representar cualquier cosa cuyo valor se vaya alterando en un conjunto de unidades. Éstas, dado que se puede recabar información en un período de tiempo o en diferentes secciones de un área, pueden ser personas, países, organizaciones, años, elecciones o décadas y, con frecuencia, alguna combinación de estas u otras unidades. Las observaciones pueden ser numéricas, verbales, visuales o cualquier otro tipo de datos empíricos.]

Supongamos, por ejemplo, que nos interesa estudiar las organizaciones internacionales desde 1945. Antes de recoger los datos hay que decidir qué resultados se quiere explicar. Podríamos intentar comprender la influencia en 1990 del factor tamaño sobre la actividad de las organizaciones internacionales (que se clasificaría por su campo de acción o por el propio tipo de organización), los cambios en el tamaño agregado de la actividad de dichas organizaciones desde 1945 o las transformaciones ocurridas en su actividad desde esa fecha que hubieran tenido que ver con la distribución de ese factor. Entre las variables que miden la actividad de las organizaciones internacionales podríamos incluir el número de países que pertenecen a ellas en un momento dado, la cantidad de tareas que desempeñan o la magnitud de sus presupuestos y de su personal. En estos ejemplos, las unidades de análisis serían las organizaciones internacionales, sus campos de acción, el número de países miembros y períodos de tiempo como años, quinquenios o décadas. En la fase de recogida de datos no se aplica ninguna regla para determinar qué variables se recogen, cuántas unidades debe haber, si éstas deben ser más numerosas que las variables o cuál es el mejor método para medir las últimas. El único criterio es que, en cada caso, nosotros pensemos que lo que hacemos es importante. Cuando tenemos una idea más clara de cómo van a utilizarse los datos, la regla es intentar encontrar tantas consecuencias observables de una teoría como sea posible. Como subrayamos en el capítulo 1, la investigación empírica puede utilizarse tanto para evaluar hipótesis previas como para plantear otras que no se habían barajado anteriormente, pero si sirve para el segundo propósito, hay que recoger nuevos datos para evaluar las nuevas hipótesis.]

A partir de nuestro análisis debe quedar claro que en la mayoría de las obras que se denominan «estudios de caso» se miden numerosas variables de muchos tipos diferentes de unidades. Aunque estas investigaciones no suelen utilizar más de un puñado de casos, el número total de observaciones es, generalmente, immense. Por lo tanto, es esencial distinguir entre número de casos y de observaciones. El primero puede tener cierto interés para ciertos fines, pero el último es importante para evaluar la cantidad de información que un estudio aporta a la resolución de una pregunta teórica. Por consiguiente, reservamos la n habitual para referirnos únicamente al número de observaciones y no al de casos. Sólo de forma ocasional, como cuando las observaciones individuales sean parcialmente dependientes, distinguiremos entre información y número de observaciones. El uso de esta última expresión procede del muestreo de encuestas, en el que n es el número de personas que hay que entrevistar; sin embargo, nosotros la utilizaremos de forma mucho más general. En realidad, nuestra definición de «observación» coincide exactamente con lo que Harry Eckstein (1975, p. 85) denomina «caso». Cómo señala este autor: «Un estudio de seis elecciones generales en Gran Bretaña puede ser, pero no tiene por qué, un estudio en el que n = 1. Podría ser otro en el que n = 6 o también n = 120.000.000.

*número de*  
*casos*  
*«dato»*

### El diseño de la investigación social

*análisis de  
muestra de  
caso*  
*Muy bueno*

Todo depende de si el objeto de estudio son los sistemas electorales, las elecciones o los votantes». La «ambigüedad sobre lo que constituye un «individuo» (de ahí el «caso») sólo puede evitarse si nos ocupamos de las medidas de entidades concretas, y no de las propias entidades. De este modo, un «caso» puede definirse técnicamente como un fenómeno del que sólo señalamos e interpretamos una única medida en cada variable pertinente». La única diferencia con nuestro uso es que, desde que Eckstein escribió su artículo, los investigadores han continuado utilizando la palabra «caso» para referirse a un estudio de caso completo, que todavía tiene una definición bastante imprecisa. Por lo tanto, siempre que sea posible utilizaremos la palabra «caso» como lo hacen la mayoría de los autores y reservaremos el término «observación» para hacer referencia a las medidas de una o varias variables en una sola unidad.

En el resto del capítulo queremos mostrar de qué manera conceptos como el de variable y unidad pueden hacer que abordemos de forma más clara el diseño de una investigación, incluso cuando no sea apropiado utilizar medidas cuantitativas para resumir la información de que dispongamos. La cuestión que planteamos es: ¿cómo podemos hacer inferencias descriptivas sobre «la historia tal como fue» sin perdernos en un mar de datos irrelevantes? Dicho de otro modo, ¿cómo separar lo esencial de lo esfímero?

## 5. El resumen de los pormenores históricos

Después de recoger los datos, el primer paso de cualquier análisis es resumirlos. Los resúmenes pueden describir una gran cantidad de datos pero no están directamente relacionados con la inferencia. Como, en realidad, lo que nos interesa es generalizar y explicar, un resumen de los datos que van a explicarse suele ser un buen punto de partida, pero no constituye un objetivo suficiente para la investigación en las ciencias sociales.

Resumir es necesario, ya que, como nunca podemos decir «todo lo que sabemos» sobre un conjunto de acontecimientos, no tendría sentido intentar hacerlo. Los buenos historiadores comprenden cuáles son los datos cruciales y, por lo tanto, en vez de perderse en digresiones, realizan estudios que subrayan lo esencial. Para entender la historia europea de los primeros quince años del siglo XIX, quizás necesitemos comprender ciertos principios de estrategia militar tal como lo hizo Napoleón e, incluso, saber qué comía su ejército si su avance dependía de los víveres que encontraba a su paso; pero puede que sea irrelevante conocer el color del pelo de Napoleón o si prefería los huevos fritos a los cocidos.<sup>1</sup> Los buenos textos históricos suelen incluir, aunque no se limiten a ello, un sucinto resumen verbal de gran cantidad de pormenores históricos.

Nuestro modelo para el proceso de resumen de los detalles históricos es un estadístico, con el que se expresan datos de forma abreviada. Su objetivo

### 2. La inferencia descriptiva

vo es exponer características apropiadas de éstos mediante un formato útil<sup>5</sup>. Por ejemplo, la media muestral, o promedio, es un estadístico:

$$\bar{y} = \frac{1}{n} (y_1 + y_2 + \dots + y_n) = \frac{1}{n} \sum_{i=1}^n y_i$$

en el que  $\sum_{i=1}^n y_i$  es una manera cómoda de escribir  $y_1 + y_2 + y_3 + \dots + y_n$ . Otro estadístico es el máximo muestral, que se etiqueta como  $y_{\max}$ :

$$y_{\max} = \text{Máximo} (y_1, y_2, \dots, y_n) \quad (2.1)$$

La media muestral de las cuatro cifras de renta del ejemplo del apartado 4 (9.000, 22.000, 21.000 y 54.292 dólares) es 26.573 dólares, mientras que el máximo muestral es 54.292. Podemos resumir los datos originales que contienen cuatro cifras mediante las dos cantidades que representan, respectivamente, la media y el máximo muestrales. También podemos calcular otras características de la muestra, como el mínimo, la mediana, la moda o la varianza.

Cada resumen de este modelo reduce todos los datos (cuatro cifras en este ejemplo simple o lo que conocemos de algún aspecto de la historia europea en el anterior) a un único número. Comunicarse mediante resúmenes suele ser más fácil y tiene más sentido para el lector que utilizar todos los datos originales. Evidentemente, si en nuestro conjunto de datos sólo hubiera cuatro números, no tendría mucho sentido utilizar cinco resúmenes diferentes y sería más sencillo presentar las cuatro cantidades originales. Interpretar un estadístico suele ser más fácil que comprender todo un conjunto de datos, pero siempre se pierde información al describir una gran cantidad de números utilizando unos pocos.

¿Qué normas rigen el resumen de los pormenores históricos? La primera es que los resúmenes deben centrarse en los resultados que queremos describir o explicar. Si lo que nos interesa es el crecimiento de la organización internacional media, no sería muy inteligente centrarse en las Naciones Unidas; pero si lo que nos importa es cómo influye el mayor o menor tamaño en la distribución de dichas organizaciones, las Naciones Unidas serían seguramente una de las unidades en las que tendríamos que centrarnos. La ONU no es una organización representativa, pero sí es importante. Para estudiar la típica organización internacional desde el punto de vista estadístico examinariamnos valores medios (de presupuestos, cometidos, número de miembros, etc.), pero para comprender el abanico de actividades tendríamnos que examinar la varianza. El segundo precepto, igualmente evidente, es que un resumen tiene que simplificar la información de que disponemos. En términos cuantitativos, esto significa que siempre hay que utilizar menos estadísticos sintetizadores que unidades hay en los datos originales, porque, de no ser así, podría ser más fácil presentar los datos originales sin

ningún tipo de resumen<sup>6</sup>. La síntesis que utilicemos también ha de ser lo suficientemente simple como para que nuestro público pueda entenderla. Ningún fenómeno puede resumirse de manera perfecta, de manera que las normas que dictan lo que es adecuado dependen de nuestros propósitos y del público. Por ejemplo, un artículo científico sobre guerras y alianzas podría incluir datos sobre 10.000 observaciones. En este trabajo, estaría justificado que hubiera resúmenes que utilizaran cincuenta números; sin embargo, incluso para un experto, cincuenta indicadores separados podrían resultar incomprensibles si no existe otro resumen. Para impartir una clase sobre esta materia a uno de los primeros cursos de licenciatura, tres cifras serían más que suficiente.

## 6. La inferencia descriptiva

La inferencia descriptiva es un proceso mediante el cual se comprende un fenómeno no observado a partir de un conjunto de observaciones. Por ejemplo, en las elecciones británicas de 1979 nos podría interesar cómo cambia en cada circunscripción el voto a los partidos Conservador, Laborista y Socialdemócrata. Lo más probable es que tengamos varias hipótesis que evaluar; sin embargo, lo que observamos realmente son las elecciones a la Cámara de los Comunes de ese año en 650 circunscripciones.

Podríamos pensar ingenuamente que, al registrar qué porcentaje de votos y escaños recibieron los conservadores en cada circunscripción, estamos observando de forma directa su fuerza electoral. Sin embargo, la política siempre conlleva un cierto margen para lo aleatorio y lo impredecible, al igual que el conjunto de la vida social y toda investigación científica<sup>7</sup>. Supongamos que por un súbito descuido (o como deferencia hacia las ciencias sociales) el Parlamento Británico hubiera decidido celebrar elecciones cada semana en 1979 y supongamos también (de forma contrafáctica) que éstas fueran independientes unas de otras. Aunque el apoyo subyacente a los conservadores se mantuviera constante, cada repetición semanal no registraría el mismo número de votos por partido en cada circunscripción. Podría cambiar la situación climatológica, surgir epidemias, la gente podría tomar sus vacaciones, y todo ello afectaría a la participación en las elecciones y a su resultado. Además, podrían producirse acontecimientos fortuitos en el contexto internacional o aparecer escándalos en los medios de comunicación; aunque estas contingencias no tuvieran importancia a largo plazo, sí podrían influir en los resultados semanales. De este modo, numerosos acontecimientos transitorios podrían producir resultados electorales ligeramente diferentes. Después de todo, nuestra observación de cualquiera de estas elecciones no sería una medida perfecta de la fortaleza de los conservadores.

Por poner otro ejemplo, supongamos que nos interesa la intensidad del conflicto entre los israelíes (policía y residentes) y los palestinos en las co-

## 2. La inferencia descriptiva

munidades de los territorios ocupados por Israel en Cisjordania. Los informes oficiales de ambos bandos parecen sospechosos o están censurados, de manera que decidimos llevar a cabo nuestro propio estudio. Quizás podamos verificar la intensidad general del conflicto en diferentes comunidades mediante entrevistas en profundidad o participando en actividades familiares o de otros colectivos. Si hacemos ambas cosas durante una semana en cada una de las comunidades, nuestras conclusiones sobre la intensidad del conflicto en ellas reflejarán en parte los acontecimientos que, por casualidad, sucedan en el período en que visitamos esos enclaves. Aunque nuestro estudio se llevara a cabo durante un año, tampoco podríamos conocer perfectamente el verdadero nivel de conflictividad, aunque si se reduciría la incertidumbre sobre él.

En estos ejemplos, puede decirse que la varianza del voto conservador en diferentes circunscripciones o la del conflicto en las comunidades cisjordanas surge de dos factores separados: de diferencias *sistemáticas* y *no sistemáticas*. En el ejemplo electoral, las primeras incluyen características fundamentales y predecibles de las circunscripciones, como las diferencias ideológicas, de renta, de organización de la campaña o las referentes al apoyo tradicional a cada partido. En hipotéticas repeticiones semanales de las mismas elecciones se mantendrían las mismas diferencias sistemáticas, pero variarían las no sistemáticas, como son los cambios producidos en la participación por las condiciones climatológicas. En el ejemplo de Cisjordania, se considerarían diferencias sistemáticas las profundas divergencias culturales entre israelíes y palestinos, su conocimiento mutuo y las pautas de segregación geográfica de la vivienda. Si pudieramos comenzar nuestra semana de observación en doce ocasiones distintas, estas diferencias sistemáticas entre comunidades seguirían influyendo en la intensidad del conflicto observado. Sin embargo, las no sistemáticas, como los atentados terroristas o los ejemplos de brutalidad policial israelí, no serían predecibles y sólo afectarían a la semana en que ocurrieran. En general, con las técnicas inferenciales apropiadas, podemos conocer la naturaleza de las diferencias sistemáticas, aunque sea con la ambigüedad que está presente en un conjunto de datos de la vida real por las diferencias no sistemáticas o aleatorias que contiene.

Por lo tanto, uno de los objetivos fundamentales de la inferencia es distinguir entre el componente sistemático y el no sistemático dentro de los fenómenos que estudiamos. El sistemático no es más importante que el que no lo es, y nuestra atención no debería centrarse en uno a costa del otro. Sin embargo, en las ciencias sociales resulta fundamental diferenciarlos. Una de las formas de abordar la inferencia es considerar el conjunto de datos recogido como uno más de los que son posibles, al igual que los resultados reales de las elecciones británicas de 1979 son sólo uno de los muchos conjuntos de resultados posibles en diferentes días hipotéticos en los que las elecciones podrían haber tenido lugar, o del mismo modo que una semana

### El diseño de la investigación social

de observación en una pequeña comunidad es sólo una de las muchas semanas posibles.

Con la inferencia descriptiva pretendemos comprender en qué medida nuestras observaciones reflejan fenómenos típicos o atípicos. Si las elecciones británicas de 1979 hubieran tenido lugar durante una epidemia de gripe que se hubiera extendido por los hogares de clase obrera sin afectar a los de los ricos, quizás nuestras observaciones habrían calibrado bastante mal la fuerza subyacente de los conservadores, precisamente porque el elemento no sistemático e imprevisto de los datos tendería a prevalecer sobre el sistemático o a distorsionarlo. Del mismo modo, si nuestra semana de observación en Cisjordania hubiera tenido lugar inmediatamente después de la invasión israelí del sur del Líbano, no cabría esperar resultados indicativos de lo que ocurre habitualmente en la zona.

En teoría, el mundo político puede producir múltiples conjuntos de datos para cada problema, pero no siempre atiende las necesidades de los científicos sociales. Lo normal es que sólo tengamos la suerte de observar un conjunto de datos. Para construir un modelo dejaremos que este único conjunto lo represente una variable  $y$  (los votos recibidos por los laboristas, digamos) que se mida en todas las unidades ( $n = 650$ ), que son las circunscripciones:  $y_1, y_2, \dots, y_n$  ( $y_1$ , por ejemplo, podría representar a las 23.562 personas que votaron a los laboristas en la circunscripción 1). El conjunto de observaciones que denominamos  $y$  es una variable realizada cuyo valor varía en cada unidad. Además, definimos  $Y$  como variable aleatoria porque cambia aleatoriamente en todas las reproducciones hipotéticas de las mismas elecciones. En consecuencia,  $y_s$  es el número de personas que han votado a los laboristas en la circunscripción 5 y  $Y_s$  es la variable aleatoria que representa los votos en muchas elecciones hipotéticas que podrían haberse celebrado en esa circunscripción en condiciones fundamentalmente similares. Los votos que obtiene el Partido Laborista en la única muestra que consideramos,  $y_1, y_2, \dots, y_n$ , varían en cada circunscripción por factores sistemáticos y aleatorios. Esto quiere decir que, para distinguir los dos tipos de «variables», solemos utilizar la expresión variable realizada para referirnos a  $y$  y variable aleatoria para referirnos a  $Y$ .

El mismo criterio se aplica a nuestro ejemplo cualitativo. No podríamos albergar esperanza o deseo de cuantificar el grado de tensión entre israelíes y palestinos, en parte porque el «conflicto» es un asunto complicado que tiene que ver con los sentimientos de numerosos individuos, la oposición de ciertas organizaciones, las tensiones ideológicas y muchas otras cosas. En esta situación,  $y_s$  es una variable realizada que representa la totalidad del conflicto que se ha observado en la quinta comunidad, El-Bireh<sup>4</sup>. La variable aleatoria  $Y_s$  representa tanto lo que observamos en esta localidad como lo que podríamos haber observado; la aleatoriedad proviene de la variación en los acontecimientos imprevistos durante las posibles semanas que podríamos haber elegido para hacer la observación<sup>5</sup>.

### 2. La inferencia descriptiva

Uno de los objetivos de la inferencia es conocer las características sistemáticas de las variables aleatorias  $Y_1, \dots, Y_n$  (hay que señalar lo contradictorio de esta terminología estándar: aunque, en general, queremos distinguir el componente sistemático del no sistemático en nuestros datos, lo que pretendemos en un caso determinado es tomar una variable aleatoria y extraer sus características sistemáticas). Por ejemplo, podríamos querer saber el número de votos que se espera que obtenga el Partido Laborista en la circunscripción 5 (el promedio de votos  $Y_s$  para este partido en un gran número de elecciones hipotéticas en esa circunscripción). Al ser ésta una característica sistemática del sistema electoral subyacente, su valor esperado tiene un interés considerable para los científicos sociales. Por el contrario, los votos para el Partido Laborista en unas elecciones observadas,  $y_s$ , tienen, a largo plazo, un valor bastante más reducido, ya que son el resultado de características sistemáticas y de errores aleatorios<sup>10</sup>.

El valor esperado (una de las características del componente sistemático) en la quinta comunidad de Cisjordania, El-Bireh, se expresa formalmente como sigue:

$$E(Y_s) = \mu_s$$

donde  $E(\cdot)$  es la operación del valor esperado, de la que surge la media en un número infinito de reproducciones hipotéticas de la semana de observación en la comunidad 5, El-Bireh. El parámetro  $\mu_s$  (la letra griega mu con el subíndice 5) representa la respuesta al cálculo del valor esperado (un nivel de conflictividad entre palestinos e israelíes) para la comunidad 5. Este parámetro es parte de nuestro modelo de característica sistemática para una variable aleatoria  $Y_s$ . Se podría utilizar el nivel de conflictividad observado,  $y_s$ , como un cálculo aproximado de  $\mu_s$ , pero como  $y_s$  contiene muchos elementos imprevistos, junto con información acerca de esta característica sistemática, suele haber mejores estimadores (véase el apartado 7).

También podríamos querer conocer otra característica sistemática de estas variables aleatorias: el nivel de conflictividad en la comunidad cisjordana media:

$$\frac{1}{n} \sum_{i=1}^n E(Y_i) = \frac{1}{n} \sum_{i=1}^n \mu_i = \mu \quad (2.2)$$

Un estimador de  $\mu$  podría ser la media de los niveles de conflictividad observados en todas las comunidades que se estudián,  $\bar{y}$ , pero también existen otros para esta característica sistemática (hay que señalar que el mismo resumen de datos que aparece en el análisis del proceso de síntesis de los pormenores históricos del apartado 5 se utiliza para hacer la estimación de una inferencia descriptiva). Entre las características sistemáticas de las va-

riables aleatorias también se incluye la varianza y diversos parámetros causales que se presentan en el apartado 1 del capítulo 3.

Todavía hay otra característica de estas variables aleatorias que podría ser de interés: la variación del nivel de conflictividad dentro de una comunidad aun cuando las características sistemáticas no cambien, es decir, en qué medida las observaciones de diferentes semanas (diversas aplicaciones hipotéticas de la misma variable aleatoria) producen resultados divergentes. Dicho de otro modo, éste es el tamaño del componente no sistemático que, formalmente, se calcula en una única comunidad utilizando la varianza (y no las expectativas):

$$V(Y) = \sigma^2 \quad (2.3)$$

en donde  $\sigma^2$  (la letra griega sigma) indica lo que se obtiene al aplicar el operador de varianza a la variable aleatoria  $Y$ . Vivir en una comunidad de Cisjordania en la que hubiera un alto grado de conflictividad entre israelíes y palestinos no sería agradable, pero vivir en un lugar en el que los valores del conflicto registraran una elevada varianza y, por tanto, fueran muy impredecibles, podría ser peor. En cualquier caso, ambas situaciones pueden tener un interés considerable para los investigadores académicos.

Para comprender mejor estas cuestiones distinguimos dos puntos de vista fundamentales sobre la variación aleatoria<sup>11</sup>, que constituyen los dos extremos de una gradación. Aunque hay un número significativo de investigadores que se siente cómodo en cada uno de esos extremos, la mayoría de los politólogos tienen puntos de vista que, de alguna manera, se sitúan entre ambos polos.

#### Punto de vista 1: un mundo probabilístico

La variación aleatoria existe tanto en la naturaleza como en los mundos social y político, y nunca podrá eliminarse. Aunque midiéramos todas las variables sin cometer errores, registráramos todos los datos (en vez de sólo una muestra) e incluyéramos toda variable explicativa concebible, nuestros análisis no podrían hacer nunca predicciones perfectas. Un investigador puede dividir el mundo en componentes aparentemente sistemáticos y no sistemáticos y mejorar con frecuencia sus predicciones, pero nada de lo que haga para analizar los datos podrá reducir de manera significativa el grado de variación no sistemática que existe en diversas partes del mundo empírico.

#### Punto de vista 2: un mundo determinista

La variación aleatoria sólo representa la parte del mundo que no podemos explicar. La división entre variación sistemática y estocástica la impone el investigador, y depende

de qué variables explicativas se encuentren disponibles y de cuáles se incluya en el análisis. Con las variables explicativas apropiadas el mundo es completamente predecible.

Estos puntos de vista diferentes producen diversas ambigüedades en las inferencias de cada área de investigación<sup>12</sup>. Sin embargo, en la mayoría de ellas *ambas perspectivas pueden considerarse equivalentes desde el punto de vista de la observación*. Esto es especialmente cierto si, con el punto de vista 2, partimos de la base de que al menos ciertas variables explicativas seguirán sin conocerse. De este modo, la equivalencia desde el punto de vista de la observación tiene lugar cuando dichas variables desconocidas del punto de vista 2 se convierten, en el punto de vista 1, en la interpretación de la variación aleatoria. Al carecerse de cualquier consecuencia observable con la que distinguir entre ellos, elegir uno u otro depende más de la fe o de la creencia que de la verificación empírica.

Por poner otro ejemplo, en ambos puntos de vista determinar si un acontecimiento político o social concreto es el resultado de un proceso sistemático o no sistemático depende de las opciones que tome el investigador. Usando el punto de vista 1 podemos clasificar provisionalmente un efecto como sistemático o como no sistemático, pero, a menos que logremos encontrar otro conjunto de datos (o incluso un solo caso más) con el que comprobar la persistencia de un efecto o pauta, será muy difícil hacer una apreciación correcta.

Si usamos la versión más extrema del punto de vista 2, no podremos más que describir los datos, ya que juzgar «incorrectamente» que un acontecimiento es estocástico o sistemático es imposible o irrelevante. Una versión más realista de esta perspectiva acepta que el punto de vista 1 puede considerar de forma correcta o incorrecta que una pauta es aleatoria o sistemática, pero deja cierta libertad para decidir cuál será objeto de examen en cada estudio concreto y cuál seguirá sin explicarse. En este sentido, al comenzar cualquier análisis, todas las observaciones son consecuencia de fuerzas «no sistemáticas». Por tanto, a nosotros nos corresponde demostrar que esos acontecimientos o procesos son el resultado de fuerzas sistemáticas. Decidir si un acontecimiento o proceso no explicado es un hecho realmente aleatorio o si procede simplemente de variables explicativas aún no identificadas queda para futuras investigaciones. *x que?*

Este argumento se aplica con igual fuerza a la investigación cualitativa que a la cuantitativa. La primera suele ser histórica, pero es de extrema utilidad para las ciencias sociales cuando también es explícitamente inferencial. Para conceptualizar las variables aleatorias de las que proceden las observaciones e intentar hacer una estimación de sus características sistemáticas —en vez de limitarse a resumir los pormenores históricos— *no* es preciso recoger datos a gran escala. De hecho, uno de los rasgos que definen a un buen historiador es su capacidad para distinguir, en la situación que describe, los aspectos sistemáticos de los circunstanciales. Por lo tanto,

### El diseño de la investigación social

esta defensa de la inferencia descriptiva no es en absoluto una crítica a los estudios de caso o al trabajo del historiador. Lo que queremos decir es que, en las ciencias sociales, cualquier tipo de investigación tendría que cumplir los principios inferenciales básicos que se analizan en este libro. Demostrar que existen características sistemáticas será más difícil con algunos tipos de datos, pero no deja de ser importante hacerlo.

Como ejemplo de los problemas de la inferencia descriptiva en la investigación histórica, supongamos que nos interesaran los resultados de los encuentros en la cumbre que mantuvieron los Estados Unidos y la Unión Soviética entre 1955 y 1990. Nuestro fin último es responder a una pregunta de tipo "causal": ¿en qué condiciones y en qué medida condujeron esas reuniones a una mayor cooperación? Para contestar a esta pregunta hay que resolver ciertas cuestiones difíciles del análisis causal, especialmente las relacionadas con la dirección que toma la causalidad dentro de un conjunto de variables relacionadas de forma sistemática<sup>13</sup>. Sin embargo, en este apartado, nos limitamos a tratar los problemas de la inferencia descriptiva.

Supongamos que hemos ideado una forma de calibrar —mediante análisis históricos, expertos en encuestas, recuento de acontecimientos «cooperativos» o «conflictivos», o combinando estas técnicas de medición— en qué medida se registró después de cada cumbre un aumento de la cooperación entre las superpotencias y que tenemos ciertas hipótesis sobre las condiciones en las que aumenta la cooperación, que se relacionan con cambios de poder, con los ciclos electorales en los Estados Unidos, con la situación económica en cada país y con el grado de cumplimiento de las expectativas previas de ambos bandos. Supongamos también que esperamos poder explicar el nivel subyacente de cooperación en cada año, relacionándolo de alguna manera con la presencia o ausencia de una cumbre en el período anterior, así como con nuestros propios factores explicativos.

Lo que observamos (aunque nuestros índices de cooperación sean perfectos) es únicamente el grado de cooperación que se da realmente en cada año. Cuando se observen niveles de cooperación altos en los años posteriores a la celebración de las cumbres, no sabremos, sin un estudio más profundo, si hay una relación sistemática entre ambos factores. Con un pequeño número de observaciones, podría ser que asociar las cumbres a la cooperación reflejara una aleatoriedad debida a una incertidumbre fundamental (buena o mala suerte, según el punto de vista 1) o a variables explicativas no identificadas por el momento (según el punto de vista 2). Entre estas últimas variables se incluyen las fluctuaciones climatológicas que podrían producir malas cosechas en la Unión Soviética, así como los cambios en el equilibrio militar o en los mandatarios, circunstancias que podrían explicar las transformaciones en el grado de cooperación. Si se identifican, estas variables serán explicaciones alternativas —variables omitidas que podrían recogerse o examinarse para evaluar su influencia en el resultado de la cumbre. Si no se identifican, pueden considerarse acontecimientos no

### 2. La inferencia descriptiva

sistemáticos que podrían explicar el alto grado de cooperación observado entre las superpotencias. Para aportar pruebas que contradigan la posibilidad de que haya acontecimientos aleatorios (variables explicativas no identificadas) que expliquen la cooperación observada, podríamos considerar muchos otros años. Será extremadamente improbable que los acontecimientos y procesos aleatorios, al ser por definición no persistentes, produzcan una cooperación diferente en los años en que haya cumbres y en los que no las haya. De nuevo, llegamos a la conclusión de que sólo mediante pruebas repetidas en diferentes contextos (años, en este caso) podremos decidir si es mejor definir una pauta como sistemática o como resultante de las consecuencias pasajeras de procesos aleatorios.

Suele ser difícil distinguir entre procesos sistemáticos y no sistemáticos. Desde el punto de vista de las ciencias sociales, una epidemia de gripe que ataque con mayor virulencia a los votantes de clase obrera que a los de clase media es un acontecimiento impredecible (no sistemático) que, en una reproducción hipotética de las elecciones de 1979, disminuiría los votos laboristas. Sin embargo, la persistencia de un componente diferenciador de clase en la incidencia de una discapacidad sería un efecto sistemático que reduciría la media de votos laborista en muchas reproducciones.

En las elecciones estadounidenses, el hecho de que un candidato venza a otro por su personalidad o por un lapsus linguae ocurrido en un debate televisado podría ser un factor aleatorio que quizás hubiera influido en las probabilidades de cooperación entre la URSS y los Estados Unidos durante la Guerra Fría. Sin embargo, si la principal promesa electoral hubiera sido la de reducir la tensión con los soviéticos, las constantes victorias de los candidatos conciliadores hubieran sido un factor sistemático que explicaría las probabilidades de cooperación.

Los factores sistemáticos son persistentes y sus consecuencias se repiten cuando tienen un valor determinado. Los no sistemáticos son transitorios: no podemos predecir su impacto. Sin embargo, esto no significa que los sistemáticos representen constantes. Las promesas electorales de una campaña pueden ser un factor sistemático a la hora de explicar el comportamiento electoral, pero esto no significa que dichas promesas no cambien. En el resultado de unas elecciones, la constante de las promesas electorales es su efecto y, si es variable, cambia de forma predecible. Cuando las relaciones soviético-estadounidenses eran buenas, puede que prometer políticas conciliadoras ganara votos en las elecciones de los Estados Unidos; cuando eran malas, quizás ocurriría lo contrario. Del mismo modo, la climatología puede ser un factor aleatorio (si los cambios intermitentes e impredecibles tienen consecuencias también impredecibles) o una característica sistemática (si el mal tiempo produce siempre menos votos para los candidatos que están a favor de políticas conciliadoras).

En pocas palabras, resumir los pormenores históricos es un importante paso intermedio en el proceso de utilización de los datos, pero también te-

nemos que hacer inferencias descriptivas que distingan los fenómenos aleatorios de los sistemáticos. Saber qué ocurrió en una ocasión concreta no es suficiente. Si no nos esforzamos por determinar las características sistemáticas de un asunto, las lecciones de la historia se perderán y nada aprenderemos de los aspectos de nuestro objeto de estudio que pueden mantenerse o ser relevantes para estudios o acontecimientos futuros.

## 7. Criterios para juzgar las inferencias descriptivas

[En este último apartado presentamos tres criterios explícitos que suelen utilizarse en estadística para juzgar los métodos inferenciales: falta de sesgo, eficiencia y coherencia. Todos se basan en el marco de las variables aleatorias que señalamos en el apartado 6 de este mismo capítulo, pero tienen consecuencias directas y profundas para la evaluación y mejora de la investigación cualitativa.] Con el fin de clarificar estos conceptos, en el presente apartado sólo damos ejemplos muy sencillos, todos ellos tomados de la inferencia descriptiva. Una versión simple de la inferencia conlleva la estimación de ciertos parámetros, incluyendo en ellos el valor esperado o la varianza de una variable aleatoria ( $\mu$  o  $\sigma^2$ ) en una inferencia descriptiva. También utilizamos estos mismos criterios para evaluar inferencias causales en el siguiente capítulo (véase su apartado 4). Reservamos para capítulos posteriores los consejos específicos que implican estos criterios para la investigación cualitativa y, en el resto del apartado, nos centraremos únicamente en conceptos.

### 7.1 Inferencias no sesgadas

[Si aplicamos una y otra vez un método inferencial obtendremos estimaciones que unas veces serán demasiado grandes y otras demasiado pequeñas. En un número elevado de aplicaciones, ¿obtenemos, en promedio, la respuesta correcta? Si la contestación es sí, entonces se dice que este método, o «estimador», no está sesgado. En un estimador, esta propiedad no nos dice en qué medida cualquiera de las aplicaciones del método está lejos del promedio, pero lo deseable es que la media sea correcta.]

[Se obtienen estimaciones no sesgadas cuando la variación que hay entre la reproducción de una medida y la que le sigue no es sistemática y empuja ese cálculo a veces en una dirección y a veces en otra. Se produce sesgo cuando, en un conjunto de reproducciones, hay un error sistemático de medida que lleva la estimación más en una dirección que en otra. Si en nuestro estudio de la conflictividad en las comunidades de Cisjordania los líderes políticos hubieran creado tensiones con el fin de influir en el resultado de la investigación (quizá para impulsar sus fines), el promedio de conflictivi-

## 2. La inferencia descriptiva

dad observado en cada comunidad estaría sesgado hacia niveles altos. Si las reproducciones de nuestras hipotéticas elecciones de 1979 se hubieran hecho todas en sábado (cuando se podrían haber celebrado cualquier otro día), las estimaciones estarían sesgadas si ese hecho ayudara sistemáticamente a uno de los partidos y no al otro (si, por ejemplo, los conservadores fueran más reacios a votar en sábado por razones religiosas). Las estimaciones reproducidas también podrían haberse basado en informes de contabilizadores de votos corruptos que favorecieran a un partido más que a otro. Sin embargo, si las elecciones reproducidas se celebraran en diversos días elegidos de una manera que no estuviera relacionada con la variable que nos interesa, los posibles errores de medida no producirían resultados sesgados, aunque uno u otro día favoreciera a algún partido. Por ejemplo, si hubiera fallos en el recuento de votos por descuidos aleatorios de los encargados de contabilizarlos, el conjunto de estimaciones no estaría sesgado.

Si, por ley, las elecciones británicas se celebraran siempre en domingo o si el sistema electoral tuviera un método de recuento que favoreciera más a uno de los partidos (por la utilización de un determinado procedimiento o quizás incluso por una persistente corrupción), necesitaríamos un estimador que variara según la media de votos que pudiera esperarse en las condiciones de estas características sistemáticas. De este modo, el sesgo depende de la teoría que se investiga, y no sólo existe en los datos. No tiene mucho sentido decir que un determinado conjunto de datos está sesgado, aunque esté plagado de errores individuales.

En este ejemplo, puede que queramos distinguir entre nuestra definición de «sesgo estadístico» en un *estimador* y la de «sesgo esencial» en un *sistema electoral*. Un ejemplo del segundo tipo es un horario de votación que dificulte el ejercicio del voto a los trabajadores; sesgo esencial bastante común en varios sistemas electorales. Como investigadores, quizás queramos hacer una estimación de la media de votos en el presente sistema electoral (el del sesgo esencial), pero quizás queramos también hacer un cálculo aproximado de este indicador en un hipotético sistema que no presente un sesgo esencial debido a los horarios de los colegios electorales. Esto haría posible que calculáramos el grado de sesgo esencial del sistema. Cualquiera que sea la media que estemos calculando, lo que queremos es un estimador estadístico sin sesgos.

Los datos de las ciencias sociales pueden sufrir un sesgo importante del que hay que ser consciente: las personas que nos proporcionan los datos brutos que utilizaremos en nuestras inferencias descriptivas suelen tener razones para hacer estimaciones que son, sistemáticamente, demasiado altas o demasiado bajas. Puede que los funcionarios quieran exagerar las consecuencias de un nuevo programa para pedir una mayor financiación o calcular el índice de paro a la baja para demostrar que están haciendo un buen trabajo. Quizás sea necesario buscar con más ahínco estimaciones menos sesgadas. En el estudio cualitativo de Myron Weiner (1991) acerca de la

## El diseño de la investigación social

educación y el trabajo infantil en la India encontramos un expresivo ejemplo. Al intentar explicar el reducido nivel de aceptación de la educación obligatoria en la India, en comparación con otros países, el autor primero tuvo que comprobar si dicho nivel era realmente bajo. Se dio cuenta de que, en uno de los estados del país, las estadísticas oficiales indicaban que el 98 por ciento de los niños en edad escolar acudía al colegio. Sin embargo, un examen más atento puso de manifiesto que la asistencia sólo se media en una ocasión, la primera vez que los niños iban a la escuela. Después, figuraban en las listas de asistencia durante siete años, ¡aunque sólo hubieran ido un día! Un escrutinio más minucioso mostró que la cifra de asistencia real era mucho más baja.

### 7.2 Eficiencia

Generalmente, no tenemos una oportunidad para aplicar nuestro estimador a un gran número de aplicaciones esencialmente idénticas. En realidad, excepto en algunos cuidados experimentos, sólo lo aplicamos una vez. En este caso, la falta de sesgo es importante, pero quisieramos estar más seguros de que la única estimación que hacemos se acerca al valor correcto. La eficiencia proporciona una manera de distinguir entre estimadores no sesgados. De hecho, este criterio también puede ayudarnos a diferenciar estimadores alternativos con un reducido nivel de sesgo (en general, un estimador muy sesgado debe descartarse sin siquiera evaluar su eficiencia).

La eficiencia es un concepto relativo que se mide calculando la varianza del estimador en las reproducciones hipotéticas. En los estimadores no sesgados, cuanto menor es la varianza, más eficiente (mejor) es el estimador. Una varianza pequeña es mejor porque nuestra única estimación probablemente estará más cerca del verdadero valor del parámetro. No nos interesa la eficiencia de un estimador muy sesgado porque, en esta situación, la presencia de una varianza reducida hará improbable que la estimación se acerque al valor verdadero (porque la mayoría de tales estimaciones constituirán un apretado conglomerado cercano al valor equivocado). Como describimos posteriormente, nos interesa la eficiencia si el sesgo es pequeño, y quizás a veces estemos dispuestos a admitir un pequeño sesgo a cambio de una mayor eficiencia.

Supongamos de nuevo que nos interesaría hacer una estimación del nivel medio de conflictividad entre palestinos e israelíes en Cisjordania y que estuviéramos evaluando dos métodos: una sola observación de una única comunidad, elegida como típica, y observaciones similares de, por ejemplo, 25 comunidades. Parece evidente que 25 observaciones son mejores que una sola, siempre que se haga el mismo esfuerzo en la primera que en cada una de las 25. Vamos a demostrar precisamente por qué es así. Este resultado explica por qué debemos observar tantas consecuencias de nues-

*que sea la  
muestra  
de lo redondo  
tiene una  
gran  
obligación*

## 2. La inferencia descriptiva

### Un ejemplo formal de falta de sesgo

Supongamos, por ejemplo, que quisieramos hacer una estimación de  $\mu$  en la ecuación (2.2) y que decidieramos utilizar la media como estimador,

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ . En un único conjunto de datos,  $\bar{y}$  es la proporción de votos la-  
boristas en la media de todas las circunscripciones,  $n = 650$  (o el nivel  
medio de conflictividad en las comunidades de Cisjordania). Sin embargo,  
la media muestral se convierte en una función de 650 variables aleatorias,

$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ , si se considera en un número infinito de reproducciones hipó-  
téticas de las elecciones en cada circunscripción. De este modo, la media  
muestra también se convierte en una variable aleatoria. En algunas repro-  
ducciones hipotéticas,  $\bar{Y}$  producirá resultados electorales que se acerquen  
a  $\mu$  y en otras ocasiones éstos se alejarán de ella. La cuestión es si  $\bar{Y}$  será  
acertada, es decir, igual a  $\mu$  en la media de estas reproducciones hipotéti-  
cas. Para hallar la respuesta utilizamos de nuevo la operación del valor es-  
perado, que nos permite determinar la media del infinito número de hipó-  
téticas elecciones. Las reglas de la expectativa nos permiten realizar los  
siguientes cálculos:

$$\begin{aligned} E(\bar{Y}) &= E\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n E(Y_i) \\ &= \frac{1}{n} n\mu \\ &= \mu \end{aligned} \quad (2.4)$$

Así,  $\bar{Y}$  es un estimador no sesgado de  $\mu$  (este ejemplo es un poco menos  
formal que los que aparecen en los manuales de estadística propiamente  
dichos, pero las características principales son las mismas).

tra teoría como sea posible, pero también demuestra el concepto general de  
eficiencia estadística, que es igualmente relevante cuando estamos deci-  
diendo cómo se pueden evaluar mejor las diferentes maneras de combinar  
en una inferencia las observaciones recogidas.

La eficiencia nos permite comparar el estimador  $\mu$  del estudio de caso  
en el que sólo hay una observación ( $n = 1$ ) con el gran estimador que tiene

un  $n$  grande ( $n = 25$ ) y que representa el nivel medio de conflictividad encontrado en 25 estudios separados de una semana de duración en diversas comunidades de Cisjordania. Ambos estimadores, si se utilizan apropiadamente, no están sesgados. Si el mismo modelo es cierto, el estimador de la observación única tendrá una varianza de  $V(Y_{\text{típica}}) = \sigma^2$ . Esto quiere decir que habríamos elegido una circunscripción que consideramos «típica», la cual, no obstante, se vería afectada por variables aleatorias. La varianza del gran estimador  $n$  es  $V(\bar{Y}) = \sigma^2/25$ , o sea, la de la media muestral. De este modo, el estimador de la observación única es veinticinco veces más variable (menos eficiente) que el que se obtiene cuando  $n = 25$ . De ahí que obtengamos el resultado evidente de que es mejor tener más observaciones.

Más interesantes son las condiciones en las que un estudio muy detallado de nuestra única comunidad pudiera arrojar resultados tan buenos como los de la investigación que utiliza un  $n$  grande. Esto quiere decir que, aunque siempre son preferibles los estudios con más observaciones (si se dispone de los medios para recogerlas), hay situaciones en las que un único estudio de caso (como siempre, con muchas observaciones) es mejor que otro que se basa en más observaciones que, por separado, no son ni tan detalladas ni tan ciertas.

En igualdad de condiciones, nuestro análisis demuestra que cuantas más observaciones haya, mejor, porque así disminuye la variabilidad (y, por tanto, la ineficiencia). De hecho, la coherencia es una propiedad que consiste en que a medida que aumenta el número de observaciones, la variabilidad se acerca a cero y la estimación es igual al parámetro que estamos intentando calcular.<sup>14</sup>

Sin embargo, es frecuente que las condiciones no sean las mismas. Supongamos, por ejemplo, que cualquier medida individual de un fenómeno que estemos estudiando se vea influida por factores que aumenten la probabilidad de que se aleje del valor verdadero (o sea, que la varianza del estimador es alta), y supongamos que tenemos alguna idea —quizá por otros estudios— de cuáles pueden ser dichos factores. Supongamos, además, que nuestra capacidad de observar y corregir esos factores disminuye sustancialmente a medida que aumenta el número de comunidades estudiadas (aunque sólo sea porque carecemos del tiempo y del conocimiento suficientes para corregir esos factores en un número elevado de observaciones). Nos enfrentamos entonces al dilema de elegir entre un estudio de caso que tiene observaciones adicionales dentro del propio caso y 25 casos que contienen cada uno sólo una observación.

Si nuestro caso único sólo tiene una observación, será evidentemente inferior al estudio de 25 observaciones. Sin embargo, los investigadores que utilizan estudios de caso disfrutan de considerables ventajas, que se entienden mejor al formalizarlas. Por ejemplo, podríamos comenzar por seleccionar cuidadosamente nuestra comunidad con el fin de asegurarnos de que es especialmente representativa del resto del país o de que entendemos la rela-

ción que tiene con las demás. Podríamos preguntar a unos pocos residentes o leer informes de prensa para comprobar si es una comunidad media o si algún factor no sistemático ha hecho que esta observación sea atípica; después podríamos ajustar el nivel de conflictividad observado para hacer una estimación del nivel medio,  $\mu$ , que se registra en Cisjordania. Ésta sería la parte más difícil del estimador del estudio de caso, y habría que tener cuidado de que no se nos colara el sesgo. Una vez que estamos razonablemente seguros de haberlo minimizado, podríamos centrarnos en aumentar la eficiencia. Para hacerlo, podríamos pasar muchas semanas en la comunidad, llevando a cabo numerosos estudios independientes. Podríamos entrevistar a líderes comunitarios, leer los periódicos, acompañar a una familia en su vida diaria y utilizar otras técnicas para recabar información. Mediante tales procedimientos, podríamos conseguir muchas más de 25 observaciones dentro de esta única comunidad y producir un estudio de caso que, al igual que el de las 25 comunidades, no estuviera sesgado y que fuera más eficiente que éste.

Consideremos otro ejemplo: supongamos que estamos realizando un estudio sobre el problema internacional que plantean las drogas y que necesitamos calcular el porcentaje de tierra cultivable de una determinada región del mundo en el que se ha plantado cocaína. Supongamos, además, que podemos elegir entre dos métodos: un estudio de caso de un solo país o una investigación a gran escala de tipo estadístico que se ocupe de todos los países de la región. Parece mejor estudiar toda la región, pero hay que señalar que, para optar por esta alternativa, es necesario (por razones prácticas) utilizar los datos que proporcionan los gobiernos de la zona a un organismo de la ONU. Se sabe que estas cifras tienen poca relación con las pautas reales de cultivo, ya que las preparan los ministerios de asuntos exteriores y se basan en consideraciones relativas a las relaciones públicas. Supongamos también que, mediante visitas que observarán de cerca un país, podríamos corregir las estimaciones de su gobierno para acercarlas más a las cifras verdaderas. ¿Qué método elegiríamos? Quizá decidamos estudiar sólo un país, o quizás dos o tres; también podríamos analizar uno en profundidad y utilizar nuestros resultados para reinterpretar, y por tanto mejorar, los datos gubernamentales de los otros países. La elección tendría que basarse en qué datos responden mejor a nuestras preguntas.

Por poner un ejemplo más, supongamos que estamos estudiando la Comunidad Europea y queremos hacer una estimación del grado de regulación que se espera produzcan en un sector de toda la Comunidad las acciones de la Comisión y del Consejo de Ministros. Podríamos recabar información en un gran número de normativas adoptadas formalmente por el sector industrial en cuestión, codificarlas según su severidad y, posteriormente, hacer una estimación de la severidad media de una normativa. Si recogemos datos sobre 100 normativas que muestren, en principio, un rigor similar, la varianza de nuestro cálculo será el resultado de dividir la de cualquiera de las

## El diseño de la investigación social

normativas por 100 ( $\sigma^2/100$ ), o por menos si las normativas están relacionadas. Sin duda, la medida resultante será mejor que la que obtendríamos si utilizáramos datos sobre una sola normativa como estimador de la severidad reguladora para el conjunto del sector.

No obstante, para utilizar este procedimiento es preciso aceptar que la normativa formal es equivalente a la auténtica actividad reguladora en el sector que se estudia. Sin embargo, una investigación más profunda de la aplicación de las normativas podría poner de manifiesto una importante variación en su grado de implantación real. En consecuencia, puede que las mediciones de las normativas formales estén sistemáticamente sesgadas y que, por ejemplo, exageren la severidad reguladora. En tal caso, nos enfrentaríamos de nuevo a la necesidad de alcanzar un equilibrio entre sesgo y eficiencia, y quizás tuviera sentido realizar tres o cuatro estudios de caso en profundidad sobre la implantación de normativas con el fin de investigar la relación que existe entre las normativas formales y la actividad reguladora real. Una de las posibilidades sería sustituir el estimador basado en 100 casos por otro, menos sesgado y también menos eficiente, que procediera de tres o cuatro. Sin embargo, sería más creativo, si es factible, utilizar el estudio en profundidad de los tres o cuatro casos para corregir el sesgo de nuestro indicador de 100 casos y usar después una versión corregida de este último como estimador. En este procedimiento estaríamos combinando las aportaciones de nuestros estudios de caso en profundidad con las técnicas que manejan un  $n$  grande, práctica que pensamos debería realizarse con mucha más frecuencia de lo que es habitual en las ciencias sociales contemporáneas.

La defensa de los estudios de caso que hacen aquellos que conocen bien una determinada parte del mundo suele ser la que está implícita en el ejemplo anterior. Puede que las investigaciones a gran escala dependan de cifras que no comprenda del todo el ingenuo investigador que trabaja con una base de datos (que quizás no sea consciente de cómo se elaboran las estadísticas electorales de un determinado lugar y presuponga, equivocadamente, que tienen una relación real con los votos emitidos). Quien trabaja de cerca con los materiales y comprende de dónde proceden puede hacer las correcciones necesarias. En los apartados siguientes intentaremos explicar cómo se puede elegir de forma sistemática en relación a este asunto.

Precisamente, el análisis formal de este problema, que aparece en el recuadro posterior, muestra cómo se puede determinar cuál es el resultado de alcanzar un equilibrio entre eficiencia y sesgo, cuestión que surgía en el ejemplo de las circunscripciones electorales británicas. En cualquier ejemplo concreto la decisión siempre será mejor si se utiliza la lógica que aparece en el análisis formal siguiente. Sin embargo, para tomar este tipo de decisión casi siempre se necesitarán también consideraciones cualitativas.

Para terminar, merece la pena plantearse de forma más específica las concesiones que a veces hay que realizar para hallar un equilibrio entre ses-

guiente  
menos  
caso  
menos  
probabilidad  
de "sesgo"  
para no  
cometer  
eficiencia  
+ sesgo,  
+ eficiencia

## 2. La inferencia descriptiva

### Comparaciones de eficiencia formales

La varianza de la media muestral  $\bar{Y}$  se expresa como  $V(\bar{Y})$ , y las reglas para calcular varianzas de variables aleatorias en el caso sencillo del muestreo aleatorio permiten que ocurra lo siguiente:

$$V(\bar{Y}) = V\left(\frac{1}{n} \sum_{i=1}^n Y_i\right)$$
$$= \frac{1}{n^2} \sum_{i=1}^n V(Y_i)$$

Además, si presuponemos que la varianza en las reproducciones hipotéticas de las elecciones en cada circunscripción es la misma que en una de cada dos y que se expresa como  $\sigma^2$ , entonces la de la media muestral será:

$$V(\bar{Y}) = \frac{1}{n^2} \sum_{i=1}^n V(Y_i) \quad (2.5)$$
$$= \frac{1}{n^2} \sum_{i=1}^n \sigma^2$$
$$= \frac{1}{n^2} n \sigma^2$$
$$= \sigma^2/n$$

En el ejemplo anterior  $n = 650$ , de manera que la varianza del estimador de este  $n$  grande es  $\sigma^2/650$ , y la del estimador del estudio de caso es  $\sigma^2$ . La estimación estadística será preferible por su eficiencia, a menos que podamos utilizar correcciones cualitativas del error aleatorio para reducir la varianza del estimador del estudio de caso mediante un factor que valga, al menos, 650.

go y eficiencia. En cualquier conjunto grande de observaciones que no esté sesgado, la media muestral de las dos primeras tampoco lo estará, y lo mismo ocurre con la media muestral de todas ellas. Sin embargo, al utilizar sólo dos observaciones se prescinde de información sustancial; esto no afecta a la falta de sesgo, pero sí reduce considerablemente la eficiencia. Si no utilizamos también el criterio de eficiencia, careceremos de reglas formales para elegir entre un estimador y otro.

### Comparaciones formales entre sesgo y eficiencia

Pensemos en dos estimadores, uno el de un estudio con un  $n$  grande que, al haber sido realizado por alguien que tiene un prejuicio, está ligeramente sesgado y otro con un  $n$  muy pequeño que creemos que no está sesgado pero que es relativamente menos eficiente y ha sido elaborado por un investigador imparcial. Como modelo formal de este ejemplo, supongamos que quisiéramos hacer una estimación de  $\mu$  y que el estudio de  $n$  grande produce un estimador  $d$ :

$$d = \left( \frac{1}{n} \sum_{i=1}^n Y_i - 0,01 \right)$$

Hacemos un modelo del estudio pequeño con un estimador diferente de  $\mu$ , que es  $c$ :

$$c = \left( \frac{Y_1 + Y_2}{2} \right)$$

en el que las circunscripciones 1 y 2 son de tipo medio, de manera que  $E(Y_1) = \mu$  y  $E(Y_2) = \mu$ .

¿Qué estimador sería preferible? Nuestra primera respuesta es que no deberíamos utilizar ninguno de ellos y que sería mejor optar por la media muestral  $\bar{y}$ , o sea, por un estudio de  $n$  grande realizado por un investigador imparcial. Sin embargo, no siempre puede aplicarse el estimador obvio o el mejor. Para solucionar este problema, recurrimos a una evaluación del sesgo y de la eficiencia.

En primer lugar, calibraremos el sesgo. Podemos señalar que el primer estimador  $d$  está ligeramente sesgado según el cálculo habitual:

$$\begin{aligned} E(d) &= E\left(\frac{1}{n} \sum_{i=1}^n Y_i - 0,01\right) \\ &= E\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) - E(-0,01) \\ &= \mu - 0,01 \end{aligned}$$

También podemos indicar, mediante un cálculo similar, que el segundo estimador  $c$  no está sesgado:

$$\begin{aligned} E(c) &= E\left(\frac{Y_1 + Y_2}{2}\right) \\ &= \frac{E(Y_1) + E(Y_2)}{2} \\ &= \frac{\mu + \mu}{2} \\ &= \mu \end{aligned}$$

Utilizando sólo estos cálculos, elegiríamos el estimador  $c$ , que representa el resultado del estudio con un  $n$  pequeño de nuestro investigador imparcial, porque no está sesgado. El investigador prejuicioso obtendría con  $d$  una respuesta que, en promedio, estaría equivocada, aunque sólo fuera levemente, en un número infinito de reproducciones. Con el estimador  $c$  se obtendría una media correcta.

El criterio de eficiencia presenta una situación diferente. Comenzamos por calcular la varianza de cada estimador:

$$\begin{aligned} V(d) &= V\left(\frac{1}{n} \sum_{i=1}^n Y_i - 0,01\right) \\ &= V\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) - V(0,01) \\ &= \sigma^2/n \\ &= \sigma^2/650 \end{aligned}$$

Esta varianza es la misma que la de la media muestral porque 0,01 no cambia (tiene una varianza cero) en las diferentes muestras. Del mismo modo, se calcula la varianza de  $c$  como sigue<sup>15</sup>:

$$\begin{aligned} V(c) &= V\left(\frac{Y_1 + Y_2}{2}\right) \\ &= \frac{1}{4} [V(Y_1) + V(Y_2)] \\ &= \frac{1}{4} 2\sigma^2 \\ &= \sigma^2/2 \end{aligned}$$

### El diseño de la investigación social

En consecuencia,  $c$  es bastante menos eficiente que  $d$  porque  $V(c) = \sigma^2/2$  es 325 veces más grande que  $V(d) = \sigma^2/650$ . Esto también debería quedar claro intuitivamente, ya que  $c$  desecha gran parte de la información que hay en la matriz de datos.

¿Cuál tendríamos que elegir? El estimador  $d$  está sesgado pero es más eficiente que  $c$ , mientras que éste no tiene sesgo pero es menos eficiente. En este caso concreto, probablemente sería preferible el estimador  $d$ . Por tanto, estaríamos dispuestos a sacrificar la falta de sesgo, ya que el sacrificio es bastante pequeño (0,01), con el fin de lograr un estimador bastante más eficiente. Sin embargo, habrá un momento en el que la mayor eficiencia no compensará el pequeño sesgo porque acabaremos garantizando que los cálculos aproximados están lejos de la verdad. Formalmente, la manera de calibrar el equilibrio entre sesgo y eficiencia es calcular el *error medio cuadrático* (EMC), que es una combinación de ambos factores. Si  $g$  es el estimador de algún parámetro, y (la letra griega gamma), EMC se definirá así:

$$\text{EMC}(g) = V(g) + E(g - \gamma)^2 \quad (2.6)$$

= varianza + sesgo al cuadrado

Por tanto, el error medio cuadrático es la suma de la varianza y del sesgo al cuadrado (véase Johnston, 1984, pp. 27-28). La idea es elegir el estimador que tenga el mínimo error cuadrático medio, ya que éste muestra

Supongamos que nos interesa saber si el Partido Demócrata va a ganar las próximas elecciones presidenciales en los Estados Unidos y que preguntamos a 20 estadounidenses adultos elegidos al azar a qué partido piensan votar (en este sencillo ejemplo de selección aleatoria elegimos a los encuestados entre todos los estadounidenses adultos y cada uno de ellos tiene las mismas posibilidades de ser seleccionado). Supongamos que alguien ha realizado un estudio similar con 1.000 ciudadanos; ¿acaso tendríamos que incorporar esas observaciones adicionales a las nuestras para constituir una tómica estimación que se basara en 1.020 encuestados? Si las nuevas observaciones hubieran sido seleccionadas aleatoriamente, al igual que las primeras veinte, sería fácil decidir incorporar esos datos a los nuestros: con las nuevas observaciones el estimador sigue sin estar sesgado y ahora es mucho más eficiente.

Sin embargo, supongamos que sólo 990 de las 1.000 nuevas observaciones fueran resultado de una selección aleatoria de la población de Estados Unidos y que las otras diez fueran congresistas demócratas que hubieran sido incluidos, sin querer, en los datos, después de haberse hecho la muestra aleatoria. Supongamos además que nos damos cuenta de que se han incorporado dichas observaciones a nuestros datos pero que desconocemos cuáles son y que, por tanto, no podemos eliminarlas. Ahora sabemos que,

### 2. La inferencia descriptiva

precisamente por qué un estimador con cierto sesgo puede ser preferible si su varianza es pequeña.

En nuestro ejemplo, los dos EMCs son:

$$\begin{aligned} \text{EMC}(d) &= \frac{\sigma^2}{650} + (0,01)^2 \\ &= \frac{\sigma^2}{650} + 0,0001 \end{aligned} \quad (2.7)$$

y

$$\text{EMC}(c) = \frac{\sigma^2}{2} \quad (2.8)$$

De este modo, en la mayoría de los valores de  $\sigma^2$ ,  $\text{EMC}(d) < \text{EMC}(c)$ , y  $d$  sería preferible como estimador de  $c$ .

En teoría, siempre hay que optar por estimadores no sesgados que sean lo más eficientes posible (o sea, que utilicen la mayor cantidad posible de información). Sin embargo, en las investigaciones reales que analizamos en capítulos posteriores, este equilibrio entre sesgo y eficiencia es bastante llamativo.

en principio, un estimador que se basara en los 1.020 encuestados exageraría ligeramente la posibilidad de que un demócrata ganara las elecciones nacionales. De este modo, incluir las 1.000 observaciones adicionales sesgaría ligeramente la estimación global pero también aumentaría su eficiencia de forma considerable. En consecuencia, el que incluyamos o no estas observaciones depende de si el aumento del sesgo se compensa con el incremento de la eficiencia estadística. De una manera intuitiva, parece claro que el estimador que se base en 1.020 observaciones producirá estimaciones bastante más cercanas a la respuesta correcta con mucha más frecuencia que el que utiliza sólo 20. El sesgo resultante será lo suficientemente pequeño como para que sea preferible el estimador de la muestra más grande, aunque en la práctica probablemente apliquemos los dos (además, en este caso, conocemos la dirección del sesgo e incluso podríamos corregirlo parcialmente).

Si se dispone de datos cuantitativos adecuados y es posible formalizar este tipo de problemas, generalmente se puede tomar una decisión clara. Sin embargo, aunque la naturaleza cualitativa de la investigación haga que evaluar este equilibrio sea difícil o imposible, entenderlo debería ayudarnos a hacer inferencias más fiables.

### 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 su 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

## El diseño de la investigación social

causales, calificando de «explicaciones» hipótesis o especulaciones no evaluadas, que se basan en diseños de investigación imprecisos<sup>1</sup>. 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 parezcan apropiadas, pero aportando al lector, al mismo tiempo, la mejor y más sincera 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 descriptiva. 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, entraremos 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<sup>2</sup>. 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

## 3. Causalidad e inferencia causal

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 «variable causal principal» (también denominada «causal» 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  $y_4^r$  (el subíndice 4 indica la cuarta circunscripción de Nueva York y el superíndice *r* el hecho de que el demócrata sea ya congresista). Por tanto,  $y_4^r$  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) recibirá mediante  $y_4^d$  (*d* se refiere a un candidato demócrata que no está en el poder)<sup>3</sup>.

Esta condición *contrafáctica* es la esencia de la definición de causalidad, y la diferencia, 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 más adelante. Hay que tener mucho cuidado al definir contrafácticos porque, 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

cambia desde «ser congresista» a «no ser congresista». Durante esta transformación hipotética se mantienen todos 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 alguno<sup>4</sup>.

*La def. anterior es sint. experimental y la de control*

De manera más formal, el efecto causal que tiene estar en el poder 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:  $(y_4^D - y_4^R)$ . Por razones que pronto quedarán claras, esta diferencia la denominamos *efecto causal realizado*, y, para expresarla con una notación más general, utilizaremos la unidad  $i$  en vez de sólo la circunscripción 4<sup>5</sup>:

$$\text{(Efecto causal realizado en la unidad } i\text{)} = y_i^D - y_i^R \quad (3.1)$$

*Diseño*

Es obvio que este efecto se define únicamente en teoría, ya que, en cualquier elección real, sólo podríamos observar  $y_i^D$  o  $y_i^R$ , 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

siempre necesitan dividir el mundo en componentes sistemáticos y no sistemáticos, mientras que la definición de Holland no establece claramente esta distinción<sup>6</sup>. Para ver la importancia de esta división, pensemos en qué ocurriría si pudieramos 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 contempla nuestro análisis teórico o contiene una variabilidad fundamentalmente inexplicable<sup>7</sup>. Definimos la variable aleatoria que representa la proporción de votos recibida por el candidato demócrata que ya es congresista como  $Y_i^D$  (obsérvese la  $Y$  mayúscula), y la proporción de los que habría recibido un demócrata no congresista en reproducciones hipotéticas, como  $Y_i^R$ .

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:

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

*Holland*

(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 pudieramos 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 realizado; 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-

### El diseño de la investigación social

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 hace 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 posibilita 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 qué significa esto retomaremos 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 medida de fracciones del voto en dichas reproducciones— para el 'no congresista':

$$E(Y_i^N) = \mu_i^N$$

y para el congresista como:

$$E(Y_i^D) = \mu_i^D$$

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 circunscripción 4):

Efecto causal  
medio en la unidad  $i \equiv \beta$  (3.3)

$$\begin{aligned} &= E(\text{Efecto causal aleatorio en la unidad } i) \\ &= E(Y_i^D - Y_i^N) \\ &= E(Y_i^D) - E(Y_i^N) \\ &= \mu_i^D - \mu_i^N \end{aligned}$$

### 2. Causalidad e inferencia causal

de modo que en la primera línea de esta ecuación,  $\beta$  (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 comparables 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 habituales 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 él identificó 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, quizás 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:

$$(\text{varianza del efecto causal en la unidad } i) = V(Y_i^D - Y_i^N)$$

en la que evitamos introducir un nuevo símbolo que represente el resultado del cálculo de la varianza,  $V(Y_i^D - Y_i^N)$ . Con seguridad, los nuevos congresistas querrán saber la variación 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.

## 1.2 Un ejemplo cualitativo

En el apartado 1 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 queríamos 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 polítólogos 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, incremente 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 dé 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 éstas 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 y también por la variedad de las demás disposiciones constitucionales (como la naturaleza del 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 bloquen y que, no cuenten con un gobierno decidido. También estas características pueden conducir a la desafección y la inestabilidad<sup>8</sup>.

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, calibráramos 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 lleváramos a cabo este experimento hipotético muchas veces. Definimos el efecto causal medio como la media de los efectos causales realizados en las reproducciones 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 éstas 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<sup>9</sup>.

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 interesaría conocer el efecto causal medio que tiene uno u otro sobre la esta-

### El diseño de la investigación social

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 varianza de los efectos causales. Por ejemplo, quizás 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 varianza 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 varianza 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-

### 3. Causalidad e inferencia causal

nismo 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 multinacionales 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 detallados». Muchos de los pormenores de los estudios de caso bien hechos coullavan 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 ése, 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 infinidad 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» tendremos que listar los vínculos causales que hay entre ambas variables. En esta definición también necesitaremos determinar una serie de enlaces causales, definir la causalidad en cada par de variables consecutivas de la secuencia y precisar qué unión existe entre cada dos 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

el régimen presidencial y una menor decisión del gobierno en el parlamentario. Estos efectos intervintentes —que genera el ordenamiento constitucional y que, a su vez, influyen en la estabilidad política— pueden observarse directamente. Podríamos hacer un seguimiento de las actitudes o comportamientos de las minorías para ver de qué modo difieren en dos situaciones experimentales o estudiar lo decididos que son los gobiernos en cada uno de los sistemas. Sin embargo, aunque los efectos causales de los sistemas presidenciales y de los parlamentarios operaran de forma diferente, nuestra definición de efecto causal seguiría siendo válida. Se puede definir dicho concepto sin comprender todos los mecanismos causales que 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 definitorio 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 él denomina «causalidad múltiple». Esto quiere decir que «El fenómeno que se investiga tiene determinantes alternativos; lo que Mill (1843) llamaba el problema de la "plurallidad 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.<sup>10</sup>

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) quizás no alerten al investigador de la existencia de una «causalidad múltiple», pero si 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

que el nivel de renta de una persona depende tanto de que culmine con éxito estíudios superiores como del hecho de que sus padres también lo hicieran. 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, quizás 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. Quizás 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.

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 quizás 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 exigen más de los análisis posteriores. El hecho de que algunas variables dependientes, y quizás todas las interesantes de este tipo que hay en las ciencias sociales, se vean influidas por muchos factores causales no hace problemática nuestra definición de causalidad. La clave para entender estas 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 pueden ser relevantes (véanse los apartados 2 y 3 del capítulo 5)<sup>11</sup>.

### 2.3 Causalidad «simétrica» y «asimétrica»

Stanley Lieberson (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).

Como ejemplo de la afirmación de Lieberson, 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. Lieberson definiría 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 «simétrico» 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 «asimétrico» 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 Lieberson 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 teórico 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 Lieberson 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 contrario —de congresista a no congresista— en otros dos comicios. Cualquier 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 Lieberson en el contexto de la inferencia causal. Sin embargo, no deben confundirse con la definición teórica de causalidad, que ya hemos dado en el apartado 1 de este capítulo.

### 3. Supuestos necesarios para la estimación de efectos causales

¿Cómo 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 íntegramente a dar una respuesta completa a esta pregunta, pero aquí proporcionamos una visión general de los dos supuestos que pueden evitarnos 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 conllevan 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

*modelos  
estimativos*

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,  $\mu_1^n = \mu_2^n$  y  $\mu_1^t = \mu_2^t$ ). 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  $\beta$ , 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 dependiente. Es evidente que esto es sólo un supuesto y que puede ser erróneo: quizás ambas circunscripciones sean diferentes por alguna razón desconocida que sesgile nuestra inferencia causal. De hecho, cualquier par de circunscripciones reales diferirá en ciertos sentidos, y para aplicar este supuesto ambas tendrán que ser iguales en cuanto a la media de muchas reproducciones hipotéticas de la campaña electoral. Por ejemplo, la pluviosi-

dad (que puede reducir la asistencia a las urnas en algunas áreas) no variará, 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 (definido 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 aplique  $t$  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, ¡al menos hasta que fuera lo suficientemente ingenioso 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 períodos, 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 períodos, 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 incomprobable: que pulsar el encendido del interruptor en el primer período 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

### El diseño de la investigación social

diferentes unidades si también lo es el de la variable explicativa, sólo tendremos que presuponer que el efecto causal es constante. Esta es una versión más endeble 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 comprobable; sin embargo, es importante que se haga explícita su naturaleza. ¿En qué gama de unidades esperamos que tenga sentido nuestro supuesto de que estar en el poder tenga 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-

temporáneos  
de acuerdo  
con el efecto  
uniforme  
de las  
diferencias  
entre

la complejidad  
de la  
singularidad

### 3. Causalidad e inferencia causal

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,  $Y_1$  y  $Y_2$ . 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 cono-

ciendo el proceso e imponiendo una medida del mismo a nuestras variables de control. Por ejemplo, supongamos 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 deciden 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 quizás 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. (Y vamos para controlar el efecto)

(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 sistemáticas, 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

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  $\mu$  y que conceptualizamos como un número fijo pero desconocido. Se dice que un estimador de  $\mu$  no está sesgado si es igual a la media de  $\mu$  en muchas reproducciones hipotéticas del mismo experimento.

Al igual que entonces, seguimos conceptualizando el valor esperado de un efecto causal aleatorio, denominado  $\beta$ , como un número fijo pero desconocido. Por tanto, la falta de sesgo se define de manera análoga: un estimador de  $\beta$  carece de sesgo si es igual a la media de  $\beta$  en muchas reproducciones hipotéticas del mismo experimento. La eficiencia también se define de forma parecida a la variación que se registra de una reproducción hipotética a otra. Éstos 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 cuadros 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 estemos enfascados 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 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. Cualquier 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.

### 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 demócrata no congresista en ese reparto (la media de todas las reproducciones hipotéticas) en la circunscripción  $i$  es  $\mu_i^u$ . 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:

$$\beta = E(Y_i | X_i = 1) - E(Y_i | X_i = 0) = \mu_i^l - \mu_i^u \quad (3.4)$$

e incorporarla al siguiente modelo formal simple:

$$\begin{aligned} E(Y_i) &= \mu_i^u + X_i(\mu_i^l - \mu_i^u) \\ &= \mu_i^u + X_i\beta \end{aligned} \quad (3.5)$$

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:

$$\begin{aligned} E(Y_i | X_i = 0) &= \mu_i^u + (0)\beta \\ &= \mu_i^u \end{aligned}$$

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

#### 5.1 Regla 1: elaborar teorías falsables

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-

De la misma manera, cuando un congresista demócrata se presenta en la circunscripción  $i$ , el valor esperado es  $\mu_i^l$ :

$$\begin{aligned} E(Y_i | X_i = 1) &= \mu_i^l + (1)\beta \\ &= \mu_i^l + \beta \\ &= \mu_i^l + (\mu_i^l - \mu_i^u) \\ &= \mu_i^l \end{aligned}$$

De este modo, la ecuación (3.5) proporciona un modelo útil de inferencia causal, siendo  $\beta$  —la diferencia entre las dos proporciones teóricas— nuestro efecto causal. Para terminar, simplifiquemos por última vez esta ecuación para poder remitirnos a ella en el futuro. Si presuponemos que la media de  $Y_i$  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:

$$E(Y_i) = X_i\beta \quad (3.6)$$

El parámetro  $\beta$  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.

nárlas 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 falsa ni cierta. Subrayar la necesidad de que las teorías sean falsables nos obliga a mantener una perspectiva adecuada sobre la incertidumbre y garantiza 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

### 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 aparecía 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<sup>12</sup>:

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

Nuestra estimación de  $\beta$  es simplemente la de regresión por mínimos cuadrados:

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

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

$$\begin{aligned} E(b) &= E\left(\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}\right) \quad (3.8) \\ &= \frac{\sum_{i=1}^n X_i E(Y_i)}{\sum_{i=1}^n X_i^2} \\ &= \frac{\sum_{i=1}^n X_i \beta}{\sum_{i=1}^n X_i^2} \\ &= \beta \end{aligned}$$

Lo cual demuestra que  $b$  es un estimador no sesgado de  $\beta$ .

Preguntar 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, calibraremos la eficiencia del estimador estándar que representa el parámetro del efecto causal  $\beta$  que aparecía en el apartado 1 de este capítulo. En la ecuación (3.8) hemos demostrado que este estimador no está sesgado, y ahora calcularemos su varianza:

$$\begin{aligned} V(b) &= V\left(\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}\right) \quad (3.9) \\ &= \frac{1}{\left(\sum_{i=1}^n X_i^2\right)^2} \sum_{i=1}^n X_i^2 V(Y_i) \\ &= \frac{V(Y_i)}{\sum_{i=1}^n X_i^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^n X_i^2} \end{aligned}$$

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  $\sigma^2$ ), más variable será nuestro estimador  $b$ , y esto no debería sorprendernos. Además, cuanto mayor sea la varianza observada en la variable explicativa ( $\sum_{i=1}^n X_i^2$ ), 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 registran al colocar la regla serán mucho mayores que si estuvieran más separados (situación en la que  $X$  tiene una varianza grande).

### El diseño de la investigación social

una teoría (verificación) y desmentirla (falsación). Lo primero es prácticamente irrelevante, mientras que lo segundo es la clave de la ciencia. Popper cree que una teoría, una vez que se enuncia, pasa a convertirse de inmediato 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. 232). 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 admite 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 (véanse 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.

### 3. Causalidad e inferencia causal

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 *explicaciones 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 ésta 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 aún 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 planteando el mejor cálculo (una estimación) y midiendo su incertidumbre. La conclusión a la que lleguemos respecto a si la inferencia concuerda o no concuerda con nuestra teoría, tendrá el mismo efecto sobre la fe que tenemos en ésta. Tanto la coherencia como la incoherencia aportan información sobre la verdad de la teoría y tienen que influir en la certidumbre de nuestras creencias<sup>13</sup>.

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 predecir 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-

que ciertos errores de pronóstico (y pocos investigadores dudan de ello). Además, la predicción puede haber sido fruto de la suerte o quizás, en los años en que se han recogido los datos, las campañas no hayan incorporado ninguna táctica innovadora (y, por tanto, impredecible).

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 opinión no cambia de forma reseñable, de manera que la teoría de que las campañas presidenciales no tienen impacto se mantendría. De hecho, si hicieramos mil comprobaciones similares 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.

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 reconoció que era imposible), aprenderíamos más de esta manera que con mil comprobaciones que concordaran con ella.

Para nosotros, no es así como se hacen las ciencias sociales o como deberían hacerse. Después de hacer mil comprobaciones a favor o en contra, no prescindiríamos de la teoría que afirma que las campañas no tienen impacto, sino que la modificaríamos para decir, quizás, 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 seguiría 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.

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, propugnaremos que se intente dar a ambas una aplicación amplia. Ésta 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

pretensión—, o bien se presentan de forma tautológica (en cuyo caso ni son verdaderas ni falsas) o de manera que es posible desmentirlas empíricamente (en cuyo caso descubriremos que hacen predicciones incorrectas). Las teorías sociales más útiles son válidas en determinadas condiciones (en campañas electorales en las que no hay pruebas concluyentes de comportamiento inmoral por parte de ningún candidato) o en ciertos contextos (en países industrializados, pero no en los menos desarrollados; en las campañas para elegir la Cámara de Representantes, pero no en las del Senado). Siempre hay que intentar concretar los límites de aplicación de la teoría o hipótesis. El paso siguiente es preguntarse: ¿por qué existen esos límites?, ¿qué tienen las elecciones al Senado que invalidan las generalizaciones aplicables a las de la Cámara de Representantes?, ¿qué hace que la industrialización transforme los efectos causales?, ¿qué variable falta en nuestro análisis que podría generar una teoría de aplicación más general? Al hacernos estas preguntas traspasamos los límites de nuestra teoría o hipótesis para mostrar qué factores hay que considerar con el fin de ampliar su campo de acción.

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 tiramos la teoría inmediatamente por la borda. Podemos hacer varias cosas: llegar a la conclusión de que quizás los datos sólo son pobres por casualidad o ajustar lo que consideramos que es el ámbito de aplicación de una teoría o hipótesis aunque no se sostenga en un caso determinado y, en virtud de ese ajuste, seguir aceptando esa elaboración teórica. La ciencia avanza mediante tales ajustes, pero éstos pueden ser peligrosos. Si los llevamos demasiado lejos haremos que nuestras teorías sean invulnerables al desmentido. 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 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 comprender<sup>14</sup>. Además, este enfoque puede producir una «teoría» que no sea más que una mezcolanza 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 concisión, 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

entus  
curti  
cor.  
"obr  
j'ay

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 utilizaríarnos una teoría poco concisa, ya que ésta maximiza mejor el control que la más concisa<sup>15</sup>.

## 5.2 Regla 2: construir teorías que tengan coherencia interna

Una teoría sin coherencia interna no sólo es falsable; 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 sostenerla. 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 desarrollada está la *creación de modelos formales*, aunque es cada vez más habitual en sociología, psicología, ciencia política, antropología y en otras disciplinas (véase Ordehook, 1986). Los polítólogos 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 simplificadores y a veces no tienen gran valor

empírico. Sólo 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. Estos 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 tales 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 complementarlas 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 segundos. 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 de los partidos. 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

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 quizás 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 explicativa como 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. Quizás 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 cometan 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 manifiestamente endógenas.

En segundo lugar, no hay que seleccionar observaciones que estén basadas en la variable dependiente para que ésta 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-

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, supongamos que se seleccionan observaciones en dos categorías de una variable explicativa y que resulta que la dependiente es constante en los dos grupos. Éste 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 variación que queremos explicar. Aunque esta indicación parezca evidente, el asunto es bastante delicado, como pone de manifiesto Stanley Lieberson (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, imaginémonos un estudio en el que se dejen caer 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 falta en nuestros procedimientos si podemos abordar este problema sin siquiera referirnos a la gravedad.

En este ejemplo, los procedimientos del investigador sólo 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 Lieberson 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 supiéramos todo de la gravedad, este experimento nos seguiría dando una información valiosa, pero si, como presupone el autor, lo que nos interesa realmente es hacer una inferencia sobre el efecto causal de la gravedad, necesitaríamos 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-

## El diseño de la investigación social

teresa 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 definimos (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 ése 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 seleccionemos observaciones a partir de nuestras variables explicativas, dejando a la dependiente 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.)

### 5.4 Regla 4: maximizar lo concreto

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 diferenciadoras. La acción de un individuo o de una nación puede explicarse a partir de un deseo de maximizar la utilidad, materializar unas intenciones o trabajar en beneficio del interés nacional; pero la prueba de que esa acción ha logrado tales fines residirá 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.

## 3. Causalidad e inferencia causal

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, quizás 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», y observa que las instituciones federales estadounidenses se diferencian entre sí por el periodo medio de permanencia de los trabajadores en sus empleos y con este dato mide la «institucionalización» de dichos 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 decisarios» 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-

tado lo recomendamos). La ciencia se basa en estas clasificaciones abstractas, si no estaríamos volviendo al resumen de los pormenores 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 justificar 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 llevan a cabo nuevas investigaciones, que utilizan otros indicadores específicos, con el fin de reforzar el supuesto de que los nuestros 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 dónde 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 teóricas inobservables, 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 significa institucionalización, 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» ampliando 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, explicitamente, las consecuencias observables de la teoría e incluso los posibles proyectos de investigación que podríamos realizar.

Cuanto más impreciso sea nuestro lenguaje, menos posibilidades habrá de que nos equivoquemos, 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 específicas<sup>16</sup>.

### 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 instáramos 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.

### **El diseño de la investigación social**

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 Eckstein, 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 fijación 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 quizás 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 inferencia 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. Estas decisiones tienen gran importancia, ya que una mala selección de casos puede impedir que, en estudios posteriores, la más imaginativa de las iniciativas logre extraer inferencias causales válidas. En el capítulo 5 señalamos algunas de las principales causas de sesgo e ineficiencia que habría 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-

### El diseño de la investigación social

tes. 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 cuantitativistas 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 limitaciones en investigaciones con un  $n$  pequeño. Si estas estrategias no resultan apropiadas, podemos lograr la homogeneidad de las unidades seleccionando intencionadamente 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 las 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 cumpla 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, abordaremos el problema del sesgo de selección que surge del no cumplimiento de los supuestos de independencia condicional y de homoge-

neidad de las unidades. En el apartado 2 estudiamos los límites que tienen la selección de observaciones y la asignación de valores de forma aleatoria a la hora de lograr la independencia condicional. En el apartado 3 subrayamos los peligros que conlleva seleccionar casos intencionadamente a partir de los valores de las variables dependientes y damos ejemplos de trabajos en los que este sesgo de selección ha invalidado las inferencias causales. Para terminar, en el apartado 4 proponemos de manera sistemática formas de lograr la homogeneidad de las unidades a través de una selección de casos intencionada, no sólo con el fin de dar consejos sobre cómo hacer diseños de investigación ideales, sino para indicar cuáles son las alternativas «posibles» cuando no se puede lograr esa situación óptima.

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» suelta 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 combate 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 conllevan sus propias unidades y variables. Un investigador podría 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 cualquiera 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-

## El diseño de la investigación social

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

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, quizás 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, quizás 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, quizás de forma tardía, podría optar por utilizar cada uno de los estados indios como unidad de análisis; quizás también dividiendo su período de estudio en cuatro o cinco subperiodos. 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 caso» 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 si lo están, al menos, en cada investigación concreta. Sin embargo, algunos diseños son imprecisos, o sea, que no puede saberse casi 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<sup>1</sup>. Sin embargo, los programas informáticos no siempre funcionan como de-

### 4. Precisar qué va a observarse

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 pormenores históricos no puede ser impreciso, a menos que directamente se reúnan 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 recojamos 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 ilimitado, y resulta que la regla precisa es que un hecho (o consecuencia observable) sólo puede ofrecer información independiente acerca de otro hecho. De manera más general, cada observación puede ayudarnos a extraer, como

(4) máximos, una inferencia; un número de observaciones nos ayudará a hacer una cantidad de inferencias menor que 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 desarrollar 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 habernos 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 recogieramos los datos de la manera más cuidadosa, no nos librariamos 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 cuál de las hipótesis es correcta (si es que hay alguna que lo sea).)

Al enfrentarnos 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

#### 4. Precisar qué va a observarse

incluirlos quizás cueste un tiempo y unos recursos de los que no disponemos o quizás 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 desarrollar 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, no será necesario cambiar en absoluto el propósito del proyecto. De este modo, si nuestra teoría se especifica con propiedad podrá tener muchas consecuencias observables y nuestros datos, sobre todo los cualitativos, contendrán, en general, observaciones referentes a muchas de esas consecuencias. Si es así, cada estudio de caso puede convertirse en muchas observaciones 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 dispondremos de esta última información (por la aleatoriedad inherente a  $Y$ ), de manera que si, en este caso, tenemos problemas al hacer una estimación de  $\beta_1$  y de  $\beta_2$ , 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 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.

#### 4. Precisar qué va a observarse

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 no 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 cualquiera de las 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 interese 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.

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.<sup>2</sup> 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 politológica, hemos de abandonar la aleatoriedad, habrá que hacerlo con cuidado.

En las ciencias sociales los experimentos controlados sólo se realizan ocasionalmente.<sup>3</sup> 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 recibirlas de los procesos sociales) y dar tratamientos (valores de 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 uni-

## 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:

$$\begin{aligned} E(Y) &= X_1\beta_1 + X_2\beta_2 \\ &= X_1(\beta_1 + \beta_2) \end{aligned} \quad (4.3)$$

La segunda línea de esta ecuación debería dejar claro que, independientemente de lo que sean  $E(Y)$  y  $X_1$ , numerosos valores de  $\beta_1$  y de  $\beta_2$  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 quisieramos 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

### El diseño de la investigación social

circunscripciones electorales para el Congreso. [Podríamos juntar listas procedentes de varias fuentes, pero siempre correríamos el riesgo 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, quizás no pudieramos 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 vivieran 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 del sesgo 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 en investigaciones con un  $n$  reducido la selección aleatoria de 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 1 se refiera a una unidad seleccionada para el análisis y que 0 indique una omitida, sólo serán posibles tres criterios de selección: (1,1,0), lo cual significa que tomamos 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<sup>4</sup>. 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 utilizará el conocimiento previo de los posibles valores de la variable dependiente, optara por el tercer criterio de selección, evitando así el sesgo.

### 4. Precisar qué va a observarse

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 húngaro, 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 varianza 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 realizar, 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, ésta podría señalar, partiendo de los casos checo y polaco, que los movimientos antigubernamentales masivos surgen cuando los regímenes comunistas tienen un carácter indulgente, relativamente

mente poco represivo, y no cuando son fuertes y represivos. La investigadora tendría que saber que, aunque Hungría tenía el régimen comunista más benévolos 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 mucho menos pormenorizados que se basaran en fuentes secundarias y, quizás, 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 hubieran 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-

plo 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 responden 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 «cómodamente» de lado. El sesgo de selección casi nunca es tan flagrante como en este ejemplo, pero como en la investigación cuantitativa 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<sup>3</sup>.

### 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<sup>6</sup>.

En el capítulo I 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 cualitativistas, 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-

### El diseño de la investigación social

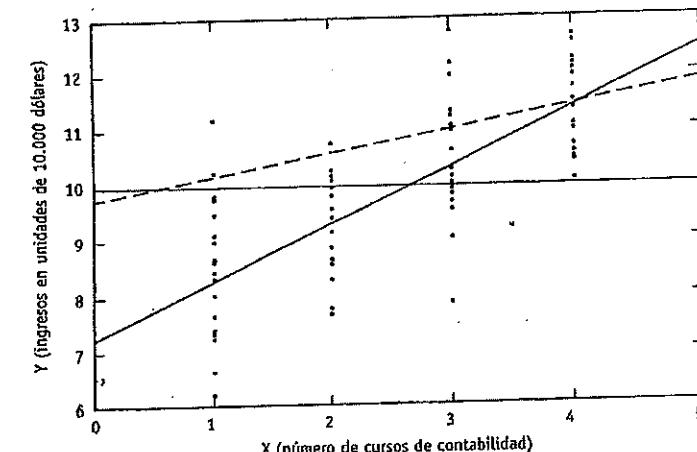
gen a partir de un determinado valor de la variable dependiente, no se puede conocer ninguna de las causas de ésta sin tener en cuenta otros ejemplos en los que adopte valores diferentes. Por ejemplo, en su investigación, Theda Skocpol (1979) soluciona en parte este problema incluyendo unos pocos datos 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, cualquier criterio de selección que se correlacione con la variable dependiente atenuará la media de estimaciones de los efectos causales (véanse Achen, 1986; King, 1989, capítulo 9). En la investigación cuantitativa, esto supondrá 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 él habría que tener pruebas detalladas o, por lo menos, cierta idea.

Los casos de sesgo de selección extremo —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 quizás no tengamos acceso a todas las observaciones que queremos. Sin embargo, por fortuna, los efectos de este sesgo no son tan devastadores, porque algo sí se puede aprender; quizás nuestras inferencias

### 4. Precisar qué va a observarse

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 sí 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 dependiente (por ejemplo, el salario inicial en el primer trabajo con dedicación exclusiva, cifrado 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-

mine sus estudios. Como no le han enseñado lo que es el sesgo de selección, opta por utilizar una muestra de anteriores estudiantes que sólo se compone de aquellos a los que les fue bien en su primer empleo: los que consiguieron trabajos como los que él quiere. Puede parecer que si desea aprender cómo ganar más dinero lo mejor es centrarse únicamente en los que tienen salarios altos, pero este razonamiento es falaz. Para una mayor sencillez, supongamos que en su selección incluya sólo a aquellos que están ganando, como mínimo, 100.000 dólares. En el gráfico 4.1, estos criterios de selección de la muestra se expresan mediante una línea continua horizontal en la que  $Y = 10$ , de manera que en la investigación del estudiante sólo se contemplan los puntos que aparecen por encima de ella. Ahora, en lugar de hacer una línea de regresión que se ajuste a todos los puntos, traza otra (la discontinua) que sólo se ajusta a los de su muestra. El sesgo de selección se manifiesta en que la inclinación de esta línea es menor que la de la continua. A consecuencia de este sesgo, el estudiante llegará a la conclusión errónea de que cada curso de contabilidad sólo representa un aumento de unos 5.000 dólares.

*Este es un ejemplo concreto de cómo podemos subestimar un efecto causal cuando seleccionamos en función de la variable dependiente. Por fortuna, sí 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. Está 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 cambiaron sus valores.*

### 3.1.1. Ejemplos de sesgo de selección producido por el investigador

El problema que acabamos de describir es habitual en la investigación cualitativa (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

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 cualitativa. Supongamos que en el estudio anterior nos interesaría 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, quizás basadas en dos docenas de observaciones, si 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 conllevan 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 ésta necesitábamos 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 era aún más intensa de lo que parecía en nuestro primer análisis.

El problema del 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

## El diseño de la investigación social

las tácticas utilizadas por los grupos de disidentes podría obtener resultados sesgados, especialmente si sólo estudiara los lugares en los que se le ha permitido la entrada, ya que las razones por las que el régimen le ha concedido este acceso probablemente se correlacionen con la variable dependiente, 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.<sup>7</sup> 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 quizás 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 observada 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. Éste 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

### 4. Precisar qué va a observarse

autor no puede determinar si la ausencia de las hipotéticas variables causales se relaciona con el fracaso en términos competitivos. De este modo, no hay forma de saber si las condiciones que él relaciona con el éxito no se asocian también con el fracaso. En su provocativo trabajo, Porter plantea un conjunto fascinante de hipótesis que se basa en los casos en los que si ha habido éxito competitivo, pero al no elaborar una gradación de éxitos o fracasos en este sentido (ni una selección fundamentada en otra cosa que no sea su variable dependiente), no puede saber si tiene toda la razón, ninguna, o si está en un punto intermedio<sup>7</sup>.

En los escritos sobre política exterior que se ocupan de la disuisión se encuentra un ejemplo notable de sesgo de selección: «el uso de amenazas para inducir a los oponentes a que se comporten de la forma deseada» (Achen y Snidal, 1989, p. 151). Los que investigan este asunto analizan con frecuencia «crisis agudas», o sea, aquellas en las que la disuisión no se ha producido en un primer estadio del proceso que lleva del cálculo político a la acción, pasando por el de emisión de ciertas señales. Para describir mejor este enfoque, hay que caracterizarlo bien, al menos inicialmente: al igual que ocurre cuando Porter hace hincapié en el éxito competitivo, el observador puede describir los episodios de interés más significativos sin quizás lograr formular hipótesis sobre las causas de los resultados observados. Sin embargo, para fundamentar inferencias (y sin una corrección adecuada), este sesgado conjunto de observaciones presenta graves problemas porque, de forma sistemática, se han excluido del análisis aquellas en las que la disuisión funcionó (en estadios preliminares del proceso). «En consecuencia, cuando se utilizan mal los casos para hacer una estimación del índice de éxito de la disuisión, el diseño genera un «sesgo de selección» como el que es frecuente en los estudios de evaluación de políticas» (Achen y Snidal, 1989, p. 162).

#### 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, supongamos 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 éste parece ser el caso), se producirá un sesgo de selección cuando elijamos circunscripciones en las que

si haya avalado a algún candidato. En este ejemplo, la selección utilizada forma parte del proceso político que estudiamos, pero sus consecuencias 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 incluimos 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 utilizar madera, 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? o ¿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 debería 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 guerras o fueron incorporadas a entidades mayores, mientras que otras habrán sobrevivido. Por lo tanto, mediante una categorización cuidadosa, podríamos obtener una variable dependiente que indicara 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 quinientas 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-

4. Precisar qué va a observarse

pecto a este punto, las historias de Francia, Alemania, España, Bélgica e Inglaterra (o las de cualquier otro grupo de países en las épocas moderna y contemporánea) para arrojar luz sobre los procesos de configuración del Estado-nación inclina la balanza de la investigación hacia un cierto resultado 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, supongamos 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, infiriendo 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 Seleccionar 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<sup>8</sup>. 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

### El diseño de la investigación social

cifra 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<sup>9</sup>.

[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, ésta 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 ocurrirá 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 intentonas 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 existía 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. Precisar qué va a observarse

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 estudiáramos 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 quizás 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. Éste 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 una observación a otra. Esta posibilidad es real, aunque no nos hayamos centrado en ella. En el apartado 1 del capítulo 3 definimos el efecto causal en una sola unidad y dejamos que variara en cada una de ellas. Supongamos, por ejemplo, que nos interesa 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 exageraremos 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.)

#### 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 «naturaleza» 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.

*Bueno*

Seleccionar de manera intencionada las observaciones implica que, como mínimo, conocemos 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 de ser peligroso, ya que sin darnos cuenta podemos introducir un sesgo de selección que quizás 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 interesaran 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-

##### 4. Precisar qué va a observarse

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 interesarán 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 (ó 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 trascendental variable de control, Snyder no pudo hacer inferencias sobre las consecuencias que tenía el poderío militar sobre el exceso de expansión, pero si 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 eficiente 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:

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 ocuparse 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 Karnatake, que representaba un punto intermedio entre los anteriores. Los tres estados fueron seleccionados íntegramente 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 también de la variable dependiente. En estas circunstancias, el diseño era impreciso y no daba información sobre la hipótesis causal. Es decir, ésta 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 su 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 *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.

#### 4. Precisar qué va a observarse

##### 4.4 Seleccionar 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, Inkeles 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 hubiera resultado apropiado hacer una firme inferencia causal, pero sí era razonable llegar a una conclusión más provisional que hacia 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 achacado 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 Laitin (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 él mismo señala, la religión, que es su variable explicativa, es una constante en toda*

Somalia y, además, tiene una relación multicolineal (véase el apartado 1 de este mismo capítulo) con otras variables, lo cual hace imposible que se aísle 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 son 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 religiosa» (1986, p. 186). Laitin supera este problema centrando su atención 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.<sup>1</sup>

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 algunas plantaciones de Virginia sí utilizaban fertilizantes, podría tener sentido observar ambos estados para explicar la variación respecto al uso del producto; al menos hay una diferencia entre los estados que constituirán 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 recurriían a él podría constituir una considerable aportación a lo escrito sobre el tema. Por lo menos, 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 polítólogos, las armas nucleares no se han utilizado como arma de guerra desde 1945. Sin embargo, aun-

#### 4. Precisar qué va a observarse

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 había, al menos, dos grandes potencias nucleares) se ha mantenido constante durante todo el período. 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 sugestivas, no hay que considerar que confirman nuestra hipótesis; únicamente que indican 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

#### El diseño de la investigación social

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 investigador 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.

## 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 reducirlas 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, podríamos 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

tiempo. La falta de sesgo tiene que ver con el hecho de centrar un intervalo cerca de la estimación correcta, mientras que la eficiencia se relaciona con el estrechamiento de un intervalo que se ha centrado correctamente.

Estas definiciones de falta de sesgo y de eficiencia son siempre aplicables, independientemente de que se quieran hacer inferencias descriptivas —como en el ejemplo de la edad— o causales. Si, por ejemplo, queremos hacer una estimación del efecto que tienen los estudios en la renta (los ingresos monetarios que se perciben por cada año de estudio), tendremos un punto de estimación del efecto que estará flanqueado por un intervalo que refleja nuestra incertidumbre respecto a la cantidad exacta. Necesitaremos un intervalo que sea lo más estrecho posible (para una mayor eficiencia) y cuyo centro esté cerca de la estimación correcta (para evitar sesgos); también será necesario que el cálculo aproximado del ancho del intervalo presente sinceramente nuestra incertidumbre.

En este capítulo nos centramos en las causas del sesgo y de la ineficiencia, desde ese primer momento del estudio en el que pretendemos mejorar la calidad de la información hasta cuando llegamos a extraer inferencias causales. En el apartado 1 analizamos los errores de medida que pueden sesgar nuestros resultados y hacerlos también menos eficientes. Posteriormente, en el apartado 2, pasamos a considerar el sesgo que puede aparecer en nuestras inferencias causales cuando hemos eliminado del análisis variables explicativas que se deberían haber incluido. En el apartado 3 abordamos el problema contrario: controlar las variables irrelevantes que reducen la eficiencia de nuestro análisis. A continuación, estudiaremos el problema que se plantea cuando nuestra variable «dependiente» afecta a las «explicativas». Este problema se conoce con el nombre de «endogeneidad» y se aborda en el apartado 4. Para terminar, en los apartados 5 y 6 analizaremos, respectivamente, la adjudicación aleatoria de valores en las variables explicativas, así como diversos métodos de control no experimental.

### 1. Error de medida

Una vez que hemos seleccionado nuestras observaciones, tenemos que medir los valores de las variables que nos interesan. Como en las ciencias sociales todas las observaciones y medidas son imprecisas, nos encontraremos enseguida con el problema del error de medida.

En la investigación social la mayoría de los análisis intentan calcular el nivel de error y reducirlo lo más posible. Los estudios cuantitativos disponen de medidas más precisas (numéricas), pero no necesariamente más certeras. La fiabilidad (el que diversas medidas del mismo fenómeno tengan el mismo resultado) a veces se obtiene sacrificando la validez (que las medidas reflejen lo que el observador trata de calibrar). Los cualitativistas intentan obtener medidas certeras que, sin embargo, suelen ser un poco menos precisas.

En ciertos rasgos esenciales, la observación cualitativa y la medición cuantitativa son muy parecidas. Es cierto que en la primera las categorías se suelen etiquetar con palabras, mientras que en la segunda se asignan valores numéricos a las categorías y medidas; sin embargo, los investigadores de una y otra tendencia utilizan criterios nominales, ordinales y de intervalo. Con los de tipo nominal, se conforman grupos de observaciones en los que no se presupone que las categorías tengan un determinado orden. Las que son de relevancia pueden basarse en el ordenamiento legal o institucional; por ejemplo, en la política comparada, pueden ser de interés las pautas de los sistemas presidenciales, parlamentarios o autoritarios en diversos países. Las categorías ordinales dividen los fenómenos siguiendo algún tipo de orden. Por ejemplo, un investigador cualitativo puede clasificar los países en tres o cuatro categorías, teniendo en cuenta su grado de industrialización o el tamaño de sus fuerzas armadas. Para terminar, la medición a través de intervalos utiliza variables continuas, como se hace en los estudios de flujos comerciales internacionales.

Las diferencias que hay entre la medición cuantitativa y la cualitativa afectan a la presentación de los datos, no a la posición teórica de dicha medida. Los cualitativistas utilizan en sus medidas palabras como «más» o «menos», «mayor» o «menor», «fuerte» o «débil», mientras que los cuantitativistas usan números.

Por ejemplo, la mayoría de los investigadores cualitativos del campo de las relaciones internacionales es muy consciente de que el «número de muertos en combate» no tiene por qué ser un buen indicador de la importancia que tienen las guerras para las tendencias posteriores de la política mundial. En la teoría del equilibrio de poderes se considera que, para calibrar la estabilidad, el concepto teórico relevante no es la gravedad de la guerra sino el cambio que se produce en los principales actores «a consecuencia» de ella (véanse Gulick, 1967, y Waltz, 1979, p. 162). Sin embargo, al querer evitar la falta de validez, el cualitativista corre el riesgo de que los errores de medida le hagan no ser fiable. ¿Cómo podemos saber lo que se produce «a consecuencia» de algo si no definimos precisamente esa expresión? De hecho, ésta parece implicar que la valoración se hace en función de un resultado sistemático (que sesgará las estimaciones posteriores de la relación, llevándolas hacia la hipótesis).

No hay ninguna fórmula que pueda determinar cómo se logra el equilibrio entre el uso de indicadores cuantitativos, que quizás no reflejen de forma válida los conceptos subyacentes que nos interesan, y el de valoraciones cualitativas siempre imprecisas y sujetas a sesgos inconscientes. Sin embargo, los investigadores de una y otra tendencia tienen que hacer una estimación de la incertidumbre en sus inferencias. Los cuantitativistas han de presentar, junto a sus medidas numéricas, cuáles son los errores típicos, mientras que los cualitativistas tienen que hacer una estimación de la incertidumbre, redactando cuidadosamente lo que piensan de sus observaciones.

### El diseño de la investigación social

La diferencia entre las mediciones de unos y otros estriba en su forma de presentar ideas que, en esencia, son las mismas.

Las mediciones cualitativas y cuantitativas también se parecen en otro sentido. En ambas, las categorías suelen ser mecanismos creados por el investigador y no vienen «dadas» por la naturaleza. La clasificación de las naciones en regímenes democráticos y autocráticos o parlamentarios y presidenciales depende de elaboraciones intelectuales, al igual que ordenarlas según criterios como el de mayor o menor industrialización.

Es evidente que no existen respuestas con validez universal: todas las mediciones dependen del problema que el investigador pretenda comprender. Lo mejor es que los criterios que forman las categorías estén lo más cerca posible de las primeras ideas teóricas y empíricas del investigador; sin embargo, esta misma pretensión recalca todavía más el hecho de que dichas categorías son mecanismos al servicio de los propósitos del investigador. El número de regímenes parlamentarios en los que la proporcionalidad es el principal sistema de representación depende de cómo se clasifiquen esos «regímenes parlamentarios» y de lo que se entienda por sistema de representación proporcional. En el campo de las relaciones internacionales es posible investigar los flujos monetarios registrados, pero el uso de una medida continua depende de qué transacciones se haya decidido computar, de las reglas que determinan lo que es una transacción y de cómo se definen las fronteras nacionales. Del mismo modo, en unas elecciones al Congreso estadounidense, la proporción de voto demócrata en una circunscripción depende de que, en la clasificación que haga el investigador, para su estudio, se presuponga que las etiquetas «demócrata» y «republicano» tienen el mismo significado en las 435 circunscripciones.

Incluso los criterios de categorización que hemos utilizado en este apartado para hacer mediciones (nominales, ordinales o de intervalo) dependen de para qué fines teóricos se utilicen éstas. Por ejemplo, podría parecer evidente que la etnia es un prototipo de variable nominal que, en los Estados Unidos, podría codificarse como negro, blanco, latino, indio y asiático. Sin embargo, existe una gran diversidad dentro de estos grupos nominales, en función de la intensidad con que los miembros de cada grupo se identifican con él. Por lo tanto, podríamos categorizar a los grupos étnicos según una escala ordinal, basándonos, por ejemplo, en la proporción de miembros de cada uno de esos grupos que está muy identificado con aquel al que pertenece. Quizá nos interese también el tamaño de estos grupos: por lo tanto, podríamos utilizar la etnia como criterio de medida para determinar ciertos intervalos. Lo principal es utilizar la medida más apropiada para nuestros fines teóricos.

Los problemas de medición más frecuentes aparecen cuando se mide sin hacer referencia explícita a ninguna estructura teórica. Por ejemplo, a veces los investigadores toman una variable que es continua de forma natural y que puede medirse bien, como la edad, y establecen categorías de juventud,

### 5. Entender qué debe evitarse

mediana edad y vejez. Para ciertos fines, esta clasificación puede ser suficiente, pero, como presentación teórica de la edad de una persona, el procedimiento resulta innecesariamente impreciso. El error de agrupación que se producirá será considerable y hay que evitarlo. El hecho de librarse de este tipo de error nos remite al principio de que no hay que descartar datos sin que haya razón para ello.)

Sin embargo, podemos cometer el error contrario: asignar valores continuos y numéricos de intervalo a variables que por naturaleza son discontinuas. En general, la medición mediante intervalos no es mejor que la ordinal o nominal. Por ejemplo, en una encuesta se puede preguntar sobre la filiación religiosa y sobre el grado de compromiso en este sentido. La intensidad del segundo indicador podría medirse —si las preguntas están bien planteadas— mediante una variable ordinal, incluso utilizando intervalos, dependiendo de cuál sea la naturaleza del instrumento de medición. Sin embargo, no tendría mucho sentido hacer una lista numérica con las religiones a las que pertenece cada individuo. En este caso, probablemente no exista una variable ordinal o continua, y este procedimiento generaría un error de medida.

Elegir entre categorías nominales, por un lado, u ordinales o de intervalo, por otro, puede implicar un intercambio entre riqueza descriptiva y facilidad a la hora de comparar. Pensemos, por ejemplo, en las normas de voto de las organizaciones internacionales. El ordenamiento institucional que las dicta es importante porque refleja una concepción de la soberanía de los estados y porque tiene consecuencias sobre la clase de resoluciones que pueden aprobarse, los recursos que se destinan a la organización y las expectativas de que se cumplan las resoluciones.

Un conjunto de categorías nominales podría distinguir entre aquellos sistemas en los que un único miembro tiene la facultad de vetar cualquier resolución (como ocurría en el Consejo de la Sociedad de Naciones, según el artículo 15 de su Acta de Constitución), aquellos en los que sólo pueden hacerlo ciertos miembros (como en el Consejo de Seguridad de las Naciones Unidas), en los que se decide en función de algún tipo de mayoría cualificada (como en las decisiones que afectan al mercado interno de la Unión Europea) y en los que es una mayoría simple la que toma las decisiones (como ocurre en la mayoría de las votaciones de la Asamblea General de las Naciones Unidas). Cada sistema puede generar sus propias dinámicas de negociación, y, si lo que pretendemos es estudiar cómo funciona uno de ellos en particular (aquel en el que un miembro puede ejercer el veto), es esencial que definamos nuestras categorías para no incluir en el análisis otros tipos de sistemas. Las categorías nominales serían apropiadas para este propósito.

Sin embargo, también podríamos ver estas categorías de forma ordinal, yendo de las más restrictivas (en las que se requiere unanimidad) a las menos (en las que se precisa mayoría simple). Esta categorización sería ne-

saria si tuviéramos que comprobar proposiciones teóricas que tuvieran que ver con la relación que existe entre lo restrictivas que son las normas de voto y las pautas de negociación o con el carácter distributivo de los resultados típicos. No obstante, al menos dos de nuestras categorías —las que contemplan el voto de ciertos miembros y la decisión por medio de mayorías cualificadas— son bastante confusas, ya que suponen la existencia de un abanico de diversos arreglos. La primera categoría permite que un solo miembro ejerza un voto total —práctica que raya en lo dictatorial— y también que veten todos los miembros, a excepción de los pocos que apenas tienen peso; en la segunda se incluye la reglamentación de la Unión Europea, que impide que dos estados constituyan una minoría de bloqueo en decisiones que afectan al mercado interno. La fórmula que utiliza el Fondo Monetario Internacional es, desde el punto nominal, un ejemplo de decisión por mayoría cualificada, pero, en la práctica, concede a los Estados Unidos y, desde hace poco, a la Unión Europea, que actúa como un bloque, la capacidad de constituir este tipo de minoría de bloqueo. De ahí que pueda pertenecer a cualquiera de las dos categorías.

En consecuencia, podríamos intentar dar un paso adelante y hacer una medida intermedia que se basara en la proporción de estados (o de recursos: según el producto nacional bruto, las aportaciones a la organización o la población) que se precisa para aprobar resoluciones, clasificando las organizaciones internacionales según lo restrictivos que sean sus procedimientos de voto.

Sin embargo, con diferentes baremos —por ejemplo, si se calibran los recursos en función de la población o del producto nacional bruto— se obtendrán resultados diferentes, por lo que las ventajas que pueda tener la precisión de tales medidas se verán limitadas por la arbitrariedad de los criterios de medición o por la complejidad de los datos agregados. Cada categoría tiene ventajas e inconvenientes: los fines del investigador deben determinar qué opción se toma.

En los dos siguientes subapartados vamos a analizar las consecuencias específicas del error de medida para la investigación cualitativa y llegaremos a conclusiones que pueden parecer sorprendentes: Casi nadie discutiría que un error sistemático de medida, como el que supone exagerar de manera constante el valor de ciertas unidades, causa sesgo y también incoherencia, ya que dicho sesgo no puede desaparecer si se incorporan más observaciones cargadas de errores. Sin embargo, si analizamos el problema con más atención, veremos que sólo ciertos tipos de error de medida sistemático sesgan nuestras inferencias causales. Además, las consecuencias del error de medida *no sistemático* están menos claras. Vamos a analizar este último tipo de error en dos partes: en la variable dependiente y, después, en la explicativa. Tal como vamos a demostrar, el error en la variable dependiente produce ineficiencias que pueden generar resultados incorrectos en cualquier situación y dificultar el rastreo de pruebas constantes de la exis-

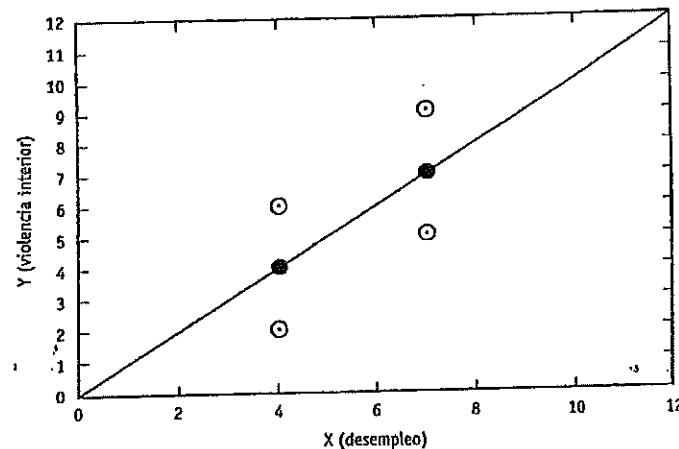
tencia de efectos sistemáticos. Dicho de otro modo, el error de medida no sistemático en la variable dependiente no produce sesgo, pero puede aumentar considerablemente la ineficacia. Más interesante resulta la presencia de este tipo de error en la variable causal clave, pues siempre sesga las inferencias de forma predecible. Comprender la naturaleza de estos sesgos nos ayudará a atenuarlos o quizás a evitarlos.

### 1.1 Error de medida sistemático

En este apartado nos ocupamos de las consecuencias del error de medida *sistemático*, que consiste en exagerar de manera constante el valor de ciertos tipos de unidades y que, a veces, puede sesgar y hacer incoherente la estimación de los efectos causales. Nuestro cometido es determinar qué tipos de sesgo produce cada error de medida sistemático. Tanto en la investigación cuantitativa como en la cualitativa, este error puede derivarse del hecho de que los investigadores tomen decisiones que hagan que los datos favorezcan la constatación de sus expectativas. En el trabajo cuantitativo, puede que utilicen datos sesgados porque son las únicas series numéricas disponibles, mientras que en el cualitativo el error de medida sistemático puede provenir de evaluaciones subjetivas realizadas por investigadores que ya tengan hipótesis formadas y que quieren demostrar que son correctas.

Debería ser evidente que *cualquier error de medida sistemático sesgará las inferencias descriptivas*<sup>1</sup>. Pensemos, por ejemplo, en el caso más simple que podamos imaginar, en el que sin darnos cuenta añadimos 1.000 dólares a la estimación de renta anual de cada una de las personas que responde a una encuesta. Es obvio que en nuestro cálculo de la renta media anual de toda la muestra habrá también una sobreestimación por ese mismo valor. Si nos interesa hacer un cálculo aproximado del efecto causal que tienen los estudios universitarios en la media de renta anual, el error de medida sistemático no tendría consecuencias sobre nuestra inferencia causal. Si, por ejemplo, el grupo de personas con estudios universitarios ganara realmente una media de 30.000 dólares, pero el grupo de control que no fue a la universidad ingresara una media de 25.000, nuestra estimación del efecto causal que tienen los estudios universitarios sobre la renta anual mostraría un incremento de 5.000 dólares. Si la renta de cada persona en ambos grupos se hubiera sobreestimado en la misma cantidad (digamos, 1.000 dólares de nuevo), nuestro efecto causal —que ahora procede de restarle 26.000 dólares a 31.000— seguiría siendo 5.000. Por lo tanto, *el error de medida sistemático que afecta a todas las unidades por medio de una misma cantidad constante no sesga la inferencia causal* (esto puede apreciarse mejor si nos centramos en la versión relativa a los efectos constantes del supuesto de homogeneidad de las unidades que se expuso en el apartado 3.1 del capítulo 3).

Gráfico 5.1 Error de medida en la variable dependiente



En el gráfico 5.1, los dos círculos negros pueden representar un estudio sencillo en el que no hay error de medida en ninguna variable. Podemos imaginar que disponemos de un gran número de observaciones y que todas coinciden con esos densos círculos negros, de manera que se conoce bastante bien la posición de ambos. Por el contrario, podemos imaginar que sólo tenemos dos observaciones y que en ellas hay poco error no sistemático de cualquier tipo. Evidentemente, no es probable que en la realidad ocurra ninguno de estos casos, pero el modelo subraya los problemas esenciales que plantea el error de medida cuando aparece en la variable dependiente de un caso más general y complicado.

Imaginemos ahora otro estudio en el que los delitos violentos se midan con un error no sistemático. Para recalcar el hecho de que la media de esas medidas es correcta, marcamos con un punto cada uno de los círculos blancos que flanquean simétricamente, por arriba y por abajo, los primeros círculos negros<sup>3</sup>. Una nueva línea que represente los seis círculos de datos será exactamente igual a la primera que hemos trazado. De nuevo, hay que señalar que esta línea se traza minimizando los errores de predicción, las desviaciones verticales de la línea.

Sin embargo, la nueva línea será más incierta en varios sentidos. Por ejemplo, los puntos estarán casi tan bien representados tanto si la línea está un poco más empinada como si es ligeramente más plana. Además, su posición vertical también será más incierta, y ella misma pronosticará peor dónde habría de estar cada uno de los círculos de datos. El resultado es que

## 5. Entender qué debe evitarse

el error de medida en la variable dependiente reduce la eficiencia de las estimaciones. Aunque éstas no tengan sesgo —es decir, en la media de numerosos estudios similares—, sí podrían estar muy alejadas de la realidad en cada uno de ellos.

### 1.2.2 Error de medida no sistemático en una variable explicativa

Como hemos señalado antes, el error no sistemático en la variable explicativa tiene las mismas consecuencias para las estimaciones del valor de esa variable —en las inferencias descriptivas— que para las que afectan al valor de la variable dependiente: a veces las medidas serán demasiado altas, a veces demasiado bajas, pero su media será correcta. Como ocurre con el error no sistemático en la variable dependiente, el error aleatorio en la explicativa también puede hacer que las estimaciones de los efectos causales sean inciertas e inefficientes. Sin embargo, cuando este error se produce en las variables explicativas, tiene una consecuencia bastante diferente de las que se registran cuando aparece en la dependiente. Cuando es la variable explicativa la que se mide con un error aleatorio, en las estimaciones de la relación causal aparece un sesgo sistemático que las inclina hacia un valor nulo, hacia la ausencia de relación. Dicho de otro modo, cuando hay un verdadero vínculo causal entre la variable explicativa y la dependiente, un error aleatorio en la primera puede ocultar este hecho al reducir el valor de la relación. Si queremos contrastar nuestra hipótesis en diversos conjuntos de datos, no sólo encontraremos una gran variedad de resultados, como ocurre cuando hay un error aleatorio en la variable dependiente, sino que también hallaremos un sesgo sistemático en ellos que nos inducirá a pensar que la relación que existe es más débil que la que hay en realidad.

Al igual que ocurre cuando el error de medida se produce en la variable dependiente, aunque nos demos cuenta de que tal error existe en la explicativa, sus *consecuencias* no se reducirán con un análisis más cuidadoso de las variables que se han medido con errores, a menos que se sigan nuestros consejos. Evidentemente, con mejores medidas mejorará la situación.

Pensemos de nuevo en nuestro estudio de los efectos del desempleo en la delincuencia que se registra en varias comunidades de un país subdesarrollado. Sin embargo, supongamos que la situación que reflejan los datos sea la opuesta a la que mencionábamos anteriormente: en el país que estamos estudiando las cifras de delincuencia son precisas y es fácil que las autoridades las proporcionen, pero el desempleo es un problema político y, en consecuencia, no puede medirse con exactitud. Como no se permiten las encuestas con muestras sistemáticas, decidimos calibrar el paro mediante la observación directa (al igual que antes hicimos con la delincuencia). Inferimos el índice de desempleo del número de personas que están

## Un análisis formal del error de medida en $y$

Consideremos un sencillo ejemplo lineal con una variable dependiente medida con errores y una explicativa que carece de ellos. Nos interesa hacer una estimación del parámetro de efecto  $\beta$ :

$$E(Y^*) = \beta X$$

También especificamos un segundo rasgo de las variables aleatorias, la varianza:

$$V(Y_i^*) = \sigma^2$$

la cual presuponemos que es igual en todas las unidades  $i = 1, \dots, n$ .

Aunque estas ecuaciones definen nuestro modelo, por desgracia no observamos  $Y^*$  sino  $Y$ , donde

$$Y = Y^* + U$$

Es decir, que la variable dependiente observada  $Y$  es igual a la verdadera variable dependiente  $Y^*$ , más algún error de medida aleatorio,  $U$ . Para formalizar la idea de que  $U$  contiene sólo un error de medida no sistemático, necesitamos que éste desaparezca en la media de reproducciones hipotéticas,  $E(U) = 0$ , y que no se correlacione con la verdadera variable dependiente,  $C(U, Y^*) = 0$ , ni con la explicativa,  $C(U, X) = 0$ . Además, presuponemos que el error de medida tiene una varianza  $V(U) = \tau^2$  en cada una de las unidades  $i$ . Si  $\tau^2$  vale cero,  $Y$  no contendrá ningún error de medida y será igual a  $Y^*$ ; cuanto más grande sea la varianza, más error contendrá nuestra medida  $Y$ .

¿Cómo afecta el error de medida aleatorio en la variable dependiente a la estimación de  $\beta$ ? Para saberlo, utilizamos nuestro estimador habitual con  $Y$  en vez de con  $Y^*$ :

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

y después calculamos la media en las reproducciones hipotéticas:

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

$$\begin{aligned} &= \frac{\sum_{i=1}^n X_i E(Y_i)}{\sum_{i=1}^n X_i^2} \\ &= \frac{\sum_{i=1}^n X_i E(Y_i + U)}{\sum_{i=1}^n X_i^2} \\ &= \frac{\sum_{i=1}^n X_i^2 \beta}{\sum_{i=1}^n X_i^2} \\ &= \beta \end{aligned}$$

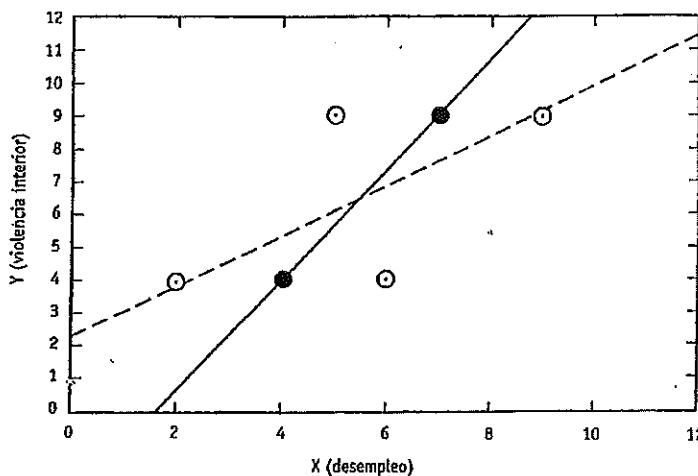
Este análisis demuestra que, incluso cuando haya un error de medida en la variable dependiente, el estimador típico no estará sesgado (será igual a la media de  $\beta$ ), tal como mostramos que ocurría —en la ecuación (3.8)— cuando una variable dependiente no tenía error de medida.

Sin embargo, para completar este análisis, tenemos que comprobar la eficiencia de nuestro estimador en presencia de una variable dependiente que se ha medido con errores. Utilizamos el procedimiento habitual:

$$\begin{aligned} V(b) &= V\left(\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}\right) \quad (5.1) \\ &= \frac{1}{\left(\sum_{i=1}^n X_i^2\right)^2} \sum_{i=1}^n X_i^2 V(Y_i + U) \\ &= \frac{\sigma^2 + \tau^2}{\sum_{i=1}^n X_i^2} \end{aligned}$$

Hay que señalar que aquí este estimador es *menos eficiente* que cuando se aplica a datos que carecen de error de medida en la variable dependiente (comárese con la ecuación [3.9]), por lo que respecta al nivel de error de medida que aparece en dicha variable  $\tau^2$ .

Gráfico 5.2 Error de medida en la variable explicativa



de pie sin hacer nada en el centro de varios pueblos cuando pasamos en coche. Como la hora y el día en que observamos los pueblos varían, al igual que lo hace la situación meteorológica, habrá un alto grado de error aleatorio en nuestras estimaciones del índice de paro. En un gran número de pueblos nuestros cálculos no serán sistemáticamente altos o bajos. Una estimación que se basara en cualquier pareja de pueblos sería bastante inefficiente; quizás las observaciones para cualquiera de ellas se hayan hecho en domingo (cuando puede que mucha gente esté largo rato en la calle) o en un día de lluvia (cuando pocos lo harían). Sin embargo, muchas observaciones de parejas de pueblos que se hayan hecho en diversos momentos de diferentes días, con sol o con lluvia, producirán una media correcta de las estimaciones del efecto. No obstante, como se ha señalado anteriormente, la consecuencia será muy diferente de la que producía un error similar en nuestra medida de la variable dependiente denominada delitos violentos.

El gráfico 5.2 exemplifica esta situación. Los dos círculos negros representan un estudio en el que no hay error de medida en ninguna variable<sup>6</sup>. Por consiguiente, la inclinación de la línea continua representa la estimación correcta del efecto causal que tiene el desempleo sobre la delincuencia. Para mostrar las consecuencias de este error, añadimos dos círculos más (los blancos) a derecha e izquierda de cada uno de los negros, con el fin de representar un error de medida en la variable explicati-

### 5. Entender qué debe evitarse

va cuya media es correcta (o sea, igual a la del círculo negro). La línea discontinua se ajusta a los círculos blancos, y la diferencia entre las dos líneas es el sesgo debido al error de medida aleatorio que hay en la variable explicativa. De nuevo recalcamos que las líneas se trazan de manera que minimicen los errores que se generan al predecir la variable dependiente (que aparecen en el gráfico como desviaciones *verticales* de la línea que pasa por los círculos), según cada uno de los valores de las variables explicativas.

De este modo, el efecto estimado del desempleo, que aquí se ha calculado con un considerable error aleatorio, será mucho más pequeño que el verdadero (ya que la línea discontinua está menos inclinada). Como sabemos que existe un error de medida en la variable explicativa, podríamos inferir que el auténtico efecto que tiene el desempleo sobre la delincuencia es mayor que la correlación observada en este proyecto de investigación.

El análisis de las consecuencias del error de medida en una variable explicativa nos lleva a dos directrices de tipo práctico:

1. Si, en principio, un análisis indica que no hay ningún efecto, el efecto verdadero será muy difícil de calibrar, ya que no se sabe la dirección del sesgo; por consiguiente, el análisis será bastante impreciso y así habrá que describirlo. Puede que el auténtico efecto sea nulo, negativo o positivo, y que nada en los datos nos señale cómo es.
2. Sin embargo, si un análisis indica que la variable explicativa con error de medida aleatorio tiene un pequeño efecto positivo, tendremos que utilizar los resultados de este apartado como justificación para concluir que el auténtico efecto probablemente sea aún mayor que el que hemos encontrado. Del mismo modo, si lo que hallamos es un pequeño efecto negativo, los resultados de este apartado podrán utilizarse como prueba de que el efecto verdadero probablemente indique la existencia de una relación negativa aún mayor.

Como el error de medida es un rasgo fundamental de todos los estudios cualitativos, estas directrices tendrían que poder aplicarse a todos ellos.

Es preciso matizar de alguna forma estas conclusiones para que los investigadores sepan exactamente cuándo son aplicables y cuándo no lo son. En primer lugar, el análisis del recuadro inferior, en el que se basan nuestras recomendaciones, puede aplicarse a modelos en los que sólo existe una variable explicativa. Hay resultados similares que sí son aplicables a muchas situaciones en las que existen múltiples variables explicativas, pero no a todas. Del mismo modo, el análisis también es aplicable si el investigador tiene muchas variables explicativas, pero sólo una con un considerable error de medida. Sin embargo, si se están analizando de forma simultánea las consecuencias de muchas variables y cada una de ellas parece formas diferentes de error de medida, sólo podremos calibrar los tí-

pos de sesgo que pueden aparecer si extendemos el análisis formal que aparece más abajo. Resulta que, a pesar de que los cualitativistas suelen tener muchas variables explicativas, lo más habitual es que estudien el efecto de cada una de ellas de forma secuencial más que simultánea. Por desgracia, como se describe en el apartado 2, este procedimiento puede causar problemas como el del sesgo de la variable omitida, pero hace que resultados similares a los analizados aquí sean bastante aplicables en la investigación cualitativa.

## 2. Exclusión de variables relevantes: el sesgo

La mayoría de los cualitativistas se dan cuenta de la importancia de controlar los posibles efectos falsos de otras variables cuando se hace una estimación de las consecuencias que tiene una variable sobre otra. Entre las formas de ejercer este tipo de control se encuentran los métodos de la diferencia y de la similitud de John Stuart Mill (1843) (a los que, irónicamente, Przeworski y Teune [1982] hacen referencia calificándolos, respectivamente, como los sistemas de diseño más similares y más diferentes); la «comparación de casos disciplinada-configuradora» de Verba (1967) (que es parecida a la «comparación estructurada-centrada» de George [1982]), así como diversas formas de utilizar supuestos en los que se mantienen casi todos los rasgos constantes y otros contrafácticos parecidos. Estas expresiones se mencionan con frecuencia, pero a los investigadores les suele resultar difícil aplicarlas de manera eficaz. Por desgracia, los cualitativistas disponen de pocas herramientas para expresar las consecuencias precisas de no tener en cuenta más variables en determinados momentos de una investigación, o sea, de ocuparse del «sesgo de la variable omitida». En este apartado, aportamos esas herramientas.

Comenzamos por analizar este problema explicando con palabras las consecuencias del sesgo de la variable omitida y continuamos después con un análisis formal. Posteriormente, nos centraremos en cuestiones más amplias del diseño de investigación que suscita este tipo de sesgo.

### 2.1 Cómo calibrar el sesgo de las variables omitidas

Supongamos que queremos hacer una estimación del efecto causal que tiene nuestra variable explicativa  $X_1$  sobre la dependiente  $Y$ . Si estamos haciendo un análisis cuantitativo, este efecto causal de  $X_1$  sobre  $Y$  lo expresaremos mediante  $\beta_1$ , que se puede estimar mediante una ecuación de regresión  $\beta$ , o utilizando otra forma de análisis que nos dé una estimación  $b_1$  de  $\beta_1$ . Si estamos realizando investigación cualitativa también intentaremos obtener esa estimación del efecto causal; sin embargo, en este caso, el

### 5. Entender qué debe evitarse

cálculo dependerá de un argumento verbal y de la valoración del investigador, que se basan en la experiencia y el criterio.

Supongamos que después de haber hecho las estimaciones (cuantitativa o cualitativamente) un colega eche un vistazo a nuestro análisis y plantea la objeción de que hemos omitido una importante variable de control,  $X_2$ . Hemos estado haciendo una estimación del efecto que tienen los gastos en las campañas electorales sobre la proporción de voto que recibe un candidato al Congreso estadounidense. Para este colega, puede que nuestra conclusión sea falsa por el «sesgo de la variable omitida». Es decir, lo que sugiere es que nuestra estimación  $b_1$  de  $\beta_1$  es incorrecta porque no hemos tenido en cuenta otra variable explicativa  $X_2$  (como medir si los candidatos son ya congresistas). El verdadero modelo tendría que controlar el efecto de la nueva variable.

¿Cómo podemos evaluar la afirmación de nuestro colega? Especialmente, ¿en qué condiciones influiría sobre nuestra estimación de las consecuencias que tiene el gasto en los votos el haber omitido la variable que calibra el hecho de ostentar ya el cargo de congresista, y en qué condiciones no influiría? Es evidente que omitir un término que mida la influencia de ser ya congresista no tendrá importancia si este hecho no influye en la variable dependiente; o sea, que si  $X_2$  es irrelevante, porque no tiene ningún efecto sobre  $Y$ , no producirá sesgo. Éste es el primer caso especial: las variables omitidas irrelevantes no producen sesgos. De este modo, si el hecho de estar ya en el poder careciera de consecuencias electorales, podríamos olvidarnos de haberlo omitido.

El segundo caso especial, que tampoco produce sesgos, se da cuando la variable omitida no se correlaciona con la explicativa incluida. De este modo, tampoco hay sesgo si el hecho de ser ya congresista no se correlaciona con nuestra variable explicativa, el gasto en la campaña electoral. De forma intuitiva, podemos decir que cuando una variable omitida no se correlaciona con la principal variable explicativa que nos interesa, someterla a control no cambiará la estimación del efecto causal que tiene sobre nuestra variable principal, porque lo que controlamos es la parte de variación que las dos variables tienen en común, si es que hay alguna. Por lo tanto, podemos omitir variables de control con tranquilidad, aunque tengan una gran influencia en la variable dependiente, siempre que no cambien con la variable explicativa incluida en el análisis<sup>1</sup>.

Si estos casos especiales no se sostienen a causa de alguna variable omitida (por ejemplo, si ésta se correlaciona con la variable explicativa incluida y tiene consecuencias en la dependiente), el no controlarla sesgará nuestra estimación (o percepción) del efecto de la variable incluida. En el caso que nos ocupa, las críticas de nuestro colega serían correctas, ya que el hecho de ser ya congresista está relacionado con la variable dependiente y también con la independiente: los que ya están en el poder reciben más votos y gastan menos.

### Un análisis formal del error de medida aleatorio en $X$

Definimos primero un modelo como sigue:

$$E(Y) = \beta X^*$$

en el que no se observa la verdadera variable explicativa  $X^*$  sino que se ve  $X$ , donde

$$X = X^* + U$$

y el error de medida aleatorio  $U$  tiene propiedades similares a las de antes: su media es cero,  $E(U) = 0$ , y no se correlaciona ni con la verdadera variable explicativa,  $C(U, X^*) = 0$ , ni con la dependiente,  $C(U, Y) = 0$ .

¿Qué ocurre cuando utilizamos el estimador típico de  $\beta$  con una  $X$  plagada de errores, en vez de con la  $X^*$  no observada? Esta situación es habitual en la investigación cualitativa si hay un error de medida y no se hace ningún ajuste especial en los resultados que se obtienen. Para analizar el efecto de este procedimiento evaluamos el sesgo, que va a ser la principal consecuencia de este tipo de problema de medida. Por lo tanto, comenzamos por aplicar el estimador típico de la ecuación (3.7) a las  $X$  e  $Y$  observadas en el modelo anterior.

$$\begin{aligned} b &= \frac{\sum_{i=1}^n Y_i X_i}{\sum_{i=1}^n X_i^2} \\ &= \frac{\sum_{i=1}^n (X_i^* + U_i) Y_i}{\sum_{i=1}^n (X_i^* + U_i)^2} \\ &= \frac{\sum_{i=1}^n X_i^* Y_i + \left( \sum_{i=1}^n U_i Y_i \right)}{\sum_{i=1}^n X_i^{*2} + \sum_{i=1}^n U_i^2 + \left( 2 \sum_{i=1}^n X_i^* U_i \right)} \end{aligned} \quad (5.2)$$

Esta idea puede plantearse en términos formales, centrándose en la última línea de la ecuación (5.5) que aparece en el recuadro inferior:

$$E(b_1) = \beta_1 + F\beta_2 \quad (5.3)$$

### 5. Entender qué debe evitarse

Debería quedar claro que  $b$  estará sesgado,  $E(b) \neq \beta$ . Además, la media de los dos términos que están entre paréntesis en la última línea de la ecuación (5.2) será cero porque presuponemos que  $U$  e  $Y$ , así como  $U$  y  $X^*$ , no se correlacionan (o sea, que  $C(U, Y_i) = E(U, Y_i) = 0$ ). Por lo tanto, esta ecuación se reduce aproximadamente a<sup>6</sup>:

$$= \frac{\sum_{i=1}^n X_i^* Y_i}{\sum_{i=1}^n X_i^{*2} + \sum_{i=1}^n U_i^2}$$

Esta ecuación con el estimador de  $\beta$  aplicado al modelo anterior sólo se diferencia de la que utiliza el estimador típico por el término adicional que

hay en el denominador,  $\sum_{i=1}^n U_i^2$  (compárese con la ecuación [3.7]), que representa el grado de error de medida en  $X$ , la varianza muestral del error  $U$ . Si no hay error de medida, este término será nulo y la ecuación se reducirá al estimador típico de la (3.7), ya que en realidad habríamos observado los auténticos valores de la variable explicativa.

En el caso general en el que hay algún error de medida  $\sum_{i=1}^n U_i^2$  será una suma de términos cuadráticos y, por tanto, siempre será positiva. Como este término se suma al denominador,  $b$  se acercará a cero. Si el estimador correcto produjera un número positivo alto, el error de medida aleatorio en la variable explicativa haría que el investigador pensara erróneamente que  $b$  tiene un valor positivo pero más pequeño. Si la estimación que se basa en  $X^*$  tuviera un valor alto negativo, un investigador que analizara datos con error de medida aleatorio pensaría que dicha estimación es un número negativo más pequeño.

Sería fácil utilizar este análisis formal para demostrar que el error de medida aleatorio en las variables explicativas también causa ineficiencias, pero el sesgo suele ser un problema más grave y nos ocuparemos de él en primer lugar.

Ésta es la ecuación que se utiliza para calcular el sesgo en la estimación del efecto que tiene  $X_1$  sobre la variable dependiente  $Y$ . En esta ecuación,  $F$  representa el grado de correlación que hay entre las dos variables explicativas,  $X_1$  y  $X_2$ . Si el estimador que se calcula utilizando sólo  $X_1$  como variable explicativa ( $b_1$ ) no estuviera sesgado, su media sería igual a  $\beta_1$ ; o sea, que se podría decir que  $E(b_1) = \beta_1$ . Este estimador no está sesgado en nin-

guno de los dos casos especiales en los que el término que expresa el sesgo,  $F\beta_2$ , es igual a cero. Resulta fácil apreciar que esto formaliza las condiciones de falta de sesgo que antes hemos señalado. Es decir, que podemos omitir una variable de control si:

- La variable omitida no tiene un efecto causal sobre la dependiente (o sea,  $\beta_1 = 0$ , independientemente del tipo de relación que exista entre la variable incluida y la excluida,  $F$ ).
- La variable omitida no se correlaciona con la incluida (es decir,  $F = 0$  cualquiera que sea el valor de  $\beta_2$ ).

Cuando sospechamos que hay una variable omitida que puede estar sesgando los resultados, nuestro análisis no podrá terminar aquí. Si es posible, hay que controlar esa variable omitida y, aunque no podamos hacerlo por falta de una buena fuente de datos acerca de ella, nuestro modelo podrá ayudarnos a determinar la dirección del sesgo, lo cual puede ser muy útil. Una sobreestimación puede reforzar sustancialmente un argumento existente y una subestimación puede debilitarlo.

Supongamos, por ejemplo, que estamos estudiando unos pocos estados subsaharianos y que nos damos cuenta de que los golpes de Estado son más frecuentes en regímenes que practican la represión política (aquellos en los que  $\beta_1$ , el efecto que tiene la represión en la probabilidad de golpe, es positivo). Esto significa que la variable explicativa es el grado de represión política y que la dependiente corresponde a la probabilidad de que haya un golpe de Estado. La unidad de análisis son los países africanos del área-subsahariana. Incluso, podríamos incorporar a la muestra otros países africanos y llegar a la misma conclusión. Sin embargo, supongamos que no hemos considerado los efectos que puede tener la situación económica en los golpes de Estado. Aunque quizás no dispongamos de datos sobre este indicador, sería razonable plantear la hipótesis de que probablemente el desempleo incrementa la probabilidad de golpe ( $\beta_2 > 0$ ) y también parece plausible que haya una correlación positiva entre desempleo y represión política ( $F > 0$ ). También presuponemos, porque es útil en este ejemplo, que las condiciones económicas son previas a nuestra variable causal clave: el nivel de represión política. Si así fuera, el grado de sesgo de nuestro análisis podría ser considerable. Como el desempleo se correlaciona positivamente con la variable dependiente y con la explicativa ( $F\beta_2 > 0$  en este caso), excluir ese factor supondría que, sin darnos cuenta, estaríamos haciendo una estimación del efecto que tienen la represión y el desempleo en la posibilidad de que haya un golpe de Estado, en vez de hacerla únicamente de la represión ( $\beta_1 + F\beta_2$  en vez de  $\beta_1$ ). Además, al ser el impacto conjunto de ambos factores mayor que el de la represión por sí sola, ( $\beta_1 + F\beta_2$  es mayor que  $\beta_1$ ), la estimación del efecto medio de esta última variable ( $\beta_1$ ) sería demasiado grande. Por lo tanto, el análisis demuestra que, al excluir las consecuencias del desempleo, sobreestimaremos las que tie-

## 5. Entender qué debe evitarse

ne la represión política (el impacto de esta exclusión no es el mismo que el del error de medida en las variables explicativas, ya que el sesgo de la variable omitida a veces puede tener como consecuencia que se haga una estimación positiva de una relación negativa).

Omitir variables relevantes no siempre redundará en una sobreestimación de los efectos causales. Por ejemplo, sería razonable hacer una hipótesis que señalara que en otros países (quizás esto sería tema para otro estudio) la represión política y el desempleo tienen una relación inversa ( $F$  es negativo). En estos países, la represión política podría permitir al gobierno controlar a las facciones en lucha, imponer la paz desde arriba y dar trabajo a la mayoría de la población. A su vez, esta situación conllevaría que el efecto del sesgo que introduce la relación negativa entre desempleo y represión ( $F\beta_2$ ) también sería negativo, siempre que siguiéramos estando dispuestos a presuponer que con un mayor desempleo aumentara la probabilidad de golpe en esos países. La principal consecuencia es que el efecto estimado de la represión en la probabilidad de golpe [ $E(\beta_1)$ ] ahora sería menor que el efecto real ( $\beta_1$ ). De este modo, si se excluye la situación económica,  $\beta_1$  tenderá a subestimar el efecto de la represión política. Si  $F$  es lo suficientemente negativo y  $\beta_2$  lo suficientemente grande, podríamos considerar que  $\beta_1$  siempre será negativo y llegar a la conclusión incorrecta de que una mayor represión política disminuye la posibilidad de golpe de Estado! Aunque no dispusiéramos de información suficiente sobre las tasas de desempleo para incluirla en el estudio original, un análisis como éste seguiría ayudándonos a llegar a conclusiones razonables e importantes.

Estos ejemplos tendrían que dejar claro que, en realidad, no es preciso hacer una regresión para calcular parámetros, ni para determinar el grado y dirección del sesgo, ni para llegar a conclusiones relativas a estos asuntos. Las estimaciones de tipo cualitativo e intuitivo padecen el mismo tipo de sesgos que las estrictamente cuantitativas. Este apartado demuestra que, en ambas situaciones, la información que no se refiere a los datos existentes puede ser de gran ayuda para hacer una estimación del grado de sesgo y de su dirección.

Si pensamos que quizás nuestro diseño de investigación se ha visto perjudicado por la omisión de ciertas variables pero no sabemos cuáles son, es muy posible que tengamos conclusiones imperfectas (que probablemente, en el futuro, otros investigadores detectarán). Los incentivos para profundizar más en el problema son evidentes. Por fortuna, en la mayoría de los casos, los investigadores disponen de bastante información sobre las variables que están fuera de su análisis. A veces, la información es detallada, pero sólo se refiere a ciertas unidades secundarias; es parcial pero aplicable al conjunto de la investigación, o incluso procede de estudios anteriores. Toda información, cualquiera que sea su procedencia y aunque sea incompleta, puede ayudarnos a determinar el posible nivel de sesgo que hay en nuestros efectos causales y la dirección que sigue.

Es evidente que, incluso los investigadores que comprenden las consecuencias del sesgo de la variable omitida, pueden encontrarse con dificultades al identificar las variables que han podido eliminar de sus análisis. No hay ninguna receta para enfrentarse a este problema, pero sí recomendamos a todos los investigadores, ya sean cuantitativistas o cualitativistas, que busquen sistemáticamente variables de control omitidas y que piensen si hay que incluirlas en sus estudios. En este apartado indicamos algunas directrices para proceder a esa revisión.

Las variables omitidas pueden producir dificultades aunque se disponga de información adecuada sobre las variables relevantes. A veces los investigadores tienen esa información y, creyendo que ciertas variables tienen una relación positiva con la variable dependiente, hacen una estimación secundaria de los efectos causales de dichas variables, en análisis «bivariantes» separados. Recurrir a este enfoque resulta especialmente tentador en estudios que tienen un pequeño número de observaciones, ya que, como señalamos en el apartado 1 del capítulo 4, utilizar simultáneamente muchas variables explicativas genera estimaciones poco exactas e incluso puede llevarnos a un diseño de investigación impreciso. Sin embargo, por desgracia, cada uno de los análisis excluye las otras variables relevantes, y esto hace que aparezca el sesgo de la variable omitida en cada estimación. La solución ideal no sólo consiste en recabar información sobre las variables relevantes, sino en controlar clara y simultáneamente todas ellas. El cualitativista debe admitir que no tener en cuenta todas las variables relevantes a la vez produce inferencias sesgadas. Reconocer las causas del sesgo es importante, aunque sea imposible eliminarlo porque se dispone de un pequeño número de observaciones.

Sin embargo, la preocupación por el sesgo de la variable omitida no debería llevarnos a incluir de manera automática cualquier variable cuya omisión pudiera causar sesgo porque se correlaciona con la variable independiente y tiene consecuencias sobre la dependiente. En general, *no tenemos que controlar una variable explicativa que sea, en parte, consecuencia de nuestra variable causal clave*.

En el siguiente ejemplo, suponemos que nos interesa el efecto causal que tiene aumentar el valor de la renta (variable de tratamiento) en 10.000 dólares sobre la probabilidad de que un ciudadano vote, en los Estados Unidos, a un candidato demócrata (variable dependiente). ¿Acaso habría que controlar el hecho de que esta persona, cinco minutos antes de llegar al colegio electoral, afirme en una entrevista que va a votar a los demócratas? Esta variable de control sí que influye en la dependiente y es probable que se correlacione con la explicativa. Intuitivamente, la respuesta a la pregunta anterior es no, ya que, si realmente controláramos esta variable, el efecto estimado que tendría la renta en el hecho de votar por los demócratas se atribuiría casi por completo a la variable de control, que, en este caso, apenas puede considerarse como una explicación causal alternativa.

Si aplicáramos ciegamente las normas antes señaladas para evitar el sesgo de la variable omitida, podríamos llegar a cometer el error de controlar esta variable. Después de todo, esta posible variable de control si tiene un efecto en la dependiente (votar a los demócratas) y se correlaciona con la explicativa clave (la renta). Sin embargo, al incluirla, estaríamos atribuyendo a la variable de control parte del efecto causal que tiene la explicativa clave.

Para dar otro ejemplo, supongamos que nos interesa el efecto causal que tiene un drástico incremento de los precios del crudo sobre la actitud de la opinión pública ante la existencia de escasez de energía. Podríamos conseguir los precios del petróleo (variable causal clave) en los periódicos y utilizar las encuestas de opinión como variable dependiente para calibrar cómo percibe la opinión pública la posibilidad de que haya escasez de energía. Sin embargo, cabría preguntarse si hay que controlar la influencia que tiene la información televisiva que se refiere a los problemas energéticos. Es cierto que la atención que presta la televisión a estos asuntos se correlaciona tanto con la variable explicativa incluida (el precio del crudo) como con la dependiente (la opinión pública en relación a la escasez de energía). No obstante, como la información televisiva es, en parte, consecuencia de los precios del crudo en el mundo real, no deberíamos controlarla al valorar la influencia causal que tienen los precios del crudo en la actitud de la opinión pública ante la escasez de energía. Si, en lugar de esto, nos interesara el efecto causal de la información televisiva, habría que controlar los precios del crudo, ya que éstos son *anteriores* a la variable explicativa clave (que ahora es la información televisiva)<sup>10</sup>.

En consecuencia, para hacer una estimación del efecto total que tiene una variable explicativa, hay que enumerar todas aquellas variables que, según nuestro modelo teórico, puedan generar la dependiente. Para repetir lo que señalamos anteriormente: en general, no tenemos que controlar una variable explicativa que sea, en parte, consecuencia de la variable causal clave. Una vez que se han eliminado estas posibles variables explicativas, hay que controlar otras del mismo tipo que pudieran causar sesgo de la variable omitida (las que se correlacionan con la variable dependiente y con las explicativas incluidas en el análisis)<sup>11</sup>.

La idea de que no tenemos que controlar las variables que sean consecuencia de las variables explicativas clave tiene una importante implicación para el papel de la teoría en nuestro diseño de investigación. Si pensamos en este problema, entenderemos por qué hay que partir de un modelo con fundamento teórico —o al menos tender hacia él— en vez de «escobar en busca de datos», es decir, realizar regresiones o análisis cualitativos con cualquier variable explicativa que se nos ocurra. Sin un modelo teórico no podremos decidir qué posibles variables explicativas hay que incluir en nuestro análisis. De hecho, si no hay modelo, se pueden lograr los resultados más sólidos utilizando una variable explicativa trivial —como la inten-

ción de votar a los demócratas que se expresa cinco minutos antes de entrar al colegio electoral— y controlando todos los demás factores que se correlacionan con ella. Si no disponemos de un modelo con fundamento teórico, no podremos determinar si hay que controlar las posibles variables explicativas o hacer caso omiso de las que se correlacionan entre sí y, además, nos arriesgaremos a incurrir en el sesgo de la variable omitida o a elaborar un diseño de investigación trivial.

No es en absoluto sencillo decidir cuándo hay que incorporar más variables explicativas a nuestro análisis. El número de variables adicionales es siempre ilimitado, pero nuestros recursos no, y, sobre todo, cuantas más variables de este tipo incluyamos, menos control tendremos sobre las estimaciones de cada uno de los efectos causales. Evitar el sesgo de la variable omitida justifica la incorporación de más variables explicativas. Si se omiten variables relevantes, nuestra capacidad de hacer estimaciones correctas de las inferencias causales se verá limitada.

## 2.2 Ejemplos de sesgo de la variable omitida

En este apartado, damos algunos ejemplos cuantitativos y cualitativos, unos hipotéticos y otros tomados de investigaciones reales. Por ejemplo, el nivel de estudios es uno de los mejores predictores de la participación política. Cuanto mayor es, más alta es la probabilidad de que se vote y de que se participe de otras maneras en política. Supongamos que así parece ser en un nuevo conjunto de datos, pero que queremos profundizar en este asunto y ver si la relación entre las dos variables es causal y, si lo es, determinar de qué manera los estudios conducen a la participación.

Lo primero que podríamos hacer es comprobar si se han omitido variables previas a los estudios que se correlacionen con este factor y que, al mismo tiempo, produzcan participación. Dos ejemplos de ello podrían ser el compromiso político de los padres de cada individuo y la raza de éste. Puede que los padres activos en política hayan inculcado a sus hijos el interés por participar en ella y, al mismo tiempo, puede que sean la clase de padres que favorecen el éxito en los estudios. Si no incluyéramos esta variable, podríamos obtener una relación falsa entre estudios y actividad política o hacer una estimación demasiado grande de ella.

La raza podría tener el mismo papel. En una sociedad en la que hay discriminación racial, los negros pueden verse excluidos tanto de las oportunidades educativas como de la participación política. En tal caso, no sería el efecto aparente que tienen los estudios sobre la participación. Lo mejor sería que intentáramos eliminar todas las posibles variables omitidas que justifiquen en parte o totalmente la relación que existe entre los estudios y la participación.

## Un análisis formal del sesgo de la variable omitida

Comencemos con un modelo sencillo en el que hay dos variables explicativas:

$$E(Y) = X_1\beta_1 + X_2\beta_2 \quad (5.4)$$

Supongamos ahora que nos encontramos un importante análisis en el que se informa del efecto que tiene  $X_1$  sobre  $Y$  sin controlar  $X_2$ . ¿En qué circunstancias habría razones para criticar este estudio o justificación para buscar fondos con los que repetirlo? Para contestar a esta pregunta, evaluamos formalmente el estimador con la variable de control omitida.

El estimador de  $\beta_1$  en el que eliminamos  $X_2$  es:

$$b_1 = \frac{\sum_{i=1}^n X_{1i} Y_i}{\sum_{i=1}^n X_{1i}^2}$$

Para evaluar este estimador, tomamos el valor esperado de  $b_1$  en varias hipotéticas reproducciones, siguiendo el modelo de la ecuación (5.4):

$$\begin{aligned} E(b_1) &= E\left(\frac{\sum_{i=1}^n X_{1i} Y_i}{\sum_{i=1}^n X_{1i}^2}\right) \\ &= \frac{\sum_{i=1}^n X_{1i} E(Y_i)}{\sum_{i=1}^n X_{1i}^2} \\ &= \frac{\sum_{i=1}^n X_{1i} (X_{1i}\beta_1 + X_{2i}\beta_2)}{\sum_{i=1}^n X_{1i}^2} \\ &= \frac{\sum_{i=1}^n X_{1i}^2 \beta_1 + \sum_{i=1}^n X_{1i} X_{2i} \beta_2}{\sum_{i=1}^n X_{1i}^2} \\ &= \beta_1 + F\beta_2 \end{aligned} \quad (5.5)$$

donde  $F = \frac{\sum_{i=1}^n X_{1i} X_{2i}}{\sum_{i=1}^n X_{1i}^2}$ , el coeficiente de la pendiente en la regresión de  $X_1$  y  $X_2$ . La última línea de esta ecuación se reproduce en el texto de la (5.3) y ya se ha analizado con cierto detalle anteriormente.

### El diseño de la investigación social

Sin embargo, el hecho de que la relación entre ambos elementos disminuya o desaparezca cuando controlamos una variable previa no tiene por qué suponer que los estudios sean irrelevantes. Supongamos que nos damos cuenta de que el vínculo entre los estudios y la participación se debilita cuando controlamos la raza. Una de las razones podría ser que, como en el ejemplo anterior, la discriminación de los negros implique que la raza está relacionada, por un lado, con el éxito académico y, por otro, con la participación. En estas condiciones, no habría ningún vínculo causal real entre los estudios y la participación. Por otra parte, la raza podría influir en la participación política *a través de* los estudios, ya que la discriminación racial podría reducir el acceso de los negros a la educación. A su vez, ésta podría ser el principal factor a la hora de generar participación. En este caso, la merma de la relación entre educación y participación que se produce cuando el investigador incluye la raza en el análisis no quita importancia a la educación. En realidad, explica cómo raza y educación interactúan para influir en la participación.

Hay que señalar que estas dos situaciones son fundamentalmente diferentes. Si la menor participación de los negros fuera debida a la falta de estudios, cabría esperar que aumentara al registrarse un incremento del nivel educativo medio. Sin embargo, si la razón de la menor participación fuera la discriminación política directa que impide que los negros ejerzan sus derechos como ciudadanos, la mejora de la educación sería irrelevante para la transformación de las pautas de participación.

También podríamos buscar variables que actúen a la vez que la educación o posteriormente. Se podrían buscar variables omitidas que muestren que la relación entre estudios y participación es falsa u otras que nos ayuden a explicar de qué manera la primera favorece a la segunda. En la primera categoría podría haber una variable como la del nivel de inteligencia general del individuo (la cual podría hacer que sacara buenas notas en el colegio y llevarle a la actividad política). En la segunda categoría podría haber variables que midieran aspectos educativos como la asistencia a cursos de educación cívica, las oportunidades de formar parte de órganos de gobierno estudiantiles o el aprendizaje de técnicas básicas de comunicación. Si se apreciara que la introducción en el análisis de una o más variables del segundo grupo disminuye la relación entre éxito en los estudios y participación (cuando se controlaron las técnicas de comunicación no se apreció que el éxito en los estudios tuviera un efecto independiente sobre la participación), esto no significaría que los estudios fueran irrelevantes. Las técnicas de comunicación necesarias se aprenden en la escuela y son diferentes en cada nivel educativo. Lo que el análisis nos diría es cómo influye la educación en la participación.

Una vez más, todos estos ejemplos ponen de manifiesto por qué es necesario evaluar con un modelo teórico en mente. No hay otra manera de elegir qué variables vamos a utilizar en nuestro análisis. Una teoría que se ocupara

### 5. Entender qué debe evitarse

de cómo influyen los estudios en las actividades cívicas nos indicaría qué variables hay que incluir. Aunque en una investigación cualitativa no incorporaremos más variables a una ecuación de regresión, la lógica es bastante parecida cuando decidimos qué otros factores vamos a tener en cuenta. Pensemos en una pregunta que planteamos anteriormente: qué impacto tienen los encuentros en la cumbre sobre la cooperación entre las superpotencias. Supongamos que nos diéramos cuenta de que, entre los Estados Unidos y la URSS, este tipo de cooperación hubiera sido mayor en los años posteriores a una cumbre que en los precedentes. ¿Cómo sabríamos que ese efecto es real y que no procede de alguna variable omitida? Y, si estamos seguros de que es real, ¿cómo podríamos profundizar en su funcionamiento?

Podríamos considerar variables precedentes, relacionadas con la posibilidad de que se celebrara una cumbre y que también pudieran ser causantes directas de la cooperación. Quizá cuando los mandatarios de cada país se fían el uno del otro, las reuniones son frecuentes y los países cooperan, o puede que cuando las ambiciones geopolíticas de ambos bandos se ven limitadas por razones políticas internas, programen reuniones y cooperen. En estas circunstancias, las cumbres, en sí mismas, no favorecerían directamente la cooperación, aunque programar una podría ser un buen indicador de que las cosas van bien entre las superpotencias. También es posible que las reuniones en la cumbre sean parte de una secuencia causal, al igual que la raza podía haber influido en el nivel educativo y éste, a su vez, en la participación. Cuando los mandatarios de las superpotencias confían el uno en el otro, convocan una cumbre para reforzar esa confianza mutua, y esa reunión, a su vez, conduce a la cooperación. En este caso, la cumbre no es en absoluto irrelevante, ya que sin ella habría menos cooperación. Supongamos que tenemos en cuenta estos factores y que encontramos que las cumbres parecen tener un papel independiente: es decir, que cuando controlamos la confianza mutua previa que hay entre los mandatarios y sus ambiciones geopolíticas, la conclusión es que una cumbre parece conducir a una mayor cooperación. Aún podríamos profundizar más y preguntarnos cómo ocurre esto. Sería posible comparar las cumbres a partir de las características que pueden hacerlas más o menos exitosas y ver si esos factores tienen alguna relación con el grado de cooperación posterior. De nuevo, es preciso seleccionar los elementos que hay que tener en cuenta, y entre ellos se pueden incluir: el grado de preparación, si los problemas tratados fueron económicos más que de seguridad, el nivel de armonía interna dentro de cada nación, la situación meteorológica durante la cumbre y también la comida. La teoría habrá de guiarnos; dicho de otro modo, necesitaremos tener una idea sobre conceptos y relaciones que nos lleve a variables explicativas relevantes y que planteen hipótesis lógicas que estén en consonancia con la experiencia que se tenga de sus efectos.

A los investigadores que manejen un pequeño número de observaciones les resultará muy difícil evitar el sesgo de la variable omitida. En esta situa-

ción, la ineficiencia tiene muchos costes, entre los que se puede incluir el riesgo de que la presencia de demasiadas variables de control irrelevantes haga que el diseño de investigación sea impreciso (apartado 1 del capítulo 4). Sin embargo, omitir variables de control relevantes puede producir sesgo, y, en principio, quizás el investigador no sepa si una posible variable es relevante o no.

En este punto, podríamos caer en la tentación de afirmar que la inferencia causal es imposible con un número escaso de observaciones. Sin embargo, para nosotros, las lecciones que hay que sacar de todo esto son más limitadas y optimistas. Comprender la dificultad que comporta extraer inferencias causales válidas a partir de pocas observaciones tendría que hacernos más cautelosos a la hora de hacer aseveraciones causales. Como indicamos en el capítulo 2, las buenas descripciones e inferencias descriptivas tienen más valor que las inferencias causales erróneas. De hecho, gran parte de la investigación cualitativa mejoraría si se prestara más atención a las inferencias descriptivas válidas y menos impetu a la extracción de afirmaciones causales a partir de datos inadecuados y sin calibrar adecuadamente la incertidumbre. No obstante, es posible avanzar un poco en la comprensión de los problemas causales si los problemas teóricos que nos preocupan se plantean con la suficiente claridad y se vinculan a las consecuencias observables adecuadas. Una investigación reciente del área de las relaciones internacionales puede ayudarnos a incidir en este punto.

Lo que llevó a Helen Milner a escribir *Resisting Protectionism* (1988) fue una perplejidad: ¿por qué la política comercial de los Estados Unidos era más protecciónista en los años veinte que en los setenta, a pesar de los muchos parecidos que había entre ambos períodos? Su hipótesis era que la interdependencia internacional había aumentado entre una y otra década y que esto ayudaba a explicar la diferencia de comportamiento de los Estados Unidos. Sin embargo, en este nivel de análisis agregado, sólo disponía de las dos observaciones que le habían sorprendido, y éstas no podían ayudarle a distinguir entre su hipótesis y otras muchas explicaciones posibles que se podían dar a la variación observada. Por lo tanto, el grado de incertidumbre de su teoría habría sido mucho mayor si se hubiera detenido aquí. En consecuencia, tuvo que buscar en otro sitio más consecuencias observables de su teoría.

Lo que Milner hizo fue describir el proceso que pensaba que producía su efecto causal. Planteó la hipótesis de que la interdependencia económica entre las democracias capitalistas afectaba a las preferencias nacionales porque influía en las opciones de los sectores industriales y empresas, los cuales, a base de presionar, conseguían que se llevaran a cabo las políticas que querían. En consecuencia, Milner estudió diversos sectores industriales estadounidenses de los años veinte y setenta, así como otros franceses en este segundo período, y se dio cuenta de que aquellos que más invertían a escala multinacional y dependían más de las exportaciones eran los menos protec-

cionistas. Estas conclusiones le ayudaron a confirmar su teoría general acerca de las diferencias que había, en conjunto, en la política estadounidense de los dos períodos. Por lo tanto, su procedimiento fue coherente con la parte principal de nuestros consejos metodológicos: hay que especificar cuáles son las consecuencias observables de la teoría, aunque no se refieran a nuestra preocupación principal, y diseñar la investigación de manera que se puedan extraer inferencias causales sobre tales consecuencias y que sea posible utilizarlas para evaluar la teoría. De ahí que el estudio de Milner sea ejemplar en muchos sentidos.

En cuanto al diseño de investigación, el problema más importante al que se enfrentó Milner era el relativo a las posibles variables omitidas. La variable de control más evidente era el grado de competencia de las importaciones extranjeras, ya que si éste es grande suele producir tendencias más protectionistas en las empresas. Es decir, era probable que la competencia de las importaciones se correlacionara con la variable dependiente de Milner y que, en la mayoría de los casos, fuera anterior a sus variables explicativas o que ocurriera a la vez que ellas. Si esta variable de control también se correlacionaba con sus variables explicativas clave —la inversión a escala multinacional y la dependencia respecto a las exportaciones—, sus resultados habrían estado sesgados. De hecho, según los principios de la ventaja competitiva, habría sido probable una correlación negativa entre la competencia de las importaciones y la dependencia respecto a las exportaciones, de manera que este sesgo hipotético se hubiera materializado si el primero de estos dos factores no se hubiera utilizado como control.

Milner se enfrentó a este problema seleccionando para su estudio únicamente aquellos sectores industriales a los que afectaba de manera determinante la competencia extranjera. En consecuencia, mantuvo constante la intensidad de la competencia de las importaciones y eliminó, o al menos redujo considerablemente, el problema del sesgo de la variable omitida. Podría haber mantenido constante esta variable de control clave en otro nivel —como el de aquellos sectores en los que se registraba una moderada penetración de importaciones—, siempre que hubiera sido realmente constante en sus observaciones.

Sin embargo, una vez que había controlado la competencia de las importaciones, Milner tuvo que enfrentarse a otros problemas relacionados con las variables omitidas. Las dos que barajó más seriamente —a partir de una revisión de los escritos empíricos y teóricos del área— fueron: (1) que los cambios registrados en el poder de los Estados Unidos explicaban los diferentes resultados de los años veinte y setenta y (2) que la explicación se basaba en las transformaciones registradas en los procesos políticos internos del país. El intento de controlar la primera explicación ya estaba incluido en su primer diseño de investigación: como la proporción del comercio mundial en la que participaban los Estados Unidos en la década de los setenta era prácticamente igual a la de los veinte, la investigadora controló

esta dimensión del poder estadounidense en el nivel agregado de sus políticas, así como en el de los sectores industriales y empresariales. Sin embargo, no controló las diferencias que había entre el aislacionismo político de los Estados Unidos en los veinte y su posición hegemónica como líder de una alianza en los setenta; se podría profundizar más en estos factores para calibrar su potencial capacidad de sesgo.

Milner controló los procesos políticos internos comparando sectores industriales y empresas en los veinte y en los setenta, puesto que unos y otras se habían enfrentado a las mismas estructuras gubernamentales y procesos políticos. Evidentemente, su nuevo estudio de seis sectores industriales franceses que competían en el ámbito de las importaciones durante los años setenta no le ayudó a mantener constantes los procesos políticos internos, pero sí le ayudó a descubrir que el efecto causal que tenía la dependencia de las exportaciones sobre la tendencia al proteccionismo no variaba cuando lo hacían los procesos políticos internos. Al pensar cuidadosamente en las razones potenciales que había para incurrir en el sesgo de la variable omitida y diseñar su estudio para responder a esa necesidad, Milner redujo considerablemente este posible sesgo.

Sin embargo, no controló de forma explícita otras posibles variables omitidas. Su estudio se centró «en las preferencias de los grandes grupos comerciales y no examina directamente la influencia de la opinión pública, la ideología, los sindicatos, la estructura política del país ni otros posibles factores» (1988, pp. 15-16). Desde el punto de vista teórico, la decisión de no controlar estas variables omitidas se podría haber justificado alegando que no estaban relacionadas con las principales variables causales (dependencia de la exportación e inversión a escala multinacional), que, en parte, eran consecuencia de estas últimas o que no tenían efectos sobre la variable dependiente (preferencia por el proteccionismo en un nivel agregado, compuesto de empresas y sectores industriales). Sin embargo, si se podían relacionar tales variables omitidas tanto con las explicativas del estudio como con la dependiente, y, desde el punto de vista causal, eran anteriores a las explicativas, Milner tendría que haber diseñado una investigación que las hubiera controlado explícitamente<sup>12</sup>.

Para terminar, con su procedimiento de selección de sectores industriales Milner corrió el riesgo de hacer ineficientes sus inferencias causales. Como hemos señalado, su forma de elegir los casos le permitió controlar la causa más importante de sesgo de la variable omitida mediante el mantenimiento en valores constantes de la competencia de las importaciones, variable que, desde un punto de vista teórico, se esperaba que fuera causalmente anterior a la causal clave, que se correlacionara con ella y que influyera también en las variables dependientes. Seleccionó los sectores que sufrían una mayor competencia de las importaciones y no los estratificó en función de ninguna otra variable. Posteriormente, en cada uno de los sectores de su muestra, y también en muchas empresas, estudió el grado de preferencia

#### 5. Entender qué debe evitarse

por el proteccionismo (su variable dependiente) e investigó el nivel de dependencia económica del exterior (su variable explicativa).

Este procedimiento de selección resultó ineficiente para sus inferencias causales porque sus variables causales clave cambiaron menos de lo que se hubiera deseado (Milner, 1988, pp. 39-42). Aunque, en este caso, la ineficiencia no resultó ser un grave problema, sí tuvo como consecuencia que la investigadora tuviera que realizar más estudios de caso de los necesarios para llegar al mismo nivel de certidumbre en sus conclusiones (véase el apartado 2 del capítulo 6). Dicho de otro modo, si hubiera trabajado con el mismo número de casos, pero eligiéndolos de modo que, en general, variaran con la variable explicativa, podría haber extraído inferencias causales más sólidas. Es decir, que el diseño habría sido más eficiente si se hubieran seleccionado, por una parte, algunos sectores industriales y empresas que carecían de vínculos con el extranjero y, por otra, industrias y empresas en las que había un alto nivel de participación extranjera; todas ellas habrían sufrido constantes apuros económicos y también la entrada de importaciones.

Los investigadores nunca pueden rechazar del todo la posibilidad de que las variables omitidas hayan sesgado sus análisis. Sin embargo, Milner, al haber intentado controlar algunas evidentes fuentes de sesgo de la variable omitida, pudo defender su hipótesis con más ahínco y de modo más convincente. Su riguroso estudio demuestra que los científicos sociales que trabajan con elementos cualitativos no tienen por qué desesperarse ante la perspectiva de extraer inferencias causales limitadas. La perfección es inalcanzable, quizás incluso indefinible, pero relacionar con cuidado la teoría y el método puede hacer que diseñemos estudios que aumenten la plausibilidad de nuestros argumentos y que reduzcan la incertidumbre de las inferencias causales.

#### 3. Inclusión de variables irrelevantes: la ineficiencia

A causa de los problemas potenciales que presenta el sesgo de la variable omitida —que hemos descrito en el apartado anterior—, podríamos pensar ingenuamente que es esencial recoger todas las posibles variables explicativas, estimando a la vez sus efectos causales. Desde el principio hay que recordar que esto no es lo que se desprende del apartado 2 de este capítulo, en el que hemos señalado que omitir una variable explicativa que no se correlaciona con las explicativas incluidas no genera sesgo —aunque esa variable tenga un fuerte impacto causal en la dependiente— y que es un error controlar variables que son consecuencia de las explicativas. En consecuencia, nuestro argumento no debería llevar a los investigadores a recabar información sobre todas las inferencias causales posibles o a criticar los estudios que no lo hagan.

Por supuesto, puede darse el caso de que el investigador siga sin estar seguro de qué variables de control previas tienen un impacto causal en las incluidas o cuáles se correlacionan con éstas. En tal situación, habrá investigadores que intenten incorporar todas las variables de control cuya correlación con las variables explicativas utilizadas sea concebible, así como todas aquellas que, desde un punto de vista teórico, se pueda esperar que afecten a la variable dependiente. Es probable que el resultado de esto sea una larga lista de variables y que muchas de ellas sean irrelevantes. Este enfoque, que a primera vista parece una manera cuidadosa y prudente de evitar el sesgo de la variable omitida, de hecho nos haría correr el riesgo de elaborar un diseño de investigación cuyos resultados no podrían ser más que imprecisos. Cuando se utiliza un número relativamente escaso de observaciones, la imprecisión, tal como se analizó en el apartado 1 del capítulo 4, es un problema bastante grave para el que este «cuidadoso» diseño sería en realidad perjudicial. En el presente apartado analizamos los costes que tiene incluir variables explicativas irrelevantes y mostramos reservas fundamentales ante la pretensión de «incluirlo todo». Incorporar variables irrelevantes puede ser muy dañino. Nuestra idea principal es que aunque la variable de control no tenga ningún efecto causal en la dependiente, *cuenta mayor correlación haya entre la variable explicativa y la irrelevante de control, menos eficiente será la estimación del principal efecto causal*.

Vamos a ilustrar esta afirmación centrándonos en dos diferentes procedimientos (o «estimadores»), con el fin de calcular el efecto causal de una variable explicativa cuya inclusión sí es apropiada. La primera estimación de este efecto procede de un análisis realizado sin variables de control irrelevantes; en el segundo hay una que sí lo es. El análisis formal que figura en el recuadro posterior llega a las siguientes conclusiones respecto al valor relativo de estos dos procedimientos y al del método antes mencionado. En primer lugar, *ninguno de los estimadores está sesgado*. Esto quiere decir que, aunque se controle una variable explicativa irrelevante, el estimador habitual seguirá dándonos una respuesta media correcta. En segundo lugar, *si la variable de control irrelevante no se correlaciona con la explicativa principal, la estimación del efecto causal de la segunda no sólo estará sesgada, sino que resultará tan ineficiente como si no se hubiera incluido la variable irrelevante*. En realidad, si estas variables no están correlacionadas, se obtendrá precisamente la misma inferencia. Sin embargo, si la variable de control irrelevante está muy correlacionada con la explicativa principal, se registrará una ineficiencia considerable.

Por lo tanto, los costes de controlar variables irrelevantes son altos. Si lo hacemos, es más posible que nuestros estudios hagan estimaciones alejadas de los auténticos efectos causales. Cuando reproduzcamos un estudio, en un conjunto de datos que presenta una alta correlación entre la variable explicativa clave y una de control irrelevante incluida en el análisis, probablemente obtengamos resultados diferentes que nos lleven a otras inferencias

causales. De este modo, aunque controlemos todas las variables explicativas irrelevantes (sin cometer otros errores), obtendremos la respuesta media correcta, pero quizás en algunos de los proyectos, o en todos, nos alejaremos de ella. La repetición del análisis tendrá, como promedio, el mismo efecto, pero la variable irrelevante aumentará la ineficiencia, tal como habría ocurrido si hubiéramos descartado alguna de nuestras observaciones. Las consecuencias tendrían que ser claras: al incluir una variable irrelevante ejercemos más presión sobre un conjunto de datos limitado, lo cual hace que dispongamos de menos información para cada inferencia.

Como ejemplo, pensemos de nuevo en el estudio de los golpes de Estado en los países africanos. Una investigación preliminar indicaba que el grado de represión política, la variable explicativa principal de interés, aumentaba la frecuencia de dichos golpes. Supongamos que otro investigador señalara que el primer estudio era incorrecto porque no había tenido en cuenta si cada país había logrado la independencia librándose del gobierno colonial de manera violenta o negociada. Supongamos que creemos que este investigador está equivocado y que el tipo de proceso descolonizador no influye en la variable dependiente, que es la frecuencia de los golpes (después de utilizar la principal variable explicativa, se controla la represión política). ¿Cuáles serían las consecuencias de controlar esta otra variable irrelevante?

La respuesta depende de la relación que exista entre la variable irrelevante, que evalúa el tipo de proceso descolonizador, y la explicativa principal, que mide la represión política. Si la correlación entre ambas es alta —como parece posible—, incluir estas variables de control hará que las estimaciones del efecto de la represión política sean bastante ineficientes. Para entender esto, hay que señalar que, para controlar de qué manera se accedió a la independencia, el investigador podría dividir sus categorías de regímenes represivos y no represivos teniendo en cuenta si el proceso descolonizador había sido violento o negociado. Se podría medir la frecuencia de los golpes de Estado en cada categoría para evaluar los efectos causales de la represión política, mientras que se mantiene bajo control la forma de librarse del gobierno colonial. Aunque este tipo de diseño sea una forma razonable de evitar el sesgo de la variable omitida, puede tener muchos costes: cuando la nueva variable de control no influya en la dependiente pero se correlacione con una variable explicativa incluida, se reducirá el número de observaciones de cada categoría y la estimación del principal efecto causal se realizará de forma mucho menos eficiente. Este resultado supone que se desperdicia gran parte del trabajo del investigador, ya que reducir innecesariamente la eficiencia equivale a descartar observaciones. La mejor solución es siempre recoger más observaciones, pero, si esto no es posible, los investigadores ya saben que tienen que identificar cuáles son las variables irrelevantes para no tener en cuenta sus consecuencias.

### Un análisis formal de las ineficiencias de la variable incluida

Supongamos que el verdadero modelo sea  $E(Y) = X_1\beta_1 + X_2\beta_2$  y  $V(Y) = \sigma^2$ . Sin embargo, pensamos erróneamente que la variable explicativa  $X_2$  también ha de estar en la ecuación; de manera que hacemos la siguiente estimación:

$$E(Y) = X_1\beta_1 + X_2\beta_2 \quad (5.6)$$

sin saber qué, en realidad,  $\beta_2 = 0$ . ¿Qué consecuencias tiene una estimación simultánea de ambos parámetros para nuestra estimación de  $\beta_1$ ?

Si definimos  $b_1$  como el estimador correcto —que se basa únicamente en la regresión entre  $Y$  y  $X_1$ — y  $\hat{\beta}_1$  como el primer coeficiente de  $X_1$  en una regresión entre  $Y$ , por una parte, y  $X_1$  y  $X_2$ , por otra, es fácil mostrar que no podemos distinguir entre estos dos estimadores utilizando la falta de sesgo (una media correcta en muchos experimentos hipotéticos), ya que ambos carecen de él:

$$E(b_1) = E(\hat{\beta}_1) = \beta_1 \quad (5.7)$$

Sin embargo, los estimadores sí difieren en cuanto a su eficiencia. El estimador correcto tiene una varianza (calculada en la ecuación [3.9]) de

$$V(b_1) = \frac{\sigma^2}{\sum_{i=1}^n X_{1i}^2} \quad (5.8)$$

mientras que la del otro estimador es

$$\begin{aligned} V(\hat{\beta}_1) &= \frac{\sigma^2}{(1 - r_{12}^2) \sum_{i=1}^n X_{1i}^2} \\ &= \frac{V(b_1)}{(1 - r_{12}^2)} \end{aligned} \quad (5.9)$$

donde la correlación entre  $X_1$  y  $X_2$  es  $r_{12}$  (véase Goldberger, 1991, p. 245).

En la última línea de la ecuación podemos ver la relación precisa que existe entre las varianzas de los dos estimadores. Si la correlación entre las dos variables explicativas es nula, no tendrá importancia si se incluye la variable irrelevante o no, ya que ambos estimadores tendrán la misma varianza. Sin embargo, cuanto más se correlacionen dos variables, mayor será la varianza y, por tanto, menor la eficiencia de  $\hat{\beta}_1$ .

### 5. Entender qué debe evitarse

#### 4. Endogeneidad

En las ciencias políticas la investigación experimental no es frecuente. No solemos tener la oportunidad de manipular las variables explicativas; sólo las observamos. Una de las consecuencias de esta falta de control es la endogeneidad, es decir, que los valores de nuestras variables explicativas a veces provienen de nuestra variable dependiente en vez de ser una de sus causas. Si hay una auténtica manipulación experimental, la dirección de la causalidad es inequívoca. Sin embargo, en muchas áreas de la investigación cualitativa y cuantitativa la endogeneidad es un problema habitual y grave<sup>13</sup>.

Si el investigador no controla los valores de las variables explicativas, la dirección de la causalidad siempre será un problema difícil. En la investigación no experimental —ya sea cuantitativa o cualitativa— las variables explicativas y dependientes varían por la influencia de factores que escapan al control del investigador (y que, a menudo, no se ven). Los países invaden, los oficiales traman golpes de Estado, la inflación cae, las políticas de los gobiernos se implantan, los candidatos deciden presentarse a las elecciones, los votantes eligen entre ellos y el investigador tiene que intentar elaborar un argumento para determinar causas y consecuencias.

Pensemos en un ejemplo tomado de la bibliografía sobre las elecciones al Congreso estadounidense. Muchos académicos han señalado que el drástico aumento que se registró a finales de los sesenta en las ventajas electorales de que disponían los candidatos que ya estaban en el poder se debió principalmente al incremento de los servicios que prestaban estos congresistas a su electorado. Es decir, que el franqueo gratuito, los presupuestos para viajar a la propia circunscripción, el personal de que se dispone en ella para ocuparse de las peticiones concretas de los electores, la financiación de proyectos locales con fondos públicos para lograr objetivos electorales, así como otros incentivos de los congresistas, han hecho posible que éstos recaben apoyos en sus circunscripciones. Muchos ciudadanos les votan teniendo en cuenta estas cuestiones.

La hipótesis del servicio al electorado parece perfectamente razonable, pero ¿hay pruebas de ella? Numerosos investigadores han intentado presentarlas (para una revisión de esta bibliografía, véase Cain, Ferejohn y Fiorina, 1987), pero pocos son los datos que constatan tal hipótesis. El estudio modal de esta cuestión se basa en medidas del servicio que ha dado a su circunscripción cada uno de los congresistas de una muestra y en las que indican la proporción de votos que recibe esa clase de candidatos. Posteriormente, utilizando un análisis de regresión, los investigadores hacen una estimación del impacto causal que tiene ese servicio sobre el voto. Resulta sorprendente que muchas de esas estimaciones indiquen que el efecto es nulo o incluso negativo.

Es probable que el problema de la endogeneidad explique esos resultados paradójicos. Dicho de otro modo, los congresistas que corran un mayor

riesgo de perder el puesto en las siguientes elecciones (quizá a causa de un escándalo o por haber atravesado una mala época en su circunscripción) darán más servicios a sus electores, mientras que los que estén seguros de que no van a rechazarlos se concentrarán en otras vertientes de su trabajo, como la elaboración de políticas en Washington. El resultado es que los congresistas que dan más servicios reciben menos votos. Esto no significa que el servicio al electorado reduzca el número de votos, sólo que unas expectativas de voto altas reducen tal servicio. Si se hace caso omiso de este efecto de retroalimentación, las inferencias serán muy sesgadas.

David Laitin traza las líneas generales de un ejemplo de endogeneidad que toma de uno de los clásicos de las ciencias sociales de principios de siglo, *La ética protestante y el espíritu del capitalismo* de Max Weber. «Weber intentó demostrar que las enseñanzas y doctrinas protestantes provocaban (de forma inadvertida) un cierto tipo de comportamiento económico: el espíritu capitalista. Sin embargo [...] Weber y sus discípulos no pudieron responder a una de las objeciones que suscitó esta tesis: que los europeos a los que ya les interesaba romper los límites del espíritu precapitalista bien podían haberse apartado de la iglesia católica, precisamente para alcanzar ese propósito. Dicho de otro modo, se podría interpretar que los intereses económicos de ciertos grupos habían hecho que se desarrollara la ética protestante. Sin un estudio mejor controlado, se puede dar la vuelta a la línea de causalidad de Weber» (Laitin, 1986, p. 187; véase también R. H. Tawney, 1935, que fue quien originó este tipo de crítica).

En lo que queda de este apartado analizaremos cinco métodos para enfrentarse al difícil problema de la endogeneidad:

- Corregir una inferencia sesgada (apartado 4.1).
- Precisar cuáles son los componentes de la variable dependiente y estudiar sólo aquellos que sean consecuencia, y no causa, de la explicativa (apartado 4.2).
- Convertir un problema de endogeneidad en un sesgo producido por una variable omitida y controlar esta variable (apartado 4.3).
- Seleccionar con cuidado al menos algunas observaciones en las que no haya problema de endogeneidad (apartado 4.4).
- Precisar cuáles son los componentes de las variables explicativas con el fin de asegurarse de que sólo entran en el análisis los que sean realmente exógenos (apartado 4.5).

Cada uno de estos cinco procedimientos puede considerarse un método para evitar los problemas de endogeneidad, pero también pueden verse como formas de aclarar una hipótesis causal, puesto que las que no prestan atención a un problema de endogeneidad serán, finalmente, problemas teóricos que haya que redefinir para que, al menos, sea posible que la variable explicativa influya en la dependiente. Analizaremos las dos primeras soluciones que se dan a la endogeneidad mediante el ejemplo cuantitativo del

servicio al electorado y para las otras tres nos apoyaremos en ejemplos de investigaciones cualitativas ampliados.

#### 4.1 Corregir inferencias sesgadas

La última línea de la ecuación (5.13) que aparece en el siguiente recuadro nos proporciona un método para determinar la dirección exacta y la magnitud del sesgo que procede de la endogeneidad. Para una mayor comodidad, reproducimos aquí esa ecuación:

$$E(b) = \beta + \text{sesgo}$$

Esta fórmula implica que cuando hay endogeneidad no estamos haciendo la inferencia causal deseada. Esto quiere decir que, si el término del sesgo es cero, la media de nuestro método inferencial (o estimador  $b$ ) no estará sesgada (o sea, será igual a  $\beta$ ). Sin embargo, si hay un sesgo de endogeneidad, estaremos haciendo una estimación de la inferencia correcta más un factor de sesgo. La endogeneidad es un problema porque, en general, no somos conscientes de la magnitud o dirección del sesgo. Este factor de sesgo será grande o pequeño, negativo o positivo, según sea el ejemplo empírico específico. Por fortuna, aunque en principio no podemos evitar el sesgo de endogeneidad, a veces podremos corregirlo a posteriori si determinamos la dirección y quizás el nivel de sesgo.

La ecuación (5.13) demuestra que el factor de sesgo depende de la correlación que haya entre la variable explicativa y el término de error, que es la parte de la variable dependiente que no se aclara mediante la explicativa. Por ejemplo, si la hipótesis del servicio al electorado es correcta, el efecto causal que tenga dicho servicio sobre los votos ( $\beta$  en la ecuación) será positivo. Si, además, el voto esperado influye en el nivel de servicio al electorado que observamos, el término de sesgo será negativo. Esto quiere decir que, incluso después de que se tenga en cuenta el efecto del servicio al electorado sobre los votos, habrá una correlación inversa entre tal servicio y el término de error, porque los congresistas que esperan obtener menos votos darán más servicios. El resultado es que el término de sesgo será negativo y que, en este caso, las inferencias que no se hayan corregido serán estimaciones sesgadas del efecto causal  $\beta$  (o, de forma equivalente, estimaciones sesgadas de  $[\beta + \text{sesgo}]$ ). De esta manera, aunque la hipótesis del servicio al electorado sea cierta, el sesgo de endogeneidad nos llevará o bien a estimarlo con un valor positivo más pequeño del que tendría que tener, o bien a considerarlo nulo, o incluso negativo, dependiendo de la magnitud del factor de sesgo. Por lo tanto, podemos afirmar que la estimación correcta del efecto del servicio al electorado sobre los votos será mayor que la calculada en un análisis en el que no se haya corregido la endogeneidad. En consecuencia, en este

## El diseño de la investigación social

análisis no corregido, el efecto del servicio presentará un límite más bajo, lo cual hará que resulte más plausible la hipótesis del servicio al electorado.

De este manera, aunque no podamos evitar el sesgo de endogeneidad, a veces sí podremos mejorar las inferencias a posteriori haciendo una estimación del nivel de sesgo. Como mínimo, esto nos permitirá determinar su dirección y quizás nos indique el límite superior o inferior de la estimación correcta. En el mejor de los casos, podremos utilizar esta técnica para hacer inferencias que no tengan sesgo alguno.

### 4.2 Precisar cuáles son los componentes de la variable dependiente

Una de las formas de evitar el sesgo de endogeneidad es redefinir la variable dependiente para dotarla de sus propios componentes dependiente y explicativo. Este último obstaculiza nuestro análisis a través de un mecanismo de retroalimentación, es decir, influye en la variable causal (explicativa) clave. La otra parte de nuestra variable dependiente es realmente dependiente: es una función, y no una causa, de la variable explicativa. El objetivo de esta forma de evitar el sesgo de endogeneidad es identificar y medir sólo la parte dependiente de nuestra variable dependiente.

Por ejemplo, King (1991a), al estudiar la hipótesis del servicio al electorado, desgajó del total de votos recibido por un congresista estadounidense la porción que se debía únicamente al hecho de que ya estuviera en el poder. En los últimos años, la ventaja electoral de los que ya son congresistas se ha cifrado entre un 8 y un 10% de los votos; proporción que hay que comparar con el 52% aproximado que constituye la base electoral de la mayoría de los que, en este sistema bipartidista, ya ocupan un escaño en el Congreso. Posteriormente, King calculó, mediante un procedimiento estadístico, la ventaja que suponía ser congresista —elemento que era el único dependiente de la variable dependiente— y utilizó esta cifra en lugar del voto bruto para hacer una estimación de los efectos que tenía el servicio al electorado. Como la ventaja de los congresistas en número de votos, al representar sólo una pequeña porción del total, no iba a tener un gran impacto en la propensión de ese tipo de candidatos a realizar servicios para el electorado, King evitó el sesgo de endogeneidad. Sus resultados señalaban que, si se añadían 10.000 dólares al presupuesto que tenía el legislador estatal medio para (entre otras cosas) dar servicios al electorado, se estaba concediendo a este candidato una ventaja de 1,54 puntos (más o menos 0,4%) en las elecciones siguientes. De esta manera, el autor estaba dando la primera base empírica a la hipótesis del servicio al electorado.

## 5. Entender qué debe evitarse

### 4.3 Convertir la endogeneidad en un problema de variable omitida

Siempre se puede abordar la endogeneidad como si fuera un sesgo de la variable omitida, tal como demostramos a continuación con el famoso ejemplo del estudio sobre sistemas electorales comparados. Para las primeras generaciones de polítólogos, uno de los grandes enigmas del análisis político era la caída de la República de Weimar y su sustitución por el régimen nazi a principios de los años treinta. Una de las explicaciones, que se basaba en algunos rigurosos y convincentes estudios de caso de la Alemania de esa época, señalaba que la causa principal había sido la consagración en la constitución de esa república de un sistema de representación proporcional para las elecciones. En pocas palabras, la idea es que el sistema proporcional permite que estén en el parlamento los pequeños partidos que representan a determinadas ideologías, intereses o grupos religiosos. En este tipo de ordenamiento electoral los candidatos no se ven en la necesidad de negociar sus posturas para tener éxito en las elecciones, como sí ocurre en aquellos sistemas en los que sólo hay un ganador por circunscripción. En consecuencia, el parlamento se llena de pequeños grupos ideológicos que ni quieren ni pueden trabajar juntos. La parálisis y la frustración resultantes harán posible que uno de estos grupos —en este caso los nacionalsocialistas— tome el poder (para la exposición clásica de este teoría, véase Hermens, 1941).

El argumento del párrafo anterior se fue elaborando en varios importantes estudios de caso que se ocuparon de la caída de la República de Weimar. Hubo historiadores y polítólogos que achacaron este hecho al éxito electoral de los pequeños partidos y a su falta de voluntad negociadora en el Reichstag. Este planteamiento tiene muchos problemas, como siempre ocurre cuando se explica un resultado complejo a partir de un único factor, pero sólo vamos a centrarnos en el de la endogeneidad. Detrás de esta explicación subyace un mecanismo causal que se encadena de la siguiente manera: se implantó la representación proporcional y esto hizo posible que partidos pequeños que contaban con una escasa base electoral consiguieran escaños en el Reichstag (y en éstos se incluían aquellos que, como los nacionalsocialistas, querían acabar con dicha institución). El resultado fue que el funcionamiento del Reichstag alcanzó un punto muerto y la frustración se extendió entre las masas. A su vez, esto condujo a un golpe de Estado propiciado por uno de los partidos.

Sin embargo, otros estudios —tanto de Alemania como de otras consecuencias observables— señalaron que la fragmentación de los partidos no sólo era consecuencia de la representación proporcional. Algunos investigadores argüían que si dicha fragmentación condujo a la implantación de la representación proporcional, también podía haber sido su causa: Al aplicar la misma variable explicativa a otras observaciones (en consonancia con la regla que indicamos en el capítulo 1: intentar probar las hipó-

tesis con datos diferentes a los que las produjeron), los investigadores se dieron cuenta de que en aquellas sociedades en las que muchos grupos de mentalidad estrecha y arraigada se oponen a otros minoritarios, étnicos o religiosos, es más probable que se opte por la representación proporcional, ya que éste es el único sistema electoral que ambas facciones pueden aceptar. Una mirada más atenta a la situación de la política alemana antes de que se introdujera este tipo de representación confirmó esta idea, ya que se localizaron muchas facciones pequeñas. La representación proporcional no las había creado, aunque puede que sí facilitara la expresión de sus ideas en el parlamento. Tampoco las facciones eran la única causa de la representación proporcional; sin embargo, sí parecía que tanto la adopción del sistema proporcional como la gran división del parlamento eran consecuencia de la fragmentación social (véase Lakeman y Lambert, 1955, p. 155, para una de las primeras explicaciones de este argumento).

De este modo, hemos transformado un problema de endogeneidad en un sesgo de la variable omitida. Esto quiere decir que la fragmentación social previa es una variable omitida que lleva a la representación proporcional; desde el punto de vista causal es anterior a ésta y, en parte, condujo a la caída de Weimar. Al transformar el problema de esta manera, los académicos consiguieron abordar mejor el problema, ya que podían medir de manera explícita esta variable omitida y controlarla en estudios posteriores. En este ejemplo, una vez que se incluyó y controló dicha variable omitida, los investigadores se dieron cuenta de que existía una posibilidad razonable de que la aparente relación causal que había entre la representación proporcional y la caída de la República de Weimar fuera casi completamente falsa.

La relación que existe entre los sistemas electorales y la democracia sigue suscitando polémica, aunque el estudio de este tema ha avanzado mucho desde los primeros trabajos que se hicieron. Los investigadores han ampliado esta área, dejando a un lado los intensos estudios de caso, que apenas se preocupaban de la lógica de las explicaciones, para hacer trabajos que se basan en muchas observaciones de determinadas consecuencias y así resolver gradualmente algunos aspectos relacionados con la medida y, finalmente, con la inferencia. Con este cambio, han logrado separar de modo más sistemático los efectos exógenos de los endógenos.

#### 4.4 Seleccionar observaciones para evitar la endogeneidad

La endogeneidad es un problema muy común en gran parte de las obras que se ocupan del impacto que tienen las ideas en las políticas (Hall, 1989; Goldstein y Keohane, 1993). En la medida en que las ideas *reflejan las condiciones* en las que operan los actores políticos —por ejemplo, las circunstancias materiales que generan sus propios intereses en ese sentido—, el análisis del impacto que tienen sobre las políticas siempre será suscepti-

ble de incurrir en un sesgo de la variable omitida: las ideas de los actores se correlacionan con una variable omitida anterior en sentido causal —los intereses materiales— que influye en la variable dependiente —la estrategia política (véase el apartado anterior). Además, en la medida en que las ideas funcionan como *racionalizaciones* de políticas que se han llevado a cabo en otros ámbitos, también pueden ser simples *consecuencias* de esas políticas, más que sus causas. En tales circunstancias, las ideas son endógenas: puede parecer que explican las estrategias de los actores, pero, en realidad, son su resultado.

Desde el punto de vista metodológico, lo más difícil a la hora de estudiar el impacto de las ideas sobre las políticas es compensar la influencia que tienen, en un determinado problema de investigación, el problema de la endogeneidad y el de la variable omitida, que tan relacionados están. Para poder señalar la importancia causal de las ideas, hay que demostrar que algunas ideas —o un aspecto de las mismas— que mantienen los que elaboran las políticas influyen en aquellas que realmente se llevan a cabo, y que no sólo reflejan esas mismas políticas o los previos intereses materiales de sus impulsores. Los investigadores de esta área tienen que ser especialmente cuidadosos a la hora de definir el efecto causal del interés. En concreto, tienen que comparar la variable dependiente observada (las políticas) y la explicativa (las ideas de los individuos) con una situación *contrafáctica* —definida con precisión— en la que la variable explicativa adopte un valor diferente: cuando los individuos relevantes tienen ideas diferentes.

El análisis comparado es una buena forma de determinar si un conjunto de ideas concreto es exógeno o endógeno. Por ejemplo, en un reciente estudio acerca del papel de las ideas en la adopción de políticas económicas estalinistas en países socialistas, Nina Halpern (1993) hace este tipo de análisis. Su hipótesis es que la doctrina de la planificación estalinista, en la que creían los mandatarios de Europa del Este y de China, ayuda a explicar las políticas económicas que implantaron cuando llegaron al poder en estos países después de la Segunda Guerra Mundial. Esta hipótesis concuerda con el hecho de que tales líderes tuvieran ideas estalinistas y aplicaran políticas de ese tipo, pero una simple correlación no demuestra que haya causalidad. En realidad, puede que la endogeneidad esté funcionando: las políticas estalinistas podrían haber generado ideas que las justificaran o prever que habría que seguir tales políticas quizás produjera esas ideas.

Aunque Halpern no utiliza este lenguaje, sí recurre a un procedimiento similar al que se ha analizado en el apartado anterior: transformar la endogeneidad en sesgo de la variable omitida. La principal hipótesis alternativa que esta autora considera es que Europa del Este y los estados comunistas de Asia sólo implantaron economías centralizadas después de la Segunda Guerra Mundial por el poderío militar y la influencia política de la Unión Soviética. El contrafáctico que plantea esta hipótesis es que, aunque los europeos orientales y los chinos no hubieran creído en la conveniencia de las

economías planificadas que dictaba el estalinismo, de todas maneras se habrían implantado esa clase de economías y habrían aparecido ideas que las justificaran.

Posteriormente, Halpern señala que en los países de Europa del Este que fueron ocupados por el Ejército Rojo, la adopción de una economía centralizada podría explicarse más por el poderío soviético que por la creencia en la superioridad de las doctrinas estalinistas: «... la explicación alternativa de que la elección no fue más que una respuesta a las órdenes de Stalin es imposible de refutar» (1993, p. 89). En consecuencia, Halpern busca posibles observaciones en las que no aparezca esta fuente de sesgo de la variable omitida y encuentra las políticas que se implantaron en China y Yugoslavia, los dos países socialistas más grandes de los no ocupados por tropas soviéticas después de la Segunda Guerra Mundial. Como China era un país enorme que había tenido su propia revolución, Stalin no podía indicarle qué políticas tenía que practicar. Los comunistas yugoslavos también habían llegado al poder sin la ayuda del Ejército Rojo, y el mariscal Tito demostró su independencia respecto a las órdenes de Moscú desde el final de la guerra.

China implantó una economía centralizada sin estar bajo la dominación política o militar de la Unión Soviética y en Yugoslavia las medidas de tipo estalinista se adoptaron *a pesar de* las políticas soviéticas. De estos datos, Halpern infiere que, en estos casos, el poder soviético no explica por sí solo los cambios en las políticas. Además, por lo que a China se refiere, también considera, y rechaza, otra hipótesis alternativa según la cual las ideas serían endógenas: situaciones económicas parecidas habrían aconsejado transplantar los métodos de planificación estalinista a China.

Halpern, una vez que se planteó y rechazó hipótesis alternativas que mantenían ideas endógenas bien en relación al poder soviético o a las condiciones económicas, pudo indicar que la adopción por parte de China (y, hasta cierto punto y durante menos tiempo, de Yugoslavia) de la doctrina estalinista aportaba bases para el acuerdo y resolvía la incertidumbre de estos regímenes postrevolucionarios. Aunque este tipo de análisis sigue siendo bastante provisional, a causa del escaso número de observaciones de su propia teoría que contempló la autora, sí aporta razones para creer que, en la situación estudiada, las ideas no eran del todo endógenas y tuvieron una función causal.

Este ejemplo pone de manifiesto de qué manera podemos, en primer lugar, trasladar una preocupación general sobre la endogeneidad a posibles fuentes específicas de sesgo de la variable omitida, para buscar después un subconjunto de observaciones en el que podrían no estar presentes esas causas de sesgo. En este caso, al transformar el problema en una cuestión de sesgo de la variable omitida, Halpern pudo comparar otras hipótesis explicativas con resultados especialmente productivos para su hipótesis principal. Se planteó diversas hipótesis explicativas alternativas para explicar la adopción de economías planificadas y se dio cuenta de que sólo en China, y hasta cierto punto en Yugoslavia, era razonable pensar que las doctrinas

## 5. Entender qué debe evitarse

estalinistas (las ideas en cuestión) fueran en gran medida exógenas. En consecuencia, centró su atención en China y Yugoslavia. Si no hubiera diseñado su estudio con cuidado para poder enfrentarse al problema de la endogeneidad, sus conclusiones habrían sido mucho menos convincentes; ¡pensemos, por ejemplo, que hubiera intentado demostrar sus planteamientos utilizando Polonia y Bulgaria!

### 4.5 Precisar cuáles son los componentes de la variable explicativa

En este apartado presentamos el quinto y último método para eliminar el sesgo que produce la endogeneidad. Su objetivo es dividir una variable explicativa potencialmente endógena en dos componentes: uno claramente exógeno y otro que sea, al menos en parte, endógeno. Posteriormente, el investigador sólo utilizará en su análisis causal la parte exógena de la variable explicativa.

En el estudio realizado por Verba, Schlozman, Brady y Nie (1993) sobre participación política voluntaria, encontramos un ejemplo de este método. A los autores les interesaba explicar por qué los afroamericanos son mucho más activos políticamente que los latinos, teniendo en cuenta que ambos grupos están igualmente desfavorecidos. Se dieron cuenta de que hay diversos factores que contribuyen a explicar la diferencia, entre ellos el poco tiempo transcurrido desde que se ha emigrado a los Estados Unidos y el conocimiento del idioma. Una de sus principales variables explicativas era la asistencia a servicios religiosos (en iglesias o sinagogas, por ejemplo). Evidentemente, los investigadores no podían controlar si los individuos asistían a esas reuniones; de ahí que no pudiera eliminarse la posibilidad de endogeneidad. En realidad, sospechaban que algunos latinos, y en mayor medida los afroamericanos, asistían a servicios religiosos porque eran políticamente activos. Quizás los que tenían interés en política entraban a formar parte de una iglesia porque ésta les daba la oportunidad de aprender o porque estaba muy politizada. Un clero muy consciente de su labor política podía adiestrar a los feligreses para ese tipo de actividad o estimularles políticamente. Dicho de otro modo, la orientación de la causalidad podía ir desde la política a las experiencias no políticas y no al revés.

Verba y sus colegas solucionaron este problema analizando los componentes de su principal variable explicativa. Para ello, partieron de la base de que las instituciones religiosas influyen en la participación política de dos maneras. En primer lugar, en ellas los individuos aprenden habilidades sociales (como dar un discurso o dirigir una reunión). A su vez, ese aprendizaje hace que el ciudadano sea más competente a la hora de participar en la vida política y que esté más dispuesto a hacerlo. En segundo lugar, los ciudadanos reciben constantes estímulos políticos (por ejemplo, a través de discusiones o cuando otros miembros del grupo les piden que se impliquen

directamente), y esta situación también tiene que influir en su actividad política. Estos autores señalaron que el primer componente es principalmente exógeno, mientras que el segundo es, al menos en parte, exógeno: o sea, que procede parcialmente del grado de participación política de los individuos (la variable dependiente).

Posteriormente, los autores llevaron a cabo un estudio auxiliar para contrastar su hipótesis en relación a los componentes exógeno y endógeno de la participación en los servicios religiosos. Desde el principio reconocieron que la posibilidad de que el individuo adquiera habilidades sociales en su iglesia depende de la estructura organizativa de ésta. Las que tengan una estructura jerárquica, donde el clero sea elegido por órganos rectores centralizados y los feligreses apenas puedan opinar sobre el funcionamiento interno, ofrecerán menos oportunidades de aprender habilidades sociales que aquellas cuya base cooperativa concede a los miembros un importante papel en el gobierno de la organización. La mayoría de los afroamericanos pertenecen a iglesias protestantes que funcionan de manera muy participativa, mientras que los latinos son miembros de congregaciones católicas organizadas de modo jerárquico. Los autores mostraron que esta diferencia es la que explica la probabilidad de que adquieran habilidades sociales. Por ejemplo, señalaron que en estos dos grupos, así como en los estadounidenses blancos, el factor que influye en el aprendizaje de esas habilidades es el tipo de confesión religiosa y no los componentes étnicos, ni otros rasgos sociales ni, especialmente, la participación política.

Después de convencerse de que la adquisición de habilidades sociales era realmente exógena a la participación política, Verba y los otros autores calibraron este factor en los servicios religiosos y lo utilizaron como variable explicativa, en vez de usar para tal fin la asistencia a tales servicios. Este procedimiento solucionó el problema de la endogeneidad, ya que ahora se habían analizado los componentes de la variable explicativa para poder incluir únicamente los exógenos.

Este estudio auxiliar aportó más pruebas que demostraban la resolución del problema de la endogeneidad, ya que la pertenencia de los latinos y afroamericanos a las iglesias no podía explicarse de manera coherente en función de su participación política concreta, puesto que su vinculación con ellas comienza desde la infancia a través de la familia. Las razones que explican el hecho de que los afroamericanos sean en su mayoría protestantes hay que buscarlas en la historia de la esclavitud en los Estados Unidos y en la de las organizaciones que se formaron en las plantaciones del sur, mientras que el catolicismo de los latinos procede de la conquista española de Hispanoamérica. Tampoco puede atribuirse la diferente estructura institucional de las congregaciones católicas y de las protestantes al interés que tienen los dirigentes de cada una de ellas por participar en la política estadounidense actual. En realidad, hay que retrotraerse a la época de la Reforma protestante para poder explicar tal diferencia.

## 5. Asignación de valores a la variable explicativa

En el apartado cuarto del capítulo 4 señalamos que los experimentos mejor dirigidos tienen dos ventajas: el control de la selección de observaciones y de los valores que se asignan a las variables explicativas en las unidades. Sólo se analizó la selección en ese punto. Como en este capítulo ya nos hemos ocupado del sesgo de la variable omitida y de otros problemas metodológicos, podemos enfrentarnos ahora al control de la asignación de valores.

En un experimento médico, los tratamientos son un fármaco que se está probando y un placebo, los cuales se administran aleatoriamente a los pacientes. Aquí se da prácticamente la misma situación que aparece con una selección aleatoria de observaciones: la asignación aleatoria es muy útil cuando existe una gran número de observaciones, pero no suele ser la estrategia óptima si  $n$  es pequeño. Cuando es grande, asignar valores aleatorios a las variables explicativas acaba con la posible endogeneidad (porque dichas variables no pueden verse influidas por la variable dependiente) y con el error de medida (siempre que se registre exactamente qué tratamiento se administra). Quizá lo más importante sea que la asignación de valores en estudios con un  $n$  grande hace extremadamente improbable la aparición de sesgo de la variable omitida, porque una variable explicativa cuyos valores se hayan asignado aleatoriamente no se correlacionará con ninguna variable omitida, ni siquiera con las que influyan en la dependiente. Por lo tanto, en estudios de  $n$  grande, la asignación aleatoria de valores hará que las variables omitidas no sean dañinas, que no produzcan sesgo. Sin embargo, si se dispone de un número reducido de observaciones, es muy fácil que una variable asignada de forma aleatoria se correlacione con alguna variable omitida relevante y que tal correlación produzca sesgo de la variable omitida. De hecho, el ejemplo del sesgo de selección mostraba cómo una variable asignada aleatoriamente se correlacionaba con una variable dependiente observada; es muy fácil que ocurra exactamente lo mismo si el número de observaciones es pequeño.

Aunque al realizar experimentos muchas veces se pueden fijar los valores de las variables explicativas, los investigadores cualitativos no suelen tener esa suerte. Cuando el objeto de estudio determina el valor de sus propias variables explicativas, aumentan considerablemente las posibilidades de sesgo de selección, de endogeneidad y las de otros tipos de sesgo y de inefficiencia. Por ejemplo, si en un experimento se estuviera estudiando la eficacia política que tiene participar en manifestaciones, se elegiría aleatoriamente a ciertas personas para que acudieran a una manifestación y a otras para que se quedaran en casa, con el fin de calibrar después las diferencias que se registran entre los dos grupos del experimento en relación a la eficacia (o quizás para compararlos en función del cambio que registra ese indicador en una medida realizada antes del experimento y después de él). Sin embargo, en el trabajo no experimental, es frecuente que los pro-

### Un análisis formal de la endogeneidad

Este modelo formal demuestra el sesgo que aparece cuando un diseño de investigación padece endogeneidad y no se hace nada por evitarlo. Supongamos que tenemos una variable explicativa  $X$  y una dependiente  $Y$ . Nos interesa el efecto causal que tiene  $X$  sobre  $Y$ , y utilizamos la siguiente ecuación:

$$E(Y) = X\beta \quad (5.10)$$

Ésta también puede expresarse como  $Y = X\beta + \varepsilon$ , en la que  $\varepsilon = Y - E(Y)$  se denomina el «error» o «término perturbador». Supongamos, además, que hay endogeneidad, o sea, que  $X$  también depende de  $Y$ :

$$E(X) = Y\gamma \quad (5.11)$$

¿Qué ocurre si no prestamos atención a la parte recíproca de la relación expresada en la ecuación (5.11) y hacemos una estimación de  $\beta$  como si sólo fuera cierta la ecuación (5.10)? Dicho de otro modo, estimamos  $\beta$  (presuponiendo incorrectamente que  $\gamma = 0$ ) mediante la ecuación habitual:

$$b = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} \quad (5.12)$$

Para evaluar este estimador utilizamos la propiedad de falta de sesgo y, por lo tanto, calculamos su valor esperado:

$$E(b) = E\left(\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}\right) \quad (5.13)$$

pios objetos de estudio decidan si quieren participar. En estas circunstancias, hay otros rasgos individuales (como el hecho de que los individuos sean jóvenes o no, que sean estudiantes y, si lo son, la cercanía del campus al lugar de la manifestación, etc.) que influirán en la decisión de manifestarse. Además, es evidente que muchos de estos factores también pueden correlacionarse con la variable dependiente, que es la eficacia política.

Pensemos en otro ejemplo en el que las unidades de análisis sean más grandes y menos frecuentes: el clásico problema del impacto que tiene la acumulación de armamento en la posibilidad de guerra. ¿El tamaño del presupuesto nacional para esta partida aumenta las probabilidades de que el país participe en una guerra? La variable explicativa es el presupuesto para

$$\begin{aligned} &= E\left(\frac{\sum_{i=1}^n X_i (X_i \beta + \varepsilon_i)}{\sum_{i=1}^n X_i^2}\right) \\ &= \beta + \frac{\sum_{i=1}^n C(X_i, \varepsilon_i)}{\sum_{i=1}^n V(X_i)} \\ &= \beta + \text{sesgo} \end{aligned}$$

donde el sesgo =  $\sum_{i=1}^n C(X_i, \varepsilon_i) / \sum_{i=1}^n V(X_i)$ . Normalmente, la covarianza de  $X_i$  y el término perturbador  $\varepsilon_i$ ,  $C(X_i, \varepsilon_i)$  valen cero, de manera que el término de sesgo también tendrá este valor. En consecuencia, el valor esperado de  $b$  es  $\beta$  y, por tanto, carece de sesgo. Suele ocurrir que, después de que se tiene en cuenta  $X$  para predecir  $Y$ , la parte que queda ( $\varepsilon$ ) no se correlaciona con  $X$ . Sin embargo, en el presente ejemplo, después de que se contempla el efecto de  $X$ , sigue quedando alguna variación por la retroalimentación que procede del efecto causal de  $Y$  sobre  $X$ . De este modo, la endogeneidad implica que, en general, el segundo término de la última línea de la ecuación (5.13) no valdrá cero y que la estimación estará sesgada.

La dirección del sesgo depende de la covarianza, ya que la varianza de  $X$  siempre es positiva. Sin embargo, en los pocos casos en los que ésta sea extremadamente grande, arrollará a la covarianza y hará que el término de sesgo resulte insignificante. En el texto se ofrece un ejemplo con una buena interpretación de este término de sesgo.

armamento (quizá el porcentaje en relación al PNB o los cambios que ha sufrido el propio presupuesto); la variable dependiente es la presencia o ausencia de guerra en un período determinado posterior al momento en que se ha medido la variable explicativa. En un diseño de investigación ideal sería el investigador el que asignara valores a la variable explicativa, eligiendo diversos países para su estudio y determinando el presupuesto armamentístico de cada gobierno (mediante valores aleatorios o utilizando quizás una de las variables «intencionadas» que analizamos más adelante). Es evidente que esto no es factible. Lo que se hace en realidad es medir los valores de la variable explicativa (el tamaño del presupuesto armamentístico) que cada gobierno se asigna. Por supuesto, el problema es que estos valores autoa-

signados de la variable explicativa no son ajenos a la dependiente —la probabilidad de ir a la guerra—, como lo hubieran sido si hubiéramos podido elegirlos nosotros. En este caso, hay un claro problema de endogeneidad: el valor de la variable explicativa está influido por los valores que se cree tendrá la dependiente: la amenaza de guerra percibida. La endogeneidad también plantea problemas para estudiar la relación causal entre las alianzas y la guerra. Las naciones eligen sus alianzas; los investigadores, sin adjudicar países a esas alianzas, estudian el impacto que tienen sobre las guerras. Por lo tanto, las alianzas no tendrían que considerarse variables explicativas exógenas en los estudios bélicos, ya que suelen formarse antes que la propia guerra.

Estos ejemplos demuestran que la endogeneidad no siempre es un problema que haya que precisar, sino que a veces constituye una parte integral del proceso por el que el mundo genera observaciones. Con frecuencia, resulta bastante difícil averiguar cómo se han determinado los valores de las variables explicativas, y no suele ser posible recurrir a ningún procedimiento automático para resolver los problemas que plantea. No obstante, al investigar no puede eludirse esta tarea.

Como la probabilidad de que, en cualquier prueba de un experimento hipotético, el uso de criterios aleatorios de selección o de asignación de valores produzca sesgo disminuye rápidamente a medida que aumenta el número de observaciones, es útil recurrir a tales procedimientos aunque no se disponga de un gran número de unidades. Si la cantidad de unidades es «suficientemente grande» —lo cual definimos con precisión en el apartado 2 del capítulo 6—, la selección aleatoria de unidades cumplirá automáticamente el presupuesto de independencia condicional del apartado tercero del capítulo 3. Sin embargo, cuando sólo existan unos pocos ejemplos del fenómeno que nos interesa o únicamente podamos recabar información acerca de un pequeño número de observaciones, como ocurre con frecuencia en la investigación cualitativa, la respuesta no estará en la selección y asignación de valores con procedimientos aleatorios. Ni siquiera los experimentos controlados, si son posibles, serán una solución cuando no se disponga de un número suficiente de observaciones.

Los cualitativistas, al enfrentarse a estos problemas, tienen que preguntarse si es posible aumentar el número de observaciones que van a investigar, ya que, a falta de reunirlas todas, el procedimiento más fiable es recoger de forma aleatoria datos relacionados con un gran número de unidades y aleatorizar la asignación de valores para las variables explicativas. Sin embargo, si esto no es posible, no hay que seleccionar las observaciones aleatoriamente, sino que es preciso utilizar el conocimiento que ya tenemos de las observaciones disponibles —que se basa en investigaciones anteriores, suposiciones fundadas o valoraciones de otros expertos en el área— y hacer la selección de las observaciones y (si se puede) la asignación de valores a las variables explicativas de tal modo que se evite el sesgo y las ineficiencias. Si

## 5. Entender qué debe evitarse

el sesgo es inevitable, al menos habrá que intentar averiguar su dirección y su posible magnitud. Si todo esto falla —es decir, si sabemos que hay sesgo pero no podemos determinar su dirección o envergadura—, nuestra investigación mejorará si, por lo menos, describimos nuestros resultados con un mayor grado de incertidumbre. Si comprendemos los problemas inferenciales que aborda este libro, estaremos más preparados que cualquier generador de números aleatorios para enfrentarnos a todos estos dilemas. En cualquier caso, *en todos los estudios tendría que haber un apartado o capítulo en el que se explicaran cuidadosamente los procesos de selección y de asignación de valores*. Habría que mencionar en él qué reglas se han utilizado, incluir un listado detallado de todas las previsibles fuentes de sesgo ocultas y señalar si se ha hecho algo para combatirlas.

## 6. Controlar la situación que se investiga

Si se seleccionan de forma deliberada observaciones sin tener en cuenta las variables de control relevantes y tampoco otros problemas inferenciales, no se cumplirá el presupuesto de homogeneidad de las unidades. Tenemos que asegurarnos de que los valores de la variable explicativa en las observaciones elegidas se midan con el mínimo error posible, de que no se correlacionen con alguna variable explicativa omitida clave y de que no estén parcialmente determinados por la dependiente. Es decir, hay que calibrar de manera adecuada problemas como el error de medida, las variables omitidas y la endogeneidad, que ya se han analizado en este capítulo. Si éstos persisten, a pesar de todos nuestros esfuerzos, por lo menos tendremos que reconocerlos, sopesarlos e intentar corregirlos.

En estudios que tienen un *n* pequeño los controles son de por sí difíciles de diseñar, pero suele ser absolutamente esencial ocuparse de ellos para evitar el sesgo. Por desgracia, hay muchos investigadores cualitativos que apenas utilizan controles o que no lo hacen en absoluto. Por ejemplo, Bollen, Entwistle y Alderson (1993) han señalado, en un estudio sobre libros y artículos sociológicos, que alrededor de un cuarto de los investigadores no utiliza ningún método de control.

Supongamos, por ejemplo, que nos interesaría el efecto causal que tiene un año de cárcel en el grado de radicalismo político que muestran las personas. Lo ideal sería un estudio auténticamente experimental en el que se utilizaría una muestra aleatoria de un gran número de ciudadanos y que a la mitad de ellos se les adjudicara aleatoriamente un año de prisión, para medir después la radicalidad de las ideas políticas de uno y otro grupo. El efecto causal estimado sería la diferencia media entre las posturas de los dos grupos al final del año. Con un *n* grande, cabría presuponer la existencia de independencia condicional y probablemente la inferencia causal fuera sensata. No hace falta decir que tal estudio es imposible.

## El diseño de la investigación social

Sin embargo, para continuar con nuestro argumento, supongamos que se llevara a cabo el experimento, pero sólo con pocas personas. Teniendo en cuenta los problemas que se han analizado en el segundo apartado del capítulo 4, esta escasa muestra, aunque se hubiera seleccionado aleatoriamente y la asignación de valores también se hubiera hecho con este método, es probable que no cumpliera la independencia condicional y, por lo tanto, sería necesario aplicar algún control explícito. Una sencilla forma de hacerlo sería medir las ideas políticas radicales antes del experimento. Después, nuestra estimación causal sería la diferencia que hubiera entre los dos grupos en relación al *cambio* operado en las ideas políticas radicales. Este procedimiento nos permitiría controlar una situación en la que ambos grupos no son idénticos respecto a esta única variable antes de realizarse el experimento. Para comprender cómo hay que calcular el efecto causal en este contexto conviene recordar el problema fundamental de la inferencia causal. Lo ideal sería que tomáramos a un solo individuo, esperáramos durante un año controlando cuidadosamente su medio para mantenerlo constante, excepto en lo relativo al paso del tiempo y a los acontecimientos del mundo exterior, y que después midiéramos la radicalidad de sus ideas políticas. Simultáneamente, tomariamos a ese individuo en el mismo momento y lo meteríamos en la cárcel durante un año, para medir después el mismo indicador. La diferencia entre estas dos medidas es la definición del efecto causal que tiene la cárcel sobre las ideas políticas de esta persona<sup>14</sup>. El problema fundamental consiste que sólo podemos observar las ideas de esta persona en una de las situaciones. Es evidente que el mismo individuo no puede estar, a la vez, dentro y fuera de la cárcel.

Mediante el control se intenta evitar el problema fundamental de la manera más directa. Con lo que no podemos observar las ideas de esta persona en ambas situaciones, buscamos dos individuos (o, más probablemente, dos grupos de individuos) que sean lo más parecidos posible, excepto en lo tocante a la variable explicativa (que hayan ido o no a la cárcel). Tampoco hacemos la selección teniendo en cuenta el grado de radicalidad de las personas. En principio, podríamos seleccionar una muestra de individuos que hubieran sido puestos en libertad recientemente y buscar después personas *equiparables* a cada ex prisionero (lo más parecido posible a éste, pero sin haber estado en la cárcel). Quizás, en primer lugar, podríamos entrevistar a alguien que hubiera salido de la cárcel y, partiendo del conocimiento de su historia y características, ir en busca de personas equiparables (que tuvieran un perfil demográfico similar y que quizás fueran del mismo barrio y hubieran ido al mismo colegio).

Por definición, las variables con las que equiparamos a los individuos son las mismas en todos los grupos y las controlaremos cuando calculemos el efecto causal de la excarcelación. Ejercer controles es un proceso difícil porque hay que aplicarlos a todas las variables que puedan ser confusas. Si no equiparamos en función de una determinada variable, si no podemos

## 5. Entender qué debe evitarse

controlarla de ninguna otra manera e influye en la variable dependiente a la vez que se correlaciona con la explicativa (afecta a la radicalidad de las ideas y no es la misma en los prisioneros y en los que no lo son), la estimación de nuestro efecto causal estará sesgada.

Cuando se comparan países desde el punto de vista político, ejercer controles para lograr la homogeneidad de las unidades es difícil: cualquier pareja de países diferirá en innumerables dimensiones. Por ejemplo, Bélgica y Holanda podrían parecerle a un observador poco preparado países «muy parecidos», en el sentido que señalan Przeworski y Teune (1982): son pequeñas democracias europeas con economías abiertas y sus vecinos no les amenazan. Por lo tanto, en muchos sentidos, sería factible compararlos (Katzenstein, 1985). Sin embargo, son muy diferentes desde el punto de vista lingüístico o religioso y en cuanto a sus recursos fundamentales, al inicio de su industrialización y a muchos otros factores políticos relevantes. En consecuencia, cualquier diseño de investigación comparado que pretenda estudiar el conjunto de la vida política en estos dos estados y que sólo se centre en ellos correrá el riesgo de ser impreciso.

Si lo que pretendemos es hacer una comparación general entre Bélgica y Holanda, se podrá evitar tal imprecisión. Sin embargo, supongamos que el investigador tenga un objetivo más específico: estudiar el impacto de ser una potencia colonial sobre las estrategias políticas que siguen los gobiernos de las pequeñas democracias europeas. En ese caso, sería posible comparar las políticas de Bélgica, Holanda y Portugal con las de pequeños estados no coloniales como Austria, Suecia, Suiza y Noruega. Éste podría ser un diseño de investigación valioso, pero no estaría controlando innumerables factores —aparte de la historia colonial— que hacen que estos países sean diferentes. Un investigador sensible al problema de la homogeneidad de las unidades podría pensar en otro diseño —quizá alternativo, aunque sería mejor que fuera un complemento para el anterior— en el que estudiaría las políticas belgas, holandesas y portuguesas antes y después de la pérdida de sus respectivas colonias. En este diseño, Bélgica no sería una «única observación», sino el escenario de un análisis controlado, antes y después de que se concediera la independencia a sus colonias a principios de los años sesenta. Muchos de los factores que diferencian a Bélgica de Portugal y de Holanda —y no digamos de los países que carecen de un pasado colonial— se controlarían automáticamente en este diseño de series temporales. En realidad, ambas comparaciones —entre países y entre diversos momentos de cada país— presentan problemas en cuanto a la homogeneidad de las unidades. Las diversas naciones difieren en muchas formas no controladas y no medidas, que podrían ser relevantes para el tema de la investigación, y una nación que sea observada en diferentes momentos también presenta estas disparidades. Sin embargo, las *diferencias* no serán las mismas. Ninguna comparación (ya sea en el tiempo o en el espacio) constituye, ni mucho menos, un experimento perfectamente controlado, pero si juntamos ambos

enfoques podremos lograr más pruebas de nuestra hipótesis que si sólo utilizamos uno de ellos.

La estrategia de la selección intencionada conlleva algunos peligros ocultos de los que los investigadores tendrían que ser conscientes, especialmente al intentar equiparar observaciones con el fin de controlar variables potencialmente relevantes. El principal peligro se encuentra en un tipo de sesgo de la variable omitida especialmente engañoso. Imaginémosnos el siguiente diseño de investigación en el que se utiliza la equiparación. Con el fin de animar a los países africanos que parecen encaminarse hacia una mayor democratización, el gobierno de los Estados Unidos decide enviarles asistencia —textos educativos sobre la democracia y materiales similares— a través de un programa denominado «ayuda a la democracia». El investigador quiere estudiar si, en cada país, esa ayuda aumenta el nivel de democracia, si lo disminuye o si no influye en él, pero no puede conceder o suspender la ayuda al mismo país simultáneamente. En consecuencia, opta por un enfoque prospectivo-comparativo: compara los países que están a punto de recibirla con los que no lo están. También decide, con razón, buscar unidades en ambos grupos que sean equiparables en cuanto a los valores de todas las variables de control relevantes, excepto en aquella que le preocupa: el programa de ayuda estadounidense.

El tiempo y los idiomas que conoce condicionan su investigación, de manera que sólo puede estudiar dos países (aunque los problemas que van a mencionarse seguirían apareciendo en un estudio que tuviera un mayor —aunque todavía escaso— número de unidades). Elige un país que recibe bastante ayuda del citado programa norteamericano y otro que apenas la recibe. Acierta al situar su variable dependiente en el incremento del nivel de democracia que se registra en el país desde que comienza el programa de ayuda hasta que se hace el estudio dos años más tarde. Como hay muchas otras variables que podrían correlacionarse tanto con la variable explicativa como con la dependiente, el investigador intenta elegir dos países que sean muy equiparables al respecto, con el fin de evitar el sesgo de la variable omitida.

Podría haber dos variables de control: el nivel educativo del país y el de violencia guerrillera contra el régimen. Ambas pueden producir sesgo si no se controlan, porque las dos se correlacionan tanto con la variable explicativa como con las dependientes (recuérdese el apartado segundo de este capítulo, que versaba del sesgo de la variable omitida). Es probable que los Estados Unidos den más ayuda a aquellos países que tienen un buen sistema educativo (quizá porque éstos podrán establecer mejores relaciones con Washington o porque Estados Unidos impulsa la educación), y, a veces, la enseñanza funciona como una fuerza democratizadora. Del mismo modo, los Estados Unidos prefieren ayudar a los países donde la actividad guerrillera es escasa, y, por supuesto, una amenaza de este tipo hace que disminuya la probabilidad de democratización. Al equiparar en función de estas variables, el investigador espera controlar sus efectos engañosos.

5. Entender qué debe evitarse

Sin embargo, siempre hay otras variables omitidas que podrían causar sesgo porque se correlacionan tanto con la variable explicativa clave como con la dependiente (y porque, desde el punto de vista causal, son anteriores a aquélla). El problema está en que intentar equiparar dos unidades, si se hace de forma incorrecta o incompleta, puede aumentar la probabilidad de que haya otra variable omitida significativa que se correlacione tanto con la variable explicativa como con la dependiente.

¿Por qué es así? Hay que señalar que, para equiparar países, el investigador tiene que encontrar uno que reciba bastante ayuda y otro que reciba poca. Supongamos que elige dos países que son similares respecto a las otras dos variables, es decir, que tienen un alto nivel educativo y pocas amenazas internas. El resultado sería el siguiente:

País A: mucha ayuda, mucha educación, pacífico.

País B: poca ayuda, mucha educación, pacífico.

Es posible que el país B tenga algo «especial». ¿Por qué no recibe ayuda si tiene condiciones tan favorables? Quizá lo «especial» sea una variable omitida que causará sesgo porque se correlaciona con la variable explicativa y con la dependiente. Un ejemplo de ello podría ser la existencia en B pero no en A de un poderoso ejército que impulse la educación y acabe con los movimientos guerrilleros. Como la fuerza de los militares se correlaciona con la variable dependiente y con la explicativa principal, omitir este factor producirá sesgo. Se puede apreciar que habría existido el mismo problema si la equiparación hubiera partido del otro extremo de las gradaciones de educación y de paz interna. En ese caso, la anomalía sería el país que registra un nivel de educación bajo y mucha violencia y que recibe mucha ayuda. El problema se podría mitigar si la equiparación se realizará utilizando los valores medios de las variables que se basan en la educación y la paz interna. Sin embargo, incluso en este caso, el investigador dispondría de dos países anómalos, cada uno en una dirección opuesta. Lo importante aquí es que, a veces, la equiparación nos induce a buscar observaciones que, en cierto sentido, se apartan de lo que cabría esperar si se dan unos determinados valores en las variables de control y que esa desviación puede deberse a variables omitidas especialmente significativas.

Nótese de qué manera esto tendría que funcionar en el ejemplo de la cárcel. Podríamos intentar equiparar a los prisioneros según ciertas observaciones: que fueran parecidos en cuanto a la procedencia social y económica, la historia de su familia, el expediente escolar y otros factores similares, pero no en el hecho de estar en la cárcel. La mejor equiparación sería la que se basara en personas sin pasado carcelario pero con muchas probabilidades de haberlo tenido, por proceder de un barrio muy pobre; haber dejado el colegio, estar en contacto con el mundo de las drogas, proceder de una familia rota, etc. Cuanto mejor sea la equiparación, más podremos fiarnos

de la relación entre encarcelamiento e ideas políticas. Sin embargo, aquí está de nuevo el problema. Si todo estaba en contra de los no encarcelados, puede que éstos tengan algo especial que les ha mantenido fuera de la prisión —quizá una arraigada práctica religiosa— y que se correlaciona tanto con la variable explicativa (encarcelamiento) como con la dependiente (ideología política).

Hay otra forma de abordar este peligro de la equiparación. Recordemos los dos puntos de vista respecto a la variabilidad aleatoria que se describieron en el apartado sexto del capítulo 2. En la equiparación, tal como la hemos trazado hasta ahora, es posible que plantea problemas una variable omitida identificable. Sin embargo, aún podríamos sospechar que las dos variables equiparadas en una larga lista de variables de control son «especiales» en algún sentido que no podemos determinar; es decir, que existe una variable omitida desconocida. En esta situación, lo único que podemos hacer es ocuparnos de cómo la aleatoriedad inherente a nuestra variable dependiente influirá en esta observación. Cuanto más se vaya alejando nuestra medida de su verdadero valor, a causa de la variación aleatoria, con más ahínco buscaremos observaciones «ínnusuales» con el fin de lograr una mayor equiparación entre los grupos y, de este modo, nos arriesgaremos a caer en un sesgo de la variable omitida.

Estas matizaciones *no* tendrían que llevarnos a evitar diseños de investigación que utilicen la equiparación. De hecho, este procedimiento es una de las estrategias más valiosas cuando  $n$  es pequeño. Simplemente, habrá que ser conscientes de que con la equiparación, al igual que con todas las estrategias que se utilizan cuando  $n$  tiene un tamaño reducido, se corren riesgos que se hubieran evitado con la aleatorización y con una muestra más grande. Un procedimiento muy fructífero es utilizar la equiparación para elegir estudios de caso pero seleccionar con otros criterios las observaciones de cada uno de ellos.

Si lo que se quiere es evitar el sesgo de la variable omitida, la equiparación tiene que ver con el debate que se mantiene actualmente en la bibliografía sobre política comparada acerca de la conveniencia de seleccionar observaciones lo más parecidas posible (Lijphart, 1971) o lo más diferentes que se pueda (Przeworski y Teune, 1970). Aquí recomendamos un enfoque diferente. Este debate sobre el diseño de investigación, que enfrenta a los partidarios de lo «más parecido» con los de lo «más diferente», se ocupa poco, o nada, de precisar «a qué hay que parecerse». Las etiquetas suelen ser confusas y la polémica no resulta concluyente en estos términos: ninguno de los dos enfoques es siempre preferible. Para nosotros, lo principal a la hora de recoger datos es identificar observaciones que puedan maximizar el control sobre las hipótesis causales. A veces nuestra estrategia conducirá a un plan de investigación que podría calificarse de «diseño de sistemas lo más parecidos posible», otras a lo que se podría denominar «diseño de sistemas lo más diferentes posible». Sin embargo, a diferencia del debate que

#### 5. Entender qué debe evitarse

contrapone ambas posibilidades, nuestro procedimiento siempre generará datos relevantes para responder a las preguntas del investigador.

En la equiparación, los posibles efectos de las variables omitidas se controlan mediante la selección de observaciones que tienen el mismo valor en tales variables. Por ejemplo, el deseo de mantener constante el mayor número posible de variables contextuales subyace en las opciones de Seymour Martin Lipset (1963, p. 248) al comparar el desarrollo político de los Estados Unidos con el de otras ex colonias británicas de habla inglesa. Lipset indica que los Estados Unidos, Canadá y Australia «... son ex colonias de Gran Bretaña que fijaron fronteras relativamente abiertas y que hoy son estados federales que abarcan continentes». Además, señala muchos otros rasgos comunes que se mantienen constantes: el nivel de desarrollo, el régimen democrático, la similitud de valores, etc.

El estudio que realizó David Laitin (1986) sobre los efectos de las creencias religiosas en la política utiliza una técnica de equiparación especialmente cuidadosa. Eligió un país, Nígeria, en el que eran muy acentuadas tanto la tradición musulmana como la cristiana porque quería comparar las consecuencias que tenían ambas en la política. Sin embargo, las zonas musulmana y cristiana de Nígeria difieren en muchos otros rasgos, además de en el religioso, y hacer caso omiso de ellos habría sido arriesgarse a incurrir en un sesgo de la variable omitida. «En Nígeria, los centros islámicos dominantes se encuentran en los estados del norte, que han tenido durante siglos contacto directo con el mundo musulmán, una historia como países islámicos que precede al dominio británico y una memoria del renacer de la guerra santa a principios del siglo XIX que unificó una zona muy amplia bajo una doctrina islámica ortodoxa. [Por el contrario], hasta finales del siglo XIX las comunidades cristianas no arraigaron [...]. Las escuelas de las misiones trajeron la educación occidental y los empresarios capitalistas animaron a la población a introducir cultivos comerciales y a entrar cada vez más en la economía capitalista mundial» (Laitin, 1986, p. 187).

Laitin se preguntó cómo «se podían controlar las diferencias nacionales, las económicas, las relativas al número de generaciones que se habían visto influidas por una cultura abierta al mundo, las que tenían que ver con las motivaciones para convertirse o las de tipo ecológico; todas ellas diferentes en los baluartes cristianos y musulmanes (1986, pp. 192-193). Optó por elegir una determinada localidad de la región nigeriana de Yoruba, en la que las dos religiones se habían introducido en el mismo grupo nacional más o menos a la vez y donde ambos credos atraían por razones parecidas a los posibles conversos.

La equiparación no es completa ni en el estudio que realizó Kohli sobre los tres estados de la India ni en el análisis de Lipset acerca de las tres ex colonias británicas, ni tampoco en la investigación de Laitin sobre los musulmanes de la tierra yoruba, y nunca podría serlo. Para equiparar es necesario prever y especificar qué posibles variables omitidas relevantes van a

## 6. Aumentar el número de observaciones

A lo largo de este libro hemos venido recalculando la importancia crucial que tiene maximizar el control sobre los problemas de la investigación. La principal forma de hacerlo es encontrar tantas consecuencias observables de nuestra teoría como sea posible y hacer observaciones acerca de ellas. Tal como hemos subrayado, lo que puede parecer un estudio de un solo caso, o de unos pocos, quizás contenga en realidad muchas posibles observaciones que sean relevantes para la teoría que se está contrastando y que se encuentren en diferentes niveles de análisis. Al aumentar el número de observaciones, incluso sin recoger más datos, el investigador suele poder transformar un problema insoluble que tiene un diseño de investigación impreciso en algo manejable. En este último capítulo se dan consejos sobre cómo aumentar el número de observaciones relevantes en un estudio social de tipo científico.

Comenzaremos por analizar los problemas inherentes a las investigaciones que se ocupan de una sola observación: el problema que se plantea cuando  $n = 1$ . Demostraremos que, si realmente hay una sola observación, resulta imposible evitar el problema fundamental de la inferencia causal. Incluso en supuestos ejemplos de comprobación de un único caso, el investigador tiene que examinar, al menos, el pequeño número de observaciones que existen dentro de los «casos» y compararlas. Sin embargo, una inferencia causal puede llegar a apoyarse en una comparación disciplinada que se centre en un reducido número de estudios de caso que ofrezcan observaciones equiparables.

aparecer. Después se controla la investigación seleccionando observaciones cuyos valores no cambian en dichas variables. Es evidente que no podemos saber si hemos cubierto toda la lista de posibles factores de sesgo. Sin embargo, si tenemos ciertos objetivos analíticos —y al evaluar si es adecuado un procedimiento de selección que tenga como fin equiparar hay que relacionarlo con algún objetivo de este tipo—, el control que genere la equiparación aumentará la probabilidad de obtener inferencias válidas.

En suma, el investigador que intente hacer inferencias causales puede seleccionar los casos de dos maneras. La primera es la que utiliza criterios aleatorios para seleccionar y adjudicar valores, que es útil en estudios con un  $n$  grande. En tales investigaciones, la aleatoriedad cumple automáticamente la independencia condicional y es un procedimiento mucho más fácil que el de seleccionar observaciones a propósito para cumplir la homogeneidad de las unidades. La aleatoriedad nos garantiza que no se omiten variables relevantes y que la selección de observaciones no se está realizando con criterios que se correlacionan con la variable dependiente (después de controlar las explicativas). Este procedimiento también nos garantiza que los sesgos del investigador no entran en el criterio de selección y que, por tanto, no influyen en los resultados. El segundo método es el de la selección intencionada de observaciones, que recomendamos en estudios que tengan un  $n$  reducido. Este tipo de investigaciones, que parten de una selección intencionada para hacer inferencias causales razonables, casi siempre arrojarán inferencias más arrriesgadas que las que aplican criterios aleatorios a grandes muestras, y dependerán más de las opiniones previas que tenga el investigador sobre el mundo empírico; en ellas los controles pueden introducir sesgos diversos y sutiles. No obstante, por las razones que se han subrayado, los controles son necesarios en un estudio de  $n$  pequeño, y, si son adecuados —si las variables de control se mantienen constantes, quizá mediante la equiparación—, quizás sólo haya que calcular el efecto causal de una única variable explicativa, con lo que aumentaremos el control que tenemos sobre un problema.

## 7. Observaciones finales

Esperamos que los consejos que hemos dado en este capítulo y en el anterior sean útiles para los investigadores cualitativos, aunque no sean recetas que siempre puedan aplicarse fácilmente. En la realidad, los problemas no suelen venir solos. Por ejemplo, supongamos que un investigador detecta un sesgo de selección menor, un cierto error de medida en la variable dependiente y una importante variable de control que sólo puede medirse ocasionalmente. Si sigue los consejos que se han dado anteriormente para cada caso, tendrá cierta idea sobre lo que hay que hacer. Sin embargo, en este y en otros casos complejos, los que realizan estudios cualitativos tienen que

## 5. Entender qué debe evitarse

reflexionar con cuidado sobre los problemas metodológicos específicos que aparecen en sus investigaciones. Puede serles de utilidad remitirse a modelos formales de investigación cualitativa parecidos a los que hemos presentado aquí, pero sólo si están en consonancia con los problemas de su investigación. Gran parte de las ideas que respaldan los modelos formales más sofisticados están en los escritos estadísticos, de manera que no siempre es necesario desarrollar un modelo propio.

El cualitativista, independientemente de que se ayude o no con modelos formales, tiene que atender de manera explícita esas cuestiones metodológicas. Éstas son tan importantes para los investigadores cualitativos que pretenden hacer inferencias causales como para los de orientación cuantitativa.

El análisis que hacemos en el apartado 1 de este capítulo de los diseños que se centran en una sola observación puede parecerle pesimista al investigador que utilice el estudio de caso. Sin embargo, como un solo caso puede contener en realidad muchas posibles observaciones, este pesimismo no está justificado, aunque si está garantizada la búsqueda persistente de más observaciones. Después de criticar los diseños que trabajan con una única observación y, en consecuencia, dar sólidas razones para que se aumente el número de observaciones, pasaremos a analizar cuántas son suficientes para alcanzar niveles de incertidumbre satisfactorios (apartado 2). Para terminar, en el apartado 3 demostraremos que casi todos los diseños de investigación cualitativos pueden reformularse para darles muchas observaciones y que, si el investigador conceptualiza adecuadamente las consecuencias observables que ya tiene, esto suele poder hacerse sin que sea necesario llevar a cabo una costosa recogida de datos adicionales.

### 1. Inferencias causales con diseños de una sola observación

En una investigación el problema más difícil se plantea cuando el investigador sólo dispone de una unidad para evaluar una teoría causal, o sea, cuando  $n = 1$ . En este apartado comenzaremos a analizar esta cuestión señalando que es muy improbable que se pueda salir airoso de ella. En primer lugar, estudiaremos el argumento de Harry Eckstein en el artículo clásico que dedicó a los estudios de caso cruciales (apartado 1.1). Después, en el apartado 1.2, nos ocuparemos de un ejemplo concreto, el razonamiento analógico.

#### 1.1 Estudios de caso «cruciales»

Eckstein ha señalado con contundencia que, cuando no se aclaran las condiciones en las se espera que aparezcan ciertas pautas de comportamiento, resulta imposible comprobar bien o mal esas teorías (Eckstein, 1975). Estamos de acuerdo con él en que los investigadores tienen que esforzarse por encontrar teorías que sean precisas en sus pronósticos y por contrastarlas con datos del mundo real.

Sin embargo, Eckstein va aún más lejos y afirma que si nuestra teoría hace pronósticos precisos se podrá utilizar para fines explicativos un estudio de «caso crucial» que, para él, es una investigación que sólo se basa en una «única medida de cualquier variable pertinente» (lo que nosotros denominamos «observación única»). Lo principal en el capítulo de Eckstein es su idea de que «los estudios de caso [...] [son] más valiosos cuando [...] se llega al punto en el que se "ponen a prueba" otras posibles teorías» (1975, p. 80). En concreto, señala (1975, p. 127) que «un solo caso crucial puede dejar realmente fuera de combate a una teoría». Para Eckstein, los estudios

### 6. Aumentar el número de observaciones

de caso cruciales pueden refutar teorías suficientemente precisas con una sola observación. En particular, si un investigador elige un estudio de caso que, en principio, no es probable que concuerde con predicciones teóricas —una observación «muy poco probable»— y resulta que, a pesar de todo, la teoría es correcta, ésta habrá superado una prueba difícil y tendremos razones para suscribirla con una mayor confianza. Por el contrario, si los pronósticos de una teoría que no parece plausible coinciden con las observaciones de otra observación «muy probable», la teoría no habrá superado un examen riguroso pero si una «prueba de plausibilidad», con lo que quizás merezca la pena seguir analizándola.

El argumento de Eckstein es bastante valioso, especialmente cuando indica a los investigadores que comprendan si han de evaluar sus teorías mediante observaciones «poco probables» o «muy probables». La consistencia de nuestra inferencia, en relación a la validez de la teoría, dependerá en gran medida de lo difícil que haya sido la prueba que ésta haya logrado o no superar. Sin embargo, la defensa que hace Eckstein de una comprobación que se base en una observación crucial no es coherente con el problema fundamental de la inferencia causal. Por lo tanto, si define el «caso» igual que nosotros la observación única, creemos que está equivocado en este punto<sup>1</sup>.

Hay tres razones por las que dudamos de que un estudio que se base en una observación crucial pueda servir para los fines explicativos que Eckstein le concede: (1) hay muy pocas explicaciones que sólo dependan de una variable causal; para evaluar el impacto de más de una variable explicativa el investigador necesita haber observado más de una consecuencia; (2) medir la observación crucial es difícil y no del todo fiable; y (3) no es razonable abordar la realidad social partiendo de la base de que es una consecuencia de procesos deterministas, de manera que, aunque las medidas fueran perfectas, habría errores aleatorios.]

#### 1. Explicaciones alternativas

Supongamos que comenzamos un estudio de caso con la hipótesis de que un determinado factor explicativo aclara el resultado observado. Sin embargo, durante la investigación descubrimos una posible explicación alternativa. En esta situación, hay que hacer una estimación de dos efectos causales —el de la primera hipótesis y el de la otra explicación— pero sólo disponemos de una observación y, por tanto, está claro que el diseño de la investigación será impreciso (capítulo 4, apartado 1). Además, aunque hiciéramos una equiparación (lo cual suele ser una estrategia valiosa), no podríamos comprobar las explicaciones causales con una única observación. Supongamos que pudieramos hacer una equiparación perfecta utilizando todas las variables relevantes (circunstancia que es muy improbable en las ciencias sociales). Para poder observar cualquier variación que se produjera en la variable explicativa seguiríamos necesitando comparar, como mínimo, dos unidades. En consecuencia, sería imposible extraer una inferencia causal válida si las hipótesis alternativas sólo se comprobaran con una única comparación.

## 2. Error de medida

Aunque dispusiéramos de una teoría que hiciera predicciones consistentes y precisas, aún tendríamos que enfrentarnos al hecho de que nuestra medida de esa predicción, como todas las medidas, pudiera contener un error (véase el apartado 1 del capítulo 5). Si sólo tenemos una observación, un error de medida bien podría llevarnos a rechazar una hipótesis verdadera, o a lo contrario. Las teorías precisas pueden necesitar medidas más exactas que las que permite el estado actual de nuestras inferencias descriptivas. Si tenemos muchas observaciones podemos reducir la magnitud e influencia del error de medida mediante la agregación, pero en una única observación siempre existirá la posibilidad de que este tipo de error sea determinante y de que nos lleve a una conclusión falsa.

## 3. Determinismo

La última razón, y quizás la más decisiva, para considerar inadecuados los estudios que sólo se basan en una consecuencia observable tiene que ver con el grado de determinismo que se da en el mundo. Si éste fuera determinista y la observación produjera una medida que no se correspondiera con la teoría, podríamos afirmar con certeza que dicha teoría es falsa. Sin embargo, en toda teoría social de interés siempre existe la posibilidad de que se omitan variables desconocidas, que podrían conducir a un resultado impredecible aunque el modelo fundamental de dicha teoría fuera correcto. Si sólo existe una consecuencia observada de la teoría causal, careceremos de fundamentos para determinar si la observación confirma o desmiente una teoría o si proviene de algún factor desconocido. Aunque tuviéramos dos observaciones en un experimento perfecto, variáramos sólo un factor explicativo y produjéramos únicamente una observación para diferenciar dos observaciones que, de otro modo, serían idénticas respecto a la variable dependiente, tendríamos que considerar la posibilidad de que, en nuestro mundo probabilístico, existiera algún factor no sistemático, fruto del azar, que cambiara el efecto causal que se está observando. No importa que el mundo sea inherentemente probabilístico (en el sentido del apartado 6 del capítulo 2) o que, simplemente, no podamos controlar todas las posibles variables omitidas. En cualquier caso, nuestras predicciones sobre las relaciones sociales sólo podrán ser precisas desde un punto de vista probabilístico. En realidad, Eckstein admite que los factores fortuitos influyen en todas las investigaciones:

«La posibilidad de que un resultado se deba al azar no puede descartarse en ningún tipo de estudio; incluso en las grandes investigaciones comparadas es más o menos probable [...] Por lo tanto, la auténtica diferencia entre los estudios de la observación crucial y los comparados es que en estos últimos, pero no en los primeros, se pueden respetar ciertas convenciones para adjudicar un determinado valor a la posibilidad de que haya resultados fortuitos (por ejemplo, "significativo en el nivel de 0,05")».

Eckstein tiene razón cuando afirma que sólo es habitual admitir que existe una determinada posibilidad de que se produzcan resultados fortuitos en los estudios que uti-

## 6. Aumentar el número de observaciones

lizan un  $n$  grande. Sin embargo, admitirlo tiene la misma importancia en todos los estudios, tanto en los que barajan muchas observaciones como en los que utilizan pocas<sup>2</sup>.

En general, podemos concluir que utilizar una sola observación no es útil para la comprobación de hipótesis o teorías. Sin embargo, es preciso hacer una matización. Aunque tengamos un estudio «puro» de una sola observación, en el que sólo haya realmente una para todas las variables relevantes, ésta podrá ser útil para evaluar explicaciones causales si forma parte de un programa de investigación. Si hay otras observaciones únicas, quizás recogidas por otros investigadores, con las que se pueda comparar, ya no estaremos utilizando una sola observación: esto es precisamente lo que queremos señalar. [No hay que confundir la lógica de la explicación con el proceso de realización del estudio.] Si dos investigadores llevan a cabo estudios de una sola observación, puede que —si presuponemos que recogen los datos de manera sistemática y equiparable y que comparten de alguna forma sus resultados— se obtenga una comparación entre dos y una inferencia causal válida. También está claro que los estudios de una sola observación pueden hacer importantes aportaciones al resumen de los pormenores históricos o a la inferencia descriptiva, incluso si no hay comparación (véase el apartado 2 del capítulo 2). Evidentemente, un estudio de caso que, como la mayoría de ellos, contenga muchas consecuencias observables no planteará los problemas que aquí se han señalado.]

## 1.2 Razonamiento analógico

Los peligros de los diseños que no utilizan más que una observación quedan patentes cuando se hace referencia a la forma de equiparación habitual que usan los que elaboran políticas y algunos analistas de esa área cuando intentan comprender acontecimientos políticos, es decir, estamos hablando del razonamiento analógico (véase Khong, 1992). El uso adecuado de la analogía es fundamentalmente el mismo que el de la equiparación que mantiene otras variables constantes. Nuestra hipótesis causal es que si dos unidades son las mismas en todos los aspectos relevantes (esto quiere decir que hemos logrado equipararlas o que, dicho de otro modo, hemos encontrado una buena analogía), unos valores parecidos en las variables explicativas relevantes harán que también se asemejen los de la dependiente. Si nuestra equiparación fuera perfecta y no hubiera error aleatorio en el mundo, sabríamos que la situación de crisis a la que se enfrenta en la actualidad el país B, que es equiparable a la del país A al año pasado, tendrá las mismas consecuencias que las que se observaron en este último. Si lo expresamos de este modo se verá que el razonamiento analógico puede ser apropiado.

Sin embargo, este tipo de argumentación nunca es mejor que el análisis comparado que conlleva. Al igual que suele ocurrir en los estudios comparados, siempre será mejor (o, en último extremo, no peor) partir de más observaciones para nuestra generalización. Por ejemplo, lo que ocurrió en el país A podría haber sido consecuencia de factores estocásticos que habrían quedado anulados si nuestras predicciones se hubieran basado en las crisis ocurridas en otros cinco países equiparables. Como ocurre en todos los estudios que utilizan equiparaciones, la analogía sólo es tan buena como la propia equiparación. Si ésta es incompleta —si se han omitido variables relevantes—, puede que al calcular los efectos causales cometamos errores. De este modo, al igual que en toda investigación social o pronóstico, es importante aclarar lo más posible el grado de incertidumbre que acompaña nuestra predicción. En general, siempre es preciso ir más allá de una sola observación análoga, al margen de lo similar que nos parezca. Es decir, el método comparado —en el que se combinan datos de muchas observaciones, aunque algunos de ellos no sean analogías cercanas a la situación que nos ocupa— es, como mínimo, siempre tan bueno como la analogía, y lo normal es que sea mejor. La razón es sencilla: la analogía utiliza sólo una observación para predecir otra, mientras que el método comparado emplea una combinación bien sopesada de un número mayor de observaciones. Éstas nos ayudarán a realizar un pronóstico más preciso y eficiente, siempre que contengan ciertos rasgos que se parezcan de algún modo —aunque sea poco— al acontecimiento que estamos pronosticando y que utilicemos la información adicional de manera razonable. Por lo tanto, si nos tienta el uso de analogías, tendremos que abordarlas desde un punto de vista comparativo y más general, tal como se analiza más adelante en el apartado 3<sup>3</sup>.

## 2. ¿Cuántas observaciones son suficientes?

Al llegar a este punto el investigador cualitativo podría plantearse una cuestión cuantitativa: ¿cuántas observaciones son suficientes? La pregunta tiene importantes consecuencias para la evaluación de los estudios existentes y para el diseño de nuevas investigaciones. La respuesta depende en gran medida del diseño del estudio, de qué inferencia causal se intente calcular y de algunos rasgos del mundo que el investigador no puede controlar.

Aquí respondemos a esta pregunta mediante otro sencillo modelo formal, tomado de la investigación cualitativa. Al utilizar el mismo modelo de regresión lineal al que nos remitimos constantemente en los capítulos 4 y 5, centraremos nuestra atención en el efecto causal de una variable ( $x_1$ ). Todas las demás se consideran controles que son importantes para evitar el sesgo de la variable omitida u otros problemas. Se puede expresar fácilmente el número de unidades que se precisan en una determinada situación mediante una fórmula sencilla:

## 6. Aumentar el número de observaciones

$$n = \frac{\sigma^2}{(1 - R_t^2)S_{xt}^2 V(b_t)} \quad (6.1)$$

cuyo contenido pasamos a explicar.

El símbolo  $n$ , por supuesto, representa el número de observaciones sobre las que hay que recoger datos, y en este modelo formal se calcula a partir de  $\sigma^2$ ,  $V(b_t)$ ,  $R_t^2$  y  $S_{xt}^2$ . El significado de cada una de estas cantidades es importante, y todas influyen en el número de observaciones que el investigador cualitativo debe recoger para llegar a una inferencia válida. Hacemos la ecuación (6.1) sin más premisas que las que ya se habían introducido<sup>4</sup>. Ahora las describimos según un orden fijado por la posibilidad de que el investigador pueda influir en ellas: (1) variabilidad fundamental  $\sigma^2$ , (2) incertidumbre de la inferencia causal  $V(b_t)$ , (3) colinealidad relativa entre la variable causal y las variables de control  $R_t^2$  y (4) varianza de los valores de la variable causal clave  $S_{xt}^2$ <sup>5</sup>.

### 1. Variabilidad fundamental $\sigma^2$

Cuanto mayor sea la variabilidad fundamental, o variabilidad no explicada de la variable dependiente (tal como se describió en el apartado 6 del capítulo 2), más observaciones habrá que recoger para hacer una inferencia causaliable. Esto debería ser relativamente intuitivo, ya que si hay más ruido en el sistema será más difícil encontrar una señal clara mediante un número fijo de observaciones. Recoger datos sobre más unidades puede incrementar nuestro control lo suficiente como para permitirnos encontrar pautas causales sistemáticas.

Mediante una analogía directa podemos decir que un estimador más eficiente también necesitará que se recojan más datos. Un ejemplo de esta situación aparece cuando en la variable dependiente hay un error de medida aleatorio (apartado 1.2.1 del capítulo 5). Desde el punto de vista del analista, este tipo de error de medida suele equivaler a una mayor variabilidad adicional, ya que no siempre se puede distinguir entre ambos fenómenos. De este modo, si hay una variabilidad fundamental mayor (o, su equivalente, estimaciones menos eficientes), habrá que recabar más datos.

Aunque el investigador no puede influir en la variabilidad fundamental del mundo, esta información es bastante importante en dos sentidos. En primer lugar, cuanto más sepamos sobre un objeto de estudio, más pequeña será (es de suponer que hasta cierto límite positivo) esta variabilidad fundamental (o no explicada), de manera que no será necesario recoger tantas observaciones para saber algo nuevo. Por ejemplo, si supiéramos mucho acerca de lo que decidió el resultado de ciertas batallas en la guerra de independencia de los Estados Unidos, necesitaríamos un número relativamente pequeño de observaciones (batallas) para hacer una estimación del efecto causal de alguna nueva variable explicativa que hubiera generado una hipótesis.

En segundo lugar, aunque comprender el grado de variabilidad fundamental no nos ayude a reducir el número de observaciones sobre las que tenemos que recoger datos, si será bastante útil para calibrar con precisión la incertidumbre de las inferencias que se

## El diseño de la investigación social

hagan. Esto tendría que quedar claro con la ecuación (6.1), ya que se puede solucionar la incertidumbre del efecto causal  $V(b_1)$  considerándola una función de las otras cuatro cantidades (si conocemos  $n$  y el resto de las cantidades, excepto en lo relativo a la incertidumbre de la estimación causal). Esto significa que mediante este modelo formal podemos calcular tanto el grado de incertidumbre de una inferencia causal, utilizando información sobre varias observaciones, como la variabilidad fundamental, la varianza de la variable explicativa causal y la relación que existe entre ésta y las de control.

### 2. Incertidumbre de la inferencia causal $V(b_1)$

*mujeres  
afroamericanas*

La presencia de  $V(b_1)$  en el denominador de la ecuación (6.1) demuestra que es obvio que cuanta más incertidumbre estemos dispuestos a tolerar, menos observaciones habrá que recoger. En aquellas áreas en las que cualquier nueva información sea crucial, quizás podamos hacer aportaciones importantes con un número relativamente pequeño de observaciones. En otros casos en los que ya se haya profundizado más en el tema y un estudio sólo pueda ser relevante con un grado de certidumbre considerable, necesitaremos un número relativamente mayor de observaciones para convencer de la existencia de un nuevo efecto causal (vease el apartado 2.1 del capítulo 1).

### 3. Colinealidad entre la variable causal y las de control $R_i^2$

Si la variable causal no se correlaciona con otras variables que estemos controlando, incluir éstas —lo cual puede ser necesario para evitar el sesgo de la variable omitida u otros problemas— no influirá en el número de observaciones que hay que recoger. Sin embargo, cuanto más alta sea la correlación entre la variable causal y cualquier otra que estemos controlando, más le estará pidiendo a los datos el diseño de investigación y, por lo tanto, mayor será el número de observaciones que haya que recabar para alcanzar el mismo nivel de certidumbre.

Por ejemplo, supongamos que estamos realizando un estudio para comprobar si las mujeres reciben igual salario por el mismo trabajo en algún sector. No tenemos acceso a datos oficiales y, por tanto, sólo podemos hacer entrevistas de tipo informal. Nuestra variable dependiente es el sueldo anual de un empleado, y la explicativa clave, el género. Una de las variables de control importantes es la raza. En un caso extremo, si todos los hombres del estudio fueran negros y todas las mujeres blancas, no tendríamos control sobre la inferencia causal: sería imposible hallar las consecuencias del género después de haber controlado el factor raza. En consecuencia, el género se hace constante en esta muestra; de ahí que éste sea un ejemplo de multicolinealidad, un diseño de investigación impreciso (apartado 1 del capítulo 4); sin embargo, hay que fijarse en lo que ocurre cuando la colinealidad es alta pero no perfecta. Supongamos, por ejemplo, que recogemos información sobre quince empleados y qué, de éstos, todos los hombres, excepto uno, son negros y todas las mujeres blancas. En esta situación, el efecto que tiene el género, al controlar el factor raza, se basa completamente en la única observación que no es perfectamente colineal.

### 6. Aumentar el número de observaciones

Por lo tanto, en la situación general, al igual que en este ejemplo, cuanta más colinealidad haya entre la variable causal explicativa y las de control, más observaciones desperdiciaremos. En consecuencia, serán necesarias más para alcanzar un determinado nivel de incertidumbre. Este asunto tiene una gran importancia práctica para el diseño de investigaciones, ya que suele ser posible seleccionar observaciones para mantener baja la correlación entre la variable causal y las de control. En el ejemplo actual, para reducir esta correlación sólo necesitaríamos entrevistar a un número suficiente de mujeres negras y de hombres blancos.

### 4. Varianza de los valores de la variable explicativa causal $S_{x1}^2$

Para terminar, cuanto mayor sea la varianza de los valores de la variable explicativa causal, menos observaciones necesitaremos recabar para alcanzar un determinado nivel de certidumbre en relación a la inferencia causal.

Este resultado, al igual que el último que hemos mencionado, tiene consecuencias prácticas, ya que, si seleccionamos apropiadamente las observaciones, podremos reducir la necesidad de que tengan que ser muchas. Sólo necesitamos centrarnos en elegir aquellas que presenten una amplia gama de valores en relación a la variable causal clave. Si nos interesa saber el efecto que tiene el nivel de estudios medio de una comunidad sobre la delincuencia, lo mejor será elegir algunas comunidades que registren niveles muy altos de este indicador y otras que los tengan muy bajos. Si se hace así, será menor el trabajo necesario para extraer una inferencia causal que tenga un determinado nivel de certidumbre porque habrá que recoger menos observaciones.

El presente modelo formal presupone que el efecto que estamos estudiando es lineal. Es decir, cuanto mayores sean los valores de las variables explicativas, más alto (o más bajo) será el valor esperado de la variable dependiente. Lo mismo ocurrirá si la relación no es lineal pero si bastante uniforme (o sea, no decreciente). Si, por el contrario, el efecto es claramente no lineal, puede que los valores medios de la variable explicativa presenten un resultado completamente diferente. Por ejemplo, supongamos que un estudio que sólo se basara en los valores extremos de la variable explicativa llegara a la conclusión de que no existe ningún efecto: el nivel educativo de la comunidad no influye en la delincuencia. Sin embargo, de hecho, podría ser que sólo los niveles educativos medios redujeran el grado de delincuencia de una comunidad. En la mayoría de los problemas esta matización no es cierta, pero, al diseñar la investigación, hay que procurar especificar exactamente nuestras premisas.

Si prestamos atención a la variabilidad fundamental, a la incertidumbre, a la colinealidad y a la varianza de los valores de la variable causal, podremos lograr más control con pocas unidades. Sin embargo, no deja de ser razonable hacerse la pregunta que da título a este apartado: ¿cuántas observaciones son suficientes? No podemos responder a esta cuestión con una

respuesta precisa que siempre sea aplicable. Como hemos demostrado con el último modelo formal, la respuesta depende de cuatro informaciones diferentes, que variarán en cada diseño de investigación. Además, la mayoría de las situaciones que son objeto de investigación cualitativa no se ajustarán exactamente a este modelo formal, aunque, en un sentido más general, las intuiciones fundamentales sí sean aplicables.

**Cuántas más observaciones, mejor, pero ¿cuántas son necesarias?** En la situación más sencilla —en la que hay niveles de variabilidad escasos, varianza alta en la variable causal, ninguna correlación entre ésta y las de control y se requieren niveles bastante bajos de certidumbre— se necesitarán pocas observaciones: probablemente más de cinco pero menos de veinte. De nuevo, la respuesta precisa depende de la exacta especificación del modelo formal y del valor exacto de cada uno de sus componentes. Lamentablemente, la investigación cualitativa, por definición, no suele ser precisa y, en consecuencia, no siempre podemos reducir este aspecto a una sola respuesta.

Por fortuna, sí suele ser posible evitar estos problemas aumentando el número de observaciones. A veces esto supone recoger más datos, pero, como señalamos en el apartado siguiente, suele ser posible reconceptualizar un diseño de investigación cualitativo para sacarle muchas más observaciones y darle así una estructura más consistente. A continuación nos ocupamos de este asunto.

### 3. Aumentar el número de observaciones a partir de unas pocas

Ya hemos hecho hincapié en las dificultades que conlleva investigar con un pequeño número de observaciones y hemos señalado ciertas medidas que se pueden tomar para mejorar el diseño en esos casos. Sin embargo, quizás el lector se haya dado cuenta de que consideramos todos estos consejos «soluciones de recambio», que son útiles cuando la cantidad de observaciones es limitada pero que no lo son tanto como estrategia para aumentarla<sup>6</sup>. Tal como señalamos, estas soluciones de recambio son valiosas porque a veces no podemos recoger más observaciones del tipo que queremos analizar: quizás sólo haya unos pocos ejemplos del fenómeno que nos interesa o puede que sea demasiado caro o arduo investigar más observaciones que las pocas que ya tenemos. En este apartado analizamos diversas formas de aumentar el número de nuestras observaciones que resultan útiles cuando nos encontramos con lo que parece un número escaso de ellas y no disponemos del tiempo o de los recursos para recoger más. Señalamos varias formas de aumentar el número de observaciones relevantes para nuestra teoría mediante la redefinición de las mismas. Estas estrategias de investigación aumentan *h* a la vez que mantienen la atención en la búsqueda de pruebas

que constaten o refuten la teoría. Tal como hemos recalcado anteriormente, suelen ser útiles incluso después de que se ha terminado la recogida de datos.

Como señalamos en el apartado cuatro del capítulo 2, para Harry Eckstein (1975) un caso es «un fenómeno del que sólo señalamos e interpretamos una única medida en cada variable pertinente». Como la palabra «caso» se ha utilizado de muchas formas diferentes en las ciencias sociales, preferimos centrarnos en las observaciones, que hemos definido como cada una de las medidas de una variable dependiente en una unidad (y de tantas variables explicativas como se disponga para esa misma unidad). Las observaciones son el componente fundamental de la investigación social empírica: las reunimos con el fin de presentar pruebas que consideramos fiables para evaluar nuestras teorías. En realidad, como señalamos en el capítulo 2, en los proyectos de investigación no se estudian fenómenos globales como Francia, la Revolución francesa, las elecciones estadounidenses de 1992 o la decisión iraquí de invadir Kuwait, sino que lo que se hace es seleccionar aspectos de esos fenómenos—conjuntos de variables explicativas y dependientes—y concretarlos en nuestras teorías. Identificamos unidades en las que funcionan tales variables y hacemos observaciones de éstas en las primeras<sup>7</sup>.

En consecuencia, lo que utilizamos para evaluar nuestras teorías es un conjunto de observaciones referidas a ciertas unidades que, a su vez, tienen que ver con unas variables relevantes. Aquí lo que nos interesa es aumentar el número de observaciones, y todas las formas de hacerlo parten de la teoría o hipótesis que estamos comprobando. Tenemos que preguntarnos cuáles son las posibles consecuencias observables de nuestra teoría o hipótesis y cuántos ejemplos podemos encontrar para contrastarlas. Si necesitamos más observaciones para realizar esta comprobación, hay tres formas de conseguirlas: podemos observar más unidades, medirlas otra vez de diferente manera o hacer ambas cosas (es decir, observar más unidades y utilizar, a la vez, otras medidas). Dicho de otro modo, podemos hacer medidas similares en otras unidades (lo cual describimos en el apartado 3.1), utilizar las mismas unidades cambiando las medidas (apartado 3.2) o cambiar tanto las medidas como las unidades (apartado 3.3). El primer método puede considerarse una reproducción completa de nuestra hipótesis: utilizamos las mismas variables explicativas e independientes con nuevos ejemplos. El segundo conlleva una reproducción parcial de nuestra teoría o hipótesis, en la que se utiliza una nueva variable dependiente con las mismas variables explicativas. Para terminar, en el tercero se plantea una hipótesis nueva (o muy revisada)—que estaba implícita en nuestra teoría original y que utiliza otra variable dependiente—para aplicarla a nuevos ejemplos<sup>8</sup>. Si se utilizan estos métodos se pueden observar muchas consecuencias diferentes de nuestra teoría, incluso en aquellas investigaciones que se denominan convencionalmente «estudios de caso» único. De hecho, en un único caso sue-

lamente  
en el  
Anexo  
Glossary

## El diseño de la investigación social

le haber múltiples medidas de las variables principales; de ahí que, por definición, contenga múltiples observaciones".

### 3.1 Las mismas medidas con otras unidades

Obtener más observaciones a partir del mismo método de medida es la forma habitual de incrementar el número de observaciones. Se aplica la misma teoría o hipótesis, más o menos con las mismas variables, a más ejemplos del proceso que describe dicha teoría. Las dos formas principales de encontrar más ejemplos observables del proceso que aborda nuestra teoría se basan en transformaciones «en el espacio» o en el tiempo.

El método fundamental para obtener más observaciones «en el espacio» es buscar otras unidades similares; añadir Pakistán, Bangladesh y Sri Lanka a la India en la propia base de datos. Si se tiene suficiente tiempo, dinero y conocimientos, esta forma de proceder es válida. El trabajo llevado a cabo por Kohli en la India (que se analizó en el apartado 4.3 del capítulo 4) es un ejemplo de esta situación y también pone de manifiesto una de las maneras que él utilizó para superar el problema que planteaba estudiar tres estados indios, que se habían seleccionado en función de valores conocidos de la variable dependiente y de las explicativas. El autor se ocupó de otras dos unidades nacionales. Una de ellas era el Chile de Allende, en el que los programas de ayuda a los pobres fracasaron. Kohli señaló que a este fracaso había contribuido la ausencia de una de las tres características que, para él, llevan al éxito de tales programas (en el caso chileno, la falta de un partido político reformista bien organizado)<sup>10</sup>. El otro país que consideró era el Zimbabue de Robert Mugabe, que había llegado al poder, cuando Kohli estaba escribiendo su libro, con un programa de reducción de la pobreza parecido al de Bengala occidental. Los resultados, aunque provisionales, parecían constatar la teoría de Kohli. Estos dos casos los trató por encima, pero no se equivocó al considerarlos consecuencias observables de su teoría.

Sin embargo, no es necesario que rebasemos los límites de la unidad que hemos estado estudiando. Una teoría que, en principio, se centre en un estado-nación puede contrastarse en las unidades geográficas que lo componen, ya sean estados, condados, ciudades, regiones, etc. Evidentemente, esto aumenta el margen de variabilidad tanto de las variables explicativas como de la dependiente. Supongamos que quisieramos comprobar una teoría de la inestabilidad social que relacionara este problema con los precios agrícolas. Una unidad podría ser la nación denominada «India», pero, si tenemos en cuenta las diferentes partes del país, este caso aportará numerosas observaciones sobre la relación que nos interesa. Sin salir de él podemos aumentar el número de observaciones, buscando reproducciones del proceso que se está estudiando.

### 6. Aumentar el número de observaciones

Los que estudian las políticas sociales suelen ocuparse de unidades administrativas que, en realidad, son subunidades del estado-nación en el que les quieren comprobar sus hipótesis acerca del origen de diversos tipos de políticas. El análisis que llevó a cabo Kohli de tres estados de la India es un ejemplo de una tendencia habitual en este tipo de estudios, que consiste en comparar estados, ciudades o regiones. Sin embargo, para Kohli, el conjunto de observaciones original eran los tres estados indios. Como hemos señalado, el autor los seleccionó de forma que no se podían utilizar para comprobar su hipótesis sobre el efecto de la estructura del régimen en las políticas contra la pobreza en la India. Sin embargo, igual que utilizó otras naciones como unidades de observación, también solucionó gran parte del problema que había en su primera selección de unidades mediante el uso de subunidades. Partiendo de los tres estados indios con los que había comenzado, descendió a otro nivel de observación para aplicar su hipótesis a los panchayats (consejos de gobierno locales de los distritos, bloques y pueblos), que son subunidades de los estados. Los panchayats diferían considerablemente en cuanto al compromiso de sus dirigentes con las políticas contra la pobreza y a la estructura organizativa local. De este modo, se pudo comprobar el impacto que tenía esa diversidad en los resultados de la política que el autor utilizaba como variables dependientes.

Las subunidades que aportan más observaciones no tienen por qué ser de tipo geográfico. Las teorías que son aplicables al estado-nación también pueden contrastarse en organizaciones gubernamentales o en el marco de determinadas decisiones, lo cual puede hacerse sin tener que ir a otro país. En un trabajo de Verba y otros autores (1993) se puede apreciar cómo se buscan más consecuencias observables de la propia hipótesis en otras unidades no geográficas. En el ejemplo que presentamos en el apartado 4 del capítulo 5, estos investigadores explicaban que el hecho de que los afroamericanos aprendan más habilidades sociales que los latinos se explica por el tipo de iglesia a la que acuden unos y otros; es probable que el primer grupo vaya a iglesias protestantes que tienen una organización de tipo comunitario, mientras que el segundo suele acudir a congregaciones católicas con un funcionamiento jerárquico. Los autores señalan que, si su hipótesis sobre la influencia que tiene la organización de la iglesia es correcta, el mismo tipo de diferencia que aparece entre fieles católicos y protestantes debería apreciarse al comparar otras unidades religiosas, especialmente las sectas protestantes que difieren en cuanto a su forma de organizarse. Les parece que los episcopalianos, cuya iglesia tiene una estructura jerárquica, son bastante parecidos a los católicos en cuanto a la adquisición de habilidades sociales a través de su congregación. El hecho de que los episcopalianos suelen tener más estudios y recursos que, por ejemplo, los baptistas —aunque en su iglesia demuestren menos sus habilidades sociales— concede un mayor grado de control a la confirmación de esta hipótesis causal.

Hay que tener cuidado al decidir si las nuevas unidades son apropiadas para reproducir nuestra hipótesis, es decir, si en ellas puede tener lugar el proceso que implica dicha hipótesis. La validez que tenga la aplicación de la hipótesis a otras clases de unidades dependerá de la teoría e hipótesis que se estén utilizando, así como del tipo de unidades. Si la variable dependiente es la política de protección social, los estados o provincias serán adecuados si pueden implantar ese tipo de políticas, pero si lo que estamos estudiando es la política arancelaria y todas las decisiones al respecto las toma el gobierno central, quizás las unidades estatales o provinciales no sean las indicadas. Del mismo modo, no tendría sentido estudiar los gobiernos locales de la India o de Pakistán para comprobar una teoría relativa a las condiciones en las que una unidad política decide desarrollar armamento nuclear, ya que este tipo de elección la realiza el gobierno central. Para dar otro ejemplo, podemos señalar que se puede comprobar el impacto que tiene el cambio de los precios agrícolas sobre la inestabilidad social en los estados indios, pero que no es factible utilizar varias organizaciones del gobierno indio para verificar esta relación, ya que este proceso no tiene lugar dentro de ellas. Dicho en pocas palabras, el hecho de que las subunidades sean lugares apropiados para observar una teoría «en acción» depende de esa misma teoría. Esta es la razón por la que aconsejamos que se comience por elaborar una lista de las consecuencias observables de nuestra teoría y no buscando montones de posibles unidades, independientemente de la teoría que se utilice. Solo podremos elegir las unidades de estudio una vez que hayamos especificado cuál es la teoría.

Un enfoque alternativo es considerar observaciones a lo largo del tiempo. La India actual y la de hace una década pueden ser dos ejemplos del proceso que nos interesa. De hecho, en la mayoría de las investigaciones que se consideran «estudios de caso» hay múltiples medidas de una hipótesis a lo largo del tiempo.

Creemos que aconsejar que se aumente el número de observaciones buscando más ejemplos en las subunidades o teniendo en cuenta más situaciones a lo largo del tiempo es una de las indicaciones más útiles que podemos dar a los investigadores cualitativos. Soluciona el problema que plantea un «pequeño porque lo aumenta sin que sea necesario, por ejemplo, viajar a otro país o analizar una decisión completamente nueva. Sin embargo, es un consejo que hay que seguir con cuidado: la teoría o hipótesis tiene que ser aplicable a la nueva situación, es decir, la subunidad debe contener realmente una consecuencia observable de dicha teoría. No tiene por qué ser (ni siquiera aproximadamente) la consecuencia observable que nos interese en ese preciso momento; siempre que sea una consecuencia de la misma teoría, los datos que organicemos nos darán un mayor control sobre la inferencia causal.]

Hay que ser consciente de otro problema. Queremos utilizar estos nuevos ejemplos para someter nuestra teoría a otras pruebas, pero las subuni-

#### 6. Aumentar el número de observaciones

dades o los diversos ejemplos que hemos encontrado a lo largo del tiempo quizás no sean comprobaciones *independientes* de dicha teoría. De este modo, como reconoce George (1982, pp. 20-23), cada nuevo «caso» no nos aportará tanta información sobre el problema como lo haría si las observaciones fueran independientes. El hecho de que las observaciones dependan unas de otras no descalifica las comprobaciones, a menos que esa dependencia sea perfecta, es decir, a menos que se puedan predecir totalmente los nuevos datos a partir de los existentes. Si no se da este caso improbable, al menos si habrá nueva información en los nuevos datos que nos ayude a analizarlos. Estas nuevas observaciones, que se basan en información no independiente, no aportarán tanta información como las que son completamente independientes, pero, de todos modos, pueden ser útiles.

Esta conclusión tiene dos consecuencias prácticas. En primer lugar, al barajar observaciones parcialmente dependientes tenemos que tener cuidado de no exagerar la certeza de las conclusiones. Especialmente, no hay que considerar que esos datos nos ofrecen tantas observaciones como tendríamos si las observaciones fueran independientes. En segundo lugar, es preciso analizar con precaución por qué las observaciones dependen unas de otras. Con frecuencia, la dependencia procederá de una o de algunas variables omitidas interesantes que pueden inducir a confusión. Por ejemplo, supongamos que nos interesa la participación política de los ciudadanos en los condados de los Estados Unidos. Quizás los condados contiguos no sean independientes porque la gente va a trabajar cada día de unos a otros, por la movilidad residencial o por los valores socioeconómicos similares que tienen las personas que viven en ellos. Recabar datos en condados vecinos si aportaría más información al estudio, aunque no tanto como se podría esperar de condados que fueran completamente independientes de aquellos en los que ya habíamos recogido datos.]

Para dar otro ejemplo, pensemos en la relación que existe entre el cambio de los precios agrícolas y la inestabilidad social. Podríamos comprobar si se da esta relación en varios estados indios, calculando estos precios y la inestabilidad social. Sin embargo, los estados no son unidades experimentales aisladas. Los valores de la variable dependiente pueden verse afectados no sólo por los que adopten las variables explicativas dentro de cada unidad, sino por los de las omitidas que están fuera. En un estado el detonador de la inestabilidad social pueden ser los precios agrícolas (tal como predice la teoría), pero esa alteración del orden social puede influir directamente en que se genere una situación parecida en un estado contiguo (lo cual hará que ésta sólo sea una demostración parcialmente independiente de nuestra teoría). Sería posible enfrentarse al problema controlando adecuadamente esta propagación. Algo similar puede ocurrir con la influencia que ejerce cierto período sobre otro posterior. Podríamos reproducir nuestro análisis en la India una década después, pero la inestabilidad social del primer período quizás tuviera un efecto directo en la del segundo.)

## El diseño de la investigación social

Estos ejemplos ponen de manifiesto que la reproducción de un análisis sobre nuevas unidades no siempre produce un nuevo estudio importante. Se pueden utilizar las observaciones que estén dentro del estudio actual y que tengan la misma forma que las ya utilizadas para comprobar la hipótesis. De este modo, el investigador del «estudio de casos» podrá darse cuenta de que existen muchas más observaciones de las que pensaba<sup>11</sup>.

### 3.2 Nuevas medidas con las mismas unidades

*Se pueden obtener más ejemplos para comprobar una teoría o hipótesis con la misma unidad de observación pero cambiando la variable dependiente. Para utilizar este método hay que buscar muchos efectos de la misma causa, lo cual constituye una técnica muy convincente para comprobar una hipótesis. Una vez más, partimos de una teoría o hipótesis, que suponemos es correcta, y nos preguntamos: aparte de la variable dependiente actual, ¿en qué otras cosas cabe esperar que influyan nuestras variables explicativas? Este ejercicio nos puede sugerir indicadores alternativos de la variable dependiente. En el capítulo 1 señalamos que una determinada teoría acerca de la extinción de los dinosaurios estaba relacionada con la composición química de las rocas. De ahí que, incluso una teoría causal que explique un acontecimiento prehistórico singular, tenga múltiples consecuencias observables que puedan evaluarse.*

*En el ejemplo de las fluctuaciones de los precios agrícolas y la inestabilidad social puede que hayamos medido el segundo factor utilizando la cantidad de desórdenes públicos. Además de la inestabilidad social, podríamos preguntarnos qué más cabría esperar si la teoría es correcta. Quizá haya otras medidas válidas de ese desorden social (comportamientos desviados de uno u otro tipo). Esta investigación podría llevarnos a la hipótesis de que hay otras variables que se verían afectadas, como son el comportamiento electoral, la inversión empresarial o la emigración. El mismo proceso que hace que la fluctuación de los precios agrícolas genere inestabilidad podría vincular el primer factor con los otros resultados mencionados.*

*La obra de Putnam (1993) sobre el impacto de los recursos sociales en el comportamiento de los gobiernos locales en Italia participa de un enfoque similar. El comportamiento de las regiones no es una medida única. En realidad, Putnam utiliza una amplia gama de variables dependientes para explicar de dónde surge la eficiencia en el comportamiento democrático de las regiones italianas. Dispone de doce indicadores de comportamiento institucional que, ocupándose de las políticas, pretenden medir su proceso de elaboración, las declaraciones que suscitan y su puesta en práctica. Además, utiliza encuestas para calibrar cómo evalúan los ciudadanos la actuación del gobierno. Cada una de estas medidas representa una consecuencia observable de su teoría.*

### 6. Aumentar el número de observaciones

Como hemos indicado anteriormente, la utilización de unidades administrativas subnacionales para estudiar una determinada política arancelaria no sería apropiada si este tipo de impuestos los fijara el gobierno central. Aunque las variables explicativas —por ejemplo, el tipo de sector industrial o de producto agrícola— varíen en diferentes estados o provincias, el proceso por el que se fijan los niveles arancelarios (que es lo que le interesa a la hipótesis que se está comprobando) no tiene lugar dentro de las unidades subnacionales. Sin embargo, si la variable dependiente pasa a ser el voto de los representantes de diversos estados o provincias en cuestiones comerciales o arancelarias, sí será posible estudiar el problema. De este modo, podemos aumentar el número de ejemplos en el que opera el proceso teórico.

### 3.3 Nuevas medidas con nuevas unidades

También podemos salirnos del conjunto de variables explicativas y dependientes que se han aplicado a un determinado conjunto de variables para buscar otras consecuencias observables que manejen nuevas variables y unidades. Puede que las medidas que utilicemos para comprobar hipótesis, que son fundamentalmente nuevas aunque se deriven de las primeras, sean bastante diferentes de las utilizadas hasta el momento. El proceso que describe la nueva teoría quizás no sea aplicable al tipo de unidad que se estudia, sino a otro (con frecuencia, a una unidad que está en un nivel de agregación inferior o superior). La hipótesis general sobre la relación que existe entre los precios agrícolas y la inestabilidad social puede sugerir otras relativas a la incertidumbre y a la inestabilidad en tipos de unidades diferentes, como son las empresas o los entes públicos. También puede llevarnos a hipótesis que se ocupen del comportamiento de los individuos. En este mismo ejemplo podríamos preguntarnos: «si nuestra teoría acerca del efecto que tienen las fluctuaciones de precios en la estabilidad social (que ya hemos comprobado en diversas unidades políticas) es correcta, ¿qué es lo que esto supone para el comportamiento de las empresas, de las cooperativas agrícolas o de los individuos (quizás en el mismo conjunto de unidades políticas)? ¿Qué consecuencias tendría —si tiene alguna— sobre cómo deciden las organizaciones gubernamentales la asignación de recursos? ¿Qué reacción psicológica cabría esperar que suscitara la incertidumbre en los individuos y qué influencia podría tener en sus posibles comportamientos desviados?».

Este enfoque resulta especialmente útil cuando no hay casos observables de un proceso social potencialmente significativo. En el estudio de la guerra nuclear hay un ejemplo de ello. Como nunca ha tenido lugar un enfrentamiento con este tipo de armamento entre dos potencias que dispongan de él, no podemos observar los efectos de las variables explicativas en el establecido de este tipo de guerra. Supongamos que, según nuestra teoría, la presencia de las armas nucleares en ambos bandos ha evitado una guerra

abierta. Aunque no haya ejemplos observables de nuestra hipótesis fundamental, plantear otra más específica podría conllevar otras posibles observaciones. Por ejemplo, podríamos pensar que nuestra teoría implica que el hecho de que ambos bandos posean armas nucleares tendría que evitar la existencia de amenazas serias de guerra abierta. Posteriormente, al estudiar la frecuencia y severidad de las amenazas entre parejas de países con capacidad nuclear y entre los que carecen de ella, y al analizar esas amenazas a medida que la probabilidad de guerra pareciera aumentar durante las crisis, podríamos encontrar, y comprobar, otras consecuencias observables de nuestra teoría.

Para llegar a la nueva teoría o hipótesis, que es diferente de la original pero que parte de ella, suele ser necesario desplazarse a un nivel de agregación más bajo y a un nuevo tipo de unidad: esto no supone trasladarse desde una unidad política como la nación a otro nivel de agregación inferior como es la provincia, sino desplazarse desde esas naciones o provincias hasta los individuos que viven en las unidades o hasta las decisiones que toman dentro de ellas. Cada teoría puede implicar diversas conexiones entre variables que conducen a un determinado resultado; es decir, diversas formas de producirse un fenómeno (Dessler, 1991, p. 345). Antes de diseñar comprobaciones empíricas, quizás tengamos que describir un «mecanismo causal» en el que haya series de hipótesis causales relacionadas que indiquen de qué manera se conectan las variables. Definir esos mecanismos causales diversos y buscarlos puede hacer que encontremos una gran variedad de nuevas consecuencias observables de una teoría (en el apartado 2.1 del capítulo 3 señalamos las diferencias que hay entre el concepto de mecanismo causal y nuestra definición fundamental de causalidad, que es más general).

Este desplazamiento hacia un nuevo tipo de «observación» —otra clase de unidad social, un individuo, una decisión— puede conllevar que se introduzcan variables explicativas no aplicables a la unidad original. Con frecuencia, plantear una hipótesis o teoría sobre unidades políticas supone hacer otra elaboración teórica acerca de cómo se produce el resultado concreto que se observa en el nivel de la unidad. En concreto, la hipótesis que se refiere a ésta puede entrañar otras hipótesis referidas a las actitudes y comportamientos que se registran en el nivel de los individuos que viven dentro de las unidades. Estas segundas hipótesis pueden comprobarse utilizando datos sobre esos individuos. Si nos desplazamos al nivel individual, podemos centrarnos en variables psicológicas o en aspectos de la experiencia o la posición social que carecen de sentido si se aplican a las unidades políticas.]

Pensemos en el ejemplo de la relación entre precios agrícolas e inestabilidad social. Podríamos plantear una hipótesis referida a una unidad administrativa, como es una nación o provincia. El ejemplo podría ser el siguiente: cuanto mayor es la fluctuación de los precios agrícolas en una

unidad, más probabilidades hay de que se registre inestabilidad social. A su vez, esta hipótesis sugiere otras acerca de los individuos que viven dentro de esas unidades. Por ejemplo, podríamos plantear que hay una mayor probabilidad de que participen de comportamientos sociales perturbadores los más vulnerables a la fluctuación de los precios (los que cultivan ciertos productos o aquellos que sólo pueden abastecerse adecuadamente si los alimentos son baratos). Para comprobar esta hipótesis quizás tuviéramos que medir factores psicológicos como la alienación o el comportamiento individual desviado.

Los estudios que dan explicaciones culturales a los fenómenos políticos dependen a menudo de este tipo de análisis que se hace en el nivel individual<sup>12</sup>. El estudio que realizó Weiner sobre las políticas de educación y de trabajo infantil en la India se basa en una explicación cultural: que la razón de que este país sea casi el único del mundo que no tenga una legislación eficiente para imponer la educación universal y que también carezca de leyes que prohíban el trabajo infantil tiene que ver con valores sociales que comparte el ciudadano medio con las élites gobernantes (Weiner, 1991). La India es un único país, y puede decirse que en el estudio de Weiner *n* es sólo 1. El autor evita este problema de diversas maneras. Para empezar, compara la India con otros países que han desarrollado un sistema de educación universal. También realiza comparaciones menores entre los estados indios; dicho de otro modo, cambia las unidades. Sin embargo, la hipótesis sobre la cultura y las políticas de la India implica, a su vez, otras hipótesis relativas a los valores de los individuos y a lo que piensan de las políticas. En este sentido, los más importantes son los miembros de las élites que participan en la elaboración de las políticas de educación y de trabajo infantil. De este modo, la principal comprobación a que se somete la hipótesis de Weiner tiene lugar en el plano individual. El autor entrevistó en profundidad a las élites para extraer información sobre lo que pensaban acerca de los valores relativos a la educación y el trabajo infantil. Estas opiniones, además de expresar lo que pensaban los entrevistados de las políticas, constituyan las consecuencias observables de la hipótesis global de Weiner.

Esta forma de hacerse con más consecuencias observables de una teoría a partir de unidades que están en un nivel de agregación inferior también puede aplicarse a los análisis de las decisiones. George y McKeown mencionan un enfoque denominado «rastreo de procesos», en el que el investigador se fija con atención en «el proceso decisivo por el que varias condiciones iniciales se transforman en resultados» (George y McKeown, 1985, p. 35)<sup>13</sup>. En lugar de considerar que el resultado final (por ejemplo, de una crisis internacional) es la variable independiente, se construyen otras nuevas. De manera que, por ejemplo, se convierte en una nueva variable cada decisión dentro de una secuencia o cada conjunto mensurable de percepciones que los encargados de decidir tengan de acciones o intenciones ajenas. Este enfoque llega con frecuencia al nivel del actor individual. Una teoría

### El diseño de la investigación social

que vincule las condiciones iniciales con los resultados a menudo implicará que en esos actores se dé un determinado conjunto de motivaciones o percepciones. Por consiguiente, el rastreo de procesos conllevará una búsqueda de pruebas —coherentes con la teoría causal general— relativas a los procesos decisivos que han producido los resultados. Para utilizar este procedimiento quizás se necesite entrevistar a los actores o leer lo que han escrito para explicar las razones de sus actos.

Por ejemplo, en las relaciones internacionales, la cooperación entre estados puede producirse por muchas razones: porque se espere obtener beneficios de la reciprocidad, porque funcione una disuasión con amenazas de destrucción o porque haya unos intereses comunes en cierto conjunto de resultados. En cada uno de estos mecanismos causales participarán muchas variables explicativas, pero en cada uno de los mecanismos posibles el conjunto de variables será diferente, y también lo serán las relaciones entre ellas. Un estudio atento del proceso que lleva a las naciones a cooperar quizás nos permita determinar qué mecanismo causal es más posible que esté funcionando. Esto puede conllevar el estudio, entre otros asuntos, de las motivaciones expresadas por los actores o del tipo de comunicación que hay entre ellos.

Desde nuestra perspectiva, el rastreo de procesos y otras formas de elaborar mecanismos causales aumentan el número de observaciones relevantes desde el punto de vista teórico<sup>14</sup>. Estas estrategias vinculan el trabajo teórico y el empírico porque utilizan consecuencias observables de una teoría para indicar nuevas observaciones que habría que hacer para evaluarla. Este método, al aportar más observaciones relevantes para las consecuencias de una teoría, puede ayudarnos a superar los dilemas que plantean los estudios de *n* pequeño y hacer posible que tanto los investigadores como los lectores tengan más confianza en los hallazgos de las ciencias sociales. En cada sucesión de acontecimientos, el rastreo de procesos produce muchas observaciones. Así ocurre también, dentro de cada unidad política, cuando se analizan actitudes o comportamientos individuales. Además, el investigador controla las variables que son aplicables a todas las observaciones porque se relacionan con la sucesión de acontecimientos o con el conjunto de la unidad. En general, un enfoque que sólo preste atención al resultado final hará que el investigador no disponga más que de un número muy escaso de observaciones para resolver el dilema que se presenta cuando se encuentra o bien sesgo de la variable omitida o imprecisión. Al revisar múltiples observaciones relativas a las actitudes o comportamientos individuales, el investigador puede calibrar qué mecanismos causales se han activado.

No es probable que este tipo de análisis produzca inferencias causales consistentes porque se puede poner en marcha más de un mecanismo y, dentro de cada uno de ellos, quizás no esté clara la fuerza relativa de las variables explicativas. Sin embargo, si que servirá, en cierta medida, para contrastar hipótesis, ya que una elaboración teórica que explique resultados

### 6. Aumentar el número de observaciones

también puede tener consecuencias para el proceso que lleva a que éstos se produzcan. Por lo tanto, buscar mecanismos causales aporta observaciones que podrían refutar la hipótesis. Quizás este enfoque también permita al investigador hacer alguna generalización descriptiva sobre la frecuencia con que se activa cada posible mecanismo causal y puede que tales generalizaciones descriptivas sirvan de base para futuros análisis de otros mecanismos causales relacionados y de las condiciones en las que cada uno de ellos se pone en marcha.

Para nosotros, el rastreo de procesos y la búsqueda de los fundamentos psicológicos que sustentan una hipótesis desarrollada en un nivel de agregación superior son enfoques muy valiosos. Sin embargo, constituyen extensiones de la lógica analítica fundamental que hemos venido utilizando y no formas de evitarla. Los estudios de este tipo, si quieren hacer alguna aportación a la inferencia causal, tendrán que enfrentarse a todos sus problemas: la homogeneidad de las unidades, la endogeneidad y el sesgo. Para alcanzar inferencias causales válidas es preciso que, en el nivel del individuo que toma las decisiones, nos planteemos las preguntas que surgen al diseñar una investigación y que las respondamos. Habrá que calibrar con precisión las razones que nos dé ese individuo, seleccionar las observaciones de manera que sean independientes del resultado que se ha producido (si no, tendremos problemas de endogeneidad) y evitar que haya variables omitidas relevantes. También es importante recalcar aquí que los mecanismos causales que se rastreen de este modo tendrán que hacer que nuestra teoría sea más, y no menos, restrictiva: este tipo de técnicas ha de aumentar, y no evadir, las posibilidades de refutación de una teoría. En suma, el rastreo de procesos y otros análisis de subunidades son útiles para encontrar hipótesis plausibles sobre mecanismos causales que, a su vez, pueden favorecer la elaboración de generalizaciones descriptivas y allanar el camino hacia la inferencia causal. Sin embargo, este enfoque tiene que enfrentarse a todos los problemas del análisis causal.

### **4. Observaciones finales**

En principio y en la práctica, tanto en los estudios cuantitativos como en los cualitativos se plantean los mismos problemas inferenciales. Las investigaciones que se diseñan para ayudarnos a entender la realidad social sólo pueden lograr su propósito mediante la lógica de la inferencia científica. Esta máxima es aplicable a los estudios cualitativos, cuantitativos, con *n* grande o pequeño, experimentales, de observación, históricos, etnográficos, a los que utilizan la observación participante y a todos los demás que investigan la sociedad de manera científica. Sin embargo, como este capítulo tendría que haber dejado claro, los problemas fundamentales de las inferencias descriptiva y causal suelen ser más difíciles de evitar cuando se utiliza

un *n* pequeño que cuando éste es grande. En este libro se han señalado formas de incrementar el número de observaciones para un estudio y también cómo hacer inferencias a partir de una cantidad de observaciones relativamente pequeña.

Tanto los investigadores cuantitativos como los cualitativos pueden hacer más eficiente un estimador aumentando la cantidad de información que se utiliza para resolver un problema —con frecuencia mediante el incremento del número de observaciones (apartado 7.2 del capítulo 2)— y, a veces, recurrir a procedimientos aleatorios de selección y adjudicación de valores con el fin de evitar automáticamente el sesgo. Gran parte de los análisis de este libro pretendían ayudar a los cualitativistas a tener estimadores más precisos, pero las técnicas que hemos señalado son diversas y, con frecuencia, hay que llegar a un equilibrio entre varios objetivos de investigación válidos; de ahí que sea difícil concentrar nuestras indicaciones en enunciados concisos que se correspondan con las ecuaciones formales que prefiere la investigación cuantitativa.

Los investigadores que, en el análisis de los fenómenos sociales, deciden prescindir de los procedimientos cuantitativos formales no pueden hacer caso omiso de los sesgos e inefficiencias que crean los diseños de investigación metodológicamente irreflexivos. Sus temas de estudio son tan importantes, y a veces más, que los que analizan los cualitativistas. Las inferencias descriptivas y causales que hacen los cualitativistas merecen ser tan sensatas como las de cualquier otro investigador, y para que sean válidas sus autores van a necesitar acercarse más que antes a las cuestiones metodológicas. También deben ser más contenidos al diseñar sus estudios y más explícitos al plantear resultados importantes. Los lectores no tienen por qué reformular los estudios cualitativos publicados para hacerlos científicamente válidos. Si un autor conceptualiza un proyecto de investigación de numerosas consecuencias observables utilizando sólo dos de ellas y doce hipótesis causales, no será responsabilidad de quien lo lea o reseñe el explicar que el diseño implícito del autor era mejor que el explícito. Lo que es aún más fundamental, los investigadores que comprendan y expliquen la lógica de sus análisis producirán estudios más valiosos. Por fortuna, las cuestiones metodológicas que los cualitativistas tienen que comprender son, precisamente, las mismas que el resto de los investigadores científicos debe respetar. Una inferencia válida sólo es posible si se entiende y respeta la lógica interna que subyace en todas las investigaciones sociales de tipo científico.

## Notas

### Capítulo 1

1 Rechazamos el concepto de «semieperimentos» o, al menos, la expresión. Un diseño de investigación o bien conlleva un control por parte del investigador de las observaciones y valores de las principales variables causales (en cuyo caso es un experimento) o no tiene este control (entonces es una investigación de carácter no experimental). Ambos tipos de investigación tienen sus ventajas e inconvenientes, y no puede decirse que uno sea siempre mejor que el otro.

2 Aunque nos ocupemos de casi todas las reglas importantes de la inferencia científica, no puede decirse que éstas sean completas. De hecho, casi todos los filósofos están de acuerdo en qué, incluso como principio, es imposible llegar a una lógica inductiva completa y exhaustiva.

3 Sin embargo, una hipótesis alternativa, que postula que la extinción fue producida por erupciones volcánicas, también puede explicar la presencia del iridio, y parece que concuerda mejor con el descubrimiento de que todas las extinciones de especies no ocurrieron simultáneamente.

4 Este dilema no es otro que al que se enfrentan los que estudian las ciencias naturales a la hora de decidir si llevan a cabo investigación aplicada o básica. Por ejemplo, la aplicada que tenga como objeto un determinado fármaco o enfermedad puede mejorar, a corto plazo, la atención médica sin aportar gran cosa al conocimiento general de los mecanismos biológicos subyacentes. La investigación básica puede tener consecuencias opuestas. La mayoría de los investigadores señalarían, al igual que se hace en las ciencias sociales, que esta dicotomía es falsa y que la investigación básica conduce finalmente a eficaces resultados prácticos. Sin embargo, todos estarían de acuerdo en que el mejor diseño de investigación es aquel que de alguna manera se las arregla para participar directamente en la resolución de problemas del mundo real y profundizar en los objetivos de lo escrito en una determinada área científica.

5 Éste es el principio de falsabilidad (Popper, 1968), asunto sobre el que hay diversas opiniones en la filosofía de la ciencia. Sin embargo, pocos cuestionan la idea de que las teorías deban enunciarse con la mayor claridad posible, de forma que pueda revelarse que están equivocadas.

6 Probablemente, ésta sea la pregunta que se plantea más a menudo en las entrevistas de trabajo de nuestro departamento y de muchos otros.

7 Esta frase es conocida como el «postulado de simplicidad de Jeffreys-Wranch». El concepto es similar al de la navaja de Occam.

8 Si hemos elegido un asunto de importancia para el mundo real y/o éste hace alguna aportación a lo escrito en un área de estudio, la naturaleza social de la academia corrigeá cualquier posible error: alguien reproducirá nuestro estudio con otro conjunto de datos y demostrará que estábamos equivocados.

9 Nos hemos dado cuenta de que muchos estudiantes de doctorado tienen un miedo innecesario a compartir los datos y la información precisa para reproducir sus resultados. Temen que alguien les robe su preciado trabajo o que incluso demuestre que están equivocados. Son temores habituales, pero pocas veces están justificados. Publicar (o, al menos, enviar copias de lo escrito sobre la investigación a otros académicos) y compartir los datos es la mejor forma de darle credibilidad a nuestra aportación. Además, compartir los datos no hará más que facilitar que otros sigan el camino trazado con la investigación que uno ha comenzado. Una vez que ésta se publique, citarán nuestro esfuerzo e incrementarán nuestra visibilidad y reputación.

10 Podemos comprobar nosotros mismos la fiabilidad midiendo una misma cantidad dos veces y fijándonos en si el resultado es igual. A veces esto parece fácil, como en el caso de hacer exactamente la misma pregunta en diferentes ocasiones durante una entrevista; sin embargo, plantearla una vez puede influir en el entrevistado para que en la segunda ocasión responda de una forma coherente con la anterior, de modo que hay que tener cuidado de que las dos mediciones sean realmente independientes.

11 Ejemplo de ello es la utilización de más de un codificador para la extracción de información sistemática de las transcripciones de entrevistas en profundidad. Si dos personas utilizan las mismas normas de codificación, podemos observar con cuánta frecuencia emiten el mismo juicio. Si no se obtienen medidas fiables, se pueden elaborar normas de codificación más precisas e intentarlo de nuevo. Al final, suele ser posible crear un conjunto de normas para que la aplicación de un mismo procedimiento por diferentes codificadores arroje el mismo resultado.

12 La expresión «falacia ecológica» es confusa porque el razonamiento que lleva de los niveles agregados a los individuales no es ni ecológico ni falaz. No es muy afortunado utilizar el término «ecológico» para describir un análisis de tipo agregado. Aunque Robinson (1990) llegó a la conclusión, en su primer artículo sobre este asunto, de que utilizar el análisis agregado para hacer razonamientos sobre individuos es una falacia, muchos cuantitativistas de las ciencias sociales y también estadísticos reconocen ahora que en niveles agregados de análisis sí existe cierta información sobre los individuos, y se han desarrollado muchos métodos de inferencia «ecológica» no sesgados.

## Capítulo 2

1 Tampoco podemos rechazar a Caro por pertenecer a otra profesión: la de periodista/biógrafo, cuyos objetivos son diferentes a los del científico social. Su obra se centra en algunos problemas de los que podría ocuparse un político, como ¿qué lleva al éxito o al fracaso en una campaña electoral?, ¿cuál es el papel del dinero y de la finan-

cación de la campaña en el éxito electoral?, ¿qué motiva a los que dan dinero para la campaña? El análisis se centra en una determinada candidatura de una circunscripción concreta, pero el problema y las cuestiones que plantea se solapan con lo que es habitual en la ciencia política.

2 Con el fin de completar esta explicación, merece la pena señalar que podríamos imaginar una teoría completamente diferente en la que una contracción del párpado no fuera un guiño pero tuviera, a pesar de todo, un efecto causal en otros actores. Por ejemplo, el tic podría haberse malinterpretado. Si también nos interesaría si la persona que contrae el párpado *pretendía* hacer un guiño, tendríamos que buscar otras consecuencias observables de esta misma teoría.

3 La bibliografía sobre los estudios de caso comparados es enorme. Para complementarla, algunas de las mejores obras son las de Eckstein (1975), Lijphart (1971) y Collier (1991).

4 Coombs (1964), por ejemplo, demostró que casi todas las colecciones de datos útiles requieren o implican un cierto grado de teoría, o de «miniteorías». Sin embargo, hay muchos datos cuantitativos, e históricos de tipo cualitativo, que se recogen con el propósito explícito de animar a futuros investigadores a que los utilicen para fines no previstos anteriormente. Quince minutos con el *Statistical Abstract of the United States* (Resumen estadístico de los Estados Unidos) convencerá a la mayoría de este hecho. Las labores de recogida de datos también difieren por el grado de rigidez con que los investigadores siguen sus presupuestos previos.

5 Formalmente, para un conjunto de unidades  $n$  en el que se mide la variable  $y$  ( $y_1, \dots, y_n$ ), un estadístico  $h$  es una función con un valor real que se define como:  $h = h(y) = h(y_1, \dots, y_n)$ .

6 Este asunto está muy relacionado con el concepto de diseño de investigación impreciso, que analizamos en el apartado 1 del capítulo 4.

7 Véase Popper (1982) para un libro entero dedicado a la defensa del indeterminismo.

8 Evidentemente, lo mismo puede aplicarse a todas las comunidades que podríamos estudiar.

9 Hay que señalar que la aleatoriedad no tiene que ver exactamente con diferentes semanas reales, ya que tanto los acontecimientos imprevistos como las diferencias sistemáticas podrían explicar las divergencias observadas. Por lo tanto, creamos una situación lo más ideal posible en la que podamos imaginar que ponemos en marcha el mundo de nuevo con características sistemáticas constantes y dejando que varíen los factores imprevistos.

10 Por supuesto,  $y$ , puede tener mucho interés para los habitantes de la circunscripción 5 en ese año y, por tanto, puede valer la pena estudiar los componentes aleatorio y sistemático de ese acontecimiento. No obstante, siempre tendríamos que intentar distinguir lo aleatorio de lo sistemático.

11 Véase King (1991b) para una elaboración de esta distinción.

12 Los economistas tienden a estar más cerca del punto de vista 1, mientras que los estadísticos están más próximos al 2. El primero también es especialmente común en la rama de la ingeniería que se denomina «control de calidad». Los físicos han llegado a debatir esta distinción dentro de la mecánica cuántica. En esta área, los primeros que propusieron el punto de vista 2 fueron los partidarios de la «teoría de la variable oculta». Sin embargo, las obras más recientes tienden a constatar el punto de vista 1: el mundo físico parece intrínsecamente probabilístico. Todos esperamos la resolución de las numerosas contradicciones que quedan en esta importante teoría y en sus consecuen-

cias para la naturaleza del mundo físico. Sin embargo, esta disputa dentro de la física, aunque solía justificar gran parte de la filosofía de las ciencias sociales, probablemente no influirá en la lógica inferencial de éstas o en la práctica de sus investigaciones.<sup>1</sup>

13 En nuestro lenguaje, como analizaremos más adelante en el apartado 5 del capítulo 3, éste es un problema de *endogeneidad*. El prever la cooperación podría conducir a que se convocaran reuniones en la cumbre, de modo que no serían estas cumbres las que explicaran la cooperación, sino que ésta parecería el resultado de tal previsión, lo cual no sería muy sorprendente si los actores son racionales!

14 Hay que señalar que el estimador puede no estar sesgado pero sí ser incoherente. Por ejemplo,  $\bar{Y}_t$  es un estimador no sesgado de  $\mu$ , pero es incoherente porque no mejora cuando va disminuyendo el número de unidades ( $t$ , en realidad, no cambia en absoluto). Un estimador también puede ser coherente pero estar sesgado. Por ejemplo  $\bar{Y} - 5/n$  está sesgado, pero es coherente porque  $S/n$  se convierte en cero a medida que  $n$  se acerca al infinito.

15 En la segunda línea de este cálculo y en los siguientes presuponemos que no existe una correlación espacial entre los distritos.

### Capítulo 3

1 A la vista de la preferencia que muestran algunos científicos sociales por la explicación, en vez de por la «simple descripción», no resulta sorprendente que los que estudian acontecimientos complicados pretendan envolver su trabajo con los atavíos de la jerga explicativa. Temen que si no lo hacen así se considerará que su trabajo es inferior. En esencia, una auténtica explicación siempre se basa en inferencias causales. Creemos que los argumentos de la bibliografía sobre «explicación no causal» tienen una terminología confusa y que, en casi todos los casos, se refieren en realidad a la explicación causal o carecen de coherencia interna. Si la incapacidad de explicar que tienen los científicos sociales no se debe a investigaciones débiles o a una falta de imaginación sino a la naturaleza de los difíciles pero significativos problemas que estudian, esos sentimientos de inferioridad no están justificados. Describir bien acontecimientos importantes es mejor que explicar mal cualquier cosa.

2 Nuestro punto de partida en este apartado es el artículo de Holland (1986) sobre causalidad y lo que él denomina «modelo de Rubin». Holland se basa en la obra de numerosos autores. En esta materia, la de Donald Rubin (1974, 1978) fue la que tuvo una relevancia más inmediata, pero Holland también cita, entre otros, a Aristóteles, Locke, Hume, Mill, Suppes, Granger, Fisher y Neyman. Aquí extendemos la definición de efecto causal de Holland utilizando algunas ideas, que expresaron claramente Suppes (1970) y otros autores, en relación a la «causalidad probabilística». Esta ampliación nos ha parecido necesaria porque ninguno de los enfoques que existen puede definir la causalidad en relación a una sola unidad, permitiendo, *a la vez*, que se dividan los efectos causales en componentes sistemáticos y no sistemáticos.

3 Para más información sobre este ejemplo, véase Gelman y King (1990). De forma más general,  $I$  y  $N$  pueden representar el grupo de «tratamiento» y el de «control», o cualquiera de los dos tratamientos que se han experimentado, de hecho o en teoría. Evidentemente, si se utiliza este lenguaje, denominar, dentro de una variable explicativa, a un valor «tratamiento» y al otro «control» es una decisión completamente arbitraria.

4 Jon Elster (1983, pp. 34-36) ha señalado que existen muchas situaciones, como en las que hay un tercer factor que explica tanto la variable aparentemente explicativa como la dependiente, en que «el significado de la causalidad no puede expresarse me-

diantes enunciados *contrafácticos*». Según nuestro lenguaje, Elster no hace más que mencionar problemas habituales en las *inferencias* que, en cierta medida, son siempre inciertas. Sin embargo, estas dificultades de la inferencia no invalidan una *definición* de causalidad que se base en contrafácticos. A pesar de sus objeciones, Elster reconoce que este tipo de enunciados «tiene un importante papel en el análisis causal» (Elster, 1983, p. 36); de ahí que, para nosotros, su argumento sea más convincente si se considera como un conjunto de valiosas advertencias contra la utilización descuidada de los contrafácticos que como una crítica de su importancia fundamental para la definición de conceptos en el razonamiento causal.

5 Podemos limitar el sentido de la siguiente ecuación a esta circunscripción sustituyendo « $i$ » por « $4$ ».

6 Probablemente, la razón sea que Holland es un estadístico que está muy próximo a una versión extrema de la variación aleatoria según el «punto de vista 2» que se ha descrito en el apartado 6 del capítulo 2. Al explicar la «solución estadística» para el problema de la inferencia causal, se acerca mucho a nuestra definición de efecto causal, pero su planteamiento tiene que ver sobre todo con la utilización de diferentes unidades para resolver el problema fundamental, más que con preservar la definición de causalidad en una sola. En concreto, su operador de valor esperado hace una media de las unidades, mientras que nosotros (como se describe posteriormente) hacemos una media de las reproducciones hipotéticas del mismo experimento en una sola unidad (véase Holland, 1986, p. 947).

7 Como indicamos con más detalle en el apartado 2 del capítulo anterior, esta expresión puede llevar a engaño. Una «variable aleatoria» tiene un cierto componente sistemático y, por tanto, no siempre es del todo impredecible. Por desgracia, este lenguaje tiene un significado específico en estadística, y los conceptos que implica son importantes. La razón primordial para utilizar esta terminología es que la aleatoriedad no significa que «todo vale» o que «puede pasar cualquier cosa». En realidad, hace referencia a un proceso probabilístico muy bien especificado que se extrae de entre los muchos que son posibles. Por ejemplo, el proceso aleatorio que determina qué cara de una moneda queda mirando hacia arriba después de arrojarla al aire es muy diferente al que afecta al crecimiento de la burocracia de la Unión Europea o a las inciertas consecuencias políticas de un cambio en el sistema electoral italiano. La clave de nuestra representación es que todos estos procesos «aleatorios» tienen componentes sistemáticos y probabilísticos.

8 Estas mismas distinciones son objeto de debate. Hay autores que señalan que un sistema presidencial puede ser más representativo, y otros, que los regímenes parlamentarios son más decididos.

9 La distinción entre característica sistemática y no sistemática no está en absoluto clara en todas las ocasiones. La repentina enfermedad de un presidente parece un rasgo no sistemático del sistema presidencial. Por otra parte, la vulnerabilidad general de este tipo de ordenamiento ante las fluctuaciones de salud y personalidad de un único individuo es un efecto sistemático que posibilita la aparición de *alguna* característica no sistemática.

10 Con frecuencia, esta idea se explica afirmando que no hay variable explicativa que sea necesaria o suficiente para que una dependiente tenga un determinado valor. Sin embargo, ésta es una terminología engañosa porque la distinción entre *condiciones necesarias y suficientes* desaparece en gran medida cuando concedemos a las causas la *posibilidad de ser probabilísticas*. Como explica Little (1991, p. 27): «Pensemos en la afirmación de que una mala comunicación entre las superpotencias durante una crisis aumenta la posibilidad de guerra. Ésta es una afirmación probabilística, ya que identifi-

## El diseño de la investigación social

ca una variable causal (la mala comunicación) y sostiene que con ella aumenta la probabilidad de que se produzca un determinado resultado (una guerra). Sin embargo, no puede convertirse en una afirmación acerca de las condiciones necesarias y suficientes para que haya una guerra, por lo que es irreduciblemente probabilística».

11 El insistir en la necesidad de diferenciar los componentes sistemáticos de los no sistemáticos en las observaciones que son objeto de inferencia causal refleja nuestra idea general de que el mundo, al menos tal como lo conocemos, es más probabilístico que determinístico. De ahí que tampoco estemos de acuerdo con la premisa de Ragin (1987, p. 15), según la cual las «explicaciones que surgen de aplicar el método comparado no se conciben de forma probabilística porque, si es posible, cada uno de los ejemplos de un fenómeno se examina y explica». Aunque se pudiera hacer una especie de censo con la información de cada uno de los ejemplos de un fenómeno, así como con cada permutación y combinación de los valores de las variables explicativas, el mundo, de todas formas, produciría esos datos mediante algún proceso probabilístico (tal como se definió en el apartado 6 del capítulo 2). Esto parece invalidar el enfoque del álgebra booleano de Ragin como forma general de diseñar explicaciones teóricas o extraer inferencias. Para aprender de los datos es necesario utilizar la misma lógica de la inferencia científica que analizamos en este libro. Sin embargo, el enfoque de este autor sí puede ser valioso como un tipo de teoría formal (véase el apartado 5.2 del capítulo 3), ya que posibilita que el investigador concrete una teoría y sus consecuencias mucho más fácilmente que si se careciera de él.

12 Con el fin de evitar la utilización de un término constante, presuponemos que todas las variables tienen una media igual a cero. Esto simplifica la presentación, pero no limita en modo alguno nuestras conclusiones.

13 Se nos podría llamar (¡a acusar de!) «justificacionistas» o incluso «justificacionistas probabilísticos» (véase Lakatos, 1970), pero si nos han de poner una etiqueta, preferimos la más coherente y filosófica de «bayesianos» (véanse Learner, 1978; Zellner, 1971; y Barnett, 1982). De hecho, nuestros objetivos son lo que más nos diferencia de Popper. Si el objetivo es concretamente el suyo, estamos de acuerdo con el procedimiento; si es el nuestro, quizás él estuviera de acuerdo con nosotros. Sin embargo, creemos que nuestros objetivos están más cerca de los que se plantean en las ciencias sociales y también de los que es más probable alcanzar.

14 Como siempre, cuando modificamos una teoría para que concuerde con los datos que hemos recogido, hay que evaluar dicha teoría (o la parte a la que se refieren nuestros datos) en un contexto diferente o en otro conjunto de datos.

15 Otra formulación de las ideas de Popper es la de que «no es posible demostrar una negación». Para este autor, no es posible porque el hecho de que un resultado concuerde con la hipótesis puede proceder, simplemente, de una comprobación equivocada. Quien intente demostrar negaciones siempre se encontrará con este problema. De hecho, las dificultades no sólo serán teóricas, sino también profesionales, ya que las revistas académicas tienden a publicar más resultados positivos que negativos.

Esto ha llevado a lo que se denomina *problema del archivador*, que donde más claramente se manifiesta es en la bibliografía cuantitativa. Supongamos que no existen pautas en el mundo. De manera que cinco de cada cien comprobaciones a las que se someta cualquiera de ellas se apartarán del 95% que constituye el intervalo de confianza y que, por tanto, generarán inferencias incorrectas. Si partimos de la base de que las revistas académicas publican resultados positivos en vez de negativos, sólo publicarán ese 5% que es «significativo», o sea, únicamente conclusiones equivocadas, por lo que

## Notas

juegues archivadores estarán llenos de artículos que llegan a conclusiones acertadas! (véase Iyengar y Greenhouse (1988) para una revisión de la bibliografía estadística que se ocupa de este problema). De hecho, los investigadores conocen bien estos incentivos y es probable que también influyan en su trabajo. Aunque el índice de aceptación en las principales publicaciones académicas sea más o menos del 5%, la situación no es tan grave, aunque no deja de ser un importante problema. Desde nuestro punto de vista, el problema del archivador podría solucionarse si todos adoptaran nuestra posición alternativa. *Un resultado negativo es tan útil como uno positivo; los dos pueden aportar la misma cantidad de información sobre el mundo*. Estaremos sobre seguro siempre que presentemos nuestras estimaciones y midanlos la incertidumbre.

16 La regla para hacer las mejores preguntas en una entrevista es casi la misma que se utiliza al explicar: ser lo más concreto posible. No hay que preguntar a un estadounidense blanco y conservador: «¿es usted racista?», sino más bien «¿le importaría que su hija se casara con un hombre negro?». No se debe preguntar a la gente si sabe de positiva; hay que pedirle que nos diga los nombres del secretario de Estado y del presidente de la Cámara de Representantes. En general, y siempre que sea posible, *no hay que pedir al entrevistado que nos haga el trabajo*. Es mejor no pedir la estimación de un efecto causal; tenemos que preguntar por medidas de las variables explicativa y dependiente para hacer la estimación nosotros. No tenemos que pedir motivaciones, sino hechos.

Esta regla no implica que no haya que preguntarle a alguien por qué hizo tal cosa. De hecho, preguntar las motivaciones suele ser muy útil para generar hipótesis. Las motivaciones que uno admite también pueden constituir un valioso conjunto de consecuencias observables. Sin embargo, hay que interpretar las contestaciones como una reacción del entrevistado a las preguntas del investigador, y no necesariamente como una respuesta correcta. Para que las preguntas de este tipo sirvan de algo, hay que diseñar el estudio de manera que una determinada respuesta obtenida (independientemente de las justificaciones, adornos, mentiras o recuerdos selectivos que nos encuentren) sea una consecuencia observable.

### Capítulo 4

1 En econometría y estadística a la bibliografía sobre la «identificación», le interesa determinar cuándo es impreciso el diseño de investigación en la investigación cuantitativa y cómo se puede corregir el modelo o recoger diferentes tipos de datos para afrontar el problema. Véanse Hsiao (1983) y King (1989, apartado 8.1).

2 Subrayamos de nuevo que no hay que confundir aleatoriedad con azar. En este contexto, hacer una selección aleatoria significa que cada unidad potencial tiene la misma probabilidad de ser seleccionada para nuestra muestra y que las sucesivas elecciones son independientes; como cuando extraemos nombres de un sombrero y después volvemos a introducirlos en él. Ésta es sólo la forma más simple de aleatoriedad, pero todas ellas dependen de procesos probabilísticos.

3 Algunos ejemplos pueden encontrarse en: Roth (1988), Iyengar y Kinder (1987), Fiorina y Plot (1978), Plot y Levine (1978) y Palfrey (1991).

4 El criterio de selección (1,1,0) omite el extremo inferior de la escala (la unidad baja) y el segundo (0,1,1) prescinde de la unidad que está en el extremo superior (unidad alta). Sólo el tercer caso, en el que no se selecciona «media», no se correlaciona con la variable dependiente.

5 Éste es un buen ejemplo de lo que caracteriza a la ciencia. Cuando introducimos este sesgo para confirmar nuestra conclusión, no nos comportamos como tiene que ha-

cerlo un científico social, sino como suele ser habitual cuando, al discutir de política, defendemos nuestras profundas convicciones seleccionando con frecuencia ejemplos que refuerzan nuestros argumentos. Al investigar, si es posible, hay que intentar tener todas las observaciones. Si se precisa una selección, hay que procurar obtener las observaciones que sean esenciales para decidir cuál es la pregunta que tiene interés, y no sólo las que concuerdan con nuestra posición.

6 En este apartado, no contemplamos la posibilidad de que ciertos proyectos de investigación que se diseñan para no dejar que cambie en absoluto la variable dependiente sean parte de un programa de estudio mayor y que, por tanto, puedan aportar información útil sobre las hipótesis causales. Explicaremos esta cuestión en el apartado 4 del capítulo 4.

7 Porter afirma que tiene muchos ejemplos de países que no han alcanzado el éxito competitivo; sin embargo, éstos entran en su análisis mediante una elección selectiva de anécdotas y no se estudian con métodos parecidos a los de sus primeras diez naciones. Cuando los ejemplos que corroboran nuestro argumento se extraen de forma no sistemática de un infinito abanico de posibilidades que incluye elementos que nos favorecen y otros que no, resulta demasiado fácil hacer el ridículo encontrando relaciones que no existen. No variaría a valorar si las hipótesis de Porter son correctas; sólo queremos señalar que la información necesaria para decidir sobre este punto debe recabarse de manera más sistemática.

8 En general, el sesgo de selección tiene lugar cuando ésta se realiza a partir de la variable dependiente, después de tener en cuenta (o controlar) las variables explicativas. Como una de ellas es el método de selección, lo controlamos y no introducimos sesgo.

9 La inferencia también será un poco menos segura si el abanico de valores de las variables explicativas se limita mediante la selección. Véase el apartado 2 del capítulo 6.

10 Dicho de otra forma diferente, si seleccionamos en función de las distribuciones marginales de la variable dependiente y de las explicativas, podremos seguir aprendiendo algo sobre su distribución conjunta al hacer el estudio.

## Capítulo 5

1 Una excepción a esta regla se produce cuando los errores sistemáticos positivos anulan los negativos, pero es más correcto definir este caso inusual como un tipo de error de medida no sistemático.

2 Este error, independientemente de que se produzca por nuestra incapacidad de medir el mundo real con precisión o por la aleatoriedad de la naturaleza, siempre nos planteará una pregunta filosófica que puede responderse de varias maneras (apartado 6 del capítulo 2). Cualquiera que sea la posición que aceptemos, la consecuencia será la misma.

3 Imaginamos de nuevo que los cuatro círculos blancos representan o bien un gran número de observaciones que coinciden exactamente con ellos o bien una escasa variación estocástica.

4 Los lectores versados en estadística reconocerán que esto corresponde a la propiedad de homocedasticidad, o varianza constante.

5 Estos supuestos de error implican que el valor esperado de la variable dependiente observada es igual al esperado para la verdadera variable dependiente:

$$E(Y) = E(Y^* + U) = E(Y^*) + E(U) = E(Y^*) = \beta X$$

6 Seguimos presuponiendo aquí que cada círculo representa o bien datos en los que apenas hay variación estocástica o bien una situación en la que numerosos puntos coinciden en el mismo lugar. Al igual que en el apartado 1 de este mismo capítulo, el propósito de este supuesto es centrarnos en el problema.

7 Es preciso señalar la diferencia que hay entre los dos casos en los que resulta aceptable omitir una variable. En el primero, donde la variable omitida no tiene relación con la dependiente, no existe sesgo y no perdemos ninguna capacidad de predecir los valores futuros de esta última variable. En el segundo caso, en el que la variable omitida no tiene relación con la independiente pero sí con la dependiente, tampoco aparece sesgo en nuestra estimación de la *relación* entre la variable explicativa incluida en el análisis y la dependiente, pero sí perdemos cierta capacidad de pronosticar con exactitud los valores futuros de la segunda. Por lo tanto, si el estar ya en el poder no tuviera relación con los gastos de la campaña electoral, omitir este factor no sesgaría nuestra estimación de la relación que hay entre ese gasto y los votos. Pero si nuestro objetivo fuera pronosticar, tendríamos que representar todas las variaciones sistemáticas de la variable dependiente, y omitir el hecho de estar ya en el poder nos lo impediría, porque estaríamos dejando de lado una importante variable causal. Sin embargo, aunque a largo plazo nuestro objetivo fuera explicar el voto de la forma más completa y sistemática, podría resultar difícil confiar en varios efectos causales que estuvieran encuadrados dentro de un solo estudio. Por lo tanto, quizás necesitáramos centrarnos en un único efecto causal (o sólo en unos pocos), cuaquiera que fuera nuestro objetivo último.

8 Para decirlo de una forma más precisa,  $F$  es la estimación del coeficiente que se produce cuando se hace una regresión entre  $X_1$  y  $X_2$ .

9 Como esta ecuación sólo es exactamente aplicable cuando hay grandes muestras, en realidad estamos analizando la coherencia en vez de la falta de sesgo (apartado 7.1 del capítulo 2). Para ser más exactos, los términos que aparecen entre paréntesis en la ecuación (5.2) se desvaneцен al dividirlos por  $n$ , a medida que este valor se acerca al infinito.

10 Merece la pena considerar lo que supone ocuparse del efecto causal estimado de los precios del crudo sobre la actitud de la opinión pública ante la escasez de energía, controlando a la vez la cantidad de información televisiva al respecto. Pensemos en dos descripciones que son importantes porque posibilitan análisis y estudios más profundos de los procesos causales. En primer lugar, este efecto estimado sólo representa la influencia que tiene el elemento de los precios del crudo que afecta *directamente* a la actitud de la opinión pública ante la escasez de energía, y que es diferente del que influye en la audiencia de manera indirecta cuando cambia la información televisiva. En resumen, es el efecto directo que tiene el crudo sobre la opinión pública y no el indirecto. El efecto total puede encontrarse si no se controla en absoluto la cantidad de información televisiva que hay acerca de la escasez de energía. Este efecto también se puede describir como el que tienen los precios de la energía sobre la variable «opinión pública en relación a la escasez de energía si se registran unos niveles fijos de información televisiva al respecto». Como ejemplo de esta posibilidad, imaginemos un experimento en el que controláramos la información que dan las cadenas de televisión sobre la escasez de crudo, manteniendo esta variable constante mientras varían, lógicamente, los precios del petróleo. Como la cantidad de información es constante en este experimento, se controla este factor sin recurrir a ningún otro procedimiento explícito. Aunque no pudieramos llevar a cabo un experimento, sí podríamos hacer una estimación del efecto condicional que tienen los precios del crudo sobre lo que piensa la opinión pública de la escasez de energía, controlando la información televisiva.

11 Además, podría interesarnos únicamente el efecto directo o indirecto de una variable o, incluso, qué efecto causal tiene alguna otra en una ecuación. Si es así, es perfectamente razonable realizar varios análisis diferentes con los mismos datos, siempre que entendamos las diferencias que hay en la interpretación.

12 Milner se enfrentó al posible sesgo de la variable omitida, pero su razonamiento era incorrecto: «Al considerar diferentes sectores industriales, en diversos países, [el diseño de investigación] permite que cambien [las variables de control omitidas], a la vez que muestra que el argumento fundamental sigue teniendo base» (1988, p. 15). En realidad, la única manera de mantener constantes las variables de control<sup>1</sup> es impedir realmente que varíen. Si, mediante otras teorías plausibles, Milner hubiera detectado la importancia de esas variables, habría podido incorporar un conjunto de observaciones que cambiaban en su variable explicativa clave (grado de dependencia económica del país o sector industrial respecto al exterior), pero no en esas variables de control.

13 Los cualitativistas a veces sí manipulan variables explicativas mediante la observación participante. Incluso las entrevistas en profundidad pueden ser un tipo de experimento si plantean de manera sistemática diferentes cuestiones o cambian ciertas condiciones en diversas entrevistas. De hecho, experimentar puede llegar a ser un problema para los que realizan este tipo de entrevistas, ya que quizás se sientan más cómodos al aplicar «tratamientos» experimentales (hacer determinadas preguntas) a ciertas personas que no han sido seleccionadas de forma aleatoria. Los experimentadores tienen sus propios problemas, y son muchos, pero la endogeneidad no se cuenta entre ellos.

14 Para seguir estrictamente los métodos del capítulo 3, habría que llevar a cabo este experimento varias veces y considerar que la media de ellas es el efecto causal medio del tratamiento aplicado. También podría interesarnos la varianza del efecto causal en este individuo.

## Capítulo 6

1 Sin embargo, como señalaremos más adelante, Eckstein parece reconocer la debilidad de un argumento que, en realidad, no le lleva a refutar la única observación, sino a buscar muchas observaciones.

2 El estudio de sociología comparada de Bollen, Entwistle y Alderson (1993) demuestra que prácticamente todos los libros y artículos que ellos analizaron concedían algún papel al azar, incluso aquellos que explícitamente utilizaban el método de la diferencia de Mill.

3 Kahneman, Slovic y Tversky (1982) describen una falacia psicológica del razonamiento que tiene lugar cuando los decisores sometidos a incertidumbre eligen analogías que se basan en hechos recientes o disponibles, lo cual hace que sus valoraciones estén sistemáticamente sesgadas. A este fenómeno lo denominan la «heurística de la disponibilidad». Véase también Keane (1988).

4 Las premisas son que  $E(Y) = X_1\beta_1 + X\beta$ ,  $V(Y) = \sigma^2$ , no existe multicolinealidad e implicitamente todas las expectativas dependen de  $X$ .

5 Técnicamente,  $\sigma^2$  es la varianza de la variable dependiente, que está condicionada por todas las variables explicativas  $V(Y|X)$ ;  $V(\hat{\beta}_j)$  es el cuadrado del error típico de la estimación del efecto causal de  $X_j$ ;  $R_j^2$  es  $R^2$  calculada a partir de una regresión auxiliar de  $X_j$  y de todas las variables de control, y  $S_{\hat{\beta}_j}^2$  es la varianza muestral de  $\hat{\beta}_j$ .

6 En la bibliografía relativa al método comparado se plantea habitualmente la conveniencia de aumentar el número de observaciones. Lijphart (1971) insiste de forma especial en este punto.

7 Estamos de acuerdo con lo que señala William Baumol (1990, p. 1715) en relación a la historia económica: «Muchos historiadores económicos se tienden a sí mismos una absurda trampa cuando pretenden explicar completamente una determinada evolución histórica. Quien intente describir las «cinco causas principales» del punto crítico al que llegó Gran Bretaña a finales del siglo XIX o las de la depresión económica europea de 1847 estará embarcándose en una empresa imposible. Las ciencias naturales, a pesar de sus muchos avances y conocimiento acumulado, siguen basándose en gran medida en experimentos controlados que, como tales, se centran en la influencia simultánea de una o varias variables. En realidad, los científicos buscan derivaciones parciales, más que explicaciones globales para fenómenos complejos de la realidad».

8 También podemos mantener la misma variable dependiente pero cambiando las explicativas. Sin embargo, en la mayoría de las situaciones, esta estrategia se utiliza para evitar el error de medida, utilizando medidas múltiples de la misma variable explicativa subyacente.

9 A veces, los investigadores realizan estudios que se describen como reproducciones de otros anteriores y en los que no se incluyen nuevas observaciones. Fundamentalmente, lo que reproducen —o intentan reproducir— son estudios ajenos, con el fin de comprobar si los resultados se repiten. Un cualitativista intentará repetir el análisis de un estudio anterior utilizando los mismos datos. Un historiador revisará las fuentes utilizadas por otro historiador. Un etnógrafo escuchará las entrevistas grabadas para comprobar si las conclusiones originales eran sensatas. Este tipo de actividad es muy útil porque las pruebas científicas deben ser reproducibles; sin embargo, no es lo que planteamos en este apartado, ya que no conlleva la utilización de nuevas observaciones.

10 Ciertas fuerzas externas también favorecieron el fracaso de Allende, pero Kohli concede más importancia a las internas.

11 Los cuantitativistas han desarrollado una enorme colección de elaboradas técnicas estadísticas para analizar datos en los que aparecen las denominadas propiedades de las series temporales o de la autocorrelación espacial. No sólo consiguen corregir estos problemas, sino que han logrado sacar una extraordinaria información de los datos. Véanse Giringer y Newbold (1977), Anselin (1988), Beck (1991) y King (1989, 1991c).

12 La utilización de la «cultura» como variable explicativa en las investigaciones sociales es objeto de un gran debate, pero no es el tema de este libro. Sólo señalaremos que las explicaciones de este tipo deben someterse a las mismas comprobaciones lógicas y de medida que se aplican al resto de las investigaciones.

13 Donald Moon denomina *explicación de los fundamentos* a una versión de este enfoque o, por utilizar una expresión de otros autores, análisis de las razones (Moon, 1975).

14 Lo que George y McKeown denominan «explicación dentro de la observación» constituye, siguiendo a Eckstein, una estrategia de redefinición de la unidad de análisis, con el fin de aumentar el número de observaciones. George y McKeown (1985, p. 36) afirman que en los estudios de caso «el comportamiento del sistema no se resume en un solo punto de datos, sino en una serie de puntos o de curvas que se van trazando a través del tiempo». En nuestra terminología, que tomamos de Eckstein (1975), este método consiste en aumentar el número de observaciones, ya que una sola observación se define como «un fenómeno del que sólo señalamos e interpretamos una única medida en cada variable pertinente».

## Bibliografía

- Achen, Christopher H. (1986): *Statistical Analysis of Quasi-Experiments*. Berkeley, University of California Press.
- , y Duncan Snidal (1989): «Rational Deterrence Theory and Comparative Case Studies», *World Politics* 41, núm. 2 (enero), pp. 143-169.
- Álvarez, Walter, y Frank Asaro (1990): «An Extraterrestrial Impact». *Scientific American* (octubre), pp. 78-84.
- Anselin, Luc (1988): *Spatial Econometrics: Methods and Models*. Boston, Kluwer Academic Publishers.
- Barnett, Vic (1982): *Comparative Statistical Inference*. 2.ª ed., Nueva York, Wiley.
- Baumol, William J. (1990): «St. John versus the Hicksians, or a Theorist Malgré Lui?». *The Journal of Economic Literature* 28, núm. 4, pp. 1708-1715.
- Beck, Nathaniel (1991): «Alternative Dynamic Structure». *Political Analysis* 3, pp. 51-87.
- Becker, Howard S. (1966): «Whose Side Are We On?». *Social Problems* 14, pp. 239-247.
- , y Charles C. Ragin (1992): *What Is a Case? Exploring the Foundations of Social Inquiry*. Nueva York, Cambridge University Press.
- Blainey, Geoffrey (1973): *The Causes of War*. Nueva York, Free Press.
- Bollen, Kenneth A., Barbara Entwistle y Arthur S. Alderson (1993): «Macrocomparative Research Methods». En Judith Blake (ed.), *Macrocomparative Research Methods*. Palo Alto, Calif. Annual Reviews, Inc.
- Cain, Bruce, John Ferejohn y Morris Fiorina (1987): *The Personal Vote: Constituency Service and Electoral Independence*. Cambridge, Harvard University Press.
- Caplow, Theodore, Howard M. Bahr, Bruce A. Chadwick y Dwight W. Hoover (1983a): *All Faithful People: Change and Continuity in Middletown's Religion*. Minneapolis, University of Minnesota Press.

### El diseño de la investigación social

- (1983b): *Middletown Families: Fifty Years of Change and Continuity*. Nueva York, Bantam Books.
- Caro, Robert (1983): *The Years of Lyndon Johnson*. Nueva York, Vintage Books.
- Collier, David (1991): «The Comparative Method: Two Decades of Change». En Dankwart A. Rustow y Kenneth Paul (eds.), *Comparative Political Dynamics: Global Research Perspectives*. Nueva York, Harper Collins.
- (1993): «The Comparative Methods». En Ada W. Finifter (ed.), *Political Science: The State of the Discipline*. Washington, D.C., American Political Science Association.
- Cook, Karen Schweers, y Margaret Levi (eds.) (1990): *The Limits of Rationality*. Chicago, University of Chicago Press.
- Coombs, Clyde H. (1964): *A Theory of Data*. Nueva York, Wiley.
- Courtillot, Vincent E. (1990): «A Volcanic Eruption». *Scientific American* (octubre), pp. 78-84.
- Dahl, Robert (1961): *Who Governs? Democracy and Power in an American City*. New Haven, Yale University Press.
- Dessler, David (1991): «Beyond Correlations: Toward a Causal Theory of War». *International Studies Quarterly* 3, núm. 35 (septiembre), pp. 337-355.
- Dewald, William G., Jerry G. Thursby y Richard G. Anderson (1986): «The Journal of Money, Credit and Banking Projects». *American Economic Review* 76, núm. 4 (septiembre), pp. 587-603.
- Diamond, Larry, y Marc F. Plattner (eds.) (1993): *The Global Resurgence of Democracy*. Baltimore, Johns Hopkins University Press.
- Duneier, Mitchell (1993): *Slim's Table*. Chicago, University of Chicago Press.
- Easton, David (1965): *A Systems Analysis of Political Life*. Nueva York, Wiley.
- Eckstein, Harry (1969): «Authority Relations and Governmental Performance». *Comparative Political Studies* 2, pp. 269-325.
- (1975): «Case Study and Theory in Political Science». En Fred I. Greenstein y Nelson W. Polsby (eds.), *Handbook of Political Science*, vol. 1, *Political Science: Scope and Theory*. Reading, Mass., Addison-Wesley.
- Elster, Jon (1983): *Explaining Technical Change: A Case Study in the Philosophy of Science*. Nueva York, Cambridge University Press [ed. castellana: *El cambio tecnológico: investigaciones sobre la racionalidad y la transformación social*. Barcelona: Gedisa, 1992].
- Fearon, James D. (1991): «Counterfactuals and Hypothesis Testing in Political Science». *World Politics* 43, núm. 2 (enero), pp. 169-195.
- Fenno, Richard F. (1978): *Home Style*. Boston, Little, Brown.
- Ferejohn, John (1993): «Structure and Ideology: Change in Parliament in Early Stuart Englands». En Judith Goldstein y Robert O. Keohane (eds.), *Ideas and Foreign Policy: Beliefs, Institutions and Political Change*. Ithaca, Cornell University Press.
- Ferguson, Yale H., y Richard W. Mansbach (1988): *The Elusive Quest: Theory and International Politics*. Columbia, University of South Carolina.
- Feynman, Richard P. (1965): *The Character of Physical Law*. Cambridge, Mass., MIT Press [ed. castellana: *El carácter de la ley física*. Barcelona: Orbis, 1987].
- Fiorina, Morris, y Charles R. Plott (1978): «Committee Decisions under Majority Rule». *American Political Science Review* 72, núm. 2 (junio), pp. 575-598.
- Fisher, Ronald A. (1935): *The Design of Experiments*. Nueva York, Hafner Publishing.
- Fogel, Robert William (1989): *Without Consent or Contract: The Rise and Fall of American Slavery*. Nueva York, W. W. Norton.

### Bibliografía

- Friedrich, Carl J. (1958): «Political Philosophy and the Science of Politics». En Roland Young (ed.), *Approaches to the Study of Politics*. Chicago, University of Chicago Press.
- Fudenberg, Drew, y Jean Tirole (1989): «Noncooperative game theory for industrial organization: an introduction and overview». En Richard Schmalensee y Robert D. Willig (eds.), *Handbook of Industrial Organization*, vol. 1. Amsterdam, North Holland.
- Garfinkel, H. (1964): «Studies of the Routine Grounds of Everyday Activities». *Social Problems* 11, pp. 225-250.
- Geddes, Barbara (1990): «How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics». *Political Analysis* 2, pp. 131-152.
- Geertz, Clifford (1973): *An Interpretation of Cultures*. Nueva York, Basic Books [ed. castellana: *Interpretación de las culturas*. Barcelona: Gedisa, 1988].
- (1983): «Local Knowledge: Fact and Law in Comparative Perspective». En Clifford Geertz (ed.), *Local Knowledge: Further Essays in Interpretive Anthropology*. Nueva York, Basic Books [ed. castellana: *Conocimiento local: ensayos sobre la interpretación de las culturas*. Barcelona: Paidós, 1994].
- Gelman, Andrew, y Gary King (1990): «Estimating Incumbency Advantage without Bias». *American Journal of Political Science* 34, núm. 4 (noviembre), pp. 1142-1164.
- (1993): «Why Are U.S. Presidential Election Polls So Variable When the Vote Is So Predictable». *British Journal of Political Science*.
- George, Alexander L. (1982): «Case Studies and Theory Development». Trabajo presentado en el segundo simposio anual sobre el proceso de la información en las organizaciones, Carnegie-Mellon University (octubre), pp. 15-16.
- , y Timothy J. McKeown (1985): «Case Studies and Theories of Organizational Decision Making». *Advances in Information Processing in Organizations* 2, pp. 21-58.
- , y Richard Smoke (1974): *Deterrance in American Foreign Policy*. Nueva York, Columbia University Press.
- Gigerenzer, Gerd, Zeno Swijtink, Theodore Porter, Lorraine Daston, John Beatty y Lorenz Kruger (1989): *The Empire of Chance: How Probability Changed Science and Everyday Life*. Nueva York, Cambridge University Press.
- Gilpin, Robert (1981): *War and Change in World Politics*. Nueva York, Cambridge University Press.
- Goldberger, Arthur (1991): *A Course in Econometrics*. Cambridge, Mass., Harvard University Press.
- Goldstein, Judith, y Robert O. Keohane (eds.) (1993): *Ideas and Foreign Policy: Beliefs, Institutions and Political Change*. Ithaca, Cornell University Press.
- Gould, Stephen J. (1989a): *Wonderful Life: The Burgess Shale and the Nature of History*. Nueva York, Norton [ed. castellana: *La vida maravillosa*. Barcelona: Mondadori, 1995].
- (1989b): «The Horn of Triton». *Natural History* (diciembre), pp. 18-27.
- Granger, G. W. J., y P. Newbold (1977): *Forecasting Economic Time Series*. Nueva York, Academic Press.
- Gulick, Edward V. (1967): *Europe's Classical Balance of Power*. Nueva York, Norton.
- Hall, Peter A. (ed.) (1989): *The Political Power of Economic Ideas: Keynesianism Across Nations*. Princeton, Princeton University Press.
- Halpern, Nina (1993): «Stalinist Political Economy». En Judith Goldstein y Robert O. Keohane (eds.) (1993), *Ideas and Foreign Policy: Beliefs, Institutions and Political Change*. Ithaca, Cornell University Press.

- Hermans, F. A. (1941): *Democracy or Anarchy: A Study of Proportional Representation*. South Bend, Ind., University of Notre Dame Press.
- Hirschman, Albert O. (1970): «The Search for Paradigms as a Hindrance to Understanding». *World Politics* 22, núm. 3 (abril), pp. 329-343.
- Hoffmann, Stanley (1960): *Contemporary Theory in International Relations*. Englewood Cliffs, Prentice-Hall [ed. castellana: *Teorías contemporáneas sobre las relaciones internacionales*. Madrid: Técnicos, 1979].
- Holland, Paul (1986): «Statistics and Causal Inference». *Journal of the American Statistical Association* 81, pp. 945-960.
- Horowitz, Donald (1993): «Comparing Democratic Systems». En Larry Diamond y Marc F. Plattner (eds.), *The Global Resurgence of Democracy*. Baltimore, Johns Hopkins University Press.
- Hsiao, C. (1983): «Identification». En Zvi Griliches y Michael Intriligator (eds.), vol. I, *Handbook of Econometrics*. Amsterdam, North-Holland.
- Huth, Paul (1988): «Extended Deterrence and the Outbreak of War». *American Political Science Review* 82, núm. 2 (junio), pp. 423-443.
- , y Bruce Russett (1990): «Testing Deterrence Theory: Rigor Makes a Difference». *World Politics* 42, núm. 4 (julio), pp. 466-501.
- Inkeles, Alex, y Peter Rossi (1956): «National Comparisons of Occupational Prestige». *American Journal of Sociology* 61, pp. 329-339.
- Iyengar, Satis, y Joel B. Greenhouse (1988): «Selection Models and the File Drawer Problem». *Statistical Science* 3, núm. 1 (febrero), p. 109-135.
- Iyengar, Shanto, y Donald Kinder (1987): *News That Matters*. Chicago, University of Chicago Press.
- Jeffreys, Harold (1961): *Theory of Probability*. Oxford, Clarendon Press.
- Jervis, Robert (1976): *Perception and Misperception in International Politics*. Princeton, Princeton University Press.
- , Richard Ned Lebow y Janice Gross Stein (1985): *Psychology and Deterrence*. Baltimore, Johns Hopkins University Press.
- Jones, E. L. (1981): *The European Miracle: Environments, Economies, and Geopolitics in the History of Europe and Asia*. Cambridge, Cambridge University Press [ed. castellana: *El milagro europeo*. Madrid: Alianza, 1994].
- Johnston, J. (1984): *Econometric Methods*. 3.ª ed., Nueva York, McGraw Hill [ed. castellana: *Métodos de econometría*. Barcelona: Vicens-Vives, 1994].
- Kahneman, Daniel, Paul Slovick y Amos Tversky (eds.) (1982): *Judgement under Uncertainty: Heuristics and Biases*. Nueva York, Cambridge University Press.
- Katzenstein, Peter J. (1985): *Small States in World Markets: Industrial Policy in Europe*. Ithaca, Cornell University Press [ed. castellana: *Pequeños estados en mercados mundiales: política industrial en Europa*. Madrid: Ministerio de Trabajo y Seguridad Social, 1987].
- Keane, Mark T. (1988): *Analogical Problem Solving*. Chichester, Ellis Horwood.
- Kennedy, Paul (1987): *The Rise and Fall of the Great Powers*. Nueva York, Random House [ed. castellana: *Auge y caída de las grandes potencias*. Barcelona: Plaza y Janés, 1997].
- Keohane, Robert O. (1980): «The Theory of Hegemonic Stability and Changes in International Economic Regimes, 1967-1977». En Ole R. Holsti, Randolph M. Siverson y Alexander L. George (eds.), *Change in the International System*. Boulder, Westview Press.
- (1984): *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, Princeton University Press.
- (1988): «International Institutions: Two Approaches». *International Studies Quarterly* 32, p. 379.
- (1989): *International Institutions and State Power: Essays in International Relations Theory*. Boulder, Westview.
- , y Joseph S. Nye, Jr. (1977): *Power and Interdependence: World Politics in Transition*. Boston, Little, Brown.
- Khong, Yuen Foong (1992): *Analogy at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965*. Princeton, Princeton University Press.
- King, Gary (1989): *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Nueva York, Cambridge University Press.
- (1993): «The Methodology of Presidency Research». En George Edwards III, John H. Kessel y Bert A. Rockman (eds.), *Researching the Presidency: Vital Questions, New Approaches*. Pittsburgh, University of Pittsburgh Press.
- (1991a): «Constituency Service and Incumbency Advantage». *British Journal of Political Science* 21, núm. 1 (enero), pp. 119-128.
- (1991b): «Stochastic Variation: A Comment on Lewis-Beck and Skalaban's 'The R-Square»». *Political Analysis* 2, pp. 185-200.
- (1991c): «On Political Methodology». *Political Analysis* 2, pp. 1-30.
- Kohli, Atul (1987): *The State and Poverty in India: The Politics of Reform*. Nueva York, Cambridge University Press.
- Kreps, David M. (1990): «Corporate Culture and Economic Theory». En James E. Alt y Kenneth Shepsle (eds.), *Perspectives on Positive Political Economy*. Nueva York, Cambridge University Press.
- Laitin, David D. (1986): *Hegemony and Culture: Politics and Religious Change among the Yoruba*. Chicago, University of Chicago Press.
- Lakatos, Imre (1970): «Falsification and the Methodology of Scientific Research Programs». En I. Lakatos y A. Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge, Cambridge University Press [ed. castellana: *La crítica y desarrollo del conocimiento*. Barcelona: Grijalbo-Mondadori, 1974].
- Lakeman, Enid, y James D. Lambert (1955): *Voting in Democracies*. Londres, Faber and Faber.
- Leamer, Edward E. (1978): *Specification Searches: Ad Hoc Inference With Nonexperimental Data*. Nueva York, Wiley [ed. castellana: *Búsquedas de especificación: inferencia ad hoc con datos no experimentales*. Bilbao: Editorial Españoña Desclée de Brouwer, 1983].
- (1983): «Let's Take the Con Out of Econometrics». *American Economic Review* 73, núm. 1 (marzo), pp. 31-43.
- Levy, Jack S. (1985): «Theories of General War». *World Politics* 37, núm. 3 (abril), pp. 344-374.
- «Quantitative Studies of Deterrence Success and Failure». En Paul C. Stern, Robert Axelrod, Robert Jervis y Roy Radner (eds.), *Perspectives on Deterrence*. Nueva York, Oxford University Press.
- Lieberson, Stanley (1985): *Making It Count: The Improvement of Social Research and Theory*. Berkeley, University of California Press.
- (1992): «Einstein, Renoir and Greeley: Some Thoughts about Evidence in Sociology». *American Sociological Review* 56 (febrero), pp. 1-15.
- Lijphart, Arend (1971): «Comparative Politics and Comparative Method». *American Political Science Review* 65, núm. 3 (septiembre), pp. 682-698.
- Lindberg, Leon N., y Stuart A. Cheingold (1970): *Europe's Would-Be Polity: Patterns of Change in the European Community*. Englewood Cliffs, Prentice-Hall.

### El diseño de la investigación social

- (1971): *Regional Integration: Theory and Research*. Cambridge, Harvard University Press.
- Linz, Juan J. (1993): «The Perils of Presidentialism». En Larry Diamond y Marc F. Plattner (eds.), *The Global Resurgence of Democracy*. Baltimore, Johns Hopkins University Press, pp. 108-126.
- Lipset, Seymour Martin (1963): *The First New Nation: The United States in Comparative and Historical Perspective*. Nueva York, Basic Books.
- Little, Daniel (1991): *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*. Boulder, Westview.
- Longino, Helen E. (1990): *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, Princeton University Press.
- Lowenthal, Abraham F. (1972): *The Dominican Intervention*. Cambridge, Harvard University Press.
- Mankiw, N. Gregory (1990): «A Quick Refresher Course in Macroeconomics». *Journal of Economic Literature* 28, núm. 4 (diciembre), pp. 1645-1660.
- Martin, Lisa L. (1992): *Coercive Cooperation*. Princeton, Princeton University Press.
- Merck & Co., Inc. (1989): *Annual Report*. Rayway, New Jersey, Merck & Co., Inc., 1989.
- Merton, Robert K. [1949] (1968): *Social Theory and Social Structure*. Nueva York, Free Press.
- Mill, John Stuart (1843): *A System of Logic* S. e. [ed. castellana: *Sistema de lógica induktiva y deductiva*. Madrid: Juan Pueyo, 1917].
- Miller, David (1988): «Conjectural Knowledge: Popper's Solution of the Problem of Induction». En Paul Levinson (ed.), *In Pursuit of Truth*. Atlantic Highlands, Humanities Press.
- Milner, Helen V. (1988): *Resisting Protectionism: Global Industries and the Politics of International Trade*. Princeton, Princeton University Press.
- Moe, Terry M. (1990): «The Politics of Structural Choice: Toward a Theory of Public Bureaucracy». En Oliver Williamson (ed.), *Organization Theory: From Chester Barnard to the Present and Beyond*. Nueva York, Oxford University Press.
- Moon, Donald J. (1975): «The Logic of Political Inquiry: A Synthesis of Opposed Perspectives». En Fred I. Greenstein y Nelson W. Polsby (eds.), *Handbook of Political Science*, vol. 1./*Political Science: Scope and Theory*. Reading, Mass., Addison-Wesley.
- Neustadt, Richard E., y Earnest R. May (1986): *Thinking in Time: The Uses of History for Decision-Makers*. Nueva York, Free Press.
- Nye, Joseph S. (1971): *Peace in Parts*. Boston, Little, Brown.
- O'Hear, Anthony (1989): *Introduction to the Philosophy of Science*. Oxford, Clarendon Press.
- Orlitzky, Peter C. (1986): *Game Theory and Political Theory: An Introduction*. Nueva York, Cambridge University Press.
- Palfrey, Thomas R. (ed.) (1991): *Laboratory Research in Political Economy*. Ann Arbor, University of Michigan Press.
- Pearson, Karl (1892): *The Grammar of Science*. Londres, J. M. Dent & Sons, Ltd.
- Plot, Charles R., y Michael E. Levine (1978): «A Model of Agenda Influence on Committee Decisions». *American Economic Review* 68, núm. 1 (marzo), pp. 146-160.
- Popper, Karl R. (1968): *The Logic of Scientific Discovery*. Nueva York, Harper and Row [ed. castellana: *Lógica de la investigación científica*. Madrid: Tecnos, 1985].

### Bibliografía

- (1982): «The Open Universe: An Argument for Indeterminism». En W. W. Bartley III (ed.), *The Postscript to the Logic of Scientific Discovery*. Totowa, N. J., Rowman and Littlefield.
- Porter, Michael E. (1990): *The Competitive Advantage of Nations*. Nueva York, Free Press [ed. castellana: *Ventaja competitiva de las naciones*. Barcelona: Plaza y Janés, 1991].
- Przeworski, Adam, y Henry Teune (1982): *The Logic of Comparative Social Inquiry*. Malabar, Florida, Krieger Publishing Company.
- Psathas, George (1968): «Ethnomethods and Phenomenology». *Social Research* 35, pp. 500-520.
- Putnam, Robert D., con Robert Leonardi y Raffaella Y. Nanetti (1993): *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, Princeton University Press.
- Ragin, Charles C. (1987): *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Berkeley, University of California Press.
- Rivers, Douglas, y Morris P. Fiorina (1989): «Constituency Service, Reputation, and the Incumbency Advantage». En Morris P. Fiorina y David Rohde (eds.), *Home Style and Washington Work*. Ann Arbor, University of Michigan Press.
- Robinson, William S. (1990): «Ecological Correlations and the Behavior of Individuals». *American Sociological Review* 15, pp. 351-357.
- Rogowski, Ronald (1987): «Trade and the Variety of Democratic Institutions». *International Organization* 41, núm. 2 (primavera), pp. 203-224.
- Rosenau, Pauline (1990): «Once Again into the Fray: International Relations Confronts the Humanities». *Millennium: Journal of International Studies* 19, núm. 1 (primavera), pp. 83-110.
- Rosenstone, Steven R. (1983): *Forecasting Presidential Elections*. New Haven, Yale University Press.
- Roth, Alvin E. (1988): «Laboratory Experimentation in Economics: A Methodological Overview». *The Economics Journal* 98 (diciembre), pp. 974-1031.
- Rubin, Donald B. (1974): «Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies». *Journal of Educational Psychology* 66, pp. 688-701.
- (1978): «Bayesian Inference for Causal Effects: The Role of Randomization». *The Annals of Statistics* 6, pp. 34-58.
- Russett, Bruce (1978): «The Marginal Utility of Income Transfers to the Third World». *International Organization* 32, núm. 4, pp. 913-928.
- Sanday, Peggy Reeves (1983): «The Ethnographic Paradigm(s)». En John Van Maanen (ed.), *Qualitative Methodology*. Sage, Beverly Hills.
- Schumpeter, Joseph A. [1936] (1991): «Can Capitalism Survive?». En Richard Swedberg (ed.), *The Economics of Sociology and Capitalism*. Princeton, Princeton University Press.
- Shepsle, Kenneth A. (1986): «Institutional Equilibrium and Equilibrium Institutions». En Herbert F. Weisberg (ed.), *Political Science: The Science of Politics*. Nueva York, Agathon Press.
- Shively, W. Phillips (1990): *The Craft of Political Research*. 3.ª ed., Englewood Cliffs, Prentice-Hall.
- Simon, Herbert A. (1985): «Human Nature in Politics: The Dialogue of Psychology with Political Science». *American Political Science Review* 79, núm. 2 (junio), pp. 293-305.
- Skocpol, Theda (1979): *States and Social Revolutions*. Cambridge University Press.

- (1984): «Emerging Agendas and Recurrent Strategies in Historical Sociology». En Theda Skocpol (ed.), *Vision and Method in Historical Sociology*. Nueva York, Cambridge University Press [ed. castellana: *Los estudios y las revoluciones sociales: un estudio comparativo de Francia, Rusia y China*. México: Fondo de Cultura Económica, 1984].
- Snyder, Glenn H., y Paul Diesing (1977): *Conflict among Nations: Bargaining, Decision Making and System Structure in International Crises*. Princeton, Princeton University Press.
- Snyder, Jack (1991): *Myths of Empire: Domestic Politics and International Ambition*. Ithaca, Cornell University Press.
- Sober, Elliot (1988): *Reconstructing the Past: Parsimony, Evolution and Inference*. Cambridge, MIT Press.
- Suppes, Patrick C. (1970): *A Probabilistic Theory of Causality*. Amsterdam, North-Holland.
- Tawney, R. H. (1935): *Religion and the Rise of Capitalism*. Nueva York, Harcourt, Brace & Co.
- Tilly, Charles (ed.) (1975): *The Formation of National States in Western Europe*. Princeton, Princeton University Press.
- Verba, Sidney (1967): «Some Dilemmas of Political Research». *World Politics* 20 (octubre), pp. 111-128.
- , Kay L. Schlozman, Henry Brady y Norman Nie (1993): «Race, Ethnicity, and Political Resources: Participation in the United States». *British Journal of Political Science* 23, pp. 453-497.
- Waltz, Kenneth N. (1979): *Theory of International Politics*. Reading, Mass., Addison-Wesley.
- Webb, Eugene J., D. T. Campbell, R. D. Schwartz y L. Sechrest (1966): *Unobtrusive Measures*. Chicago, Rand McNally.
- , y Karl E. Weick (1983): «Unobtrusive Measures in Organizational Theory: A Reminder». En John Van Maanen (ed.), *Qualitative Methodology*. Sage, Beverly Hills.
- Weber, Max [1905] (1949): «Critical Studies in the Logic of the Cultural Sciences». En Max Weber (ed.), *The Methodology of the Social Sciences*. Traducido y editado por Edward A. Shils y Henry A. Finch. Nueva York, Free Press.
- Weiner, Myron (1991): *The Child and the State in India*. Princeton, Princeton University Press.
- Wendt, Alexander (1992): «Anarchy is What States Make of It: The Social Construction of Power Politics». *International Organization* 64, núm. 2 (primavera), pp. 391-426.
- Wolfinger, Raymond, y Steven Rosenstone (1980): *Who Votes*. New Haven, Yale University Press.
- Woods, John, y Douglas Walton (1982): *Argument: The Logic of the Fallacies*. Nueva York, McGraw-Hill Ryerson Ltd.
- Zelditch, Morris Jr. (1971): «Intelligible Comparisons». En Ivan Vallier (ed.), *Comparative Methods in Sociology*. Berkeley y Los Ángeles, University of California Press.
- Zellner, Arnold (1971): *An Introduction to Bayesian Inference in Econometrics*. Nueva York, Wiley.
- (1984): *Basic Issues in Econometrics*. Chicago, University of Chicago Press.

## Índice analítico

- Achen, Christopher, 140, 145, 255  
acontecimientos únicos, estudio de los, 21-22, 53-55, 104-5  
Alderson, Arthur, 211, 252 n. 2, 255  
Alemania de Weimar, 201-2  
álgebra booleana, enfoque del, 248 n. 11  
alianza, 210  
Álvarez, Luis, 22  
Álvarez, Walter, 22, 255  
análisis bivariante, 184  
análisis contrafáctico, 21, 89-90, 99-100, 203  
análisis de regresión, 109, 140-42, 178-79  
análisis racional, 253 n. 13  
Anderson, Richard, 256  
Anselin, Luc, 253 n. 11, 255  
aportación a los escritos científicos, 27-28  
área de las relaciones internacionales, necesidad de describir en el, 55-56  
armas nucleares, 158-59  
Asaro, Frank, 255  
asignación aleatoria de valores, 105, 126, 207-11, 242  
asignación de valores, 105, 125-26, 207, 242  
aumento de las observaciones, 230-42  
Australia, 217  
autocorrelación espacial, 253 n. 11  
autocorrelación, 253 n. 11  
Bahr, Howard, 255  
Barnett, Vic, 248 n. 13, 255  
Baumol, William, 253 n. 7, 255  
Beatty, John, 257  
Beck, Nathaniel, 253 n. 11, 255  
Becker, Howard, 255  
Blainey, Geoffrey, 255  
Bollen, Kenneth, 211, 252 n. 2, 255  
Cain, Bruce, 197, 255  
Campbell, D. T., 262  
Canadá, 217  
Caplow, Theodore, 37, 255  
carnes rojas en Japón, 43  
Caro, Robert, 47  
casos, 63-64, 127-28, 221-25  
causalidad asimétrica, 100-1  
causalidad múltiple, 98-100

## El diseño de la investigación social

- causalidad probabilística, 246 n. 2  
causalidad simétrica, 100-1  
causalidad, 88-101  
Chadwick, Bruce, 255  
Chile, 232  
ciencia como labor social, 19-20  
ciencias naturales, 21  
Cisjordania, 67-70, 106  
codificación de entrevistas, 244 n. 11  
codificación, 168  
coherencia interna, 107, 112, 116-18  
colinealidad, 227-29  
*Véase también* multicolinealidad  
Collier, David, 256  
complejidad de los acontecimientos, 20-21, 54, 104-5  
componente estocástico, 66-74, 91-93, 95, 247 n. 9  
componente no sistemático, 66-74, 91-93, 95, 247 n. 9  
componente sistemático, 45-46, 55-56, 66-74, 91-93, 95, 247 n. 9  
*Véase también* componente estocástico  
comprobación de un único caso, 221-25  
*Véase también* casos; estudios de caso; observaciones  
concisión, 30, 40, 115-16  
concretar, 122  
consecuencias observables, 39, 52, 107, 120-23, 190-91, 236-37  
subunidades de las, 234  
constantes, selección de, 157-59  
control de calidad, 245 n. 12  
control, 134, 178, 197, 207-18  
de lo que se investiga, 211-18  
y problemas del *n* pequeño, 207-11  
Cook, Karen, 112, 256  
Coombs, Clyde, 245 n. 4, 256  
cooperación entre las superpotencias, 189  
cooperación internacional, 15  
Courtillot, Vincent, 256  
creencias religiosas, 217  
crisis agudas, 145  
cristianismo en Nigeria, 158, 217  
cultura, 120, 239  
política, 203, 239  
*Véase también* inferencia descriptiva
- cumbre soviética, 72-73  
cursos de contabilidad, 141-42  
Dahl, Robert, 256  
Daston, Lorraine, 257  
datos, recogida de, 62-63  
cuantificación en la, 55  
directrices para mejorarla, 33-36, 57-58  
fiabilidad como objetivo, 36  
garantizar la fiabilidad en la, 36, 162  
nivel de agregación, 60, 128  
para compensar la variabilidad fundamental, 227  
para maximizar el control, 216-17  
proyectos piloto antes de la, 33  
registrar el proceso de, 33-34  
reproducción como objetivo, 36-37  
utilización de modelos formales para la, 62-64, 116-17  
y eficiencia, 227  
y el problema del *n* pequeño, 226-30  
datos, utilización de los, 37-38  
al resumir, 64-65  
con error de medida, 176-77  
descartar, 165, 195  
para aumentar el control, 41, 115-16, 133, 216-17  
para aumentar el número de observaciones, 230-32  
para comprobar teorías, 24, 30-33, 39-41, 57-60, 108-16  
para garantizar la falta de sesgo, 37-38, 74-76, 106, 110  
para maximizar la eficiencia, 38, 76-85, 106-7, 111  
delincuencia, 171-73, 176-77  
demonstración, 210  
descripción detallada, 48-53, 54  
desempleo, consecuencias sobre la delincuencia, 171-73, 176-77  
desfase entre concepto e indicador, 121-22  
Dessler, David, 238, 256  
detalles de los hechos, 48-54  
*Véase también* inferencia descriptiva

## índice analítico

- determinista, punto de vista, 70-71, 224-25, 248 n. 11  
Dewald, William, 36, 256  
Diamond, Larry, 256  
Diesing, Paul, 262  
dinosaurios, extinción de los, 21-22, 53, 236  
diseño de investigación impreciso, 128-34, 156  
diseño de la investigación, 24, 28, 128-34, 144, 185, 226-30, 241  
disusión, teoría de la, 34-35  
Duneier, Mitchell, 48, 256  
Easton, David, 124, 256  
Eckstein, Harry, 48, 63-64, 124, 222-24, 231, 233 n. 14, 256  
educación obligatoria en la India, 76  
efecto causal aleatorio, 91-93  
efecto causal realizado, 90-92, 95  
efecto causal, 92-93, 95-96, 99-100  
modelo formal del, 108-9  
sobreestimarlo, 149  
efecto renta/ingresos, 184-85  
eficiencia, 38, 76-85, 106-7, 111, 161-62, 192-96  
comparada con el sesgo, 80-85  
modelo formal de, 81, 111  
y error de medida, 170, 227  
y recogida de datos, 227, 242  
Einstein, Albert, 17  
elecciones presidenciales, 84  
élites de la política exterior, 135-36  
Elster, Jon, 246-47 n. 4, 256  
empaparse y hurgar, 48-53, 54  
encuestas, 16, 42  
endogeneidad, 105, 118, 197-206, 241, 246 n. 13  
como problema de la variable omitida, 201-2  
como producto natural de los procesos políticos, 210  
modelo formal de, 208  
y elección de observaciones, 202-5  
enfoque comparado, 226  
enfoque globalizador (incluirlo todo), 194  
entrevistas, 249 n. 16, 252 n. 13  
Entwistle, Barbara, 211, 252 n. 2, 255  
equifinalidad, 98  
error aleatorio, 66-74, 91-93, 169-78, 180-81, 95  
error de agrupación, 165  
error de medida, 162-78, 180-81  
modelos formales de, 174-75, 178, 180-81  
no sistemático, 169-78, 180-81  
sistemático, 167-69  
y los estudios de un solo caso, 224  
error medio cuadrático (EMC), 84-85  
errores de predicción, 141  
escrbar en busca de datos, 185  
especificación, 198  
estadísticos, definición de los, 64-65  
estados europeos, 146-47  
Estados Unidos, 217  
estimadores, 194  
estrategias en campañas electorales, 113-15  
estudio de control de caso, 152  
estudios de caso cruciales, 222-25  
estudios de caso, utilización de, 16, 55-57, 63-64, 71-72, 78-79  
cruciales, 221-25  
de la República de Weimar, 201-2  
de un solo caso, 221-25, 253 n. 14  
medidas múltiples a lo largo del tiempo, 234  
y efectos causales constantes, 104  
y homogeneidad de las unidades, 104  
y mecanismos causales, 96-98, 238  
y sesgo de selección, 126  
estudios,  
efecto sobre la renta, 99-100  
en la India, 76  
*Ética protestante y el espíritu del capitalismo*, 198.  
Europa del Este, 137-38  
Europa occidental, 146-47  
exclusión de variables relevantes, 72-73, 100, 105, 118, 133-34, 178-93  
modelo formal de, 179-83, 186-87  
y asignación aleatoria de valores, 207  
y endogeneidad, 201-2

- y estudios de un solo caso, 223-36  
y selección intencionada, 214-15  
explicación de los fundamentos, 253 n. 13  
explicación dentro de una observación, 253 n. 14  
explicación, 87-88
- falacia ecológica, 40, 244 n. 12  
falsabilidad, 30, 108-10, 112-16, 241  
falta de sesgo, véase sesgo  
Fearon, James, 256  
Fenno, Richard, 49, 256  
Ferejohn, John, 48, 255, 256  
Ferguson, Yale, 256  
fertilizantes, uso de, 158  
Feynman, Richard, 256  
Fiorina, Morris, 197, 249 n. 3, 256  
Fisher, Ronald, 255, 256, 261  
Fogel, Robert, 257  
Friedrich, Carl, 257  
Fudenberg, Drew, 35, 257  
fuentes históricas, 146
- Garfinkel, H., 257  
Geddes, Barbara, 142, 257  
Geertz, Clifford, 48, 50-51, 257  
Gelman, Andrew, 113, 246 n. 3, 257  
generalización, 21-22, 46-47, 53-54,  
57-60, 104-5, 241  
Véase también inferencia causal; inferencia descriptiva  
George, Alexander, 56-57, 98, 178,  
235, 239, 253 n. 14, 257  
Gigerenzer, Gerd, 257  
Gilpin, Robert, 257  
Goldberger, Arthur, 257  
Goldstein, Judith, 48, 202, 257  
Gould, Stephen J., 21, 257  
Granger, G. W. J., 253 n. 11, 257  
Greenhouse, Joel, 249 n. 15  
guerra, 20  
guiñar y parpadear, 50-51  
Gulick, Edward, 163, 258
- hacha, ejemplo del, 17  
Hall, Peter, 202, 258  
Halpern, Nina, 203-4, 258

- Hermens, F. A., 201, 258  
heurística de la disponibilidad, 252 n. 3  
hipótesis causal, 51  
hipótesis, 29
- aplicación a la recogida y análisis de datos, 23, 29-30, 39-40, 56, 57-60, 185, 203-4, 239-41
  - comprobarlas con datos, 23, 30-33, 39-41, 57-60, 113-16
  - concisión como objetivo de las, 30
  - control de las, 40, 115
  - especificidad como objetivo de las, 30
  - hacerla más restrictiva, 31-32, 241
  - papel de las, 20-21
  - proyectos piloto al comprobarlas, 33
  - reglas para construirlas, 107-24
  - su falsabilidad, 30
  - utilización de modelos formales para evaluarlas, 116-17
  - y aumento de observaciones, 231, 234
  - y endogeneidad, 198
- Hirschman, Albert O., 21, 258  
historia económica, 253 n. 7  
Hoffman, Stanley, 258  
Holland, Paul, 93, 103, 246 n. 2, 247 n. 6, 258
- homocedasticidad, 250 n. 4  
homogeneidad de las unidades, 102-5, 126-27, 241
- Hoover, Dwight, 255  
Horowitz, Donald, 94, 258  
Hsiao, C., 249 n. 1, 258  
Huth, Paul, 34, 258
- ideas como variable explicativa, 203  
identificación, 249 n. 1  
idiomas,
- conocimiento de, 137
- incertidumbre, 170, 226-29
- en la investigación científica, 19, 93
  - informar de su existencia, 42, 88, 106, 163
- inclusión de variables irrelevantes, 193-96
- modelo formal de, 196
- independencia condicional, 102, 105-6, 125-26
- India, 155-56, 217, 232, 239
- educación en la, 76
  - y su papel en las políticas contra la pobreza, 155-56
- inferencia causal, 18, 29, 42-43, 87-124, 126-27
- con error de medida, 167
  - corrección del sesgo en la, 199-200
  - criterios para juzgarlas, 106-7, 111
  - su incertidumbre, 227-29
  - supuestos de la, 102-7
  - y los estudios de caso, 56
- Véase también inferencia descriptiva; inferencia,
- inferencia descriptiva, 18, 26, 29, 45-57, 64, 66-85
- con error de medida, 167
  - con sesgo de selección, 151
  - criterios para evaluarla, 74-85
  - e interpretación, 48-53, 54
  - importancia en la ciencia política, 55-56
- inferencia, 57-60, 241-42
- con error de medida, 177
  - corrección del sesgo en la, 199-200
  - de las relaciones sistemáticas, 45-46, 91-93, 95, 247 n. 9
  - eficiencia como criterio de la, 38, 76-85, 106-7, 161-62
  - en la investigación científica, 18
  - falta de sesgo como criterio de la, 38, 74-76, 106-7, 110, 161-62
- incertidumbre de la, 19, 88, 93, 106, 163, 170
- utilización de reglas en la, 16-17, 19, 88
- Italia, 16, 236
- política en, 16, 236
- Iyengar, Satis, 249 n. 15 cap. 3, 249 n. 3 cap. 4, 258
- Jeffreys, Harold, 30, 258
- Jervis, Robert, 258

## El diseño de la investigación social

- Johnston, J., 258  
 Jones, E. L., 54, 258
- Kahneman, Daniel, 253 n. 3, 258  
 Katzenstein, Peter, 213, 258  
 Keane, Mark, 253 n. 3, 258  
 Kennedy, Paul, 259  
 Keohane, Robert O., 48, 202, 257, 259  
 Khong, Yuen Foong, 225, 259  
 Kinder, Donald, 249 n. 3, 258  
 King, Gary, 42, 61, 113, 140, 200, 245  
     n. 11, 246 n. 3, 249 n. 1, 253 n. 11,  
     257, 259  
 Kohli, Atul, 155-56, 217, 232-33, 259  
 Kreps, David, 259  
 Kruger, Lorenz, 257
- Laitin, David, 157-58, 198, 217, 259  
 Lakatos, Imre, 22, 248 n. 13, 259  
 Lakeman, Enid, 202, 259  
 Lambert, James, 202, 259  
 Leamer, Edward, 248, 259  
 Lebow, Richard, 258  
 Leonardi, Robert, 261  
 Levi, Margaret, 112, 256  
 Levine, Michael, 249 n. 3  
 Levy, Jack, 26, 259-60  
 Lieberson, Stanley, 29, 40, 100-1, 119,  
     260  
 Lijphart, Arend, 94, 216, 252 n. 6, 260  
 Lindberg, Leon, 168, 260  
 Linz, Juan, 94, 260  
 Lipset, Seymour Martin, 217, 260  
 Little, Daniel, 96, 247 n. 10, 260  
 lógica del análisis, 16-17, 19, 88, 241  
 Longino, Helen, 260  
 Lowenthal, Abraham, 260  
 Lynd, Robert y Helen, 37
- manchas de tinta, 31  
 Mankiw, N. Gregory, 28, 260  
 Mansbach, Richard, 256  
 Martin, Lisa L., 15, 260  
 máximo muestral, 65  
 máximo, 65  
 May, Ernest, 42  
 McKeown, Timothy, 56-57, 239, 253 n.  
     14, 257
- mecanismos causales, 96-98, 238  
 media muestral, 65, 77, 92-93  
 media, 65, 77, 92-93  
 media, 65, 77, 92-93  
 medición de variables, 163-65  
 medidas con intervalos, 163-65  
 medidas nominales, 163-65  
 medidas ordinales, 163-66  
 Merck & Co., 260  
 Merton, Robert, 18, 124, 260  
 método histórico, 48-53, 54  
     *Véase también* mecanismos causales; inferencia descriptiva; investigación cualitativa  
 método,  
     de la diferencia, 178  
     de la similitud, 178  
     del compromiso, 144  
     importancia esencial para la investigación científica, 19  
 Middletown, 37  
 Mill, John Stuart, 98, 144, 178, 252  
     n. 2, 260  
 Miller, David, 112, 260  
 Milner, Helen, 190-93, 260  
 modelo de avión, 60-61  
 modelos formales, 60-64, 219  
     aplicados a la recogida de datos, 62-  
     64, 116-17  
     de eficiencia, 81, 111  
     de endogeneidad, 208  
     de ineficiencias de la variable incluida, 196  
     de investigación cualitativa, 136  
     de la falta de sesgo, 77, 107, 110  
     de multicolinealidad, 134, 135  
     del efecto causal medio, 107, 108-9  
     del error de medida, 174-75, 178,  
         180-81  
     del número de unidades, 226-27  
     del problema del *n* pequeño, 130,  
         132, 227-28  
     del sesgo de la variable omitida, 182-83, 186-88  
     relevancia para la investigación cualitativa, 61  
     su utilidad para las teorías, 116-17  
 Moe, Terry, 260

## Índice analítico

- moneda y pluma, 119  
 Moon, Donald, 48, 253 n. 13, 260  
 motivaciones personales para investigar, 25  
 muestreo por orden alfabético, 148  
 Mugabe, Robert, 232  
 multicolinealidad, 131, 133-34, 135,  
     158, 252 n. 4  
     colinealidad relativa, 227-29  
     modelo formal de la, 134-35  
 nacionalsocialistas, 201-2  
 Nanetti, Raffaella, 261  
 Neustadt, Richard, 42, 260  
 Newbold, P., 253, 257  
 Nigeria, 158, 217  
 notación matemática, 64-65, 68-70  
 Nye, Joseph, 168, 259, 261
- O'Hear, Anthony, 113, 261  
 observaciones equiparables, 106, 212-  
     18, 225-26  
 observaciones más probables, 223  
 observaciones menos probables, 223  
 observaciones, 62-63, 68, 127-28, 221-  
     25  
     crear muchas a partir de pocas, 230-  
     42  
     dependencia entre las, 235  
         y reducción de la endogeneidad,  
         202-5  
 Ordeshook, Peter, 116, 261  
 organizaciones internacionales, 63
- Palfrey, Thomas, 249 n. 3, 261  
 ¿parecerse a qué?, 216  
 parpadear y guñar, 50-51  
 participación voluntaria, 205-6  
 Pearson, Karl, 19, 261  
 persona equiparable, 212  
 Plattner, Marc, 256  
 Plott, Charles, 249 n. 3, 256, 261  
 política comparada, área de la  
     importancia de la descripción en,  
         55-56  
 Popper, Karl, 24, 110, 112-14, 244 n. 5,  
     245 n. 7, 248 n. 13 y 15, 261  
 Porter, Michael, 144-45, 261
- Porter, Theodore, 257  
 postulado de sencillez de Jeffreys-Wrinch, 30  
 postulado de sencillez, 244 n. 7  
 precios del petróleo, 185  
 precisar los componentes,  
     de la variable dependiente, 200  
     de la variable explicativa, 205-6  
 preguntas de la investigación, 24-29  
     criterios para elegirlas, 25  
     su contribución a los escritos científicos, 27-28  
     y estructuración de los estudios de caso, 56-57  
 problema de tener más inferencias que observaciones, 129-31, 132, 136,  
     155, 207, 221-42  
     modelo formal del, 132  
 número de observaciones que se necesitan para superarlo, 226-30  
     y diseño de la investigación, 226-42  
     y equiparación, 216-18  
     y subunidades, 233-34  
     y teoría de las ciencias sociales, 240  
 problema de tener más parámetros que unidades, *véase* problema de tener más inferencias que observaciones  
 problema del archivador, 248 n. 15  
 problema del *n* pequeño, 130, 132, 136,  
     155-56, 207, 221-42  
     modelo formal del, 130, 132, 227-28  
 número de observaciones necesario para superarlo, 226-30  
     y diseño de la investigación, 226-42  
     y equiparación, 216-18  
     y subunidades, 233-34  
     y teoría de las ciencias sociales, 240  
 problema fundamental de la inferencia causal, 90-91, 93, 102, 105, 134,  
     212, 221-23  
 procedimiento de la coherencia, 56-57,  
     251 n. 9  
 pronóstico, 251 n. 7  
 protecciónismo, 190-93  
 proyectos piloto, 33  
 Przeworski, Adam, 47, 178, 213, 216,  
     261  
 Psathas, George, 52, 261

público, 66  
 punto de vista probabilístico, 70  
 Putnam, Robert D., 16, 49, 236, 261

Ragin, Charles, 98, 248 n. 11, 255, 261  
 rastreo de procesos, 97-98, 238-41  
 razonamiento analógico, 225-26  
 recogida de datos, 62-63  
   cuantificación en la, 55  
   directrices para mejorarla, 33-36, 57-58  
   fiabilidad como objetivo, 36  
   garantizar la fiabilidad en la, 36, 162  
   nivel de agregación, 60, 128  
   para compensar la variabilidad fundamental, 227  
   para maximizar el control, 216-17  
   proyectos piloto antes de la, 33  
   registrar el proceso de, 33-34  
   reproducción como objetivo, 36-37  
   utilización de modelos formales para la, 62-64, 116-17  
   y eficiencia, 227, 242  
   y el problema del  $n$  pequeño, 226-30  
 régimen nazi, 201-2  
 reglas de la inferencia, 16-17, 19, 88  
 representación proporcional, 117-18, 164, 200, 201-2  
 reproducción, como objetivo de la recogida de datos, 36-37  
*Resisting Protectionism*, 190-93  
 resumir los datos, 64-66  
   estadísticos para, 64-65  
 reuniones en la cumbre, 189  
 revoluciones sociales europeas, 139-40  
 revoluciones, 21  
 Rivers, Douglas, 261  
 Robinson, William, 244 n. 12, 261  
 Rogowski, Ronald, 261  
 Rosenau, Pauline, 261  
 Rosenstone, Steven, 261, 263  
 Rossi, Peter, 157, 258  
 Roth, Alvin, 249 n. 3, 261  
 Rubin, Donald, 246 n. 2, 261  
 Russett, Bruce, 258, 261  
 Ryle, Gilbert, 50

sanciones económicas, 15  
 sanciones internacionales, 15  
 Sanday, Peggy, 262  
 Scheingold, Stuart, 168, 260  
 Schumpeter, Joseph, 17, 262  
 Schwartz, R. D., 262  
 Sechrest, L., 262  
 Segunda Guerra Mundial, 127  
 selección aleatoria, 105, 126, 242  
   e investigación con  $n$  pequeño, 136, 155-56, 207  
   sus límites, 134-38  
 selección,  
   de variables constantes, 157-59  
   e inferencia descriptiva, 152  
   en función de la variable causal clave, 157-58  
   en función de la variable dependiente, 139-47, 151-59  
   en función de la variable independiente, 147-49, 150-59  
   intencionada, 150-59, 214  
   y sesgo de la variable omitida, 214  
*Véase también* selección aleatoria  
 ser explícito, como objetivo de la investigación científica, 18-19  
 series temporales, 253 n. 11  
 sesgo de la variable omitida, 72-73, 100, 105, 118, 134, 178-93  
 modelo formal del, 182-83, 186-88  
 y asignación aleatoria de valores, 207  
   y endogeneidad, 201-2  
   y estudios de un solo caso, 223-35  
   y selección intencionada, 214-15  
 sesgo de selección, 105, 119, 127, 136-49  
   e inferencia descriptiva, 152  
   hacer ajustes para combatirlo, 142-44, 146-47  
   inducido por el mundo, 145-47  
   y fuentes históricas, 145-47  
*Véase también* selección aleatoria  
 sesgo, 38, 74-76, 106-7, 161-62, 207  
   comparado con la eficiencia, 80-85  
   ejemplo formal de, 77, 110  
   y endogeneidad, 206  
   y error de medida, 173-78  
   y sesgo sustancial, 75

Shepsle, Kenneth, 262  
 Shively, W. Phillips, 26, 262  
 Simon, Herbert, 262  
 simplificación, 20-21, 46-47, 53-55, 57-60, 104-5, 241  
 sistemas electorales, sesgo en los, 75  
 Skocpol, Theda, 15, 139-40, 262  
 Slovic, Paul, 252, 258  
 Smoke, Richard, 257  
 Snidal, Duncan, 145, 255  
 Snyder, Glenn, 151, 262  
 Snyder, Jack, 262  
 Sober, Elliot, 30, 262  
 Somalia, 157  
 Stein, Janice, 258  
 subunidades, 233-34  
 Suppes, Patrick, 246 n. 2, 262  
 supuesto del efecto constante, 103-4, 126  
 Swijtink, Zeno, 257

Tawney, R. H., 198, 262  
 temas que investigar, 24-29  
   criterios para elegirlos, 26  
   su aportación a los escritos científicos, 27-28  
   y estructuración de los estudios de caso, 56-57  
 teoría de la variable oculta, 224-25, 245 n. 12, 248 n. 11  
 teoría en las ciencias sociales, 29-30  
   aplicación a la recogida y análisis de datos, 23-24, 29-31, 39, 57-60, 185, 240-41  
 comprobarla con datos, 23-24, 30-33, 39-41, 57-60, 112-16  
 concisión como objetivo de la, 30  
 control de la, 40, 115  
   e incremento de observaciones, 231-234  
 especificidad como objetivo de la, 30  
 hacerla más restrictiva, 31-33, 241  
 proyectos piloto al comprobarla, 33  
 reglas para elaborarla, 107-24  
 su falsabilidad, 30-31  
 su función, 21  
   utilización de modelos formales para evaluarla, 116-17

y endogeneidad, 198  
 teorías causales, 107-8  
   reglas para construirlas, 107-24  
 Teune, Henry, 47, 178, 213, 216, 261  
 Thursby, Jerry, 256  
 Tilly, Charles, 146-47, 262  
 Tirole, Jean, 35, 257  
 tradiciones musulmanas en Nigeria, 158, 217  
 Tversky, Amos, 252 n. 3, 258

unidad de la ciencia, 19  
 unidad de observación, 62-63, 68, 88-89  
   y subunidades para aumentar las observaciones, 233-34  
 utilización de subunidades para aumentar las observaciones, 233-34

validez, 35  
 validez, construir la, 35  
 valores atípicos, 68  
 valores esperados, 69  
 variabilidad fundamental, 70, 224, 227-28, 248 n. 11  
 variable causal clave, 89  
   seleccionar en función de la, 157  
 variables aleatorias, 62-63, 68-69, 91, 247 n. 7  
 variables de control, 89  
 variables de resultado, 89, 118-20  
   con error de medida, 169-75  
   precisar sus componentes, 200  
   seleccionar en función de las, 139-47, 151-56  
   su variabilidad fundamental, 227-28  
 variación de las, 139-40, 144  
 variables dependientes, 89, 118-20  
   con error de medida, 169-75  
   precisar sus componentes, 200  
   seleccionar a partir de las, 139-47, 151-56  
   su variabilidad fundamental, 227-28  
 variación en las, 139-40, 144  
 variables explicativas, 89, 133-34  
   asignación de valores para las, 207-11  
   con error de medida, 169-70, 173-78, 180-81

El diseño de la investigación social

- precisar sus componentes, 205-6
- seleccionar en función de, 147-49, 150-59
- variables independientes, 89, 133
  - con error de medida, 169, 173-78, 180-81
- precisar sus componentes, 205-6
- seleccionar en función de las, 147-49, 150-60
- variables irrelevantes, 193-96
  - modelo formal para las, 196
- variables relevantes, 72-73, 100, 105, 118, 133-34, 178-93
  - modelo formal de las, 182-83, 186-88
  - y asignación aleatoria de valores, 207
  - y endogeneidad, 201-2
  - y estudios de un solo caso, 223-36
  - y selección intencionada, 214-15
- variables, 62-64, 68, 91, 247 n. 7
- variación aleatoria, 70-71
- varianza, 69-70, 92, 227-30
  - de la variable causal clave, 228-30
- de los efectos causales, 93, 95-96, 111
- ventaja competitiva, 144
- ventaja electoral del que ya está en el poder, 89-93
- Verba, Sidney, 56, 178, 205-6, 233, 262
- verstehen*, 48-49, 52
- veto, procedimientos de, 165-66
- voto en Gran Bretaña, 66-69
- Walton, Douglas, 29, 263
- Waltz, Kenneth, 163, 262
- Webb, Eugene, 262
- Weber, Max, 21, 198, 262
- Weick, Karl, 262
- Weiner, Myron, 75, 239, 262
- Wendt, Alexander, 263
- Wolfsinger, Raymond, 263
- Woods, John, 29, 263
- Zelditch, Morris, 157, 263
- Zellner, Arnold, 30, 248 n. 13, 263
- Zimbabue, 232