

Estadística y causalidad en la sociología empírica del XX

Ángeles Lizón

Universitat Autònoma de Barcelona. Departament de Sociologia
angeles.lizon@uab.es

Resumen

Las ideas que se trabajan en este artículo tienen que ver con el desarrollo de la teoría de la probabilidad y con lo que, por vía estadística, se ha llegado a establecer acerca de la causalidad dentro de algunas tradiciones reconocibles de la sociología empírica del siglo XX. Para ello, se hará una reconstrucción —esquemática aunque con pretensión informativa acerca de las distintas escuelas— de los capítulos más señalados en el desarrollo de la sociología estadística institucional. Se arrancará, así, de su inicial despegue en el ámbito estadounidense a comienzos de siglo, hasta llegar a bosquejar las expectativas actuales de la sociología estadística, dado el impacto de su creciente automatización y las alternativas que tales desarrollos pueden ofrecer a una posible conjunción de investigación y teoría.

Palabras clave: causalidad, sociología estadística, mecanismos sociales, teoría sociológica.

Abstract. *Statistics, causality and empirical sociology in the xxth century*

The ideas here presented deal with the development of probability theory and the statistical claims on causality made by some identifiable traditions in XXth century empirical sociology. With this view, the most relevant chapters in the development of institutional statistical sociology will be reconstructed in a summarized way, paying attention to the different schools. This reconstruction will start from the departure of this type of sociology in the United States at the beginning of the XXth century, and will outline the present prospects of statistical sociology given the impact of its growing automation; it will also consider the chances that these developments may offer for a possible link between theory and research.

Key words: causality, statistical sociology, social mechanisms, sociological theory.

Sumario

Matemática estadística y ciencia social

Un primer apunte sobre la sociología fundacional

Primera etapa: la estadística descriptiva heredada de la escuela histórica alemana

Segunda etapa: la influencia de la estadística inglesa

Tercera etapa: los modelos causales de regresión

El nuevo sesgo estadístico: automatización sí, causalidad no

Una propuesta alternativa: estructuras estadísticas y mecanismos causales

Referencias bibliográficas

El concepto de causa y su estatuto en la ciencia social ha preocupado tradicionalmente a los sociólogos, aunque sólo sea porque la mayoría de las preguntas de investigación que se plantean en sus estudios aplicados son de naturaleza causal: ¿qué impacto tiene la inmigración ilegal en la tasa de inseguridad ciudadana?; ¿por qué el gran público acepta las nuevas tecnologías menos en países altamente industrializados que en aquellos con menor desarrollo del sector secundario?; aún en sistemas en los que la educación es de libre acceso, ¿por qué se da una representación tan desigual en los programas de estudios superiores según rendimiento escolar y estrato social?, son sólo algunos de esos problemas candentes. Una respuesta general a la cuestión, seguramente parcial pero inteligente, la sugiere Simon (1968: 355), cuando afirma que el éxito en la intervención de las políticas sociales y la subsiguiente credibilidad de la ciencia social, dependen del conocimiento que se tenga sobre qué variable o variables intervienen sobre qué otras, dependen de nuestro dominio del tema de la causalidad. En efecto, no se puede pretender intervenir en la realidad sin comprender los mecanismos que llevan a cambios con uno u otro resultado y las condiciones bajo las cuales ocurren dichos cambios. Así, pues, si toda intervención requiere algún tipo de razonamiento causal, con seguridad la sociología tendrá intereses prácticos y teóricos para intentar dar respuestas a los «porqués» de la compleja vida social.

Matemática estadística y ciencia social

La penetración de la matemática estadística en cada una de las disciplinas no experimentales ha tenido su historia, habiendo retenido cada una de ellas sus características y matices distintivos en función de su propia trayectoria. Dentro del conjunto de las ciencias aplicadas, la biométrica fue la primera que pasó a constituir una suerte de laboratorio en el que la preocupación por el control estadístico de los datos eugenésicos y epidemiológicos daría origen a las mejores iniciativas de análisis. K. Pearson (1897), G. U. Yule (1899) y S. Wright (1921), entre las grandes figuras de la estadística moderna, todos estuvieron asociados a *Biometrika*, la publicación más emblemática del gremio.

A su lado, otro ámbito de desarrollo excepcional fue el de la econometría, algo que seguramente tuvo mucho que ver con el interés inicial del propio G. U. Yule en la investigación socioeconómica. En efecto, en «An investigation into the causes of changes in pauperism in England, chiefly during the last two intercensal decades», Yule se mete de lleno en los aspectos socioeconómicos de la estadística, facilitando claramente el tránsito de esas técnicas al trabajo econométrico inicial. Así, para 1933 el área contaba con *Econometrika*, un órgano de difusión de las ideas de H. Hotelling y T. Koopmans y otros destacados econométricos, quienes a través de esta publicación llegaron a ejercer una influencia decisiva en el desarrollo de modelos matemáticos útiles al conjunto de la ciencia social.

Por lo que toca a la psicometría, sobre todo en el campo de la educación, se vio directamente beneficiada por la influencia de Charles Spearman (un discípulo de Wilhem Wundt) y por el creciente debate entre su propuesta del

método unifactorial y la del análisis multifactorial expuesto por L. Thurstone. La fundación de *Psychometrica* en 1936 serviría de incentivo inmediato a la divulgación de la psicología estadística que, en torno a los años cincuenta, había llegado a alcanzar importantes cuotas de desarrollo e implementación en una investigación empírica ya para entonces altamente especializada.

En el caso de la sociología, el desarrollo de las estadísticas resultó bastante menos euforizante y hasta cierto punto enigmático. Atender a este relativo retraso y a las restricciones que tal anquilosamiento supuso para la evolución de la disciplina serán cuestiones a las que se intentará dar respuesta a lo largo de este escrito. Para ello, se hará una reconstrucción —pretendidamente esquemática pero suficientemente informativa— de aquellos capítulos más señalados en el desarrollo de la sociología estadística del siglo xx. Así, se arrancará de su despegue institucional inicial en el ámbito estadounidense a comienzos de siglo, para acabar bosquejando algunos de los desarrollos más interesantes debidos a la creciente automatización a la que hoy se enfrenta esta rama de la sociología empírica, y las alternativas que —en su desarrollo actual— puede llegar a ofrecer una posible conjunción de investigación y teoría.

Un primer apunte sobre la sociología fundacional

El hecho de centrarnos inicialmente en los desarrollos de la sociología estadounidense no supone la afirmación de que la sociología fundacional europea no se interesase o se interesase menos en el manejo estadístico de la información. Todo lo contrario. Hasta un punto insospechado, la «revolución probabilística» en el pensamiento científico (Krüger y otros, 1987) se inicia dentro del pensamiento social (cfr. Porter, 1986; Stigler, 1986). Desde épocas tempranas, los investigadores sociales habían intentado combinar el estudio de las regularidades de la acción social con un uso extensivo de las estadísticas descriptivas y otros elementos de teoría de la probabilidad. Así, bien puede decirse que la curiosidad estadística de la sociología data al menos desde la aparición de la idea del *homme moyen* de Quetelet, a inicios del siglo XIX¹. Pero algo pasó en el despegue de la disciplina que hizo que su

1. Lazarsfeld (1962: 761) señala cómo, aunque los orígenes de la investigación social pueden forzar una retrospectiva hasta el siglo xvii, en realidad ésta hace su aparición en la Europa de la temprana época moderna: los métodos maestres se derivaron del sondeo de Booth sobre trabajo y estilo de vida en Londres; el análisis factorial fue inventado por un inglés, Sperman; el énfasis en el análisis cuantitativo se debe al mineralogista Le Play; Tarde pretendió investigar la métrica de las actitudes, etc. En la sociología de épocas más recientes, en el periodo inmediatamente posterior a la Revolución Francesa, Condorcet abogó a favor de la aplicación de modelos matemáticos al voto ciudadano; Laplace y Lavoisier, sus contemporáneos, llevaron a cabo sondeos sociales para sus gobiernos; finalmente, el belga Quetelet siguió sus ideas e instauró un intento sistemático de investigación social empírica: la *physique sociale*. Nicéforo en Italia, y en Alemania, Weber, Tönnies, Von Wiese y otros. Para 1933 se hacía investigación empírica en Europa. No obstante, como una rama profesional específica, la sociología empírica no llega a tener arraigo en dicho continente y pasa a desarrollarse de forma casi exclusiva y muy floreciente en USA.

desarrollo institucional perdiese terreno en el viejo continente (cfr. Lazarsfeld, 1962: 761). Tratar de adentrarse en las razones que llevaron al aborto temprano de esta presunta vocación de síntesis entre estadística y teoría supondría un ejercicio que escapa a la intención de este trabajo. Goldthorpe (2000: 259-294), por ejemplo, narra con cierto detalle los tempranos trabajos cuantitativos de la época, señalando cómo esta tradición, inicialmente salida de suelo europeo², pierde garra a comienzos del siglo XIX y sólo se recupera después de la segunda guerra mundial, cuando América ya había asegurado su liderazgo en este campo.

En el contexto de esta reconstrucción histórica, resulta particularmente interesante la referencia de Goldthorpe (op. cit.) al empeño mostrado por Max Weber³ en la articulación de estadística y causalidad como las dos caras de cualquier «explicación» sociológica posible:

Si falta la adecuación de sentido nos encontramos meramente ante una probabilidad estadística *no susceptible de comprensión* (o comprensible en forma incompleta); y esto aunque conozcamos la regularidad en el desarrollo del hecho [...] con el máximo de precisión y de que ésta sea determinable cuantitativamente. Por otra parte, aún la más evidente adecuación de sentido puede considerarse como una proposición causal correcta para el conocimiento sociológico en la medida en la que pruebe la existencia de una *probabilidad* (determinable de alguna manera) de que la acción concreta tomará *de hecho*, con determinable frecuencia o aproximación, la forma que fue considerada como adecuada por el sentido. (1944, FCE: 11. Énfasis suyo)

Intención que, por lo demás, queda claramente ilustrada en el estilo de su mayor estudio histórico, *La ética protestante y el espíritu del capitalismo*, en el que el *explanandum* general —el espíritu del capitalismo— se puede entender como referido a patrones recurrentes de acción: aquéllos en los que los individuos se integran en relación con trabajo y dinero. Con gran claridad, además, señala que las estadísticas sociales empíricamente establecidas requieren ser explicadas causalmente por referencia a patrones de decisión y acción a través de los cuales aquéllas se crean y mantienen.

Pero, tanto en el caso de Weber, como en el de otros sociólogos europeos (cfr. Goldthorpe, *ibídem*), el proyecto de institucionalizar una sociología empírica en la que las estructuras estadísticas de datos y la interpretación formal aportaran el entramado deseable de la investigación aplicada se quedó en el camino. Aparte de por oposiciones políticas fortuitas en el contexto estrictamente alemán que aquí no vienen al caso, en el ámbito de la sociología germana este esbozo de proyecto fue abortado también por las dos conflagraciones mundiales que asolaron a Europa en la primera mitad del siglo XX. Así, a la postre, la normalización e institucionalización de la sociología estadística y

2. En este punto, puede resultar interesante remitirse a las biografías de Durkheim y Weber (cfr. Goldthorpe, 1997; Oberschall, 1965; Schad, 1972, y Desrosières, 1985).
3. Cfr. también Lazarsfeld (1970), Lazarsfeld y Oberschall (1965) y Oberschall (1965).

la subsiguiente estandarización del análisis causal estadístico no se daría en Europa, sino en los Estados Unidos de América.

Prácticamente desde su fecha fundacional (aproximadamente en 1895), hasta el período de su mayor apogeo en la década de 1960-1970, la sociología estadística de las escuelas de Columbia y Chicago, las instituciones más representativas de la vertiente, tuvo también en los Estados Unidos de América una historia peculiar (cfr. Lazarsfeld, 1962), toda ella cargada de circunstancias institucionales y de liderazgos específicos que resultaron claves en la definición de sus rasgos constitutivos. Tratar de describirlos es parte de este ejercicio que, al enfrentarse con un periodo tan extenso, sólo llegará a esbozar las sucesivas concepciones de estadística y ciencia empírica plausiblemente reconocibles en esta tradición sociológica, intentando al mismo tiempo reflejar los principales cambios de sensibilidad y de receptividad general en relación con las nociones de «probabilidad» y «causa».

Autores como R. C. Bannister (1987) y, señaladamente, C. Bernert (1983), ya han hecho este recorrido ofreciéndonos interesantes reconstrucciones de las tendencias más salientes. Aquí nos podremos beneficiar de ese material relativamente reciente y condensado en el que, además de consideraciones generales no carentes de interés, se da una particular versión del desarrollo de la sociología estadística, tradicionalmente rezagada en su adecuación instrumental y, como apenas se ha mencionado, a remolque de los desarrollos y procedimientos implantados en otras disciplinas afines. Y ello, irónicamente, muy a pesar de haber compartido prácticamente idénticas posibilidades, e incluso el mismo espacio físico puerta-con-puerta con colegas estadísticos, economistas y psicólogos altamente especializados en el tema.

Como no hay un acuerdo unánime sobre la cuestión de la explicación causal y menos aún acerca de su posible utilidad para la ciencia empírica moderna, es preciso mantener en mente el hecho de que las sucesivas etapas que se identifican a continuación no se corresponden a cortes reales que necesariamente suponen fronteras bien definidas de concepciones excluyentes de explicación científica o ciencia causal. Más bien al contrario, resulta evidente que se dan trasvases de una generación a otra y que, sólo por razones de claridad expositiva, se pueden reconstruir aquí como programas científicos distintos que pasan a identificarse primordialmente por sus estilos de trabajo y sus aspiraciones explicativas expresas.

Al margen de las reglas de uso propias, es además importante reconocer la influencia que en cada época han tenido también las tendencias filosóficas y, naturalmente, sus concepciones más generales sobre el uso de la probabilidad en la ciencia empírica. Como prácticamente ocurre con todas las demás disciplinas sociales, la sociología de hecho constituye una simple caja de resonancia de dichas tendencias. La diferencia en todo caso estriba en que, apartándose de muchas de las demás ciencias sociales, señaladamente de la psicología y de las ciencias económicas, la sociología empírica siempre se ha mostrado tendente a responder con retardo a la reflexión filosófica de la ciencia y reacia

a las innovaciones matemáticas, mostrando, en cambio, una cierta preferencia por salidas rápidas y una buena dosis de improvisación y diletantismo (cfr. Goldthorpe, op. cit.).

Primera etapa: la estadística descriptiva heredada de la escuela histórica alemana

Lejos aún de una clara lógica de la correlación-regresión, la generación fundacional (1890-1910) de la sociología cuantitativa norteamericana, mayoritariamente formada en Alemania (el 70% de sus estudiantes de postgrado, según Bernert 1983: 234-35), estuvo expuesta a la gran disputa del método de las ciencias sociales (*Methodenstreit*) que, en su momento, enfrentó las posiciones de Gustaf von Schmoller (la «escuela histórica» de economía) y Karl Menger (la escuela clásica de teoría económica), a propósito del inductivismo histórico y el deductivismo teórico, respectivamente⁴. En ese contexto, el impacto de la revolución probabilística (1830-1930) se dejó sentir mucho más claramente en el medio germano parlante que en el caso de las sociologías fundacionales inglesa (Spencer) o francesa (Comte y Durkheim). De hecho, la insistencia de los miembros de la escuela histórica en la diferencia entre leyes determinísticas (causales) y regularidades inductivas (que no acarrear fuerza causal), les permitió concebir una modalidad propia de ciencia social empírica, forjándose una tradición de estadística descriptiva en la que se argumentaba además a favor de un individualismo ajeno a las leyes y la concepción estándar de la macroeconomía (cfr. Porter, 1986: 247-53). Las discrepancias analíticas entre escuelas trascendieron también al plano político y, en el curso de su crítica a la teoría económica «abstracta», este grupo propició una encarnizada lucha contra los principios del *laissez-faire*, pasando muchos de sus miembros a ser parte activa del *Verein für Sozialpolitik*, una institución que, desde su fundación (1872), combatió los principios generales del utilitarismo económico, tanto en el terreno de los supuestos como en el de los valores morales. Método y práctica ética pasaron así a vulgarizarse en una mezcla de la que serían fácil pasto muchos de aquellos estudiantes estadounidenses.

Forjados en estas lides, al volver a casa, muchos de esos jóvenes procedentes de la escuela de Schmoller cayeron en una desafortunada asociación de procedimientos tomados de esa tradición, junto con una especie de vulgata local de darwinismo social heredado del pensamiento anglosajón, particularmente de Herbert Spencer. A diferencia de los psicólogos, discípulos de W. Wundt, que (quizá porque —como señala Bernert— corrieron la suerte de acudir a mejores centros) llegaron a adoptar una modalidad bastante más refinada de empirismo, estos sociólogos acabaron en algo bastante más primitivo e incon-

4. Guardando las distancias, en Inglaterra se da un debate similar entre los partidarios del método inductivo de J.S. Mill y los seguidores del método deductivo de W. S. Jevons. No obstante, a lo que estuvieron directamente expuestos estos estudiantes fue a la disputa del método en el contexto alemán (cfr. M. A. Oberschall, 1965).

trolado, en una especie de cuantitativismo «animista». Su ingenua creencia en una suerte de «magia» inductiva les llevó a sepultarse en verdaderas montañas de datos, a refugiarse en cuestiones estrictamente descriptivas y, erróneamente, a confundir «leyes sociales» y «causas» que —dado su peculiar animismo— asociaron a nociones de «producción», «fuerza», «energía», etc., generalizadas mediante analogías descriptivas evocativas de la naturaleza ontológica de las causas. De esta forma, como apunta M. Bunge (1959: XVIII), acabaron por dejar de lado la cuestión de fondo: el problema epistemológico de cómo, efectivamente, se reconoce el vínculo causal.

De esta matriz de pensamiento, salieron las ideas de A. Small (1923-25)⁵ y F. Giddings (1903), quienes, aunque continuaron remitiéndose particularmente al vocabulario de términos biológicos y organicistas como sinónimos de «causa», mostraron explícitamente una clara inclinación a aceptar un «análisis causal» sociológico apoyado en «estadísticas históricas» y fundamentado sobre la «inducción científica», la única que conduciría a la formulación de «leyes causales», el «fin primordial de toda ciencia, también de la sociología». De esta forma, Small y Giddings pasarían a convertirse en las piezas indiscutibles de la transición que daría lugar a los inicios de una investigación sociológica propiamente «científica».

Mientras Giddings se abría campo en la Universidad de Columbia, atendiendo a una invitación de su amigo Henderson, A. Small se incorpora al Departamento de Sociología de la Universidad de Chicago. Juntos, Small y Henderson pretendieron impulsar un programa de investigación de sociología empírica, algo que funcionase al menos «[...] hasta que las universidades europeas [pudiesen] realizar los reajustes de intereses en juego»⁶. Aún sabiendo que contaban con una tradición esencialmente contraria a la investigación estadística⁷, para 1914 la mancuerna Small-Henderson presentó un informe en el que se puntualizaba la necesidad de un «centro de investigación social», algo similar a lo que para entonces se sabía que estaba impulsando Giddings en el Departamento de Sociología de Columbia (Lazarsfeld, 1962: 762). No obs-

5. A. Small, por ejemplo, define así la sociología: «La sociología, en su objetivo máximo desde la perspectiva metodológica, es simplemente una filosofía moral consciente de su tarea y que sistemáticamente busca el conocimiento de causas y efectos dentro del proceso de evolución moral» (AJS, volumen 28, 1923: 715).

6. Cfr. Lazarsfeld (1962: 762).

7. Particularmente a partir de la influencia decisiva del psicólogo social G. Mead (1934) y de la versión sociológica que hace Blumer de su binomio mente-sociedad, la vena introspeccionista constituye un sello de identidad indiscutible de Chicago. Ya desde 1927, con la llegada de Ogburn procedente de Columbia, se había empezado a dar más importancia a la instrucción estadística en los postgrados, lo que favoreció cualquier proyecto que pudiera impulsar el desarrollo de la investigación por sondeo (cfr. Bryant, 1985: 136). Si a mediados de la década de 1920 Chicago había sido el lugar para estudiar sociología cualitativa, su preeminencia se había ido erosionando a lo largo de los años treinta, y para cuando llega Lundberg en 1937, aún a pesar de la corriente de interaccionismo simbólico que aún por entonces compartían Blumer, Huges, Becker y Goffman, el carácter distintivo de la escuela había perdido prácticamente toda su fuerza.

tante, Chicago declinó el proyecto y Small tuvo que buscar una alianza con la Business School, una institución dispuesta a propiciar un centro de investigación social empírica en Chicago mismo. Contrario a tal iniciativa, al Departamento de Sociología le tomaría más de dos décadas llegar a encajar la idea, y casi medio siglo más admitir la puesta en marcha de una sociología estadística profesionalizada como proyecto integrado a la estructura universitaria (cfr. Lazarsfeld, 1962: 763).

Segunda etapa: la influencia de la estadística inglesa

Naturalmente, también en Columbia tomó su tiempo desprenderse de las viejas ideas. Pero, a pesar de que los aires de renovación no parecían halagüeños, cuando, en lugar del malogrado Mayo-Smith (uno de esos inductivistas formados por estadísticos morales alemanes), llegó F. H. Giddings con sus primeras publicaciones sobre «leyes sociológicas inductivas» (*Statistics and Sociology*, 1891, e *Inductive Sociology*, 1903)⁸, se dio paso a la fundación de un laboratorio de estadística, la instancia de la Universidad de Columbia que marcaría el inicio definitivo del programa de una sociología empírica propiamente estadística.

A pesar de que Giddings siempre procuró rodearse de mediocres que no hicieran sombra a sus ideas (que, por otra parte, apenas si soportarían el simple paso del tiempo), impulsó a los jóvenes de su laboratorio a mantener contactos directos con Henry L. Moore, un economista asociado al laboratorio de Galton, Pearson y Edgeworth en Londres, a través de quien aquellos sociólogos llegaron a familiarizarse con los nuevos avances estadísticos y fueron imbuidos de una creciente confianza y un mayor reconocimiento de las ventajas asociadas a esos instrumentos. Más capacitada técnicamente, más eficiente y mucho más conciliadora que Giddings, esta nueva generación de estadísticos, entre quienes destacaron particularmente Chapin, Hankins y Ogburn, pasó a desempeñar un papel pionero en la introducción del análisis de regresión en el ámbito de la sociología aplicada.

Para 1920, F. Stuart Chapin fue el primer sociólogo en publicar un manual de técnicas (*Field Work and Social Research*) en el que se incluía ya una exposición de la regresión, los intervalos de confianza, el muestreo al azar, la preparación de protocolos de entrevista, la codificación, clasificación y tabulación de datos y otros procedimientos hasta entonces no empleados sistemáticamente en la investigación aplicada y que en el manual se proponían como «prácticas habituales de la investigación sociológica». A este manual le siguió un libro de texto bastante más actualizado, *Principles and Methods of Statistics*, en el que R. Chaddok (925) da también detalles acerca de rutinas computacionales de correlación y regresión, e introduce algunas de las primeras instrucciones elementales para la operación con tarjetas perforadas, todo un cambio en el horizonte de posibilidades

8. Publicados en 1891 y en 1901, en Nueva York: The MacMillan Co. (cfr. Oberschall, 1972: 225-226).

para la investigación cuantitativa aplicada. Fue así como el Departamento de Sociología de la Universidad de Columbia se convirtió en la puerta de entrada a la estadística inglesa en la sociología estadounidense aplicada.

La tradición tabular: La sociología del dato

A pesar de todo, el tránsito a la apenas concebida tradición estadística fue costosamente lento. De hecho, no es hasta que W. F. Ogburn pasa a Chicago e influye sobre Stouffer y Duncan, cuando la iniciación estadística cobra realmente fuerza y se generaliza a partir de 1927. Poco después, tanto en Chicago como en Columbia, los conocimientos de sociología estadística lograron estandarizarse y los estudios empíricos aplicados comenzaron a aparecer de forma regular en el *Journal of the American Sociological Association*. Sólo entonces puede hablarse de *ánimos renovados* que, bajo la dirección de Taylor, llevarían a configurar una modalidad específicamente sociológica de trabajo inductivo en la que la descripción estadística se asoció a la encuesta individualizada o *social survey*⁹.

Montado sobre las ideas de Poincaré, Mach, Pearson y J. S. Mill, Taylor lideró una visión de la sociología aplicada que presentaba las tablas de respuestas obtenidas por sondeo como instrumento de control empírico y validación de hipótesis, y cuya eficacia «[...] depende de la manipulación matemática de unidades objetivas cuidadosamente definidas, observadas y codificadas». Concebida como una estrategia para disponer conjunta y ordenadamente de sumas de frecuencias obtenidas en la ordenación de datos, las tablas servirían para sintetizar resultados cuantitativos y ofrecer una visión numérica de las relaciones entre categorías y variables (cfr. Greenwood, 1945; Chapin, 1947). En efecto, con las tablas, se llegó a concretar de forma definitiva la fase clasificatoria de la sociología estadística inicial, un rasgo tan propio de su estilo de trabajo que ha sido retenido como técnica auxiliar o complementaria a lo largo de la investigación de todos los tiempos (cfr. Fienberg, 1977; Upton, 1978; Gilbert, 1981).

Pero el problema central de estos modelos tabulares elementales iniciales, aun en el caso de que las conjeturas causales efectivamente pudiesen contrastarse, era que las comparaciones por ellos establecidas se basaban generalmente en muy pocas instancias y, en último caso, servían sólo para estimar los efectos para cada grupo de forma separada. Aunque, en la actualidad, algunos de estos escollos siguen en pie, el análisis de tablas cruzadas con modelos tales como el

9. A fines del siglo XIX, antes de la implementación del procedimiento de sondeo, las fuentes de datos para la investigación sociológica eran primordialmente provistas por estadísticas y datos institucionales contenidos en «fuentes secundarias». Estos datos aparecían suficientes para los estudios comparativos que animaron a la sociología del desarrollo de Comte, o la tesis evolutiva de Spencer. No obstante, como se muestra claramente en Durkheim, no facilitaban su inserción dentro de una narrativa de la acción (de ahí, quizá, lo de que los hechos son «hechos sociales»). En este sentido, el desarrollo del sondeo marca un gran avance (cfr. Goldthorpe, 1997: 409).

log-lineal ha acabado resolviéndose en descripciones muchísimo más eficientes. Además, a diferencia de las tempranas tabulaciones, el procedimiento para el análisis de tablas cruzadas en los modelos actuales altamente automatizados permite estimar los parámetros usando simultáneamente toda la información disponible en la tabla (Saris y Stronkhorst, 1984: 288), una ventaja adicional para la descripción sociológica.

Poco a poco, se fueron gestando las condiciones propicias para la entrada del mito inductivo radical que, a partir de 1937, llegó de la mano de Lundberg. Su idea rectora fue que, asociando los datos no experimentales obtenidos por sondeo con el procedimiento estadístico, el conjunto permitiría equiparar los modelos explicativos de la sociología con los de las demás ciencias empíricas (Lundberg, 1929). La introducción y el desarrollo del análisis de regresión¹⁰, unido a un sustantivo avance en el manejo de los procedimientos de tablas y gráficos, pasó a ser entonces el rasgo esencial del estilo de trabajo de la nueva generación de sociólogos. Su impacto llevó, en las décadas de 1920 y 1930, al desarrollo y a la estandarización definitivos del procedimiento del sondeo individualizado como la herramienta de investigación por excelencia de la sociología «científica» o estadística.

El «positivismo instrumental»: Método sin teoría

Con Lundberg y su gente, el programa del «positivismo instrumental» se extendió finalmente de Columbia a Chicago, generalizando un proyecto —afín al neopositivismo en auge¹¹— siguiendo el cual la sociología empírica pasaba a suscribir explícitamente una variante operacionalista en la introducción de sus términos observacionales.

El «operacionalismo», una versión del empirismo particularmente activa, había sido puesta en marcha por el físico G. Bridgman (1927), quien sostenía que todo término científico —métrico o no— «no significa otra cosa que un conjunto de operaciones» (op. cit.: 5), aquéllas que permiten su introducción. Los términos introducidos se presentaban entonces como sinónimos de un conjunto de operaciones y pasaban a definirse como «operacionales» u «operativos», puesto que eran tales operaciones las que les conferían entidad o significado. Hacer extensiva esta idea al campo de la sociología impuso a Lundberg

10. La «regresión» se generaliza en estadística para indicar el estudio de la naturaleza de la relación entre dos variables, esto es, la función matemática y la forma geométrica de tal relación: el análisis de regresión.

11. La filosofía del neopositivismo o empirismo lógico, que inicia su andadura a partir de las contribuciones de filósofos lógicos y matemáticos de los círculos de Viena (R. Carnap, O. Neurath, H. Feigl y otros) y Berlín (H. Reichenbach y, más tarde, K. Hempel), había impuesto por entonces la idea de la unidad del método de la ciencia. A pesar de las controversias que el tema levantaría, el fisicalismo como filosofía dominante impuso la idea de que el *lenguaje observacional* debe consistir en descripciones métricas o cuantitativas de entidades, propiedades o procesos, no referidos a experiencias privadas, sino un lenguaje unificado, *intersubjetivo*, proporcionado por la experiencia neutra en la que han de apoyarse las teorías científicas.

la elaboración de una suerte de operacionalismo autogestionado en el que se asumía que el procedimiento de introducción operacional de términos, el único criterio objetivo disponible para la unificación del lenguaje de las ciencias sociales, pasaba por la especificación de los protocolos o las operaciones que permitían la introducción de un sistema categorial intersubjetivo, único. Con tal iniciativa como obsesión propia, pronto llegó a creer que asociando las estructuras de datos correspondientes a respuestas reiteradas en los sondeos individuales con la matemática estadística se llegaría a aportar el conjunto necesario de «instrucciones» y «operaciones» mediante las cuales introducir definiciones operacionales «genuinas» y clasificaciones «objetivas». A partir de aquí, el énfasis en el debate del método científico se hizo girar entorno a las operaciones adecuadas para la introducción de términos (Lazarsfeld: 1970a, 1979) y sobre las virtudes y los vicios de estos protocolos en la construcción del lenguaje de las ciencias sociales (Rosenberg y Lazarsfeld: 1955).

Así, aunque personalmente no desarrolló ninguna técnica específica, finalmente fue Lundberg acabó institucionalizando la función de la estadística y del sondeo como «operaciones» esenciales para la construcción de conceptos (indicadores e índices) capaces de trascender las particularidades y llamados a la construcción de un sistema categorial único para la disciplina. De ahí el atributo de «instrumental», uno de los rubros con los que también ha pasado a conocerse este positivismo que, en manos de Lundberg, puso el énfasis central en ese lenguaje unificado llamado a fusionar la inducción estadística y los sondeos en un programa de investigación propio de la sociología empírica (cfr. Bryant, 1985).

Un efecto no deseado del creciente éxito experimentado por el método de sondeo a raíz de estos acontecimientos fue la creciente tendencia a convertir la técnica en un fin en sí misma, conduciendo a un inesperado espejismo del instrumento. Debido a esa desafortunada inversión de prioridades y objetivos, varias generaciones de sociólogos empíricos acabaron identificando «método» e «instrumentos» o técnicas, y se alejaron casi por completo de los problemas nucleares relativos a la explicación causal y a la sistemática teórica. En su particular modalidad de empirismo, sin apenas vacilaciones aparentes, el operacionalismo de los términos claramente les llevó a sustituir en su discurso la noción de causa por la de «correlación» o asociación probable, un término que les resultaba operacionalmente más general y a la vez más asible y menos problemático.

Ha de decirse que este fenómeno no resultó ser exclusivo de la sociología. Más bien al contrario, las constantes dudas sobre la relevancia última de la lógica correlacional basada en el análisis estadístico acerca de las causas habían llevado en el período de entreguerras a filósofos y científicos por igual a un vasto rechazo del lenguaje causal¹². Todo lo que requería la ciencia en temas

12. Bunge (1959: 333) muestra cómo el concepto tiende a desaparecer: es un *fetiché* para Pearson, una *superstición* para Wittgenstein, un *mito* para Toulmin, una *reliquia* para Russell.

de causalidad y todo lo que se podía esperar encontrar a este respecto por vía del conocimiento científico, pasaba, por aquel entonces, por el establecimiento de relaciones funcionales relativamente invariantes entre propiedades metrizadas¹³. Hasta las leyes de la física llegaron a concebirse como simples idealizaciones de datos descriptivos correlacionales. En este sentido, los científicos sociales sólo intentaron ajustar sus aspiraciones a las ideas en curso y, a su manera, dieron algunos pasos en el mismo sentido.

Así, imbuida de los ideales del fisicalismo y empirismo lógico de comienzos del siglo XX, y básicamente influida por Pearson¹⁴, esta sociología, eminentemente estadística e inductiva, acabó por erradicar de su lenguaje la noción de causa. Siguiendo el *dictum* pearsoniano de que «todo lo que podemos observar son las covariaciones», estos sociólogos pretendieron ceñirse a la restricción empírica de los observables y, al hacerlo, descuidaron otra restricción empírica igualmente importante, a saber, la de que la asociación o correlación no puede implicar consecuencias ontológicamente objetables. La tensión entre ambos principios empíricos en competencia pasó a pesar fuertemente sobre las discusiones en torno a la estadística y la imputación causal, tan frecuentes entre las nuevas promociones de sociólogos a finales de los años treinta. A comienzos de los cuarenta, la supremacía del inductivismo radical y la consiguiente exaltación del instrumento acabarían cediendo paso a una versión bastante más refinada de empirismo notoriamente modificado por una preocupación analítica en ascenso, algo que les permitiría plantearse de nuevo la línea de continuidad lógica entre investigación empírica y sistemática teórica.

La generación intermedia: Técnicas, método y teoría

Al tiempo que se alcanzaba el punto más bajo de la terminología causal, se habían ido dando también importantes mejoras en los instrumentos que impulsaron un importante volumen de producción en investigación empírica aplicada. Pero, a pesar de este éxito aparente, la práctica indiscriminada del inductivismo y el operacionalismo prevalecientes acabaron por producir una profunda y peligrosa disociación entre los lenguajes teóricos y de investigación. La llamada «generación intermedia» pasaría a desempeñar un papel crucial en un cambio de rumbo, alejándose del «imperio del método» convertido en fetiche y retomando con ahínco las cuestiones de inferencia causal y sistemática teórica.

13. La historia de los desarrollos filosóficos subsiguientes en conceptos causales es demasiado compleja y escapa por completo de los límites de esta discusión.
14. Sea como sea, los padres de la estadística moderna mostraron poca estima por el tema de la causalidad, al que se referían con gran escepticismo. Al menos para Pearson (1892), la causalidad era una herencia metafísica, un residuo típico del pensamiento precientífico, un mero fetiche que había de ceder el paso a la idea de *correlación*, entendida como representación inductiva de relaciones entre propiedades o eventos observables.

Gracias particularmente a la labor de difusión que Samuel Stouffer y Dorothie Swane (ambos procedentes de la London School of Economics) emprendieron en la Universidad de Chicago, la sociología empírica acabaría por abrirse a la mejor tecnología estadística británica de aquellos tiempos, lo que propició la implementación de matrices nuevas en las que, sin detrimento de los instrumentos, se intentaría recuperar el lenguaje causal. A partir de este esfuerzo divulgativo, el creciente interés en los progresos estadísticos de la investigación aplicada alcanzados por otras disciplinas afines (el análisis factorial de la psicometría, la teoría de modelos de ecuación estructural de la econometría y los modelos de senderos de la biometría) acabaría por consolidar una nueva atmósfera que llevaría a la sociología estadística a la etapa moderna (Goldthorpe, 2000).

Aunque la ocasión de propiciar este impulso de renovación se dio en un primer momento gracias al énfasis de Stouffer en la implementación de las novedades en inferencia estadística avanzadas por R. A. Fischer (1926; 1935), fue la ocasión del encargo de analizar los resultados de la gran macroencuesta de *The American Soldier* la que le llevó a asociarse con los mejores investigadores sociales de la época. Dentro de ese equipo, sería P. Lazarsfeld quien haría la contribución más significativa a la estadística aplicada, actuando de forma decisiva en la tarea de actualización técnica e integración sistemática del análisis causal en la investigación por sondeo. Básicamente, lo que el equipo Stouffer-Lazarsfeld pretendió impulsar fue la idea de una lógica explicativa (o causal) llamada a promover formas de control alternativas al experimento tradicional (la técnica por excelencia de análisis causal) que permitiesen parafrasear la asociación estadística en términos de acusación causal. De esta forma, las dos cuestiones centrales que acometieron fueron la de dilucidar el sentido preciso en el que las asociaciones estadísticas provenientes de sondeos pueden o no ser efectivamente susceptibles de experimentación controlada o, en su defecto, idear técnicas alternativas capaces de garantizar ciertos controles que permitan una interpretación plausible de las estructuras estadísticas resultantes de ese tipo de datos.

En lo que atañe a la primera cuestión, inicialmente, Stouffer y los suyos (Stouffer, Suchman y otros, 1949) pretendieron inspirarse en los experimentos estadísticos y, no sin ciertas reticencias, se propusieron integrar a los sondeos de la macroencuesta el recurso experimental de la comparación sistemática entre un grupo al que se añade el estímulo (grupo de *tratamiento*) y el grupo que permanece idéntico, esto es, sin el estímulo (el grupo de *control*)¹⁵. A pesar

15. En relación con estos modelos estadísticos de inferencia, puede decirse que, básicamente, el procedimiento responde al siguiente argumento: Si se manipula un factor causal, x , entonces, observando los controles apropiados (de regresión, por ejemplo), se ha de producir el efecto sistemático sobre la variable que responde, y . Así, por ejemplo, si se manipulan apropiadamente los resultados mediante la introducción de grupos de *tratamiento* a los que se provee de estímulos especiales (x), y grupos de *control* ($no-x$) desprovistos de tales estímulos, entonces se puede discriminar si la probabilidad de y dado x resulta ser mayor que la probabilidad de y dado $no-x$.

del empeño que pusieron en ello, la propuesta se vio satisfecha de forma sólo excepcional, hasta el punto que el propio Stouffer tuvo que acabar admitiendo que, aparte de pocas excepciones, en la mayoría de los casos el trabajo con este tipo de datos únicamente puede lograr aproximaciones muy burdas al experimento controlado. Pero, aún siendo pequeño el rendimiento global del intento, no cabe duda de que, más que cualquier otra investigación anterior, *The American Soldier* contribuyó a introducir nuevos rigores en el diseño de los sondeos individualizados y coadyuvó también a llevar a este tipo de investigación por sondeo a un momento histórico de gloria.

No obstante, sería la crítica de P. Lazarsfeld (1950) la que acabaría sentando la idea de que, en ausencia de la posibilidad de intervención o manipulación de los resultados (el rasgo propio de cualquier experimento, clásico o controlado), aquello a lo que justamente han de enfrentarse los modelos de regresión en las investigaciones sociológicas es a controlar estadísticamente todas aquellas variables que pueden proveer una historia causal interesante. Así, a diferencia del contexto propiamente experimental en el que el procedimiento de asignación aleatoria permitía eliminar diferencias sistemáticas significativas en la composición de los grupos de tratamiento y control (la idea inicial de Stouffer), en los estudios no experimentales *todas* las variables han de ser tomadas en cuenta de *forma explícita*, puesto que la omisión de cualquier variable importante puede llevar a conclusiones cuestionables. Sin el recurso de la intervención, la sostenibilidad de la explicación causal en la investigación social se ve así reducida a las restricciones que puedan sugerir los controles de regresión y supeditada al conocimiento precedente, intuitivo o teórico.

En su versión alternativa al diseño experimental, Lazarsfeld (Kendall y Lazarsfeld, 1950) propone una modalidad de análisis multivariante adaptada a la «elaboración» de las relaciones estadísticas con datos procedentes de sondeos individualizados: la «parcialización». Su invitación es a examinar las modificaciones que sufren las relaciones parciales entre las variables una vez que se introduce un factor de control. Si, en la versión más elemental V_{xy} representa la correlación original, el control de la variable test V_t puede ejercerse sobre V_x bien desplazándola (en cuyo caso V_x se considera espuria), o, por el contrario, simplemente especificándola (en cuyo caso V_x y V_t son complementarias). A través de estos procesos de combinación (o tablas parciales), «explicar» estadísticamente en contextos no experimentales acaba equivaliendo a intentar eliminar las relaciones entre la variable test V_t antecedente y la variable dada por independiente V_x en la correlación original (Lazarsfeld, 1946).

Aunque la lógica de la parcialización guarda semejanza con la del experimento controlado, es evidente que no se trata del mismo procedimiento, ni tampoco tiene idéntico rendimiento informativo. En este contexto, además¹⁶,

16. Para una crítica de estos planteamientos y, sobre todo, del mal uso que se ha hecho de ellos (cfr. Lieberman, 1985).

la noción de «causa» cobra un sentido altamente pragmático, en el entendido de que son los sucesivos controles los que llevan a equiparar la *explicación estadística* con una relación que *les sobrevive*, es decir, con una relación estadística *robusta* o manifiestamente *no espuria* (Kendall y Lazarsfeld, op. cit., nota 26, 158).

Siguiendo a Lazarsfeld, el argumento completo sería que, mientras la asociación estadística no implica causalidad, toda conjetura causal supone una *asociación controlada*. Y, aún en contextos no experimentales, los procedimientos de regresión pueden llevar a asociaciones estadísticas mejor controladas que sirven a su vez de pretexto a una paráfrasis causal plausible. Así, mientras «explicar» consiste «[...] en el aspecto formal de la elaboración [estadística]», «interpretar» causalmente requiere, además «[...] combinar el formalismo de la elaboración de una clasificación [con] algún principio sustantivo de ordenación» (Lazarsfeld, 1958: 124). La instancia central de la elaboración estadística, el control de espuridad, se plantea entonces como un requisito procedimentalmente necesario y que precede a cualquier posible interpretación causal de la covariación, imputación esta última a la que han de asociarse principios sustantivos ajenos a la estadística y generalmente sacados de la intuición o del conocimiento teórico acerca del tema en cuestión.

Siempre que Lazarsfeld habló de relación causal, la asoció a la «interpretación» de la asociación estadística genuina, paso que necesariamente presupone la incorporación de conocimientos sustantivos fundamentados sobre información precedente. En efecto, puesto que la causalidad no es observable, el reconocimiento de la asimetría causal (o no reversibilidad de las variables)¹⁷ es algo para lo que han de introducirse supuestos adicionales o «mecanismos». Con ello tácita pero rotundamente Lazarsfeld está implicando el límite del lenguaje estadístico, que resulta ciego en temas de causalidad. Así, al margen del control estadístico de espuridad, el hecho de afirmar que *x* es una posible *causa* de *y* exige la incorporación de supuestos ajenos a la estructura estadística de los datos del sondeo. El propio Lazarsfeld (1958: 124) reconocerá que la secuencia temporal, el supuesto que explícitamente incorpora a su esquema formal de «elaboración» de las variables (Lazarsfeld, 1946), no es el único supuesto sustantivo de especificación de la asimetría causal. No obstante, no sería él, sino el economista Herbert A. Simon (1953; 1979) quien, poco después, atendería detalladamente a esta cuestión.

17. En el enunciado causal, la relación es irreversible: si fumar causa cáncer de pulmón, la dolencia no puede causar el hecho de fumar; si el accidente fue producido por el estado de embriaguez, ese estado de embriaguez no pudo haber sido producido por el accidente, etc. A diferencia, pues, de la relación estadística que es reversible (si fumar está positivamente relacionado con el cáncer de pulmón, el cáncer de pulmón estará igualmente correlacionado con signo positivo con el hecho de fumar), la asociación causal es irreversible, es decir, supone una asimetría causa-efecto: el efecto no puede ser causa de su causa, no puede revertirse.

Tercera etapa: los modelos causales de regresión

El precedente sentado por Lazarsfeld fue rápidamente seguido por H. Simon, mejor conocido entre sociólogos por sus *Models of men*. Simon entronca en una sólida tradición econométrica que arranca de Hottelling, Tinbergen, Haavelmo, Wold y, particularmente, T. Koopmans y H. Schutz, sus maestros. Ellos fueron quienes fundamentaron los métodos que básicamente pasaron a caracterizar la econometría hasta bien entrados los años setenta. En la estela de esta tradición, H. Simon reconduce el problema de la «identificación» y la cuestión de las «relaciones estructurales» asociándolo al tema de la «correlación espuria».

Aunque en su momento Yule había examinado el problema de la asociación espuria, la cuestión había recibido también inmediata atención por parte de estadísticos sociales (Lazarsfeld, 1946; Zeisel, 1947; Kendall y Lazarsfeld, 1950) cuyos trabajos—como apenas se ha dicho—tuvieron que ver básicamente con refinamientos del análisis de regresión y su formalización. Aunque en todos esos desarrollos se indicaba la generalización del procedimiento a n variables, ninguna de las propuestas había examinado con detalle la correlación estadística en un esquema ampliado o con más de una ecuación. Desde el problema de la identificación y con un sistema lineal de ecuaciones en mente, Simon (1952; 1953; 1954) pretende atender a esta cuestión, a la vez que tender un puente al problema de la correlación espuria. Así, al intentar operacionalizar la dirección de la influencia causal, llega a una noción de «causa» que asocia la ordenación asimétrica genuina de las variables con sistemas de ecuaciones autocontenidos.

Esta posición contrastaba con la del operacionalismo inmediatamente anterior que, también en ciencias económicas, venía evitando cualquier uso de la noción de «causa» a la que se había sustituido por «relaciones funcionales» o relaciones de «interdependencia» entre variables. Tal omisión, según Simon, hacía que la observación quedara exclusivamente relegada a revelar asociación recurrente, pero, sobre todo, implicaba una terminología causal contraria a la genuina intuición. En su *Causal Ordering and Identifiability*, Simon (1953) hace notar el hecho de que la concepción de relaciones funcionales o «simétricas» contrasta poderosamente con cómo se concibe la relación causal en el sentido común u ordinario. De hecho, cuando, en el curso de una conversación real, se dice que « A causa B », ni se quiere ni se puede decir a la vez que « B causa A ». En cambio, cuando se afirma que « A y B están funcionalmente relacionadas» (o son interdependientes), se puede estar indistintamente queriendo decir que A y B , o B y A están funcionalmente conectadas. Mientras la relación de las variables en la correlación es reversible, la relación causal es esencialmente asimétrica o irreversible.

En tanto ligados a leyes funcionales, los modelos «causales» de ecuaciones estructurales habrían de integrar esta idea de relación asimétrica entre las variables y especificar además cuáles de ellas se excluye y de qué ecuación. Era evidente que la noción de causa había de estar íntimamente asociada a la cuestión estadística de la identificabilidad y a supuestos sustantivos sobre las variables que

se supone influyen en el sistema. Como, estadísticamente, «identificación» implica que el sistema de ecuaciones del modelo ha de comprender el número de ecuaciones y elementos suficientes para que la estimación de todos o parte de sus parámetros sea posible¹⁸, en la práctica, se impone la incorporación de ciertos supuestos *a priori* o principios sustantivos que (como el de secuencia u ordenación temporal) que sirven para completar el sistema de ecuaciones a la vez que indican la omisión o inclusión de una determinada variable en una determinada ecuación. Así, por ejemplo, por vía de omisión de variables particulares, entran en el modelo supuestos tales como el principio de «simplicidad»¹⁹ (o eliminación de variables redundantes); el postulado de «homogeneidad» para la identificación de coeficientes con idéntico valor²⁰; el supuesto de «no-comunicación, no-influencia» o contigüidad, etc., que permiten la omisión de nexos entre aquellas variables que no se suponen directamente conectadas dentro de un determinado sistema. Estos supuestos acerca de las distribuciones podrán omitirse sólo en casos en los que la correlación real entre variables observables resulte aporoblemática. Pero, en el caso de variables inobservables, son ellos por regla general los que proveen identificación adicional al sistema de ecuaciones que se estudia.

Así, siempre que se hable de un modelo econométrico, en términos generales se estará aludiendo a un modelo estructural recursivo, lineal, autocontenido y autónomo. *Lineal*, puesto que, matemáticamente, representa ecuaciones lineales; *autocontenido* porque está identificado, es decir, contiene tantas ecuaciones como variables; *recursivo* porque admite la determinación mutua entre las variables del sistema. Se dice también que se trata de un modelo de ecuación estructural *autónomo* porque, dentro de ese sistema de ecuaciones, cada ecuación representa un supuesto o un mecanismo independiente²¹. De

18. Se dice que el sistema está perfectamente *identificado*, cuando en el sistema existe información suficiente para resolver todos los parámetros. Estará *subidentificado* si la información que contiene sólo permite estimar parte de sus parámetros. En cambio, si dicha información no permite identificar ningún parámetro, entonces el sistema no es identificable en absoluto. También puede darse el caso en el que la información resulta excesiva, lo cual ocurre sólo cuando el número de ecuaciones es superior al de incógnitas y da lugar a una *sobreidentificación* (Simon, 1979: 76-77).
19. La noción de «simplicidad» va generalmente asociada a la idea de que entre varios sistemas de proposiciones (sistemas lógicos, matemáticos, físicos, etc.), se elige el que es, o parece ser, «más simple». También se prefiere una ley más simple a otra menos simple. No obstante, se trata de un término multívoco (cfr. Bunge, 1962: 113-135), con lo que no todas las formas de simplicidad resultan deseables o mutuamente compatibles. No obstante, hay que escapar siempre de la tentación de atender a los requisitos de la mera conveniencia estadística (Simon, 1979: 70).
20. La aplicación de tales postulados suele ser muy útil en el caso de los sistemas físicos. La idea, por ejemplo, de que la fuerza de gravedad es una constante se convierte en una verdadera fuente de identificación de modelos para la mecánica newtoniana. Aunque las cosas difieren bastante en el caso de los sistemas sociales, cognitivamente echamos mano del mismo recurso.
21. De hecho, a diferencia de las ecuaciones algebraicas, estas ecuaciones estructurales se caracterizan porque las soluciones no se dan individualmente, ecuación por ecuación, sino para el conjunto del sistema.

esta forma, cualquier subconjunto de ecuaciones estructurales pasa a ser en sí mismo un modelo válido de la realidad, el modelo de una estructura que prevalece bajo algún conjunto de intervenciones, un modelo pensado para dar con la interpretación causal que menos se aleje de los datos empíricos disponibles.

La sociología de comienzos de los sesenta acabó por incorporar estos modelos causales de la econometría (H. M. Blalock, 1962) y los modelos de sendero de la biometría (O. Duncan, 1966). Su éxito no fue inmediato, pero las propuestas, cada vez más manejables, hicieron que la idea de la causalidad como generación de modelos dominara rápidamente el trabajo empírico de la disciplina, llevando incluso a pensar que por esta vía se acabarían cerrando las viejas disputas entre investigación y teoría.

La causalidad como generación de modelos causales y de sendero

En el procedimiento basado en los modelos de ecuación estructural (el enfoque Simon-Blalock), el simple control por combinación es reemplazado por un control estadístico más eficiente que presume la especificación de una teoría causal que envuelve a todas las variables que el investigador considera importantes. Con estos modelos el tema de la causalidad-probabilidad alcanza el de mayor expansión de las técnicas de regresión en el ámbito de la sociología empírica.

Mientras hasta aquí el análisis multivariante sólo había discutido el efecto que una variable o variables pueden tener sobre otras, para determinar y contrastar los efectos ahora resultaba preciso explicitar las *relaciones* entre dichas variables. Esta operacionalización de la asimetría causal, o relación no reversible entre variables, se concebía como íntimamente ligada a la idea de traducción del lenguaje estadístico de las asociaciones simétricas entre variables a hipótesis causales que reflejasen la unidireccionalidad de la relación de influencia. Las teorías pasarían entonces a consistir en listas de hipótesis causales que establecían relaciones no simétricas entre las variables de un determinado sistema.

Como, por una parte, la idea de «producción» o «influencia» resultaba esencial a la idea de causa y, por otra, la mayor parte de las formulaciones verbales, sobre todo los informes de investigación por sondeo, formulaban meros enunciados de covariación, éstos últimos no podían sin más identificarse con hipótesis causales —el tipo de enunciados que requieren las teorías. La traducción de los enunciados verbales a hipótesis causales se convirtió entonces en el paso previo a cualquier uso de los modelos de ecuación estructural. De hecho, para Blalock (cfr. 1962; 1964; 1969; 1973; 1980), parafrasear los informes verbales en forma de enunciados causales resulta la tarea más inmediata que ha de realizar el sociólogo que se propone suscribir esta perspectiva analítica. Él mismo puso un gran empeño en esta tarea, dedicando gran atención a la construcción de teorías e invitando a la «traducción» de los reportes verbales al lenguaje causal. Aunque Blalock nunca llegó

a elaborar reglas fiables para la formulación de teorías, el procedimiento que, a partir de él, se ha seguido habitualmente consiste en situar el problema que interesa en un campo más general de conocimiento para, a partir de supuestos «probados», reconstruir las variables influyentes y transcribir las asociaciones estadísticas en forma de hipótesis causales plausibles para un determinado modelo²².

Siguiendo esta propuesta, Saris y Stronkhorst (1984: 25-38) nos ofrecen un relato detallado del procedimiento que va desde los contenidos de un informe preliminar hasta la construcción de los modelos. El problema que trabajan, un tema relacionado con la desigualdad de oportunidades de acceso a la formación superior a la que se enfrentan los hijos de las clases menos favorecidas arranca del siguiente informe:

Para el mes de enero, el director del departamento de una escuela elemental hace algunas recomendaciones concernientes al tipo de escuela secundaria que le parece la más apropiada en cada caso para los distintos alumnos. Tal recomendación se hace llegar a los padres y a la comisión de evaluación. Subsiguientemente, los padres solicitan una plaza en la escuela que ellos desean para sus hijos. Además de la recomendación del director de la escuela, la solicitud de admisión proporciona algunos datos personales sobre el género del alumno y el estatus socioeconómico de los padres. Para febrero o marzo se administra una prueba en la escuela. Al cabo de un mes, los resultados están disponibles y la escuela informa a las comisiones de las escuelas secundarias del entorno de tales resultados. Son éstas últimas las que finalmente toman la decisión respecto a la elección de escuela que ha de acabar haciéndose. (Citado en op. cit.: 25)

Al leer el informe, lo primero que salta a la vista es la poca nitidez con la que se enuncian los procesos de influencia implícitos y la escasa diferenciación entre las variables involucradas en la elección de escuela, el resultado que el informe describe verbalmente. Así, una vez emplazado en el marco más extenso de la teoría de la movilidad social, lo primero que interesa es discriminar claramente la lista de variables centrales al proceso descrito, a saber, la *influencia de los padres*, los *resultados de los alumnos* y los *comités de evaluación y decisión*, todas ellas en relación con la variable influida, la *elección de escuela secundaria*. Una vez establecidas las variables, se han de prefigurar *ordenaciones causales* o relaciones asimétricas plausibles entre ellas, bien atendiendo a indicadores de secuencia temporal implicados en el orden mismo de las cosas, a indicadores de influencia explícitamente pretendidos por los autores del informe verbal, o a otros indicadores que pueden ser directamente proporcionados

22. Entresacado de Blalock 1968, la transcripción clásica de la teoría del sistema social matricial basado en el estudio comparativo de H. Driver y W. C. Massey sobre los indios norteamericanos es reconstruida en prácticamente cualquier manual de sociología empírica. Para el desarrollo gráfico y estadístico de los modelos, ver, por ejemplo, Sierra Bravo (1991: 61-62).

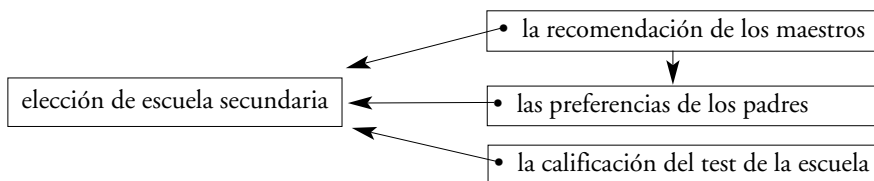


Figura 1. Diagrama causal que representa las hipótesis causales del ejemplo.

por el conocimiento procedente de la teoría. Una vez ordenadas las variables en relaciones de producción o asimétricas, el siguiente paso es la especificación de las hipótesis causales, una condición formal previa que han de cumplir los enunciados verbales para homologarse a los que integran los modelos matemáticos.

Volviendo al informe, Saris y Stronkhorst (op. cit.: 16) distinguen dos senderos de influencia. Una hipótesis (1) que hace referencia a la elección de escuela secundaria como influida por las preferencias de los padres, la recomendación de los maestros y las calificaciones obtenidas en un test impuesto por la escuela, y (2) una hipótesis causal que aludiría a la preferencia de los padres bajo la influencia del dictamen de los maestros. Gráficamente, las dos hipótesis tendrían el siguiente aspecto (figura 1).

Como, por convención, los efectos no especificados se suponen siempre iguales a cero, no pueden darse otros argumentos teóricos adicionales que puedan incluirlos, quedando de esta forma configurada la incipiente teoría.

Una vez obtenida esta traducción preliminar a hipótesis causales, el siguiente paso consiste en controlar la covariación no explicada. Con fundamento en el conocimiento teórico previo sobre movilidad social, se pueden añadir algunas variables de control tales como el *nivel de rendimiento escolar*, una variable que parece connatural al modelo. La teoría extendida incluiría entonces una tercera hipótesis causal según la cual (3) el rendimiento escolar tiene una influencia causalmente directa sobre las recomendaciones de los maestros, las preferencias de los padres y el test de evaluación de la escuela. Siguiendo la propuesta de Saris y Stonkhorst (op. cit.: 26), el diagrama correspondiente a esta transcripción causal del informe verbal ampliado sería el reflejado en la figura 2.

Se trata, pues, de un proceso bastante activo en el que se depende de la posibilidad de identificar las variables, de indicar su ordenación causal, y de poder especificar buenas hipótesis causales con fundamento en supuestos probados. Justamente, la dificultad de estos modelos radica en que los supuestos sustantivos (o no estadísticos) sobre los que se montan las hipótesis causales no resultan observables. De hecho, lo único que se puede observar es el grado de covariación descrito por los datos empíricos, dejando siempre abierta la duda de si éstos ocultan o no variables espurias o causas comunes más profundas. Y es aquí donde la propuesta de formulación

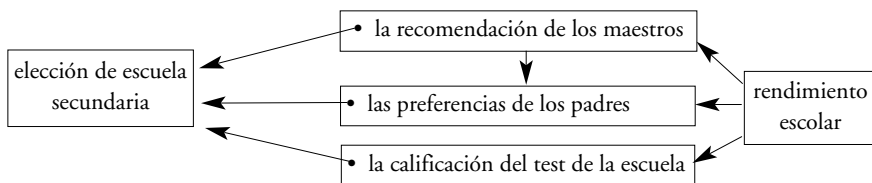


Figura 2. Diagrama causal de la teoría ampliada de la carrera escolar que incluye la variable control: rendimiento.

de teorías causales conecta directamente con la necesidad de contrastar las teorías.

De hecho, una vez construido el modelo formal, si sus predicciones teóricas (los coeficientes parciales de regresión que «miden» la fuerza de cada uno de los brazos del modelo) no se ajustan a los valores empíricos de la matriz de datos procedentes de la investigación empírica, entonces la teoría se descarta. Y, aunque nunca haya prueba concluyente de la existencia de la causalidad aducida, en general, suele decirse que «una hipótesis causal ha sido falsada si la fuerza de la relación entre las variables que interesan es igual a la covariación observada entre esas variables» (op. cit.: 28). En los casos en los que la covariación y la relación espuria no son equivalentes, entonces, o bien hay un posible efecto entre las variables (que es lo que presumiblemente se busca), o se ha omitido de la investigación alguna causa común importante²³.

R. Boudon (1965a) y particularmente O. Duncan (1966) propusieron algo más tarde otra variante de análisis multivariante vinculada a los diagramas de senderos (del genetista S. Wright, 1926). El *análisis de senderos* o análisis de dependencia pretende ponderar los efectos causales especificados en la teoría. Para ello, se estiman los parámetros que se interpretan como índices de la cantidad de cambio estandarizado en la variable dependiente y que puede atribuirse al cambio estandarizado de otra variable inmediatamente anterior. A diferencia del procedimiento apenas expuesto, más que en testar la teoría causal, el análisis de senderos se centra prioritariamente en la estimación de los efectos.

La posibilidad de control por aleatorización, la gran ventaja de la investigación experimental, se suple en estos estudios no experimentales por la comparación de las medidas derivadas de la covariación²⁴ con los valores de los datos empíricos. Pero, a diferencia de los casos experimentales, en los modelos

23. Cfr. Saris y Stronkhorst, op. cit. 28. En general, este manual constituye una buena guía para ver cómo los modelos pueden ser formulados, contrastados, corregidos, simplificados y, finalmente, constatar su grado de irresolución o de «cajanegrismo» que retienen.

24. Para mayor explicitación de estos modelos, cfr. Sierra Bravo (1981): *Ciencias sociales: Análisis estadístico y modelos matemáticos*. Madrid: Editorial Paraninfo: 285-314.

de sendero no puede excluirse ninguna variable importante. Con ello, la investigación no experimental se ve obligada a especificar teorías más completas para los campos que investiga. Es evidente que, cuando se habla de construir o generar modelos estadísticos causales o de sendero con datos no experimentales no se pueda prescindir de la discusión del tema de la contrastación; pero tampoco se puede dejar al margen la fase de traducción o formulación de enunciados o hipótesis casuales, algo para lo que se hace imprescindible el recurso al conocimiento teórico previo.

El nuevo sesgo estadístico: automatización sí, causalidad no

Con la introducción generalizada de las modernas técnicas de computación, en las décadas de 1960 y 1970 se hizo posible un análisis multivariante con datos estadísticos procedentes del sondeo a escalas antes impensables. Ello llevó a la sociología cuantitativa a una importante revolución en términos de productividad. En efecto, las tarjetas perforadas, que ya en su momento habían permitido a los investigadores contar el número de individuos que presentaban un mismo atributo para construir las entradas requeridas para sus tabulaciones cruzadas eran ahora reemplazadas por procedimientos capaces de estudiar relaciones entre un número considerable de valores en grandes conjuntos de variables simultáneas. Frente a la cantidad enorme de horas de trabajo que podía llegar a consumir la tabulación en los procedimientos anteriores, la era actual de la automatización electrónica simplificaba e incrementaba la velocidad de estas actuaciones a extremos sorprendentes. No obstante, al lado de este formidable progreso en lo que se puede hacer con los datos, la sociología cuantitativa se mantuvo teóricamente paupérrima. Así, a pesar de que las estructuras estadísticas a las que hoy se tiene acceso con estos medios son mucho más ricas en su calidad de clasificaciones o descripciones, más que coadyuvar al progreso de la disciplina, la informatización ha ido acompañada de una notoria tendencia al retroceso teórico. De hecho, en términos generales, hoy se está bastante más lejos del ideal explicativo que en los años cuarenta.

En buena medida ello responde al hecho de que, aunque la creciente automatización informática ha hecho del modelo de regresión su tipo de modelo formal favorito para el análisis de la relación entre variables, ha tendido sistemáticamente a excluir la intención interpretativa o teórica que alentó a las versiones iniciales. La justificación habitual que dan los modelos estadísticos automatizados de esta elección es prioritariamente en términos de la necesidad de una ponderación adecuada de la importancia relativa de las variables y de la significación estadística de la varianza. Pero, por regla general, se pospone o se ignora el papel fundamental del control de espuriedad (la idea central en la propuesta de codificación y elaboración de P. Lazarsfeld) y se relega al olvido el lenguaje pretendidamente causal de los modelos importados de la econometría (los supuestos no estadísticos de H. Simon). Así, centrados en los *tests* de significación y en la varianza explicada, potentes modelos

como el *log-lineal*²⁵ han llevado a ignorar prácticamente el sentido original de tales controles, esencialmente llamados a proporcionar asociaciones no espurias candidatas a la «especificación» o al establecimiento de efectos interactivos y a la «interpretación» causal o teórica del sistema de ecuaciones.

Así, en un exceso de complacencia estadística, se han llegado a ignorar los temas centrales de la causalidad y la vinculación teórica, pero ahora con el agravante de que, aún en ausencia de teoría, la nueva tecnología permite a los investigadores considerar una enorme cantidad de variables simultáneas e indagar sobre toda suerte de cuestiones alusivas a la significación relativa de variables o grupos de variables interactuando. A este respecto, Goldthorpe (2000) subraya cómo la mayoría de estas prácticas de análisis cuantitativo han acabado por hacer un «miserable» favor, un pobre trabajo en detrimento de la explicación en la sociología empírica. Es un hecho que, a pesar del enorme progreso y sofisticación de los instrumentos matemáticos y estadísticos para el análisis de los datos, la sociología cuantitativa de las últimas décadas ha permanecido en la inopia teórica, contribuyendo con ello a que semejante despliegue de medios haya acabado resultando totalmente irrelevante para la acumulación y el progreso de la disciplina.

Un resultado no buscado de la revolución informática ha sido, así, la instauración de un nuevo estilo de sociología aplicada, tanto profesional como académica. En efecto, la sociología estadística se ha convertido en una rama de la disciplina con entidad propia. Matemáticamente demandante y técnicamente muy refinada, esta sociología tiende a privilegiar el instrumento que, una vez más, crece desmedidamente revirtiéndose contra la teoría. De esta forma, en las últimas décadas, se ha visto consumado un hiato aún mayor entre investigación y teoría. Hay que decir que no toda la responsabilidad es de los sociólogos estadísticos. También es verdad que muchos de los que se autodenominan «teóricos» carecen de la pericia o preparación especializada para entender la inmensa mayoría de las investigaciones cuantitativas, u otros más ni tan siquiera muestran interés, ya que parten del supuesto de que la actividad interpretativa o hermenéutica es la meta única —técnicamente alcanzable o, simplemente, deseable— de la disciplina. Frente a tal despropósito, A. B. Sörensen (1998: 242) comenta que una de las ironías en la formación de muchos de los graduados universitarios contemporáneos es que lo que se llama «teoría» acaba

25. A partir de los años setenta, el análisis de la varianza acoplado a la investigación por sondeo ha dado lugar a modelos informatizados de asociación e interacción tales como el análisis *log-lineal*, que sirve para el estudio unitario y global de relaciones estadísticas de influencia entre variables indistintas (sin por ello plantearse diferenciar las variables dependientes e independientes); el modelo *logit*, en el que se elige una variable dependiente y se sigue el criterio de analizar sus frecuencias esperadas en función de las demás variables independientes (Sierra Bravo, 1990: 64), o los modelos multinivel, que capacitan para manejar efectivamente series de transiciones a lo largo del tiempo sufridas por individuos a través de diferentes estados discretos (carreras educacionales, profesionales, formación y disolución de familias, etc.), siendo posible examinarlos en términos de su dependencia constante o variable de tiempo (Goldthorpe, 1998).

muchas veces convirtiéndose en «[...] un simple refugio para estudiantes que o bien se muestran reacios [a la abstracción analítica] o son incapaces de aprender estadísticas».

Mal uso de los modelos de regresión

Los primeros estudios que usaron modelos estadísticos a gran escala, particularmente J. Coleman y otros (1966), en *Equality of Educational Opportunity*, y Blau y Duncan (1967), en *The American Occupational Structure*, llevaron a cabo un análisis multivariante muy elaborado, usando, a su vez, conjuntos masivos de datos. De esta forma, con una muestra inicial de 600.000 estudiantes, el primer análisis llegó a incluir unos cuantos cientos de variables en un intento por medir los efectos de la escolarización indicados en los resultados obtenidos en los tests de rendimiento. Aunque con muchas menos variables y un conjunto de datos bastante más reducido, el estudio de Blau y Duncan tuvo un impacto bastante mayor, porque introdujo de forma simultánea la práctica del análisis de senderos, un primer contacto sistemático para muchos sociólogos con el análisis de regresión (cfr. Goldthorpe, 2000: 240-241).

Ambos trabajos usaron coeficientes estandarizados de regresión o coeficientes de senderos, respectivamente, para medir el efecto en términos de la desviación estándar de las variables independientes. El razonamiento básico detrás de estos procedimientos resulta extremadamente sencillo: si la variable x tiene un coeficiente estándar mayor que la variable y , entonces x es más *significativa*. Las desviaciones estándar pasaron, así, a proporcionar una suerte de métrica común que llevó a popularizar este tipo de comparaciones, siempre en la confianza de que así se podría decir algo sobre acciones aplicadas o políticas específicas (cfr. C. Jenckes y otros, 1972 para otros ejemplos). El problema es que los tests de significación estadística están únicamente llamados a indicar si creer o no en la estimación de un parámetro, y un error, no por frecuente menos dañino, ha sido tomarlos para informar acerca de su supuesta importancia teórica. De hecho, con gran facilidad, se ha tendido a pasar de la afirmación de que algo es altamente significativo estadísticamente a la aceptación implícita de que es importante teóricamente. Visto así, con mala fortuna el advenimiento de la informática ha servido para reforzar la actitud de aquellos sociólogos poco afectos a «teorizar» y mejor dispuestos a pensar menos en lo que hacen, dejando que la técnica gobierne sus formas de análisis.

Si bien este mal uso de la significación estadística no es exclusivo de la sociología cuantitativa²⁶, el abuso de la varianza explicada y de las desviaciones estándar para evaluar efectos relativos sí que es una obsesión particular de los sociólogos, algo que no han compartido ni economistas ni psicólogos. Curiosamente, esta obsesión, que arranca de los escritos de Blau y Duncan

26. Aunque, en términos particulares, no es exactamente así. Cfr. la crítica que, a su uso entre economistas, hacen D. N. McCloskey y S. T. Ziliak (1966).

(1967), es una idea que el mismo Duncan se encargaría de rebatir personalmente más tarde: «No hay cantidad estadística que dé una respuesta significativa general a la cuestión de qué variable es teóricamente más importante que otra» (1975: 51).

La teoría como suma de variables

De esta forma, el mismo poder o potencia que han ganado las técnicas estadísticas en buena medida se ha vuelto contra la sociología fundamental y, en la práctica, la automatización ha llevado en muchos casos a sustituir a la teoría por una suerte de suma de variables. Las teorías acerca de los procesos que se estudian acaban presentándose como ideas acerca de qué variables considerar, y la verdad y la significación teórica se establece simplemente por la vía de mostrar qué variables cuentan. No es de extrañar, pues, que tales aportaciones nunca aparezcan en los cursos estándar de teoría.

No obstante, la idea de que la técnica estadística es capaz por sí sola de proveer explicaciones sociológicas, resulta simplemente insostenible. En contextos no experimentales, tales como son prioritariamente los de la sociología empírica, la idea de que pueden hacerse inferencias causales de forma directa o libre de teoría a partir de modelos estadísticos ha sido fuertemente cuestionada, incluso por los propios estadísticos. Y no cabe pensar que mejorando las estrategias se puede llegar a inferir la causa a partir de los datos de sondeos. Por potentes y sofisticados que sean los modelos de regresión de la sociología, aún en el supuesto de que se haga un buen uso de ellos, más allá de la descripción sólo podrán servir de evidencia sobre la que se monten interpretaciones teóricas, o se critiquen y se contrasten las mismas. En otras palabras, la «sociología de las variables» no puede desembocar en teoría, a menos que la teoría, tampoco autónoma o independiente de sus bases empíricas, provea los supuestos o mecanismos causales desde los que interpretar las estructuras estadísticas.

De ahí la insistencia en una renovada alianza entre investigación y teoría. Su disociación, largamente testimoniada en la historia de la disciplina, «es un escándalo» (Sörensen, 1998). A tal punto que, como afirma el sociólogo británico J. H. Goldthorpe (1997: 405), «incluso desde estándares no precisamente demasiado exigentes como los de la economía o la psicología, la sociología es un despropósito manifiesto».

Una propuesta alternativa: estructuras estadísticas y mecanismos causales

No obstante, el mismo Goldthorpe (op. cit.: 408-09) se empeña en señalar en la sociología contemporánea algunos desarrollos de la investigación y el tratamiento de datos procedentes de sondeos que resultan interesantes y dan además pie a un cierto optimismo. En primer lugar, los datos de los que hoy se dispone se extienden a largos períodos de tiempo. De hecho, los sondeos repetidamente implementados década tras década sobre poblaciones idénticas y los estudios de panel han llegado a alcanzar una madurez sin precedentes en

diferentes campos. Así, por ejemplo, cuestiones como el impacto de la industrialización sobre la igualdad de oportunidades, la descomposición de las clases sociales, la desintegración del mundo rural o la integración de la mujer al mercado de trabajo, pueden ser hoy tratadas sobre bases empíricas mucho más seguras que hace tres décadas²⁷. En segundo lugar, el imperativo —académico y profesional— de autosuperación en relación con la calidad de los sondeos, junto con una mejor comprensión y uso del análisis secundario, ha dado pie a interesantes comparaciones de carácter internacional²⁸ que, en conjunto, resultan muy informativas. Por último, a partir de los años setenta, desarrollos automatizados como los de los modelos *log-lineal* o modelos históricos multinivel (*multilevel modeling*), entre otros, han provisto a los sociólogos cuantitativistas de procedimientos poderosos para tratar datos categoriales asociados a diversos atributos de los distintos individuos sondeados. De hecho, permiten que la información relativa a atributos individuales pueda trabajarse sistemáticamente de forma conjunta con información relacionada con entidades supraindividuales, o con conjuntos de datos jerárquicamente estructurados. Algunos estudios sobre desigualdades étnicas o de clase en lo que toca a rendimientos educativos, tasas de matrimonios mixtos, asimilación de comunidades étnicas, etc. han ido arrojando así descripciones estadísticas de enorme valor para el teórico dispuesto a «interpretarlas»²⁹.

Así pues, en cuanto al progreso en la disponibilidad técnica para la recolección y el análisis de los datos, la sociología basada en las técnicas de sondeo está hoy a años luz de las expectativas de sus primeros promotores. No acaba de entenderse muy bien la extendida tendencia a analizar tales datos sin la guía de teoría, ni tampoco porqué, a pesar de tantísima información y pericia técnica en el muestreo y los procedimientos de simulación, en buena parte de la sociología contemporánea domina hoy un discurso no explicativo, tendente a oponer resistencia a cualquier tipo de compromiso entre la compenetración interpretativa, la imaginación teórica, la habilidad matemática y las herramientas disponibles a la sociología empírica contemporánea (cfr. Boudon, 1995: 233).

En el presente estadio de desarrollo, la disciplina bien podría beneficiarse de una alianza entre la sociología dedicada al análisis de grandes conjuntos de datos y la sociología analítica comprometida con matrices teóricas fecundas en supuestos interpretativos. Si, por ejemplo, como sugiere Goldthorpe (op. cit., 2000: 406-07), aquellos teóricos que suscriben matrices formales ricas en mecanismos, acabasen adoptando las estructuras estadísticas regulares resul-

27. Cfr., por ejemplo, Goldthorpe (1987); Hope (1981); Mayer y Carroll (1987); Van Leeuwen y Maas (1996), entre los citados por Goldthorpe (1997: 409).

28. Ver, por ejemplo, investigaciones sobre movilidad social, como la de Erikson y Goldthorpe (1992); estudios comparativos sobre desigualdades educativas y de movilidad, como el de Ishida y otros (1995), o el de Smeeding y otros (1990) o Atkinson y otros (1995), sobre la distribución del ingreso y las dinámicas de pobreza y exclusión social.

29. Goldthorpe (1987: 409) cita aquí a Smith y Tomlison (1989), y Jones (1994).

tantes de las investigaciones aplicadas como sus *explananda* preferidos, entonces podrían llegar a interpretarlas dentro de una narrativa general de la acción, haciéndolas inteligibles a la luz de mecanismos intencionales y acabando por presentarlas sea como consecuencias laterales u otras de acciones individuales. Frente a la mejora en la estructura de los *explananda*, los sociólogos analíticos, en efecto, podrían contribuir a cerrar el hiato entre investigación y teoría aportando mejores mecanismos a la interpretación de los mismos.

Mientras, en general, la importancia del análisis causal y la explicación han retrocedido más que avanzado en relación a las propuestas de hace tres décadas, en la actualidad se dan condiciones de madurez analítica y técnica que, al menos en lo que respecta a ciertos sectores de la sociología contemporánea, alientan la esperanza de poder engarzar una evidencia empírica mucho mejor tratada con mecanismos causales explicativos también hoy mucho mejor conocidos y sistematizados, particularmente dentro de la narrativa general de la acción-intención (cfr. Goldthorpe, 2000: 239). No faltan, pues, razones para cierto optimismo. La empresa se plantea prometedora. Queda por ver lo que de ella harán las generaciones futuras de sociólogos.

Referencias bibliográficas

- ATKINSON, J.; RAINWATER, L. y otros (1995). «Income Distribution in Advanced Economics: Evidence from the Luxembourg Income Study (LIS)». *LIS Working Paper*.
- BAIN, R. (1927). «Trends in American Sociology». *Social Forces*, 5 (1926-27): 413-422.
- BANNISTER, R.C. (1987). *Sociology and Scientism: The American Quest for Objectivity, 1880-1940*. Chapel Hill, NC: University of North Carolina Press.
- BENDIX, R. y otros (comp.) (1991). *La sociología como profesión: autobiografía intelectual de veinte sociólogos americanos*. Madrid: Ministerio de Trabajo y Seguridad Social. Ediciones de la Revista de Trabajo, 42, 1993.
- BERK, R.A. (1988). «Causal inference for sociological data». En: SEMELSER, N.J. (ed.): 155-172.
- BERKANE, M. (ed.) (1997). *Latent Variable Modeling with Applications to Causality*. Nueva York: Springer-Verlag.
- BERNERT, C. (1983). «The career of causal analysis in American sociology». *British Journal of Sociology*, 24: 230-54.
- BIRNBAUM, I. (1981). *An Introduction to Causal Analysis in Sociology*. Londres: MacMillan Press.
- BLALOCK, H. Jr. (1962). *Estadística Social*. México: FCE, 1964.
- (1964). *Causal Inferences in Non experimental Research*. Chapel Hill: University of North Carolina Press.
- (1969). «Recasting verbal theories as causal models». En: *Theory Construction*, Englewood Cliffs: Prentice Hall (cap. 3).
- (1973). *Causal Models in the Social Sciences*. Nueva York: Aldine/Atherton.
- (1980). «Measurement and conceptualization problems: The major obstacle to integrating theory and research». En BLALOCK (ed.) (1980): 31-47.
- (1984). *Theory Construction. From Verbal to Mathematical Formulations*. Englewood Cliffs, N. J.: Prentice-Hall, Inc. 1979. Edición castellana: México: Trillas, 1984.

- BLAU, P.M.; DUNCAN, O.D. (1967). *The American Occupational Structure*. Nueva York: Wiley.
- BOUDON, R. (1965a). «A method of linear causal analysis: Dependence analysis». *American Sociological Review*, 30: 365-374.
- (1968). «A new look at correlation analysis». En: BLALOCK y BLALOCK (eds.).
- (1979). *La lógica de lo social*. Traducción castellana: Madrid: Rialp, 1981.
- (1979a). *Effets pervers et ordre social*. Traducida al castellano por Ediciones Premiá, México, 1980.
- (1979b). «Generating models as a research strategy». En: MERTON, COLEMAN y ROSSI (eds.), 1979: 51-64.
- (1996). «Pourquoi devenir sociologue? Reflexions et evocations». *Revue Française de Science Politique*, 46 n° 1: 52-79.
- BOUDON, R.; LAZARSFELD, P. (ed.) (1966). *Metodología de las ciencias sociales*. Volumen II: *Análisis empírico de la causalidad*. Barcelona: Laia, 1974.
- BRIDGMAN, P.W. (1927). *The Logic of Modern Physics*. Nueva York: MacMillan.
- BRYANT, C.G.A. (1985). *Positivism in Social Theory and Research*. Houndmills, Hampshire y Londres: MacMillan Publishers LTD.
- (1975). «Positivism reconsidered». *Sociological Review*, 23: 397-412.
- BUNGE, M. (1959). *Causality and modern science*. 3a ed. Nueva York: Dover Publications, 1979.
- (1999). *The Sociology-Philosophy connection*. Traducción castellana: México: Ediciones Edaf, 2000.
- CAMIC, CH.; XIE, Y. (1994). «The statistical turn in American social science: Columbia University, 1890 to 1915». *American Sociological Review*, 59 (octubre): 773-805.
- CHAPIN, F.S. (1947). *Experimental Design in Sociological Research*. Nueva York: Harper.
- CLOGG, C.G. (1992). «The impact of sociological methodology on statistical methodology». *Statistical Science*, 7: 283-297.
- COLEMAN, J.S. y otros (1966). *Equality of Educational Opportunity*. Washington, D.C.: Government Printing Office.
- DEMPSTER, A.P. (1990). «Causality and Statistics». *Journal of Statistics Planning and Inference*, 25: 261-78.
- DESROSIÈRES, A. (1985). «Histoire de formes: Statistiques et sciences sociales avant 1940». *Revue Française de Sociologie*, 26: 277-310.
- DUNCAN, O.D. (1975). *Introduction to Structural Equation Models*. Nueva York: Academic Press.
- (1970). «Partials, partitions and paths». En: BORGATTA, E. (ed.). *Sociological Methodology*. San Francisco: Jossey Bass.
- (1966). «Path analysis: sociological examples». *American Journal of Sociology*, 72: 1-1.
- ERIKSON, R.; GOLDTHORPE, H.J. (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.
- ERIKSON, R.; JONSSON, J.O. (eds.) (1996). *Can Education be Equalized? The Swedish Case in Comparative Perspective*. Boulder, Col.: Westview Press.
- ERIKSON, R.; GOLDTHORPE, H.J. y otros (1979). «Intergenerational Class Mobility in three western European societies». *British Journal of Sociology*, 30: 415-41.
- FARIS, R.E.L. (1967). *Chicago Sociology, 1920-1932*. San Francisco: Chandler.
- FIENBERG, S.E. (1977). *The Analysis of Cross-Classified Categorical Data*. Cambridge, Mass.: MIT Press.

- FISHER, R.A. (1926). «The arrangement of field-experiments». *Journal of Ministry of Agriculture*, 33: 503-513.
- (1935). *The Design of Experiments*. Nueva York: Hafner Publishing.
- FORNELL, C. (ed.) (1982). *A Second Generation of Multivariate Analysis*. Nueva York: Praeger.
- FOX, J. (1984). *Linear Statistical Models and Related Methods*. Nueva York: John Wiley.
- GIDDINGS, F.H. (1924). *The Scientific Study of Human Society*. Chapel Hill, N.C.: University of North Carolina Press.
- (1909). «The teaching of Sociology in the United States». *American Journal of Sociology*, 15 (1909-1910): 164-95; 192-211 (discussion). Contribution to discussion of Luther L. Bernard.
- (1903). «A theory of social causation». *Publications of the American Economic Association*. 3a serie, 5: 2 (mayo de 1904).
- GILBERT, G.N. (1981). *Modelling Society: An Introduction to Loglinear Analysis for Social Researchers*. Londres: Allen & Unwin.
- GILLIN, J.L. (1927). «The development of Sociology in the United States». *Publications of the American Sociological Society*, 24: 2 (1930).
- GLOCK, C.Y. (1979). «Organizational innovation for social science research». En: MERTON, R.K. y otros (eds.) (1979).
- GOLDBERGER, A.S. (1973). «Structural equation models. An overview». En: GOLDBERGER, A.S.; DUNCAN, O.D. (eds.). *Structural Equation Models in the Social Sciences*. Nueva York: Seminar Press, 1-18.
- (1970). «On Boudon's Method of Linear Causal Analysis». *American Sociological Review*, 35: 97-101.
- GOLDBERGER, A.S.; DUNCAN, O.D. (eds.) (1973). *Structural Equation Models in the Social Sciences*. Nueva York: Seminar Press.
- GOLDTHORPE, J.H. (2000). *On Sociology. Numbers, narratives and the integration of research and theory*. Oxford: Oxford University Press.
- (2000a). «Sociology and the probabilistic revolution, 1830–1930: Explaining an absent synthesis». En: *On Sociology: Numbers, narratives and the integration of research and theory*. Oxford: Oxford University Press: 259-295.
- (1998). *The Quantitative Analysis of Large-Scale Data Sets and Rational Action Theory: For a Sociological Alliance*. Londres: JAI Press.
- (1997). «The integration of sociological research and theory. Grounds for optimism at the end of the twentieth century». *Rationality and Society*, 9, 4: 405-426.
- (1996). «The quantitative analysis of large-scale data sets and Rational Action Theory». *European Sociological Review*, 12: 109-26.
- GOLDTHORPE, J.H.; LLEWELLYN, C. y otros (1987). *Social Mobility and Class Structure in Modern Britain*. 2a ed. Oxford: Clarendon Press.
- GREENWOOD, E. (1945). *Experimental Sociology: a Study in Method*. Nueva York: King's Crown Press.
- HAAVELMO, T. (1943). «The statistical implication of a system of simultaneous equations». *Econometrica*, 11: 11-12.
- HALFPENNY, P. (1982). *Positivism and Sociology: Explaining Social Life*. Londres: Allen & Unwin.
- HEDSTRÖM, P.; SWEDBERG, R. (eds.) (1996). «Rational Choice, Empirical Research, and the Sociological Tradition». *European Sociological Review*, 12: 127-46.
- (eds.) (1999). *Causal Mechanisms. An Analytical Approach to Social Theory*. Cambridge: University Press.

- HOPE, K. (1981). «Trends in the openness of British society in the present century». *Research in Social Stratification and Mobility*, 1: 127-70.
- ISHIDA, H. y otros (1995). «Class origin, class destination, and education: A cross-national study of ten industrial nations». *American Journal of Sociology*, 60: 145-93.
- JENKS, C.; SMITH, M. y otros (1972). *Inequality. A Reassessment of the Effect of Family and Schooling in America*. Nueva York: Basic Books.
- JONES, F.L. y otros (1994). «Comparative Social Fluidity: Trends over time in father-to-son mobility in Japan and Australia, 1965-1985». *Social Forces*, 72: 775-98.
- KENDALL, P.; LAZARSFELD, P. (1950). «Problems of Survey Analysis». En: MERTON, R.; LAZARSFELD, P. *Continuities in Social Research*. Glencoe, Ill.: The Free Press: 133-196.
- KRÜGER, L. (1987). «The slow rise of probabilism: Philosophical arguments in the XIXth century». En: KRÜGER; DASTON; HEIDELBERG (eds.). *The Probabilistic Revolution, I: Ideas in History*. Cambridge, Mass.: MIT Press.
- LAZARSFELD, P. (1946). «La interpretación de las relaciones estadísticas como propiedad de investigación». Comunicación presentada al Congreso de la Sociedad Americana de Sociología (Cleveland). Inicialmente publicado en LAZARSFELD & ROSENBERG (eds.) (1955) y reproducido en *Metodología de las ciencias sociales*, vol. II, R. Boudon y P. Lazarsfeld (comp.), 1966, edición castellana en Laia.
- (1950). «Problems of Survey Analysis». En: MERTON, R. K.; LAZARSFELD, P. *Continuities in Social Research. Studies in the Scope and Method of «The American Soldier»*. Glencoe, Illinois: The Free Press.
- (1953). «Some Historical Notes on the Empirical Study of Action». Aquí citado en la versión francesa *Philosophie des sciences sociales*. Precedida de una introducción de R. Boudon: «A propos d'un livre imaginaire». París: Gallimard, 1970.
- (1958). «Evidence and inference in social research». *Daedalus*, otoño de 1958, secc. III.
- (1962). «The sociology of empirical social research». *American Sociological Review*, 27, 6, 757-767.
- (1970a). «Quelques remarques historiques sur l'analyse empirique de l'action». En LAZARSFELD (1970), p. 163-184.
- (1979). «De los conceptos a los índices empíricos», extracto de *Evidence in social research*, en *Daedalus*, 87, 4: 99-109, 1958. Citado de la edición castellana BOUDON & LAZARSFELD (eds.) (1979): 35-46, Barcelona: Laia (segunda edición).
- LAZARSFELD, P.; PASANELLA, R.A.; ROSENBERG, M. (eds.) (1955). *The Language of Social Research*. Nueva York: The Free Press.
- LAZARSFELD, P.; OBERSCHALL, A. (1965). «Max Weber and Empirical Social Research». *American Sociological Review*, 30: 185-199.
- LEWIS, J.D.; SMITH, R.L. (1980). *American Sociology and Pragmatism*. Chicago: Chicago University Press.
- LIEBERMAN, S. (1985). *Making it Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.
- LONGMAN, GREEN & CO. (1939). «Contemporary positivism in sociology». *American Sociological Review*, 4: 42-52.
- (1955). «The natural science trend in sociology». *American Journal of Sociology*, 61: 191-202.
- LUNDBERG, G.A. (1929). *Social Research: A Study in Methods of Gathering Data*. Nueva York: Longman, Green & Co.
- (1939). *Foundations of Sociology*. Nueva York: MacMillan.

- (1939a). «Contemporary positivism in sociology». *American Sociological Review*, 4: 42-52.
- (1947). *Can science save us?* Nueva York: Longmans, Green & Co.
- (1955). «The natural science trend in sociology». *American Journal of Sociology*, 61: 191-202.
- MCCLOSKEY, D.N.; ZILLAK, S.T. (1996). «The standard errors of regression». *Journal of Economic Literature*, 34: 97-114.
- MACIVER, R.M. (1931). *Society: Its Structure and Changes*. Nueva York: Long & Smith.
- MCKIN, V.R.; TURNER, S.P. (eds.) (1997). *Causality in Crisis?* Notre Dame, Ind.: University of Notre Dame Press.
- MARINI, M.M.; SINGER, B. (1988). «Causality in the Social Sciences». *Sociological Methodology*: 347-409.
- MAYER, K.U.; CARROLL, G.R. (1987). «Jobs and classes: Structural constraints on career mobility». *European Journal of Sociology*, 3: 14-38.
- MEAD, G.H. (1934). *Persona, mente y sociedad*. Varias ediciones castellanas.
- MERTON, R.K. (1948). «Discussion» [of Talcott Parsons, «The position of sociological theory»]. *American Sociological Review*, 13: 164-68.
- (1949). *Teoría y estructura sociales*. México: Fondo de Cultura Económica, edición castellana 1964.
- (1967). «On Sociological Theories of the Middle Range». En POWELL, W.W.; ROBBINS, E. (eds.). *Conflict and Consensus*. Nueva York: The Free Press: 262-83.
- MERTON, R.K.; LAZARSFELD, P. (1950). *Continuities in Social Research. Studies in the Scope and Method of «The American Soldier»*. Glencoe, Illinois: The Free Press.
- (1989). *The History of Econometric Ideas*. Cambridge: Cambridge University Press.
- MULLINS, N.C.; MULLINS, C.J. (1973). *Theories and Theory Groups in Contemporary American Sociology*. Nueva York: Harper & Row.
- NEWELL, A.; SIMON, H.A. (1976). «Computer Science as empirical inquiry: Symbols and research». *Communications of the ACM*, 19: 113-126.
- O'RAND, A.M. (1992). «Mathematizing social science in the 1950s: The early development and diffusion of game theory». *History of Political Economy*, 24: 177-204.
- OBERSHALL, A. (ed.) (1972). *The Establishment of Empirical Sociology: Studies in Continuity, Discontinuity and Institutionalization*. Nueva York y San Francisco: Harper & Row, Publishers.
- (1965). *Empirical Social Research in Germany, 1848-1914*. The Hague: Mouton.
- PEARSON, K. (1897). *The Grammar of Science*. Londres: Black.
- PORTER, T.M. (1986). «Lawless Society: Social Science and the reinterpretation of statistics in Germany, 1850-1880». En: KRÜGER, L.; DASTON, L.J.; HEIDELBERG, M. (eds.). *The Probabilistic Revolution, I: Ideas in History*. Cambridge, Mass.: MIT Press.
- ROSEBERG, M.; LAZARSFELD, P.F. (eds.) (1955). *The Language of Social Research*. Glencoe, Ill.: The Free Press. Editado en francés por R. Boudon, P. Lazarsfeld y C. Chazel. Versión castellana: *Metodología de las ciencias sociales* (3 volúmenes). Barcelona: Laia, 1973-1975.
- ROSS, D. (1991). *The Origins of American Social Science*. Cambridge: Cambridge University Press.
- SARIS, W.; STONKHORST, H. (1984). *Causal Modeling in Nonexperimental Research*. Amsterdam: Sociometric Research Foundation.
- SCHAD, S.P. (1972). *Empirical Social Research in Weimar Germany*. París: Mouton.
- SIERRA BRAVO, R. (1991). *Diccionario práctico de estadística*. Madrid: Editorial Paraninfo.

- SIMON, Herbert (1957). *Models of Man: Social and Rational. Mathematical Essays on Rational Human Behavior in a Social Setting*. Nueva York: John Wiley & Sons, Inc. Tercera reimprisión 1964.
- (1954a). «Spurious correlation: A causal interpretation», en 1957: 37-49.
- (1954b). «Some strategic considerations in the construction of social science models». En: LAZARSFELD, P. F. (ed.). *Mathematical Thinking in the Social Sciences* (capítulo 8). Glencoe, Il.: The Free Press.
- (1953). «Causal ordering and identifiability», en 1957: 11-36.
- (1952). «On the definition of causal relation», en 1957: 50-61.
- SMEEDING, T.M.; RAINWATER, L. y otros (1990). *Poverty, Inequality and the Distribution of Income in an Internacional Context*. Londres: Wheatsheaf.
- SMITH, D.J.; TOMLISON, S. (1989). *The School Effect: A Study of Multiracial Comprehensives*. Londres: Policy Studies Institute.
- SMELSER, N.J. (ed.) (1988). *Handbook of Sociology*. Beverly Hills, California: SAGE Publications.
- SOBEL, M.E. (1998). «Causal inference in statistical models of the process of socio-economic achievement». *Sociological Methods and Research*, 27: 318-48.
- SÖRENSEN, A.B. (1998). «Theoretical mechanisms and the empirical study of social processes». En: HEDSTRÖM, P.; SWEDBERG, R. (eds.).
- STIGLER, S.M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, Mass.: Harvard University Press.
- STOUFFER, S.A. (1931). «Experimental comparison of a statistical and a case history technique of attitude research». *Publications of the American Sociological Society*, 25: 154-156.
- (1949-50). «Some observations on study design». *American Journal of Sociology*, LV : 355-61.
- (1950). «Afterthoughts of a contributor». En: MERTON & LAZARSFELD (eds.): 197-211.
- STOUFFER, S.A., et al. (1949). *The American Soldier: Studies in Social Psychology in World War* (4 vols.).
- SUPPES, Patrik (1982). «Problems of causal analysis in the Social Sciences». *Epistemologia*, 5: 239-250.
- (1981). «Causal analysis and hidden variables». *PSA 1980*, 2: 563-571.
- SWAMY, P.A.V.B.; COWAY, R.K.; VON ZUR MUELHEN (1985). «The foundations of econometrics: Are there any?». *Economic Reviews*, 4:1-61.
- TAYLOR, Carl C. (1920). «The social survey and the science of sociology». *American Journal of Sociology*, 25 (1919-1920): 731-756.
- TOLMAN, E.C. (1986). *The Search of Methodology of Social Science: Durkheim, Weber and the Nineteenth Century Problem of Cause, Probability and Action*. Boston: Boston Studies in the Philosophy of Science.
- TOLMAN, E.C.; FACTOR, R.A. (1994). *Max Weber: The Lawyer as Social Thinker*. Londres: Routledge.
- TOLMAN, E.C.; TURNER, J.H. (1990). *The Impossible Science: An Institutional Analysis of American Sociology*. Newbury Park, California: Sage.
- UPTON, J.G. (1978). *The Analysis of Cross-Tabulated Data*. Nueva York: Wiley.
- VAN LEEUWEN, M.H.D.; MAAS, I. (1996). «Long-term social mobility: Research, Agenda and a case study (Berlin 1825-1957)». *Continuity and Change*, 11: 399-433.
- WEBER, M. (1922/1968). *Economy and Society*. Versión castellana del FCE (2 vols.), 1944.

- WELLS, R.H.; PICOU, J.S. (1981). *American Sociology: Theoretical and Methodological Structure*. Washington: University Press of America.
- WRIGHT, S. (1974). «Path coefficients and path regressions: Alternative or complementary concepts», en *Causal Models in the Social Sciences*. Londres: Macmillan, reimpresión.
- (1960). «Path coefficients and path regression». *Biometrika*, 16: 189-202.
- (1921). «Correlation and Causation». *Journal of Agricultural Research*, 20: 557-585.
- YULE, G.U. (1903). «Notes on the Theory of Association of Attributes in Statistics». *Biometrika*, 2: 121-134.
- (1899). «An investigation into the causes of changes in pauperism in England, chiefly during the last two intercensal decades». *Journal of the Royal Statistics Society*, 69: 87-130.
- (1897). «On the theory of correlation». *Journal of the Royal Statistics Society*, 59: 318-349.
- YULE, G.U.; KENDALL, P. (1937). *An Introduction to the Theory of Statistics*. 11a ed. Londres: C. Griffin.
- ZEISEL, Hans (1947). *Say it with Figures*. Nueva York: Harper & Brother.