

Consideraciones epistemológicas sobre algunos temas de la estadística social: una mirada desde la epistemología genética¹

Fernando Cortés

Introducción

A FINES DEL SIGLO XX LA ESTADÍSTICA, como casi la totalidad de las disciplinas científicas, ha explotado en una serie de campos especializados. Actualmente el número de personas involucradas en las actividades académicas es formidable, aunque se concentra básicamente en los países desarrollados. La combinación de estas dos tendencias arroja como resultado una voluminosa producción en cualquier campo especializado del saber, que se transmite por medio de avalanchas de artículos y libros que hacen imposible estar al día en todos los campos de una disciplina específica.

La estadística no es la excepción. Se ha desmembrado en una serie de campos particulares que normalmente asumen el nombre propio de una técnica estadística particular. Obviamente, este artículo no abarcará a la estadística en general, lo cual estaría mucho más allá de mis fuerzas, sino que los planteamientos que se presentan se referirán únicamente a la estadística aplicada a las ciencias sociales. Debo dejar claro que este recorte del campo no es poca cosa. Si bien el desarrollo de la estadística estuvo muy vinculado al de la física, especialmente en lo que se refiere al cálculo de probabilidades, no debe olvidarse que a la vuelta

¹ Agradezco los comentarios de Rosa María Rubalcava así como de los árbitros anónimos que evaluaron una versión anterior de este trabajo. Además, me beneficié de las ricas discusiones que respecto a estos temas tuvieron lugar en las sesiones del “Seminario de los jueves” de la Sección de Teoría y Metodología de la Ciencia, del Centro de Investigación y de Estudios Avanzados del IPN.

del siglo XIX, los problemas sociales vividos por el antiguo imperio condujeron a sembrar las bases del análisis de correlación y de regresión (Karl Pearson), así como del análisis de asociación (Yule) (MacKenzie, 1979:39-50). A comienzos del siglo XX los problemas de medición de la inteligencia llevaron a fundar el análisis factorial. A mediados de los cincuenta los trabajos pioneros de Goodman y Kruskal buscaron, por una parte, desarrollar instrumentos más poderosos de análisis de variables no métricas y, por otra, conectar los enunciados teóricos especificados en hipótesis estadísticas, con las distribuciones esperadas de las observaciones. Sin embargo, fue a partir de los años setenta cuando las demandas analíticas planteadas por las ciencias sociales empezaron a orientar los desarrollos estadísticos de punta, que combinados con los avances de la microcomputación de los años ochenta, hicieron posible poner en práctica modelos analíticos que ya se habían venido desarrollando en el pasado, pero con pocas posibilidades de aplicación. En la actualidad el campo de las ciencias sociales ha sido invadido en gran medida por diversas técnicas de análisis de variables múltiples, entre las que destacan el análisis loglineal, las regresiones logísticas y el análisis de correspondencias.

No es el propósito de este artículo presentar un análisis teórico-histórico de la evolución de la estadística; sin embargo, sería útil para el lector tener una perspectiva panorámica de la invasión del análisis de variables múltiples en la estadística y en la investigación social empírica.

A mediados de los años cincuenta aparece en el centro del desarrollo de la estadística aplicada el problema de las relaciones espurias o ilusorias. Tanto Paul Lazarsfeld como Hubert Simon se preocupan por desarrollar técnicas que permitan detectar si una relación observada es genuina o ilusoria. Son famosos los ejemplos de Lazarsfeld sobre la relación entre los nacimientos y las cigüeñas, y la del número de extinguidores y los daños provocados por un incendio. Ambos encuentran la misma solución al problema, aunque el primero en el campo del análisis de asociación y el segundo en el de la correlación. La solución consistió en estudiar la asociación o la correlación parcial entre dos variables controlando mediante la presencia de una tercera. En los ejemplos señalados, la variable control (también llamada variable *test*) fue el contexto rural urbano, en el primero, y en el segundo la gravedad del incendio: *si la relación observada desaparece al controlar por una tercera variable, entonces se concluye que esa relación es espuria o ilusoria*. Si no desaparece, la pregunta inmediata es ¿no habrá otra u otras variables que harían desaparecer la relación?, en cuyo caso, de acuerdo con el criterio empleado, la relación sería ilusoria. Me parece que esto fue una de las

fuerzas, nacidas en el dominio de las ciencias sociales, que empujó a desarrollar el análisis de variables múltiples.

Otra cuestión que impulsó los esfuerzos dedicados al análisis de variables múltiples fue la disparidad que existía entre las complejidades del pensamiento conceptual y la disponibilidad de instrumentos. De hecho, tal como se verá en la próxima sección, una de las maneras de utilizar la estadística consiste en expresar los enunciados teóricos en lenguaje estadístico. En los años sesenta las conceptualizaciones podían establecer conjuntos complejos de relaciones entre variables, pero los instrumentos estadísticos para tratar variables no métricas se limitaban al análisis de asociación, o en el mejor de los casos al análisis de covarianzas de Lazarsfeld, y cuando todas las variables eran métricas, al análisis de regresión. El instrumento disponible imponía un corsé a la validación empírica de la teoría. El análisis de covarianzas se desarrolló matemáticamente para considerar al mismo tiempo tres variables dicotómicas. A pesar de que Lazarsfeld planteó que la generalización a más de dos categorías y a más de tres variables era *a fortiori*, la verdad es que ése no fue el caso. Por aquella época no se disponía de una técnica de análisis de datos que considerara simultáneamente cualquier número de variables no métricas sin importar el número de categorías, sujeta sólo a la restricción impuesta por el total de observaciones. La regresión se podía extender para incluir en el modelo variables explicativas no métricas (variables ficticias), pero sólo era útil para formalizar teorías que originaban modelos en que una variable endógena se explicaba por un conjunto de variables exógenas. La camisa de fuerza que imponía la técnica de análisis de datos al contraste de las teorías fue otro motor que impulsó el desarrollo del análisis multivariado.

La tercera vertiente se desprende de la lógica. Es fácilmente comprensible desde esta disciplina y perfectamente comprobable por medio de ejemplos numéricos, que es erróneo (en el sentido de que las conclusiones del análisis son falsas) establecer una relación multidimensional entre un conjunto de variables basándose en las relaciones bidimensionales entre cada factor o variable explicativa o condicionante y la explicada, consecuente o condicionada. La explicación a este hecho es simple: toda relación bidimensional no controla la presencia de las restantes variables. Esto se traduce en que la asociación observada no sólo incluye la relación entre las dos variables en cuestión sino que, además, incorpora los vínculos que tienen con terceras, cuartas..., variables. Así, la variable explicativa aparece *apropiándose* de efectos que no le pertenecen. La única manera de resolver este problema es cerrar el campo, es decir, disponer de teorías suficientemente poderosas, que permitan identificar los factores relevantes, que éstos se puedan traducir en variables

estadísticas válidas y confiables, y proceder a realizar el análisis de datos utilizando simultáneamente todas las variables. Es decir, analizar usando alguna técnica de variables múltiples.

Desde el punto de vista analítico, las tres fuentes que habrían impulsado el desarrollo de la estadística de variables múltiples se pueden reducir únicamente a dos. La denuncia de las relaciones espurias, y su contrapartida, la identificación de las genuinas, así como la imposibilidad de establecer las relaciones multidimensionales a partir de las bidimensionales, llaman, en esencia, al control de las otras variables que puedan tener incidencia. Es decir, ambas fuentes se originan en un esquema deficiente de control de variables. La otra fuente es una expresión de la necesaria interrelación entre los avances teóricos de las ciencias sociales y en especial de la sociología, la capacidad de traducir los enunciados en proposiciones que se pueden expresar dentro del lenguaje de la estadística y los requerimientos que se derivan sobre el tipo de preguntas a las que debería dar respuesta el análisis de los datos.

Este trabajo se limita exclusivamente a considerar algunos temas epistemológicos a partir del uso de la estadística en las ciencias sociales y especialmente en la sociología. El acento está puesto en la estadística. No se trata de un análisis epistemológico de esta disciplina, sino más bien de algunos problemas epistemológicos, que a mi juicio están entretenerados en la práctica de la investigación social empírica que utiliza los métodos estadísticos para someter a prueba sus hipótesis. Puesto en otros términos, esto quiere decir que el eje del desarrollo lo constituirá la relación entre los enunciados teóricos y su validación empírica. Delimitado así el campo de este artículo, excluye consideraciones sobre el marco epistémico (García, 1986:52 y 53), sobre el efecto del paradigma del observador en la observación (Russell, 1965), o bien el papel que juega la dimensión Z (Holton, 1985:27-29).

Voy a dedicar la parte central de este artículo a tocar tres temas que considero de interés epistemológico y que se encuentran incluidos en el uso que se da a cualquier técnica estadística, desde las más elementales a las más complejas. El primero de estos temas se referirá a la perspectiva con que se usa el instrumento estadístico. El segundo, al tratamiento de la causalidad en estadística; y el tercero y último tendrá por objeto hurgar en las preconcepciones en que descansan dos de los métodos de estimación más utilizados.

El uso del instrumental estadístico: dos enfoques y dos síntesis, pero una más sintética que la otra

Existen dos enfoques dominantes para exponer el material de la estadística. Uno de ellos intenta generar conocimiento a partir de regularidades empíricas, en tanto que el otro adopta una postura de docimasia de hipótesis. Abordaremos con un poco más de detenimiento ambos enfoques.

El primero, que ha dominado ampliamente, consiste sobre todo en encontrar regularidades en los datos y elevarlas al rango de hallazgos o *findings*. Por ejemplo, muchos libros de texto plantean que el análisis de regresión debe iniciarse con un diagrama de dispersión que representa los pares ordenados de las variables en un plano de coordenadas cartesianas, para establecer si están o no relacionadas y determinar la forma de la relación.

Como la gráfica de tres variables origina una representación tridimensional, que no es de tan fácil interpretación, los libros señalan que en caso de haber tres o más variables es conveniente usar diagramas bidimensionales entre la variable explicada y las explicativas. Es así como la investigación se transforma en una cacería de relaciones significativas, para posteriormente montar un modelo que reúna en una sola expresión matemática (habitualmente algebraica) el conjunto de variables explicativas con la variable explicada, procediéndose con posterioridad al ajuste del modelo. Esta perspectiva no sólo se reduce al análisis de regresión; es aplicable tanto a la estadística descriptiva, como a los análisis de correlación, loglineal, varianza, regresión logística, etc. En los últimos años ha asumido un nombre particular: análisis exploratorio de datos, que incluso ya se encuentra incluido en algunos programas estadísticos para computadoras.

Tal vez no es necesario abundar respecto a este enfoque porque su predominio en la enseñanza de la estadística lo hace fácilmente reconocible para cualquier lector que se haya visto en la necesidad de estudiar o aplicar esta disciplina. Lo que sí me parece interesante es rescatar y discutir un par de ideas epistemológicas que sostienen esta práctica.

Detrás de esta forma de exponer el conocimiento estadístico o de “hacer ciencia” se oculta una de las ideas centrales del empirismo lógico. En efecto, al decir de Hempel, “El principio fundamental del empirismo moderno es la idea de que todo conocimiento no analítico se basa en la experiencia” (1965:115). En este caso particular, la experiencia se reduce a los datos estadísticos.² Véase al respecto la obra de Susana To-

² En aras de una comunicación fluida es necesario hacer algunas precisiones

irado, en colaboración con Miguel Acuña y Vittorio Lorenzi, *Informe sobre la población económicamente activa en América Latina: 1940-1970* (Santiago, Proelce, 1976). O la introducción de Susana Torrado a *Investigación e información sociodemográficas 2: los censos de población y vivienda en la década de 1980 en América Latina* (Buenos Aires, Clacso, 1981).³ También la obra de Emilio de Ipola y Susana Torrado, con la colaboración de Arturo León y Juan María Carrón, *Teoría y método para el estudio de la estructura de clases en Chile (con un análisis concreto: Chile, 1970)*, tres volúmenes, Santiago, Flacso-Proelce, 1976.⁴ En consecuencia, el conocimiento generado es significativo porque surge de ellos —de los datos estadísticos—, es decir, de la experiencia. Esta idea se aleja de la concepción empirista para la cual el conocimiento del mundo (en nuestro caso, del mundo social) surge de la recolección de datos y el descubrimiento de regularidades.

La idea de que los datos son dados, que constituyen el punto de partida de todo conocimiento, que se dan directamente a la percepción y que son neutros, es decir, que *son los mismos para todos los individuos y todas las disciplinas* (García, 1986:46), da el soporte para dis-

terminológicas. En estadística suelen usarse como sinónimos los términos datos y observaciones. Apoyados en el constructivismo genético definiremos los “observables” como datos de la experiencia ya interpretados, y los hechos como relaciones entre observables (García, 1986:49). Desde esta perspectiva el dato estadístico es un observable en la medida en que no es nada dado sino que son registros de la experiencia contruidos a partir de conceptos y teorías. Aún más, en América Latina esta idea dio origen a algunas instituciones que dedicaron gran parte de su actividad a cultivarla. Es el caso del programa Elas-Celade (Proelce), organismo creado en conjunto por la Escuela de Sociología de Flacso (Elas) y por el Centro Latinoamericano de Demografía, y el grupo de trabajo “sobre información sociodemográfica” dentro de la Comisión de Población y Desarrollo, de Clacso. Estas instituciones produjeron una nutrida cantidad de trabajos donde se mostraba y demostraba reiteradamente que el dato se construye.

³ Tal vez la obra culminante en esta línea sea el esfuerzo realizado por un conjunto de investigadores coordinados por Emilio de Ipola y Susana Torrado para organizar el censo chileno de 1970 de acuerdo con las categorías althusserianas.

⁴ La obra contiene tres partes que van desde la discusión conceptual a la operacionalización. El resultado consistió en una cuantificación de la estructura de clases en Chile hacia 1970. Pero no sólo hay que tomar en cuenta las determinaciones conceptuales, las técnicas y los métodos empleados en su construcción, sino también sus condicionantes políticas (Miles e Irvine, 1979:113-127). En la construcción del dato intervienen una serie de determinaciones que llevan a la conclusión, parodiando a Russell Hanson, de que “todo dato estadístico está cargado de teoría” y si se quiere ser más específico, de teoría en sentido amplio, no sólo limitada al dominio conceptual sino también a las teorías en que descansan los métodos y las técnicas de construcción de datos estadísticos, a lo que habría que agregar la “carga” política en su construcción. En este sentido los datos no están dados directamente a la percepción, ni tienen el mismo significado para todos los observadores.

criminar entre lo que es y no es científico. No es necesario agregar mayores comentarios al siguiente pasaje célebre de Hume:

[...] si tomamos en nuestras manos un volumen, de teología, metafísica escolar, por ejemplo, preguntémos: ¿contiene algún razonamiento abstracto concerniente a la cantidad o al número? No. ¿Contiene algún razonamiento experimental concerniente a la materia de hecho o cosas existente? No. Entregarlo a las llamas. Pues no puede contener más que sofistería y engaños.

La otra idea que legitima esta práctica está claramente expuesta en el estudio de Manuel Gil sobre la concepción de ciencia a comienzos del siglo xx. Al estudiar la posición de Mach, Gil concluye que: “Como ya ha afirmado, la tarea de la ciencia es, para Mach, dar cuenta de las relaciones funcionales de dependencia entre los diversos elementos” (p. 154). Un poco más adelante, en la misma página, subraya y extiende la misma idea “la propuesta de relaciones funcionales entre los diversos elementos es la tarea central de la actividad científica en todos los campos, aun en aquellos, como la economía, en los cuales no parecía ser así”.

El análisis que hace Gil de Duhem lleva a las mismas conclusiones:

[...] la parte explicativa puede variar, o incluso no existir para alguien, sin que esto afecte la certeza de las leyes descubiertas por observación de los hechos [...] la teoría del calor descansa realmente en los hechos generales que son inmediatamente dados por observación. El origen de las teorías está en los hechos y en ellos ha de tallarse para dar cuenta de la validez, con independencia de las hipótesis explicativas de las que aparentemente derivan (p. 168).

En la página 169 agrega una cita que Duhem hace de Fourier: “Este científico señala que las causas fundamentales no nos son conocidas, pero están sujetas a leyes simples y constantes que pueden ser descubiertas por observación”.

Me parece que estos antecedentes arrojan luz sobre las ideas en que se apoyan aquellos que utilizan la estadística para salir a la cacería de datos, establecer regularidades en el comportamiento de las variables y pescar sus relaciones. Poniendo un poco de atención a la discusión sostenida a comienzos del siglo, se nos hace claro que esta manera de entender y usar la estadística en la investigación exuda un fuerte tufillo de trasnochado empirismo. Digo trasnochado porque los empiristas actuales han ganado en sutileza, aunque no estoy seguro si en rigor.

Lo que llama poderosamente la atención es que, al finalizar el siglo xx, una fuerte dosis de empirismo recorre las aulas universitarias y el quehacer de la investigación. La supuesta certeza de la postura que sostiene que el único conocimiento válido es aquel que deriva del análisis de los "hechos" y que se limita a establecer relaciones funcionales entre las variables, conduce a posiciones autoritarias que cada vez dejan menos espacio para la investigación sustentada a partir de otros principios. Al parecer nunca se preguntan respecto a la construcción del dato estadístico, tampoco al porqué seleccionan unas variables y no otras, ni menos respecto a los criterios para decidir qué variables relacionar ni cómo distinguir las relaciones ilusorias de aquellas que tienen sentido.

También la estadística se puede entender y usar desde una óptica hipotético-deductiva. Esta perspectiva supone que tomando pie en la teoría se pueden derivar algunas hipótesis estadísticas para contrastarlas. Tal sistema suele dar paso a un uso ritual de la estadística y se encuentra bastante extendido entre aquellos que deben manejar datos en sus tesis de licenciatura, maestría o doctorado. El esquema que se sigue es el de la docimasia de hipótesis, que es uno de los temas centrales de la estadística elemental. La idea consiste en generar las distribuciones de probabilidades correspondientes a las hipótesis nula y alternativa, y preguntarse acerca de la probabilidad de que un estimador determinado (media aritmética, varianza, coeficiente de regresión, una interacción, etc.) pertenezca a la población definida por la hipótesis nula o a la(s) de la(s) hipótesis alternativa(s). En esta perspectiva la estadística rasura las hipótesis enviando al averno las rechazadas y manteniendo en el limbo a las no rechazadas.⁵ Lo que quiero resaltar es que en esta forma de conceptualizar la investigación empírica la estadística se convierte en el juez último que decide acerca de la validez de una hipótesis teóricamente fundada.

Queda aún por mencionar una postura intermedia. Una de las ópticas del empirismo moderno sostiene que *el enunciado cuantitativo que se refiera al grado de correlación entre dos variables se acerca ya más al estatuto de una ley*, pero que la relación observada se nos puede pre-

⁵ En el campo de la inferencia estadística predominan las ideas del falsacionismo popperiano. En los libros de texto se suele aseverar que las hipótesis no se rechazan pero que esto no quiere decir que se aceptan, sino más bien que no hubo suficiente evidencia empírica para rechazarlas. Esta idea se complementa afirmando que la actitud científica consiste en someter continuamente las hipótesis al rigor de la contrastación con el propósito de rechazarlas. En la metodología de las ciencias sociales se ha agregado el *dictum* que en tanto más resistente sea la hipótesis al rechazo, eliminando más y más hipótesis alternativas, más verosímil se hace (Stinchcombe, 1970:26-39).

sentar distorsionada por el azar, por lo que es más bien una tendencia, y agrega:

Para que esa interpretación sea razonable necesitamos disponer de algún fundamento, a poder ser teórico y empírico. Más precisamente, podemos suponer que una línea de tendencia oculta una ley sólo si *i*) los datos tienden efectivamente a fundirse en la línea y con ella cuando las perturbaciones se hacen despreciables o *ii*) se dispone de un modelo teórico que dé razón de la línea central subyacente al proceso causal (Bunge, 1979:346 y 347).

Si el “o” de esta cita fuera excluyente entonces se concluiría que Bunge concuerda con el empirismo lógico, ya que en ocasiones sería posible establecer leyes a partir de la pura empiria (condición 1) mientras que también da la posibilidad de establecer la conexión entre hipótesis teóricas, hipótesis estadísticas y datos (condición 2). Si el “o” no fuera excluyente, entonces Bunge exigiría el cumplimiento de ambas condiciones para establecer una ley. Independientemente de cuál sea la postura que sostiene, queda claro que la estadística juega el papel de instrumento de caza o bien se erige en el juez que define la validez de los hallazgos. Estas imágenes del papel de la estadística en la investigación en ciencias sociales difícilmente son reconocibles en la práctica.

En primer lugar hay que notar que el investigador define un problema de investigación que implica determinar un dominio en el cual investigará el o los fenómenos que le interesan. Este problema de investigación no surge del aire sino que se encuentra fuertemente condicionado por el conocimiento y el estadio de desarrollo de la o las teorías pertinentes al problema bajo estudio. A partir de aquí el investigador decide tomar algunos objetos y no otros, construir y estudiar solamente las variables relacionadas a los conceptos teóricos y analizar únicamente las relaciones que tienen sentido dentro del cuerpo de conocimiento acumulado (incluyendo las teorías adecuadas al análisis del problema). La estadística es un poderoso auxiliar para caracterizar y detectar relaciones significativas entre las variables bajo estudio, así como para apoyar el proceso de abstracción empírica, a partir del cual se construyen nuevas teorizaciones y conceptos que se aplican a un cierto dominio de la realidad.⁶ Este proceso de asimilación del material empírico a la con-

⁶ La abstracción empírica refiere a la lectura de los observables. Como ya hemos visto esta lectura es mediada por los conceptos y teorías, así como por los instrumentos lógicos y matemáticos a disposición del investigador. En tanto, la abstracción reflexiva es el proceso por el cual los conceptos se amplían a otros dominios (Piaget y García, 1982:190-194; Piaget, 1975:26).

ceptuación conlleva la generación de nuevos conceptos derivados de los primitivos que se aplican a nuevos dominios, utilizando para ello un proceso de abstracción reflexiva. En esta parte del proceso de investigación la estadística puede volver a jugar un papel, esta vez controlando la validez empírica de las hipótesis generadas por el proceso de abstracción reflexiva.

De lo anterior debemos destacar que esta forma de conceptualizar el proceso de investigación da un nuevo sentido a la posición empirista superándola por cuanto otorga un papel significativo a la experiencia. Pero va más allá a medida que reconoce que el recorte o el cierre del campo de investigación implica la movilización de conceptos, teorías y conocimiento acumulado. Pero también lo hace con la posición hipotético-deductiva al considerar que se generan nuevos conceptos y teorías que se aplican a otros dominios por medio del proceso de abstracción reflexiva, pero la supera en tanto reconoce la posibilidad de generación de nuevo conocimiento mediante la abstracción empírica. Tenemos entonces una conceptualización del proceso de investigación que a la vez que representa adecuadamente la actividad de los investigadores en su accionar, nos permite entender los papeles básicos que se le adscriben a la estadística dentro del proceso de investigación.

Se nos podría argumentar que no habría diferencias sustantivas entre este planteamiento y el presentado por la versión moderna del positivismo si considerásemos el "o" como no excluyente. Tal vez esta objeción podría ser aceptada, aunque habría que discutirlo; sin embargo, es claro que el planteamiento de Bunge se detiene en la contrastación. Una vez que se sometió a prueba el enunciado teórico o la hipótesis, el proceso es terminal llevando al rechazo o no de la hipótesis. Sin embargo, todos sabemos que en muchas ocasiones el rechazo de una hipótesis normalmente implica el inicio de una serie de acciones.

Suponiendo que no hay duda acerca de la selección del fenómeno a estudiar, antes de rechazar su hipótesis el investigador analiza si la relación tiene una estructura matemática diferente; por ejemplo, considera la posibilidad de que la relación sea no lineal, que su variable opere sólo en presencia de otras (interacción) y no por sí misma, que la asociación sólo sea válida para un subconjunto de la población y no para su totalidad (por ejemplo, que sólo sea válida para mujeres y no para hombres), etc. Una vez que despejó este conjunto de interrogantes y la situación no mejora, suele analizarse con detalle la validez de la medición, es decir, explicitar y sistematizar los conceptos incluidos en el dato estadístico y comparar su contenido con el concepto que se quiere medir; también se estudia la confiabilidad y se procura tener una idea acerca del proceso de construcción cuando se trata de información secundaria,

o bien se revisa el procedimiento seguido si la construcción fue propia. Una vez que se tiene la seguridad de que no hay dificultades mayores con los datos estadísticos, la atención del investigador se vuelca sobre la hipótesis.

El rechazo de una hipótesis obliga a acomodar la teoría y los conceptos, pero no necesariamente lleva a rechazar todo el planteamiento conceptual. La acumulación de contradicciones entre las consecuencias de una teoría y la experiencia termina por plantear la necesidad de pasar a otro estadio de la teorización. La nueva teoría contiene algunos conceptos de la anterior, es aplicable a los dominios de la teorización antigua y reconoce nuevos dominios. Además:

El producto (resultado) de una abstracción reflexiva con respecto a una teoría se torna la *base observable* (el punto de partida observable) de una abstracción empírica con respecto a la teoría de orden superior (Piaget y García, 1982:194).

Es claro ahora que el marco de referencia que nos ofrece la epistemología genética no sólo provee de una buena caracterización del proceso de investigación tal como acontece en la práctica, por lo menos en ciencias sociales, sino que supera la posición del empirismo lógico, la hipotético-deductiva y la versión moderna del empirismo. Adicionalmente el conceptualizar la investigación como lo que es, es decir, como un proceso, permite localizar el papel de la estadística tanto en los procesos de abstracción empírica como reflexiva.

Pasemos ahora a considerar el tratamiento de la causalidad en la estadística.

El análisis causal

En el año de 1964 Hubert Blalock publica en los Estados Unidos de Norteamérica su obra *Causal Inferences in Nonexperimental Research*; en 1968, en Francia, ve la luz el libro de Raymond Boudon, *L'Analyse Mathématique des Faits Sociaux*. Estas obras marcan el inicio de una moda que se extingue poco a poco a medida que se aproximan los años ochenta y pasa a ser una más del acervo de técnicas estadísticas, pero con el nombre de análisis de trayectorias, análisis de senderos o *path analysis*. ¿Por qué el cambio de nombre? ¿Es éste el indicio de procesos más profundos? Con el propósito de esbozar una respuesta echaremos una mirada a vuelo de pájaro a los principales hitos en la evolución del otrora llamado análisis causal.

Si bien el auge de la técnica se sitúa a mediados de los sesenta, sus bases se plantearon por primera vez en un artículo publicado por Sewall Wright en la revista *Journal of Agricultural Research*, en 1921 y en los *Annals of Mathematical Statistics* en 1934. Inexplicablemente Wright no vuelve a escribir sobre este tema hasta 1954, justamente el año en que Hubert Simon publicara su famoso "Spurious Correlation a Causal Interpretation" en el *Journal of the American Statistical Association*. Hacia el final de la década de los sesenta el influyente anuario *Sociological Methodology* dedicó su edición al análisis causal, y en los dos números siguientes el mismo tema ocupó un lugar destacado. De ahí en adelante el furor empezó a menguar hasta convertirse en una técnica estadística más.

Me parece que el apogeo del análisis causal en la sociología y la ciencia política tuvo un doble motor. Por una parte, fue el primer cuerpo de análisis estadístico producido por científicos sociales, y por lo tanto se consideró adecuado a los problemas que preocupaban a esas disciplinas en los sesenta, y por la otra, se suponía que permitía abordar empíricamente el estudio de la causalidad.

La técnica Simon-Blalock o los coeficientes de dependencia de Boudon permitían tratar estadísticamente el conjunto de relaciones entre las variables consideradas; es decir, a partir de los planteamientos de estos autores ya no era necesario limitarse a esquemas en que hubiese una variable explicada y varias explicativas; el nuevo tipo de análisis necesitaba hacer la distinción entre variables exógenas (no determinadas por el sistema de ecuaciones) y endógenas (determinadas por el sistema de ecuaciones). Los modelos admitían que cada variable endógena fuese función tanto de otras variables endógenas como de las exógenas. Se suponía que esta generalización de los tradicionales modelos de regresión y de correlación a la que habían arribado los científicos sociales permitía acercarse a la complejidad de los fenómenos que estudiaban, así como cerrar el hiato entre teoría y métodos o técnicas de investigación.

La euforia empezó a disminuir conforme esta técnica requería que las variables se midieran en escala métrica, ya fuera en escala de intervalo o de razón, en tanto que gran parte de las variables típicas en la sociología y la ciencia política son ordinales o nominales. Este hecho puso un límite claro a la universalidad de su aplicación. En esa época esta limitación podía ser tratada sólo en el caso de que las variables no métricas fueran exógenas al modelo. Sólo la investigación realizada durante los setenta, que desembocó en la regresión logística, permitió tratar el caso de variables endógenas no métricas.

El segundo motor se paró al ponerse en duda que esta técnica permitiría evaluar el ligamen causal entre las variables. El golpe más con-

tundente a la pretensión de los sociólogos y politólogos de haber creado un nuevo cuerpo estadístico se lo propinó un econometrista, Arthur Goldberger, quien en su artículo "On Boudon's Method of Linear Causal Analysis", publicado en 1970 en la *American Sociological Review*, demostró que el pretendido análisis causal no era más que un sistema de ecuaciones lineales recursivas en que se suponía que los residuos de las ecuaciones no estaban relacionados y que este tema se encontraba bien desarrollado en pocas páginas en cualquier libro elemental de econometría.

En la actualidad la estadística plantea que sólo sería posible establecer una relación causal entre dos variables si: *i*) hay una variación concomitante entre ellas, es decir, covarían o bien están correlacionadas; *ii*) presentan una asimetría temporal, o sea, es posible establecer cuál interviene primero y cuál después; y *iii*) se ha eliminado todo aquel factor causal que pueda producir la relación observada entre las variables (Asher, 1983:12). Las primeras dos condiciones parecerían no constituir obstáculos mayores. En efecto, la primera se resuelve calculando un simple coeficiente de correlación. La segunda a primera vista parece inocente, pero en muchas ocasiones no es posible establecer con claridad la prioridad temporal.

La tercera condición merece una consideración aparte. En efecto, el título del artículo de Simon hacía referencia a correlaciones espurias o ilusorias y la idea central de este trabajo estaba dirigida a utilizar el análisis causal para detectar este tipo de relaciones. Este mismo aspecto provocaba por aquella época la preocupación de Paul Lazarsfeld quien diseñó el análisis de covarianzas para reconocer las asociaciones ilusorias. Simon plantea, entonces, la posibilidad de utilizar las ideas de Wright para resolver empíricamente el vínculo causal entre las variables.

La tercera condición de Asher lleva directamente a considerar el control de variables. Para estudiar una relación entre dos o más variables, eliminando el efecto que puedan tener otras, la estadística dispone de varias estrategias. Una consiste en aleatorizar el efecto de las otras variables. Por ejemplo, si interesa establecer la relación entre los métodos alternativos para enseñar a leer y la velocidad de aprendizaje, se forman grupos con miembros asignados al azar de manera que la clase social, la inteligencia, la disponibilidad de capital cultural en el hogar, etc., estén igualmente presentes en todos los grupos, de modo que las diferencias que se observen en la velocidad de aprendizaje no se expliquen por alguna de ellas. Una idea equivalente, aunque técnicamente distinta, es la de las observaciones pareadas. En este caso se comparan dos o más observaciones que sólo difieren en las variables explicativas. Siguiendo con el ejemplo anterior, este método conduciría a comparar

puntajes de velocidad de aprendizaje en niños que tuviesen las mismas calificaciones en las otras variables. Przeworski y Teune (1970:32-34) estudiaron con detenimiento las debilidades de este método, específicamente en el campo de la investigación social. El análisis de asociación controla variables manteniéndolas constantes. Por ejemplo, para eliminar el efecto de la clase social sobre la relación entre edad y preferencia partidaria se examina ésta para los electores de, por ejemplo, la clase alta y la clase baja. En cada una de estas subpoblaciones está controlada la clase porque todos los electores pertenecen a la misma clase social. El análisis de regresión y todas las técnicas derivadas de él controlan eliminando el efecto de las otras variables. Por ejemplo, si para analizar la preferencia partidaria se ajustara una regresión lineal entre la preferencia partidaria (variable explicada) y la edad y la clase social (variables explicativas), el coeficiente de regresión parcial de la edad mide el efecto neto de esta variable, descontado el impacto lineal de la clase social.

Dejando a un lado las diferencias metodológicas y técnicas obvias entre estos cuatro métodos de control, debe notarse que los tres últimos requieren que las variables por controlar sean explícitamente incorporadas al análisis; no acontece lo mismo con la aleatorización. Como ya hemos dicho, el análisis causal se basa en un sistema de ecuaciones recursivas cuya base teórico-estadística se encuentra en el análisis de regresión. En consecuencia, para controlar el efecto de las otras variables habría que incluir en el modelo todos los factores explicativos del fenómeno bajo estudio. Esto quiere decir que se debe disponer de teorías lo suficientemente poderosas para identificar *todos* los determinantes, así como las técnicas, los instrumentos y los recursos necesarios para recabar la información. No creo que sea exagerado sostener que el desarrollo de las ciencias sociales dista mucho de satisfacer estas exigencias —es decir, no se cumple el tercer criterio de Asher—, y que por lo tanto a partir únicamente de correlaciones y asimetrías temporales no es posible establecer el sentido de las conexiones causales ni mucho menos evaluar estadísticamente su fuerza.

La única puerta abierta para establecer la dirección causal es, en palabras de Asher:

La técnica de los modelos recursivos lineales no permite determinar la relación causal entre dos variables; si hay una relación causal entre dos variables, entonces el investigador tendrá que establecerla desde fuera, presumiblemente basándose en consideraciones teóricas y sustantivas.

Creo que las conclusiones de este análisis son nítidas: hoy la estadística reconoce que en cualquier situación práctica la dirección de la causalidad se atribuye a la regularidad empírica, y en segundo lugar, queda claro el argumento sustancial que llevó a sustituir el nombre de análisis de causalidad por el de análisis de trayectoria, análisis de sendero o *path analysis*.

Dos métodos de estimación en análisis de variables múltiples

En las ciencias sociales se recurre cada vez más a utilizar técnicas de variables múltiples para analizar los datos. En el ajuste de este tipo de modelos tienden a predominar los métodos de estimación mínimo cuadrático y de máxima verosimilitud. Si bien hay razones técnicas para preferir uno u otro dependiendo del problema de que se trate, no nos detendremos en estos aspectos sino en las ideas que dan sustento a ambos procedimientos de estimación.

La estimación mínimo cuadrática ordinaria, así como sus variantes, se ha utilizado profusamente en el ajuste de modelos de regresión, ya sean lineales o susceptibles de ser linealizados. El problema que se intenta resolver, así como las ideas centrales, se pueden exponer para el caso más simple (una variable explicada y una variable explicativa) sin pérdida de generalidad respecto a los propósitos de este trabajo. Consideremos entonces que por alguna razón se postula que hay una relación lineal entre la variable explicativa X y la explicada Y , de manera que si se tiene un conjunto de n pares ordenados (X, Y) se genera un sistema de n ecuaciones con dos incógnitas: la ordenada al origen α y la pendiente β . La solución matemática a este problema se debe a Adrien Legendre, quien en el apéndice a su trabajo *Nouvelles méthodes pour la détermination des orbites des comètes*, publicado el 6 de marzo de 1805, propone para resolver este problema “la técnica mínimo cuadrática” (Stigler, 1986:12). Aun cuando en el desarrollo del modelo de regresión está presente con toda su fuerza el problema que abordó Legendre, el planteamiento moderno sostiene que los valores de la variable explicada están compuestos por una parte sistemática que varía con la variable explicativa y una parte aleatoria (ϵ), de modo que el sistema de ecuaciones responde a:

$$Y_i = \alpha + \beta X_i + \epsilon_i \quad (i = 1, 2, \dots, n)$$

A las incógnitas α y β se agregan la media y la varianza de la variable aleatoria ϵ . En la versión más simple de este modelo se supone que

la esperanza matemática y la varianza de ε son 0 y σ^2 respectivamente. Nótese que el supuesto sostiene que todas las variables aleatorias ε_i , una para cada ecuación, tienen la misma media y la misma varianza 0, puesto en otros términos, que las medias y varianzas de las variables aleatorias son independientes de las observaciones.

Esta es una pílula sintética de los planteamientos estadísticos en que se basa el modelo de regresión y tiene por propósito disponer de un lenguaje básico para detenemos en lo que me parece más relevante desde el punto de vista epistemológico: las ideas que llevan a distinguir la parte sistemática y aleatoria de la variable explicada y los argumentos que justifican la introducción del término ε .

El término estocástico se incluye en el modelo para representar:

- i) que el modelo no considera todos los factores que influyen sobre Y (sino sólo los relevantes). Para fijar las ideas, presumamos que hay razones conceptuales para suponer que entre X y Y hay una relación lineal, por ejemplo, que el consumo de los hogares (Y) depende del ingreso neto que logran acumular sus perceptores (X), sin embargo, en los hechos éste no es el único factor que influye sobre Y , ya que el consumo también estará afectado por el tamaño y la composición del hogar, por la antigüedad de la pareja, por algunas características particulares de los miembros —por ejemplo, si el padre es un aficionado a las carreras de caballos, es un bebedor, un jugador de póker o bien la madre es adicta a la moda parisina, a los automóviles importados, etc.—. Esta fuente de explicación del término aleatorio supone que el modelo incorpora todas las variables significativas que condicionan o afectan los valores de la variable explicada, y que ninguna de las no consideradas es relevante en la explicación, de modo que dado un valor de X_i , es decir, de todas las unidades que tienen el mismo nivel de ingreso, ε_i puede asumir valores por arriba o por debajo del valor sistemático dependiendo de los valores que alcancen las variables no explícitamente consideradas en el modelo.
- ii) Una segunda razón para incluir el término estocástico son los errores de medición en la variable explicada. Los valores registrados de Y no reflejan los valores dados por la teoría y representados en la ecuación. En nuestro ejemplo, esto quiere decir que los errores de medición en el nivel de consumo de los hogares se consideran incorporados en ε_i .
- iii) En ocasiones se agrega como tercer factor la existencia de un elemento impredecible en la respuesta humana (Johnston, 1984:14).

Me parece que, detrás de la primera fuente que racionaliza la incorporación de un término estocástico en la ecuación de regresión, se halla subyacente una imagen del mundo muy peculiar, según la cual para tener una explicación perfecta de la variable dependiente bastaría con in-

cluir todos y cada uno de sus factores condicionantes. El único problema que se podría presentar es que el número de incógnitas supere al de ecuaciones (o de observaciones, si se quiere) (Johnston, 1984:14). Al estirar las consecuencias que entraña esta perspectiva se podría concluir que la acumulación de datos sería un camino válido para llegar a identificar las determinantes de los fenómenos bajo estudio. La idea que se esconde detrás de este planteamiento es que lo aleatorio surge de la incapacidad de los seres humanos para tratar la naturaleza multidimensional de los fenómenos sociales. En esta concepción lo aleatorio se define como un testigo de la incapacidad humana para dar cuenta de los procesos reales. Si el carácter estocástico del fenómeno se justifica de esta manera, la tercera fuente se transforma en un caso particular de la primera. En efecto, el comportamiento humano sería aleatorio debido a que no conocemos todos los factores que lo determinan. Por lo tanto, las fuentes de la aleatoriedad se reducirían únicamente a dos, las mismas que estuvieron presentes en el origen del método de estimación mínimo cuadrático, en 1805.

El método de estimación mínimo cuadrático en su nacimiento estuvo asociado a tres problemas planteados en el siglo XIX: *i*) representar y determinar matemáticamente los movimientos de la Luna; *ii*) dar cuenta de una desigualdad no periódica en los movimientos de los planetas Júpiter y Saturno; y *iii*) determinar la forma de la Tierra. Todos estos problemas envolvían, por una parte, observaciones astronómicas, y por otra, la teoría de la gravitación (Stigler, 1986:17). Se trataba entonces de ajustar las ecuaciones derivadas de la teoría a los resultados de las observaciones astronómicas. El ajuste implicaba más ecuaciones que incógnitas en la medida en que las primeras incluían unos pocos parámetros y se tenía un número apreciable de observaciones. Parecería que en estos estudios no se dudaba acerca de la teoría; ésta era tomada por buena, las desviaciones entre los resultados que arrojaba el modelo y los datos observados tenían su origen en los errores de medición, los cuales obviamente eran considerados aleatorios.

Quiero llamar la atención sobre el hecho de que las ideas básicas que originaron la técnica mínimo cuadrática se encuentran presentes en las versiones modernas que justifican la inclusión del término estocástico en el modelo de regresión. Las teorías nos informan sólo sobre los factores relevantes que explican un fenómeno. El resto se incluye en el término estocástico que ayuda a medir qué tanto no estamos considerando en la explicación estadística. Según esta concepción existiría la posibilidad de dar cuenta exhaustiva de un fenómeno, pero nos encontramos con la limitación de que el número de factores es de tal magnitud que sería prácticamente imposible recabar los datos necesarios para

tener más ecuaciones que incógnitas. Además, habría que considerar que de todas maneras existirían obstáculos para emprender esta tarea, ya que la aleatoriedad surge también de los errores de medición. De este breve análisis se desprende la conclusión de que el poderoso paradigma newtoniano está subrepticamente presente en una parte importante de la estadística aun al finalizar el siglo xx.

El método de máxima verosimilitud supone que las observaciones fueron generadas por una distribución de probabilidades cuyos parámetros hay que estimar. Para llevar a cabo la estimación se construye la función de verosimilitud, que es una medida de la incertidumbre en la estimación de valores hipotéticos del o de los parámetros. Supongamos que el valor hipotético del parámetro θ es θ' , entonces la función de verosimilitud será:

$$L(\theta'/Y, M) = K(Y) P(Y/\theta')$$

donde la función de verosimilitud L representa la incertidumbre asociada a θ' , dadas las observaciones (Y) y el modelo (M). El método consiste en encontrar el valor de θ' que maximiza la verosimilitud de que el modelo haya generado los datos observados. Son muchos los recovecos técnicos de este método de estimación, sin embargo dejaremos su presentación en este punto para enfocarnos hacia las ideas centrales que organizan esta forma de enfrentar el problema.

Queremos destacar el hecho de que el planteamiento del método de estimación máximo verosímil introduce el supuesto de que las observaciones han sido generadas por un mecanismo aleatorio. Esto es una forma de "organizar" la experiencia. Reconoce un papel activo del sujeto: es éste quien decide si el modelo adecuado para representar un conjunto de datos es la distribución binomial, multinomial, hipergeométrica, exponencial, normal, etc. El modelo que se selecciona para representar la distribución observada de los datos no se elige arbitrariamente, sino que la decisión se apoya en el examen del procedimiento que se siguió para generar las observaciones y las características teóricas de las distribuciones de probabilidades.

Razones esencialmente técnicas han confinado al campo de la regresión al método de estimación mínimo cuadrático así como sus derivaciones. Esto a pesar de que la técnica de máxima verosimilitud aplicada a la regresión lineal arroja, en esencia, los mismos resultados. Sin embargo, las nuevas técnicas estadísticas multivariadas, de amplia aplicación en las ciencias sociales (análisis de cohortes, análisis de sobrevivencia, análisis loglineal, regresión logística), emplean el método de máxima verosimilitud. Pero ambos procedimientos no sólo poseen di-

ferentes atributos que los hacen aplicables a problemas de naturaleza distinta, sino que conciben de manera diferente la aleatoriedad. En efecto, el método de estimación mínimo cuadrático supone que la aleatoriedad deriva de la incapacidad de especificar todas y cada una de las variables explicativas de Y , y de los errores en su medición; la idea central subyacente a este procedimiento es que sería posible dar cuenta exacta de la variable en cuestión. Esto es una posibilidad teórica, ya que aun cuando se consideraran todas las variables explicativas habría que generar suficientes casos como para tener más ecuaciones que incógnitas y estar así en condiciones de estimar los parámetros; en tanto que el método de máxima verosimilitud reconoce que la variable en sí es aleatoria y que para realizar la estimación se debe postular un modelo estadístico. En este planteamiento hay un reconocimiento explícito a la construcción conceptual de la experiencia así como del dato.

En conclusión, los desarrollos estadísticos recientes, fuertemente condicionados por las demandas planteadas por las ciencias sociales, tienden a abandonar la naturaleza determinística de la experiencia, remplazándola por una concepción que reconoce el carácter aleatorio de la misma, y a reconocer el papel activo del sujeto en la construcción del objeto.

Conclusión

En una famosa conferencia, Albert Einstein recomienda a los interesados en el método de la física: “no escuches sus palabras [refiriéndose a los físicos teóricos], pon atención a lo que hacen”.⁷ Haciendo un paralelo podríamos decir: no te fijas en lo que se dice respecto a la aplicación de la estadística en la investigación empírica, sino en lo que se hace en la investigación. A partir de las concepciones de la epistemología genética el proceso de investigación se puede ver como una espiral que se eleva por la contradicción entre la teoría y las observaciones (Cortés y Rubalcava, 1987: Introducción). En este proceso la estadística juega un papel central en la investigación empírica.

Algunas técnicas estadísticas son más propias para estudiar regularidades y verificar la consistencia de las hipótesis con los datos (abstracción empírica), otras lo son para calibrar el sustento empírico de la aplicación de las hipótesis a nuevos dominios empíricos (abstracción reflexiva).

⁷ The Herbert Spencer Lecture, Oxford, 10 de junio, 1933.

El desarrollo del análisis causal ha llegado a la conclusión de que dada una relación estadística, el investigador debe establecer la causalidad desde fuera basado en consideraciones teóricas y sustantivas. Esta idea se enriquece y precisa en el conocimiento desarrollado por la epistemología genética. En efecto, esta corriente de pensamiento, basándose en minuciosas investigaciones empíricas, ha concluido que las relaciones causales se conceptúan como la atribución de necesidad lógica y coherencia teórica a la experiencia (Piaget y García, 1973; Halbwachs, 1977; García, 1986:51).

Las diferentes concepciones de la aleatoriedad tienen fuerte relación con el estudio de los sistemas dinámicos abiertos. *La equilibración de las estructuras cognitivas* (Piaget, 1975), una de las obras centrales de la epistemología genética, es el estudio de la evolución del sistema cognitivo. Una de las características centrales es que su desarrollo no es continuo, sino que evoluciona por niveles moviéndose hacia estadios de desarrollo superiores. La acumulación de contradicciones en un estadio desata el proceso de desequilibrio que lleva al sistema a un reequilibrio en otro estadio; este proceso no tiene límite, está abierto por arriba. Uno de los temas que abre esta forma de conceptuar la evolución del sistema cognitivo es el mecanismo de la equilibración. Para entenderlo se puede recurrir a la idea de autoorganización de Ilya Prigogine (1979). Lo que me interesa destacar es que la equilibración combina momentos de estructuración con otros dedesestructuración. En los primeros la aleatoriedad se puede conceptuar como desconocimiento, mientras que en los segundos, como una característica intrínseca de los fenómenos, sujeta a las leyes de la estadística.

Recibido en abril de 1996

Revisado en diciembre de 1996

Correspondencia: El Colegio de México/Centro de Estudios Sociológicos/Camino al Ajusco 20/Col. Pedregal de Santa Teresa/C. P. 10740 México, D. F./fax 645 04 64

Bibliografía

- Asher, Herbert B. (1983), *Causal Modeling*, Beverly Hills, California, Sage Publications.
- Blalock, Hubert (1964), *Causal Inference in Nonexperimental Research*, Chapel Hill, The University of North Carolina Press.
- Borgatta, Edgar y George Bohrnstedt (eds.) (1970), *Sociological Methodology 1970*, San Francisco, Jossey-Bass.
- Boudon, Raymond (1968), *L'analyse mathématique des fait sociaux*, París, Plon.
- Bunge, Mario (1979), *La investigación científica: su estrategia y su filosofía*, Barcelona, Ariel.
- Cortés, Fernando y Rosa María Rubalcava (1987), *Métodos estadísticos aplicados a la investigación social: análisis de asociación*, México, El Colegio de México.
- De Ipola, Emilio *et al.* (1976), *Teoría y método para el estudio de la estructura de clases en Chile (con un análisis concreto: Chile, 1970)*, Santiago, Flacso-Proelce, tres volúmenes.
- Einstein, Albert (1981), "On the methods of Theoretical Physics", en Carl Seeling (ed.), *Albert Einstein Ideas and Opinions*, París, Laurel Editions.
- García, Rolando (1986), "Conceptos básicos para el estudio de sistemas complejos", en Enrique Leff (coord.), *El problema del conocimiento y la perspectiva ambiental del desarrollo*, México, Siglo XXI.
- Gil, Antón Manuel (1985), *Crítica epistemológica a la concepción de ciencia en Max Weber*, México, IPN, tesis de doctorado, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional.
- Halbwachs, Fr. (1977), "Reflexiones de la causalidad física", en Mario Bunge *et al.*, *Las teorías de la causalidad*, Salamanca, Sigueme.
- Hempel, Carl G. (1965), "Problemas y cambios en el criterio empirista de significado", en A. J. Ayer, *El positivismo lógico*, México, Fondo de Cultura Económica.
- Holton, Gerald (1985), *La imaginación científica*, México, Conacyt-Fondo de Cultura Económica.
- Hume, David (1921), *An enquiry concerning human understanding and selections from a treatise of human nature*, Chicago, Open Court, Philosophical classics, Religion of Science Library, núm. 45.
- Johnston, J. (1984), *Econometric Methods*, Singapur International Student Edition.
- Lazarsfeld, Paul (1974a), "La interpretación de las relaciones estadísticas como propiedad de investigación", en Raymond Boudon y Paul Lazarsfeld, *Metodología de las ciencias sociales II: Análisis empírico y causalidad*, Barcelona, Laia.
- ____ (1974b), "El álgebra de los sistemas dicotómicos", en Raymond Boudon y Paul Lazarsfeld, *Metodología de las ciencias sociales II: Análisis empírico y causalidad*, Barcelona, Laia.

- McKenzie, Donald (1979), "Eugenics and the Rise of Mathematical Statistics in Britain", en John Irvine, Ian Miles y J. Evans (eds.), *Demystifying Social Statistics*, Londres, Pluot Press.
- Miles, Ian y John Irvine (1979), "The Critique of Official Statistics", en John Irvine, Ian Miles y J. Evans (eds.), *Demystifying Social Statistics*, Londres, Pluot Press.
- Piaget, Jean (1975), *La equilibración de las estructuras cognitivas: el problema central del desarrollo*, España, Siglo XXI.
- y Rolando García (1973), *Las explicaciones causales*, Barcelona, Barral editores.
- (1982), *Psicogénesis e historia de la ciencia*, México, Siglo XXI.
- Prigogine, Ilya e Isabelle Stengers (1983), *La nueva alianza: metamorfosis de la ciencia*, Madrid, Alianza Editorial.
- Przeworski, Adam y Henry Teune (1970), *The Logic of Comparative Social Inquiry*, Nueva York, John Wiley.
- Russell Hanson, Norwood (1965), *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*, Londres, Cambridge University Press.
- Simon, Herbert (1957), "Spurious correlation a causal interpretation", en H. A. Simon, *Models of Man; social and rational*, Nueva York, John Wiley.
- Stigler, Stephen M. (1986), *The History of Statistics: The measurement of Uncertainty before 1990*, Cambridge, Mass., Belknap.
- Stinchcombe, Arthur (1970), *La construcción de teorías sociales*, Buenos Aires, Nueva Visión.
- Torrado, Susana (1981), "Introducción", en Susana Torrado, *Investigación e información sociodemográficas 2: los censos de población y vivienda en la década de 1980 en América Latina*, Buenos Aires, Clacso.
- , Miguel Acuña y Vittorio Lorenzi (1976), *Informe sobre la población económicamente activa en América Latina: 1940-1970*, Santiago, Proelce.
- Wright, Sewall (1921), "Correlation and Causation", *Journal of the Agricultural Research*, núm. 20.