

Avatares de la Estadística Social en Ibero América o confieso que he enseñado Estadística¹

Fernando Cortés
CES, COLMEX

I.- A modo de introducción

Llegué a la Sociología, en 1966, como joven profesor de Estadística que presumía de una “vasta” experiencia docente de 4 o 5 años de enseñanza de la materia en las Escuelas de Economía y de Contadores Auditores de la Universidad de Chile y también en la Escuela Latinoamericana para graduados en Economía (Escolatina) dependiente del Instituto de Economía de la misma Universidad.

La puerta de entrada fue la Escuela Latinoamericana de Sociología de la Facultad Latinoamericana de Ciencias Sociales (FLACSO) que operaba en su única sede de Santiago de Chile desde 1960. Al crearse, en 1969, la Escuela Latinoamericana de Ciencia Política y Administración Pública dentro de FLACSO, también enseñé la asignatura, que por la época llevaba el nombre de Estadística Social.

A la sazón Santiago de Chile era un crisol intelectual de las Ciencias Sociales de América Latina avivado por la lucha sostenida entre las opciones políticas en pugna. Desde 1964 Chile era gobernado por el primer presidente democristiano de su historia,

¹ El subtítulo de este trabajo se inspira en parte en las memorias de Pablo Neruda y en parte en una anécdota ocurrida en 1972. Transcurría dicho año -era el segundo de la presidencia de Salvador Allende- y se celebraba en Santiago la primera reunión de la ALAS (Asociación Latinoamericana de Sociología). Se llevó a cabo en el edificio de la UNCTAD que sería la casa de gobierno del dictador Pinochet después de que el 11 de septiembre de 1973 destruyera La Moneda con bombas lanzadas desde aviones y cañonazos de tanques. Al dirigirme de mi oficina en el tercer piso del edificio de FLACSO, sito en José Miguel Infante 65, a la biblioteca en el segundo piso, me encontré sorpresivamente con un matrimonio de ex estudiantes extranjeros, ya recibidos, que presentaban ponencia en el Congreso; después del saludo de rigor, ella me preguntó a quemarropa “¿Cortés, aún enseñas Estadística?”, mi respuesta fue como la confesión de un pecado, un inaudible “Sí”. Para entender “mi pecado” es necesario un poco de contexto: Como veremos en este escrito, desde fines de la década de los sesenta, la Estadística, disciplina imprescindible y válida en la formación académica de los sociólogos y politólogos, había sido desterrada de los programas docentes. Imperaba el verbo predicado por jóvenes democristianos que había reemplazado a la Biblia, por Das Kapital, provocando la pérdida de las normas del rigor científico, como si ciencia y transformaciones sociales radicales fuesen incompatibles. Estoy seguro que aquella fue una época en que Marx se dio varias vueltas en su tumba. Lo paradójico es que esta “pequeña burguesía” intelectual era un desprendimiento de segundo orden de la derecha chilena: jóvenes que primero se separaron del Partido Conservador para formar el Partido Demócrata Cristiano (PDC) y después, otra vez los jóvenes, aunque de diferente cohorte, abandonaron el partido que formaron sus antecesores (Yoczelevzky R 1988).

Eduardo Frei Montalba, quien había derrotado a Salvador Allende en su tercer intento consecutivo para llegar a la presidencia del país. Chile era una isla de democracia en América del Sur y, en la década de los cincuenta, una serie de organismos internacionales habían abierto ahí oficinas. La apertura de escuelas con programas de postgrado en Ciencias Sociales en las universidades chilenas por organizaciones dependientes de la ONU, como la UNESCO y otros organismos internacionales; aunada a la libre conjugación de las ideas sin restricciones, hizo que una pequeña ciudad de un país en el rincón del mapamundi -pero con vista al mar- gozara del privilegio de contar con una concentración sin precedente de la inteligencia de la Región.

Mirado retrospectivamente, con el reposo que da el paso del tiempo, se puede decir que dos eran los grandes temas que atravesaban el quehacer de las Ciencias Sociales de la época: (i) la lucha entre el paradigma dominante de *la modernización* y el emergente de *la dependencia*, que reflejaba la disputa política que sostenían esencialmente la izquierda y la democracia cristiana y (ii) las discusiones teóricas, dentro de la izquierda, referidas a la conquista del poder, que surgieron después de la victoria de la revolución cubana, cuyo amplio espectro se extendía desde la posición que proponía la vía armada hasta la que planteaba el camino de las urnas.

Mientras los profesores enseñaban la teoría de los sistemas sociales de Talcott Parsons, las teorías de alcance medio de Robert Merton o la economía del bienestar, los estudiantes leían, debajo de sus pupitres, un mimeografiado (aún no existían las fotocopias) -que solo era legible después de un arduo trabajo de rellenado de las letras faltantes debido al desgaste de la matriz- que posteriormente sería publicado por siglo XXI bajo el título *Dependencia y Desarrollo en América Latina* de Fernando Henrique Cardoso y Enzo Faletto, chileno este último del barrio Ñuñoa en el cual nació y murió.

El enfoque² de *la dependencia* -nacido del corazón mismo de CEPAL- asiento institucional de la teoría del desarrollo económico, que predominó en la época, elaborada bajo la conducción de Raúl Presbich, -conocida en el ambiente académico

² Es habitual que se hable de *La teoría de la dependencia* cuando en realidad hubo varias. En este texto, dedicado a tratar principalmente temas estadísticos y secundariamente metodológicos, destaco el papel de la versión Cardoso/Faletto. Además, no uso la palabra 'teoría' sino 'enfoque' por respeto a las ideas de Enzo, para quien 'las teorías' deberían no sólo ser un conjunto coherente y articulado de enunciados abstractos, sino incluir también los métodos y las técnicas que permiten hacer observables los conceptos en el plano de la experiencia.

como la teoría cepalina -hizo ver, por una parte, que la comprensión del problema del desarrollo no sólo incluía un aspecto económico sino también un aspecto social y, por la otra, que se requerían nuevas metodologías para enfrentar el reto explicativo.

Surge así la discusión en torno a la investigación multidisciplinaria, circunscrita inicialmente al campo de la Economía, la Sociología y la Ciencia Política, en medio de profundos procesos de transformación social y política. Es la época en que varios investigadores sociales adoptamos a la Sociología como el centro de gravedad de nuestro interés académico sin perder nunca de vista que los problemas sociales no reconocen las fronteras disciplinarias arbitrarias impuestas por el hombre en su afán de conocer.

Pero en concordancia con las ideas de Enzo Faletto emergió también la necesidad de enfrentar los nuevos problemas metodológicos que se derivaban del enfoque de *la dependencia*, que ponía el acento sobre el análisis de la historia. La metodología de las Ciencias Sociales estaba fuertemente dominada por las técnicas de *survey* que, entre otras cosas, se caracterizaban por ser esencialmente estáticas. Si bien existía ya en la bibliografía la idea de la *encuesta de panel*, aún estábamos lejos de los desarrollos teóricos, de los métodos de análisis y de las posibilidades de procesamiento de que disponemos hoy. El análisis de series de tiempo se reducía a la descomposición de las series en tendencia, fluctuaciones estacionales y cíclicas e irregulares. Los poderosos métodos con que contamos hoy encapsulados en los programas que procesan las computadoras personales, para tratar eventos cronológicos, aún estaban en el futuro. A su salida de CEPAL, Faletto ingresó a FLACSO con la idea de transformar el enfoque en una teoría. El proyecto fue truncado el 11 de septiembre de 1973 por el golpe de estado en contra del gobierno de Salvador Allende.

Nanterre, en 1968, diseminó por América Latina el marxismo estructuralista de origen francés. *Para leer el capital* de Louis Althusser fue traducido al español, y sus alumnos, que llegaron a América Latina evitando la persecución política, trajeron una propuesta teórica y metodológica que cayó en terreno fértil.

Se propagó el *dictum* que afirmaba que “la estructura determina, en última instancia, la superestructura”, de modo que bastaba con cambiar las relaciones sociales de producción para desencadenar procesos irreversibles de transformación social³.

Desde el punto de vista metodológico tal vez son dos las contribuciones más importantes de los franceses. Por una parte, el acento sobre los modos de producción y su articulación en formaciones sociales concretas, y la idea de una cierta sucesión histórica de los modos de producción dominantes, nunca explícitamente reconocida. Esta primera idea, unida a la teoría de la dependencia, originó hacia fines de la década de los sesenta el análisis histórico estructural, que tuvo fuerte influencia sobre la mayor parte de las comisiones que trabajaban en el seno del naciente Consejo Latinoamericano de Ciencias Sociales (CLACSO) y en la formación del proyecto PROELCE impulsado conjuntamente por la Escuela Latinoamericana de Sociología de FLACSO y el Centro Latinoamericano de Demografía (CELADE), que aceleró y puso en el centro de la discusión el papel de la población en los procesos sociales.

Por otra parte, también de Francia llegó la idea de que el dato se construye. Hoy se puede afirmar, sin temor a equivocarse, que en los sesenta la Sociología y la Ciencia Política académicas tendían a no discutir sus fundamentos filosóficos. El positivismo lógico ejercía un dominio legítimo; se hallaba ampliamente extendida la idea de que investigar consistía en identificar regularidades en los datos que a su vez eran la piedra angular del conocimiento válido en la medida en que eran o representaban la realidad. Plantear en este medio que el dato se construye y que se construye a partir de conceptos teóricos debería haber provocado un sismo pues se hacía explotar una bomba en los cimientos de la fábrica de conocimientos.

Sin embargo, los cambios sociales, las nuevas correlaciones políticas que afectaban por la época a varios países del cono sur, y el desmoronamiento de la teoría, la metodología y las técnicas de investigación -incluida la Estadística-, ante los embates

³ Sin embargo, la historia de Chile alrededor de 1973 enseñó que el traspaso de la propiedad de los medios de producción a los trabajadores y la nacionalización de las riquezas naturales en manos de capital extranjero, no se tradujeron en un cambio revolucionario irreversible sino más bien en una violenta vuelta atrás en el tiempo. Este hecho obligó a la reflexión sobre lo acontecido: Algunos reaccionaron invirtiendo el acento al postular el predominio de la superestructura sobre la estructura o la autonomía relativa de la superestructura -posición que encontró aval teórico en Antonio Gramsci-, otros se plantearon la negación del peso de la teoría y la búsqueda descarnada de la realidad -posición que se aproxima en lo esencial al empirismo lógico de la década de los años 20, del siglo recién pasado.

del enfoque de *la dependencia*, el surgimiento del método histórico estructural y el predominio del marxismo académico, opacaron las consecuencias del *dictum* “el dato se construye”. Bajo el imperio de este *dictum* se organizó la comisión del Sistema de Estadísticas Sociales y Demográficas en el seno de CLACSO. En una de las investigaciones del PROELCE, realizada por Susana Torrado, Emilio de Ipola, Juan María Carrón y Arturo León se reprocesó el censo chileno de 1970 a partir de los conceptos althusserianos para generar una imagen de país organizado en clases sociales, a diferencia del tratamiento habitual de la información censal que ofrecía como resultado una estratificación⁴.

FLACSO, siguiendo los principios de la época, desarrollaba en Santiago de Chile sus programas docentes y en Río de Janeiro los de investigación, que tuvieron vida relativamente efímera. A Santiago llegaban a estudiar postgrado en Sociología -y posteriormente Ciencia Política- jóvenes de todos los países de América Latina y el Caribe. Se pretendía dotarlos con una sólida formación científica y muchos de ellos se transformaron en investigadores de vanguardia y en no pocas ocasiones en fundadores de departamentos o escuelas de Sociología en sus países. En sus aulas se formaron los sociólogos y los politólogos que investigaban los fenómenos y procesos sociales, estaba excluido el ensayo y la investigación teórica era desalentada.

Por la época la enseñanza de las Ciencias Sociales se centraba en la docencia. Los programas de postgrado incluían materias que se impartían teóricamente en el aula, aunque algunas como la Estadística contaba además con laboratorios en los que se trabajaba sobre la solución de ejercicios, y para obtener el grado se exigía la realización de una tesis, que en la mayoría de los casos terminaba siendo un estudio empírico. La politización que vivió la sociedad chilena a fines de la década de los sesenta, la posibilidad de ser actores en procesos de transformación social inéditos, los cambios introducidos en la orientación de los programas de estudio y los interesantes

⁴ La historia tiene misterios. La Estadística ayer como hoy tiene una fuerte influencia de la obra de Karl Popper, autor que ha sido calificado como neopositivista. Esta corriente filosófica ya en la década de los treinta planteaba la incapacidad del sujeto para aprehender directamente el objeto, la relación sujeto-objeto era mediada, el objeto era construido. Sin embargo, esta idea llegó a través de los desplazados de Nanterre. Tal vez ello se deba a la casi nula discusión epistemológica y la muy pronunciada preocupación por las técnicas de investigación y la Estadística que predominaban por aquella época.

procesos sociales y políticos que se desencadenaron impulsaron a poner en tela de juicio la división entre investigación y docencia.

A comienzos de los setenta, en la Escuela de Sociología de FLACSO, se instalaron a título experimental los primeros seminarios teórico-metodológicos en torno a problemas sociales críticos, como por ejemplo, la participación del movimiento obrero y de los sectores marginales en los procesos de transformación social que vivía el país. Estos seminarios, además de demandar el estudio de las teorías pertinentes y de los problemas metodológicos y técnicos de las investigaciones que se realizaban en su seno, conectaban a los estudiantes con los actores sociales involucrados⁵. Surgía así, de las entrañas del cambio revolucionario, la idea de vincular docencia e investigación en el proceso de enseñanza, y también la investigación acción.

Este es el marco general del punto de partida de nuestra exposición. En él inscribiremos la evolución histórica de los temas que fue abordando la Estadística Social desde los años sesenta hasta nuestros días. Cada vez que sea necesario se introducirán elementos de contexto para que se puedan apreciar, con la mayor nitidez posible, las fuerzas que han conducido a la Estadística Social en nuestro medio.

II.- La década de los sesenta: el auge

El predominio del estructural funcionalismo parsoniano y de las teorías de alcance medio de Karl Merton, en el plano conceptual, y de las técnicas de survey en lo metodológico se combinaban armónicamente con los instrumentos que proporcionaba la Estadística.

La información empírica que permitía contrastar las hipótesis de las investigaciones sociológicas de la época se obtenía preferentemente a través de muestras que usaban cuestionarios como instrumentos de recopilación de información, por lo que para los científicos sociales de esos años era crucial comprender la teoría de las muestras y las diversas formas de aplicarla. Sin embargo, para llegar hasta ese

⁵ Muchos profesores y estudiantes que participaban en estos seminarios fueron perseguidos o asesinados por miembros de las fuerzas armadas de Chile en los días posteriores al golpe de estado del 11 de septiembre de 1973.

punto del desarrollo de la Estadística era y es necesario disponer de un buen nivel de conocimiento de la estadística descriptiva y de la teoría de probabilidades.

Además era necesario estudiar la parte de la Estadística concerniente a la estimación de los parámetros poblacionales con base en los resultados de la muestra. En consecuencia, se hacía imprescindible aprender inferencia estadística -estimación punto y de intervalo y pruebas de hipótesis-. En concreto, el sociólogo en los sesenta debía ser capaz de emplear con soltura el material expuesto en los primeros capítulos de cualquier libro introductorio de estadística matemática, incluida la inferencia estadística.

Por la época la enseñanza de la Estadística referida a la construcción y descripción de los datos era bastante parecida tanto en contenido como en forma de impartirla en Economía, Sociología o Ciencia Política. El rigor matemático de la exposición no presentaba diferencias importantes aunque en Sociología y Ciencia Política se evitaba hacer uso de cálculo avanzado.

La diferencia entre la Estadística Social y la Estadística Económica radicaba en las técnicas de análisis de datos. La Economía, por un lado, empleaba profusamente el análisis de regresión, materia prima de la incipiente Econometría⁶, mientras que la Sociología tenía como principal recurso el *análisis de covarianzas* de Lazarsfeld (Lazarsfeld P. 1946: 115 a 125). Por otra parte, la Estadística aplicada a la Ciencia Política navegaba a medias aguas.

El argumento que daba racionalidad a la especialización disciplinaria en el uso de las técnicas de análisis de datos sostenía que la mayoría de las variables económicas se medían en escala de intervalo o de razón, mientras que, por el contrario, casi la totalidad de las variables sociales eran nominales o en el mejor de los casos ordinales⁷. Había calado profundo el libro de Sydney Siegel *Nonparametric Statistic*

⁶ A pesar de que en los sesenta aparecen varios libros de Econometría el más utilizado en América Latina fue *Econometric Methods* de J. Johnston (1964).

⁷ La idea de la diferenciación entre Sociología y Economía, según la escala de medición de las variables, muy popular en esa época se escucha a veces aunque con menos frecuencia que en el pasado, en pleno siglo XXI. Da la impresión que ella se desprende del supuesto que la naturaleza de lo económico sería diferente a la materia que trata la Sociología; la primera sería cuantitativa mientras que la segunda cualitativa. Este argumento no toma en cuenta que lo que se mide son conceptos y no la supuesta "realidad" (Bunge M. 1979, Carnap R. 1959 y Neurath Otto 1959).

(1956), de amplia circulación por esos años, que relacionaba las pruebas de hipótesis y las técnicas de análisis de datos a los niveles de medición de las variables.

La Sociología hacía uso intensivo de diversas pruebas de hipótesis de la familia ji-cuadrada, y para juzgar la fuerza de la relación se echaba mano al análisis de asociación. Las raíces de esta técnica se remontan a fines del siglo XIX y comienzos del XX y su desarrollo se dio en dos vertientes, una impulsada por Karl Pearson y la otra por Udney Yule⁸.

Los desarrollos de Pearson suponían la existencia de variables latentes continuas que seguían una distribución conjunta normal, sin embargo, sólo eran observables en escalas no métricas; con base en las frecuencias observadas el problema estadístico a resolver consistía en encontrar un buen estimador del coeficiente de correlación producto-momento de Pearson. Esta aproximación generó los coeficientes tetracórico, biserial y biserial punto (Kendall M. G. y A Stuart, 1961: 304 a 316), que tuvieron escasa aplicación en las ciencias sociales de Ibero América.

Fue el camino no paramétrico señalado por Yule el que se utilizó profusamente en la Región. Los trabajos desarrollados bajo esta óptica generaron una serie de coeficientes para medir la fuerza de la relación entre variables cualitativas, basándose únicamente en las frecuencias observadas, sin suponer distribución poblacional alguna. El análisis se hacía en dos etapas, en la primera, se efectuaba la prueba ji-cuadrada de independencia estadística, que supone normalidad para muestras pequeñas o bien un comportamiento normal asintótico para muestras grandes; y una vez que se rechazaba la hipótesis nula de que las variables eran estadísticamente independientes se procedía, en la segunda etapa, a calcular la fuerza de la relación.

Los coeficientes más utilizados en tablas de 2X2 eran Q de Yule y fi. Para tablas de R renglones por C columnas se disponía de un conjunto de coeficientes funciones de ji-cuadrada, a los cuales se fueron agregando los propuestos por Leo Goodman y William Kruskal (1954 y 1963), entre los que destacan los coeficientes λ , para variables nominales y γ para ordinales. En los años siguientes se agregaron otros

⁸ Donald MacKenzie (1979) hace un relato histórico iluminador de las diferencias entre los desarrollos estadísticos de Pearson y Yule.

coeficientes, sin embargo, no todos gozaron de popularidad entre nosotros excepto, tal vez, d de Sommers, τ_a y τ_b .

Al ampliarse la gama de coeficientes de asociación surgió el problema de decidir cuál emplear. Las escalas en que se medían las variables dejaron de ser un buen indicador para seleccionar la medida adecuada, ya que para cada combinación de ellas (nominal u ordinal) se disponía de varios coeficientes alternativos. Así se fue perfilando la idea de utilizar como criterio, para elegir tal o cual coeficiente, la correspondencia entre las estructuras lógicas del coeficiente y del enunciado o proposición teórica que se sometía a prueba (Cortés F. 1967; Cortés F. y R.M. Rubalcava, 1982). El planteamiento sostenía que para seleccionar el coeficiente adecuado era necesario identificar la estructura lógica de la hipótesis conceptual y también conocer las proposiciones o enunciados a partir de las cuales se construyeron los diversos coeficientes estadísticos. Aplicar este criterio no era tarea sencilla ya que demandaba no sólo conocer suficientemente la teoría sustantiva sino también los instrumentos estadísticos más allá de la aplicación de la mera fórmula.

El análisis multivariado para variables no métricas estaba en ciernes en la década de los sesenta. A pesar de que Lazarsfeld afirmaba que su ecuación de covarianzas se podía extender *a fortiori* a más de dos variables y a cualquier número de categorías por variable (Lazarsfeld P. 1974: 23 a 52 y Lazarsfeld P. 1974: 327 a 352), en la realidad la técnica era bastante limitada pues en esos casos las aplicaciones se volvían demasiado complejas. Por esta razón los estudios tendieron a limitarse al modelo básico.

La Ciencia Política requería en parte del análisis de asociación lo que la hermanaba a la Sociología. Sin embargo, otra parte no despreciable de su quehacer era el análisis electoral que demandaba el uso de técnicas para establecer la fuerza de la relación entre variables métricas. En efecto, para los diferentes agregados electorales, tales como comunas, municipios, departamentos, estados, etc. es posible registrar no sólo el número o la proporción de votos a favor de los diferentes partidos que participaron en la contienda electoral sino también una serie adicional de características económicas, sociales o culturales, tales como el ingreso promedio de los hogares, la importancia relativa de las principales ocupaciones, el nivel educativo, la

etnia, las preferencias en elecciones pasadas, etc. Con base en esa información y diversas teorías se solían ajustar modelos de regresión para “explicar” el voto a favor de los diferentes candidatos y partidos políticos⁹. Los instrumentos estadísticos adecuados para encarar el estudio de este tipo de problemas eran el análisis de regresión y de correlación, lo que aproximaba a la Ciencia Política, desde el punto de vista de la medición, al estilo de análisis de datos característico de la Economía.

El denominado análisis causal fue la técnica estadística de punta en la década de los sesenta. El libro de H. Blalock, *Causal Inference in non Experimental Research* publicado en 1964, tenía como antecedentes un trabajo de S Wright de 1934, el artículo en que Lazarsfeld exponía su análisis de covarianzas presentado en 1946 en un congreso de la Sociedad Americana de Sociología en Cleveland, y un trabajo de H. Simon publicado 1957. Todos estos esfuerzos estaban dedicados al problema de identificar empíricamente relaciones causales genuinas entre variables y distinguirlas de las relaciones espurias, empleando para ello métodos estadísticos.

El libro de Blalock fue un poco anterior al de R. Boudon quien en el capítulo 3 de su “L’ Analyse mathématique de la causalité” propuso los denominados coeficientes de dependencia, que permitían argumentar, medir el vínculo causal entre variables. Los esfuerzos de los estadísticos sociales se volcaron febrilmente al desarrollo de esta técnica, fue así como varios números del influyente anuario *Sociological Methodology* estuvieron expresamente dedicados al tema.

El análisis de causalidad, hoy transformado en análisis de trayectoria o de senderos (path análisis) de uso preferente en Sociología de la Educación, es en esencia un sistema recursivo de ecuaciones de regresión. El que una rama importante de la Sociología usara regresión pareciera contradecir el planteamiento de una cierta especialización instrumental que diferenciaba a la Sociología de la Economía. Sin embargo, no es así pues los sociólogos le daban la vuelta al problema; para hacerlo se basaban en que el modelo de Blalock estaba expresado en el lenguaje de la

⁹ Este tipo de estudio hizo tomar conciencia de que en la mayoría de los casos la interpretación de los resultados adolecían de falacia ecológica (Robinson W., 1954). Recientemente usando técnicas modernas Gary King (1999) establece las condiciones bajo las cuales es posible trasladar los resultados agregados al nivel individual, sin caer en el error de afirmaciones falaces.

correlación¹⁰ y que el coeficiente de asociación ϕ , para tablas de dos por dos es matemáticamente equivalente al coeficiente de correlación producto momento de Pearson.

En fin, el análisis multivariado en las ciencias sociales de Ibero América se limitaba casi en su totalidad al análisis de variables dicotómicas, ya sea en la versión de la ecuación de covarianzas de Lazarsfeld o al análisis de causalidad a través del modelo de Blalock o Simon Blalock como se le conocía coloquialmente por la época. En los casos en que se dispusiera de variables métricas se podía recurrir a los denominados modelos de causalidad o bien al ajuste de sistemas de ecuaciones recursivas.

Llama la atención, mirado a la distancia, que durante los sesenta haya persistido la idea de la diferencia esencial de “método” entre la Sociología y la Economía en función de la naturaleza distinta de las variables económicas y sociales, a pesar de que en los años posteriores a la segunda guerra mundial los economistas consideraron en sus modelos variables explicativas nominales definiendo variables ficticias (también denominadas variables mudas o *dummy*). La inclusión de variables mudas en el lado derecho de la ecuación de regresión permitió estimar el efecto de variables no métricas sobre la variable dependiente métrica.

Antes de abandonar esta sección hay que decir que la descripción que se ofrece de la Estadística Social en Ibero América en la década de los sesenta se limita a las corrientes principales. Es claro que en ocasiones se usaban otro tipo de técnicas tales como análisis de componentes principales, análisis factorial o análisis de conglomerados, sin embargo, esos estudios eran poco numerosos y contaban con un público bastante restringido, en pocas palabras, para emplear un término en boga en los sesenta, eran marginales. La gran mayoría de los trabajos sociológicos empleaban ji-cuadrada para probar independencia estadística, análisis de asociación para estimar la fuerza de la relación, la ecuación de covarianzas de Lazarsfeld si el análisis era “multivariable”¹¹ y en ocasiones análisis causal. En la economía se usaba prolíficamente

¹⁰ Expresó sus desarrollos conceptuales en el lenguaje de correlación, más familiar a los sociólogos, apoyándose en las relaciones matemáticas entre coeficientes de regresión y de correlación.

¹¹ Con todas las limitaciones que ya se han señalado.

el análisis de regresión y técnicas relacionadas, y los científicos políticos combinaban uno u otro tipo de técnica dependiendo del nivel de medición de las variables.

III.- Los años del destierro.

Entre comienzos de la década de los setenta hasta alrededor de la mitad de la década de los ochenta¹² la Estadística fue expulsada de los salones de clase y de la investigación social.

En América del sur los partidos de izquierda de ideología marxista tuvieron avances importantes en la conquista del poder político en los primeros años de la década de los setenta. En el plano de las ciencias sociales el surgimiento y predominio de la Corriente Histórico Estructural forzó cambios importantes en los programas de formación de economistas, sociólogos y politólogos. Los teóricos “funcionalistas” pasaron al cajón de los recuerdos de la mano con la economía del bienestar; se impuso en cambio el estudio del marxismo y de Marx. Es la época en que Karl Marx se viste de toga y birrete e ingresa a la Universidad. Las nuevas generaciones de científicos sociales estudiaban desde diferentes ángulos cada obra de Marx; se volvieron clásicos los estudios de los marxistas rusos entre los cuales destacaban los trabajos de Lenin y se leían y discutían con avidez las obras marxistas de tercera generación elaboradas en el seno de la escuela francesa.

La investigación, en consonancia con las disputas políticas por el poder, se concentró en el estudio del cambio estructural, en la sucesión de los modos de producción o en la evolución de las formaciones sociales concretas. El problema central de las ciencias sociales era dar inteligibilidad a los procesos sociales y políticos que estaban aconteciendo y proponer mecanismos para orientarlos hacia objetivos predefinidos. Las preguntas de investigación enviaban sobre unidades de análisis agregadas y revestían un carácter eminentemente histórico. Quedaba en el pasado la época en que interesaba el estudio del comportamiento, las actitudes, los valores, las percepciones, etc. de los individuos localizados en un punto del tiempo. El foco de la

¹² El límite superior del período no es tan nítido como en el anterior, cualquiera que se elija puede ser objeto de controversia, sin embargo como es bien sabido, toda periodización tiene cierta dosis de arbitrariedad; los procesos sociales suelen no tener límites tan marcados.

atención se había desplazado a los procesos de constitución y de cambio de los movimientos populares, de la clase obrera o de los campesinos. La naciente sociodemografía estudiaba los flujos migratorios en lugar de la decisión para migrar; la relación entre la fecundidad y las clases sociales; los vínculos entre la dinámica demográfica y las formaciones sociales concretas. Los aires de la época dictaban que el interés de la investigación social debía constreñirse al análisis de la dinámica de la estructura y sólo secundariamente al individuo que habitaba un espacio nacional o regional en una época determinada.

El avance del marxismo y del análisis histórico estructural y el enlace entre la investigación y los procesos de transformación social que estaban acaeciendo tuvieron incidencia evidente sobre los programas de estudio de la metodología y estadística social.

En el período anterior la metodología solía cubrir un amplio espectro de materias; algunas eran eminentemente técnicas, útiles en el proceso de generar datos como, por ejemplo, construcción de cuestionarios, principios para realizar entrevistas en profundidad o hacer investigación participante, elaboración de índices, codificación de variables no métricas, etc. sin embargo, el contenido de los cursos de metodología no se agotaba en la parte técnica sino incluía también el conjunto de operaciones que median entre la teoría y el material empírico: observación, medición y operacionalización y experimentación.

Los instrumentos que ponían a disposición de la investigación social de esos años la Metodología y la Estadística quedaron fuera de foco ante el cambio en el paradigma. No estaban diseñadas para ayudar a entender el cambio estructural.

Por otro lado, las nuevas preguntas de investigación también provocaron un cambio de contenido en la metodología de las ciencias sociales. Una parte de la exploración en busca de nuevos métodos que ayudaran a responder las preguntas que se formulaba la nueva investigación social se volcó hacia el estudio de la filosofía de la ciencia y de la epistemología. Otro camino que se ensayó fue la lectura metodológica de las investigaciones realizadas por los autores clásicos.

El muestreo y la inferencia se eliminaron de los programas de la enseñanza de la Estadística así como cualquier técnica de investigación. La estadística descriptiva se

enseñó como parte de cursos de “Fundamentos técnicos de la investigación social” en los cuales se estudiaba la forma cómo investigaciones calificadas como clásicas en esa época y las de nuevo cuño operacionalizaban las relaciones entre los conceptos teóricos¹³. Según el o los textos de investigación empírica seleccionados para la enseñanza de la Estadística era posible incluir, en ocasiones, algo de análisis de asociación y también de correlación lineal simple, utilizadas como medidas descriptivas.

Los golpes de estado acaecidos en América del Sur durante los primeros años de la década de los setenta y la persecución política desatada en contra de los académicos no tuvo mayores consecuencias sobre las orientaciones metodológicas ni sobre el papel de la Estadística en las Ciencias Sociales. Desde el exilio la preocupación intelectual se volcó sobre la experiencia vivida y el énfasis cambió de lo estructural a lo superestructural. La década de los ochenta estará signada por la reflexión acerca del proceso de transformación del Estado y su autonomía relativa. Sin embargo, fueron contados con los dedos de las manos los estudios concretos, más allá de la mera especulación, realizados en la Región.

Estas eran las corrientes dominantes en nuestros países, sin embargo, el avance de la Estadística Social continuaba en otras áreas del globo terráqueo. En los setenta se propone una solución al problema de qué coeficiente de asociación utilizar en cada caso concreto¹⁴; se basaba en una profundización de la idea de enlazar las estructuras lógicas de la hipótesis teórica y del índice de asociación que le correspondía. Por una parte, el estudio en profundidad de las bases lógicas de los índices de asociación permitió ir más allá del concepto de asociación basado en la lejanía respecto a la independencia estadística (sobre el que se erige ji-cuadrada) y se generó la idea de que los valores de los coeficientes son diferentes porque miden la cercanía o lejanía de la distribución de los datos a distintos conjuntos de proposiciones estadísticas que “operacionalizan” enunciados teóricos (Hildebrand, D., J. Lain y H.

¹³ Nótese que no uso el término de hipótesis teórica pues por en esos años la palabra hipótesis era “reaccionaria”, tenía un fuerte tufillo funcionalista.

¹⁴ Debe recordarse que uno de los problemas que aquejaban a la investigación social en los sesenta era la ausencia de criterios para decidir cuál de los coeficientes de asociación disponibles usar, toda vez que la proliferación de índices de asociación desarrollados en esa época proporcionaba un amplio abanico de opciones a disposición del investigador y que habitualmente se disponía de más de un coeficiente de asociación para las combinaciones de niveles de medición de las variables de la tabla; fue así como la decisión a favor de uno u otro coeficiente ya no podía basarse en las escalas de medida.

Rosenthal, 1977). Por otra parte, el desarrollo de esta idea develó las estructuras lógicas sobre las cuales se erigen los coeficientes de asociación más utilizados. Con base en este conocimiento se propuso la función generatriz de coeficientes delta-ro, donde el sufijo ro denota la proposición lógica que se debe especificar para que el coeficiente asuma una forma determinada. La definición misma de este nuevo coeficiente requiere para su aplicación establecer la distribución esperada de los datos.

En varios países de Ibero América, en los márgenes y rincones del quehacer se realizaban investigaciones aisladas que buscaban articular las técnicas disponibles con el estudio de procesos sociales concretos tales como las características y evolución de los mercados de trabajo, sector informal urbano, comportamiento electoral de mujeres y jóvenes, determinantes de flujos migratorios, etc.

La idea propuesta por Hildebrand *et al* en el ámbito de la Estadística Social cazó con añejas preocupaciones de la investigación social empírica en Ibero América. El análisis estadístico había privilegiado la independencia estadística como criterio para generar las frecuencias esperadas y de ahí medir la asociación por lejanía de las frecuencias observadas respecto a las esperadas, pero ¿Por qué no hacerlo con base en la estructura lógica contenida en las hipótesis teóricas?

Premunidos con la idea de que el enlace de la teoría y la técnica estadística, a través de la lógica, define unívocamente el coeficiente de asociación a utilizar Cortés F. y R. M. Rubalcava (1987) elaboraron un libro, basado en la experiencia de investigación de la Región, que mostraba las conexiones entre algunas discusiones teóricas y la investigación social empírica. Esta tarea se llevó a cabo en la época en que dominaban sin contrapeso las discusiones conceptuales que solían ser abstractas en la mayoría de los casos.

En la década de los sesenta la Econometría avanzó en la posibilidad de incluir, cualquier número de variables explicativas no métricas sin limitaciones respecto a la cantidad de categorías¹⁵, pero con las restricciones habituales de la regresión referidas a la ausencia de combinación lineal de variables y tener más observaciones que

¹⁵ El análisis de varianza, de amplio uso en la Psicología, se aplica en los casos que interesa estudiar la relación entre una variable dependiente métrica y una o más independientes medidas éstas en escalas ordinal o nominal. A este mismo problema se puede aplicar el modelo de regresión sustituyendo las variables explicativas por regresores o variables ficticias.

parámetros a estimar¹⁶. Durante la década de los 70 y 80 hubo una serie de progresos en el desarrollo de nuevas técnicas de análisis estadístico impulsadas por las demandas planteadas por las ciencias sociales. Entre los nuevos instrumentos que puso la Estadística a disposición de los investigadores destacan el análisis loglineal y las regresiones con variables dependientes no métricas dicotómicas.

El primero se puede ver como una técnica que contiene y supera el análisis de covarianzas de Lazarsfeld en la medida que permite estudiar las relaciones entre dos o más variables no métricas controlando el efecto de las restantes. Por fin se cumplía el sueño de Lazarsfeld de contar con un modelo de análisis de variables múltiples aplicable a casos donde las variables fuesen ordinales o nominales. La principal limitación ya no radica en cuántas variables y cuántas categorías se pueden considerar a la vez sino principalmente el número de casos¹⁷. En el futuro esta técnica se empleará profusamente para analizar la movilidad social (Goldthorpe J, 2000).

Las barreras que dividían los análisis de asociación y regresión habían empezado a caer en los años posteriores a la Segunda Guerra Mundial este proceso ha continuado a lo largo del tiempo. Durante estos años se elaboraron los primeros modelos de regresión lineal con variable dependiente binaria y el centro de atención se puso en los problemas de estimación. La discusión, bastante técnica, dejó en la penumbra el vínculo que existe entre estos modelos y el concepto de explicación en Max Weber.

En la concepción weberiana no sólo se plantea que las ciencias sociales tienen mayor sed causal que las ciencias naturales sino también que la explicación es esencialmente probabilística:

En la concepción weberiana, *el tema de la causalidad es el elemento central de la diferenciación entre las ciencias*: en las ciencias sin necesidad de interpretación, la causalidad se reduce a la probabilidad de ocurrencia de los fenómenos con arreglo a una regularidad generalizada por inducción. En las ciencias comprometidas con la interpretación, debido a la necesidad de elaborar esquemas teóricos que hagan posible *comprender el sentido de las acciones*, el sujeto no se reduce a constatar, sino que le es asignada, con claridad, la función de *imputar*, a las relaciones

¹⁶ Me refiero, en particular, a restricciones respecto a los datos y deo fuera los supuestos relativos a los errores aleatorios.

¹⁷ El análisis loglineal y el análisis de asociación, son técnicas que requieren de muchos casos; "son consumidoras de observaciones".

constatadas, una direccionalidad de tipo causal, en un contexto de regularidades comprendidas y con cierto grado de verosimilitud (Gil M. 1997: 214 y 215).

Hempel sistematiza, desde el punto de vista filosófico, la noción de explicación en Popper (1962: 57 a 60) y su análisis le lleva a concluir que la explicación social es esencialmente probabilística (1988:249 a 250); Przeworski y Teune (1972: 19) presentan una buena síntesis de cómo aterriza esta discusión filosófica en el campo de la investigación.

La importancia del modelo de regresión con variable dependiente binaria radica en que permite articular la teoría sociológica con la teoría estadística, el puente entre ambas es la explicación weberiana. Para develar este vínculo considérese, en primer lugar, el modelo de probabilidad Bernouilli:

$$b(x; p) = P^x(1 - P)^{1-x} ; x=0,1$$

que asume el valor de probabilidad P, si $x=1$ y $(1-P)$ si $x=0$. En otros términos, esta distribución modela la probabilidad de ocurrencia de un fenómeno aleatorio: el evento sucede con probabilidad P y no ocurre con probabilidad $(1-P)$.

Por otra parte, es fácil demostrar que en el modelo de regresión con variable dependiente dicotómica el valor esperado es igual a la probabilidad de ocurrencia del fenómeno. Además, el modelo mismo expresa la idea de que la probabilidad de un fenómeno depende de un conjunto de variables explicativas, lo que equivale a decir que los valores observados de la variable dependiente (0, 1) son generados por distribuciones Bernouilli con parámetro P_i y que estos parámetros varían con los valores de X.

En este argumento hay que destacar que la probabilidad resulta ser el punto de encuentro entre la teoría estadística y la teoría social sustantiva¹⁸. En efecto, es responsabilidad de la teoría social identificar los factores de los cuales dependen las probabilidades de las diferentes distribuciones Bernouilli. De este modo se vuelve a restablecer, aunque sólo en lo conceptual, el antiguo maridaje entre la Estadística y la Sociología.

¹⁸ El puente entre la teoría económica y el modelo de regresión con variable binaria lo proporciona la teoría de la elección racional (Luce R. y P. Suppes, 1965 y McFadden 1973).

En síntesis, los valores estimados en un modelo de regresión con variable dependiente dicotómica arrojan como resultado las probabilidades asociadas al fenómeno, dado el conjunto de circunstancias representadas por los valores de las variables explicativas.

Hay que agregar que antes de que se desarrollaran los modelos de regresión no lineales para variables dependientes binarias se contaba el modelo de regresión probabilístico lineal, el que desde el punto de vista teórico era una extensión del modelo de regresión lineal clásico, relativamente sencilla, pero su ajuste presentaba obstáculos serios desde el punto de vista computacional en una época en que el uso de la computadora personal aún no era muy extendido entre los científicos sociales, y el costo monetario del tiempo de uso de los grandes equipos era elevado. En adición a los problemas de estimación este modelo no garantizaba que los valores estimados estuviesen dentro del intervalo 0 a 1 (ambos extremos incluidos) y, por otra parte, no parece razonable suponer que el efecto de las variables explicativas sea constante¹⁹.

Los problemas del modelo de regresión lineal de probabilidades se superaron empleando transformaciones no lineales de la variable dependiente. Es así como surgieron los modelos de regresión logística y probit.

Por fin la Estadística ponía a disposición de los científicos sociales una técnica que permitía emplear el modelo de regresión en casos en que todas las variables eran no métricas, aunque existía la restricción de que la variable dependiente debía ser dicotómica. Se empezaba a caer así el criterio que diferenciaba en los años sesenta y setenta la aplicación de la Estadística en la Economía y la Sociología y consecuentemente se extendía el dominio del poderoso modelo de regresión.

El desarrollo de la Estadística Social se complementó con la invención de la computadora personal que no sólo resultaba ser mucho más amigable que los “mainframe” sino que también abatió enormemente los costos de las investigaciones. La nueva herramienta si bien era relativamente barata, en sus inicios presentaba claras limitaciones: los programas estadísticos se reducían a las técnicas más elementales y los dispositivos de almacenamiento permitían guardar pocos datos y unas cuantas

¹⁹ El detalle de estos y otros argumentos que ponen en tela de juicio la aplicación del modelo lineal de probabilidad se encuentran en Aldrich J. y F.Nelson (1984: 24 a 30).

variables. Sin embargo, si bien al principio las computadoras personales funcionaban con dos lectoras de disco flexible muy pronto aparecieron los discos duros de 5 MB, luego de 10 MB; hacia fines de los ochenta los discos duros ya tenían 100 MB.

El vertiginoso avance del “Hardware” vino acompañado de la disponibilidad del “Software” estadístico. Por una parte, los paquetes estadísticos más empleados por los científicos sociales en los “mainframe”, como SPSS, SAS y BMDP, desarrollaron versiones para las microcomputadoras y por otra parte, empezaron a aparecer en el mercado programas estadísticos desarrollados específicamente para “correr” en las computadoras personales.

Empieza a perfilarse así la última etapa en que lenta y penosamente la Estadística y parte de la antigua Metodología de las Ciencias Sociales vuelven por sus fueros.

IV.- El lento resurgimiento

A lo largo de la década de los ochenta tienden a desaparecer los gobiernos de facto en América Latina y a ser sustituidos por otros electos por el voto ciudadano; la democracia electoral se extiende como un *sunami* por la Región.

Este avance de la democracia se combinó con el estancamiento económico. La explosión de la crisis del petróleo en 1982 socavó las bases del crecimiento de los países de América Latina. La “década perdida” se caracterizó, entre otras cosas, porque el gasto de los gobiernos sufrió severos recortes buscando el equilibrio con los ingresos. Ganaba espacio el concepto de déficit cero en el balance contable de las cuentas del gobierno; idea que a la vuelta del tiempo sería uno de los componentes de la primera oleada del Consenso de Washington (Williamson J., 1990).

Hacia finales de los ochenta y durante los noventa, la mayoría de los países de la Región iniciaron procesos de cambio estructural tendientes a liberar las fuerzas del mercado, la consigna era “más mercado y menos estado”. A pesar de los pronósticos basados en la “teoría económica” que apostaban a los efectos beneficiosos sobre las economías de América Latina, éstas exhibieron durante los años noventa inestabilidad y

ausencia de crecimiento²⁰. Era la “segunda década perdida” consecutiva (Williamson J., 2003 1 a 6).

La situación económica general se dejó sentir sobre las instituciones de educación superior afectando la investigación y la docencia. El financiamiento de las investigaciones proveniente de fundaciones u organismos internacionales ganó en importancia en algunos países y también tendió a centralizarse en Consejos Nacionales de Ciencia y Tecnología. Hubo países que crearon becas para complementar salarios en las propias instituciones de educación superior o por organismos externos creados *ex profeso* para impulsar y orientar la investigación.

Los escasos recursos tendieron y tienden a ser canalizados al estudio de problemas sociales específicos privilegiándose la investigación social empírica. La caída en el poder adquisitivo de los salarios de los académicos y los cambios institucionales crearon un contexto cada vez más desfavorable para los estudios especulativos, (entendidos como aquellos que se refieren a los procesos y problemas sociales sólo a modo de ejemplo) y para los genuinamente teóricos que habían gozado de alta retribución y estima social a comienzos de la década de los ochenta.

En este ambiente la metodología de las ciencias sociales volvió sobre sus pasos y recuperó algunos de los temas de los años sesenta: técnicas de muestreo, construcción de cuestionarios, entrevistas, observación participante, etc. y también tópicos básicos de lógica de la investigación. Los avatares que han vivido las ciencias sociales en Ibero América han dejado sus huellas en la metodología; hacia fines del siglo XX y comienzos del XXI es un mosaico que exhibe parte de la historia de estas disciplinas. Bajo el término metodología se ofrecen cursos que cubren una serie de materias que van desde la estadística elemental y técnicas de *survey* en un extremo, hasta la filosofía de la ciencia y epistemología en el otro.

Durante el segundo quinquenio de la década de los ochenta se extiende como reguero de pólvora, fuera de la Región pero con lentitud en América Latina, el uso de modelos de regresión no lineales con variable dependiente no métrica. Son varios los factores que tienen incidencia sobre este impulso en el uso de la Estadística:

²⁰ Aunque la inestabilidad también afectó a la economía chilena, tal vez este país fue la excepción en la medida que su economía creció durante estos años.

- (i) Por un lado esta disciplina amplió el modelo de regresión no lineal de variables dicotómicas a variables dependientes con varias categorías (pluricotómicas) y a variables dependientes ordinales. Con base en este conocimiento se han desarrollado modelos para analizar historias de eventos, tratar problemas de selección, etc.
- (ii) Por otra parte, las capacidades de la computadora personal experimentaron cambios sorprendentes, el día de hoy la memoria se mide en megas, en lugar de hacerlo en k bytes, y la capacidad de almacenamiento pasó de 10 MB a 100 o más GB. Además, abundan los programas estadísticos capaces de procesar en segundos o minutos grandes bases de datos.
- (iii) Los desarrollos en teoría estadística y computación han hecho posible que los estudiantes y los investigadores aprendan a leer los resultados que arroja la computadora sin que se requiera tener un conocimiento en profundidad de inferencia estadística y de los supuestos en que descansa la técnica empleada. Esto es posible a pesar de que tanto el modelo de regresión con variables dependientes binarias (modelos logit y probit), como multinomiales y ordinales, requieren de un buen conocimiento de la distribución ji-cuadrada, de estimación máximo verosímil y de teoría de las distribuciones asintóticas.
- (iv) Por último, pero no menos importante, los modelos no lineales para variables dependientes no métricas establecen una relación intrínseca con la teoría. En efecto, en el caso de la economía se derivan a partir de la teoría de la acción racional, que enraíza en la teoría económica dominante, y se aplica cuando sólo se observan las decisiones racionales de los individuos como resultado del análisis comparativo de los factores determinantes de sus preferencias (Greene W: 2003: 663 a 674). En la Sociología, como

se ha señalado, el vínculo radica en la explicación sociológica en la versión weberiana.

Con el modelo de regresión no lineal para variables dependientes no métricas (pluricotómicas y ordinales) se cierra un círculo, finalmente se derrumba la idea de que los niveles de medición de las variables diferenciaban a la estadística susceptible de ser aplicada a la Sociología y a la Economía: ya es posible analizar tablas de contingencia empleando regresión.

Estrechamente vinculada a los problemas planteados por la Sociología de la Educación ha emergido con fuerza en los años noventa el análisis jerárquico lineal aunque sus antecedentes en la Sociología se remontan a 40 años atrás, en el contexto del análisis ecológico (Boudon R. 1974: 271 a 284); estos desarrollos de la Estadística aplicados a la Sociología proporcionan una técnica que permite analizar datos y formalizar los vínculos teóricos entre conceptos macro y micro sociales.

Ahora bien, esta técnica puede verse como una generalización del análisis de varianza o del análisis de regresión. Está diseñada para identificar los efectos de diferentes niveles de análisis sobre la variable medida al mayor nivel de desagregación, el ejemplo paradigmático, es el de identificar qué parte de la variabilidad en la calificación de los estudiantes se debe a las características del muchacho, cuál se origina en la familia y cuál corresponde a la escuela (Raudenbush S y A. Bryk: 2002). Obviamente las aplicaciones no tienen porqué circunscribirse a la Sociología de la Educación, se trata de una técnica cuya estructura lógica permite encarar problemas típicos de las ciencias sociales tales como, por ejemplo, la evaluación de los efectos de un programa de intervención estatal sobre los habitantes de localidades o comunidades (Hernández et al, 2000) o cualquier situación donde importe distinguir los efectos de agregados sociales sobre individuos o grupos al interior de dichos agregados.

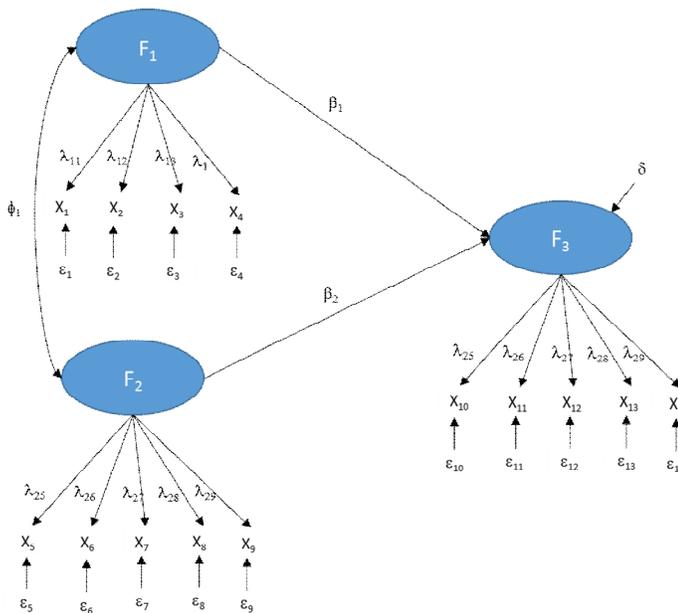
Si bien las primeras versiones de los modelos jerárquicos lineales se restringieron a variables dependiente en escala de intervalo o de razón, los últimos desarrollos consideran variables no métricas vinculándose así con los modelos de regresión con variable dependiente binaria. Además, ya hay en el mercado paquetes de

programas que permiten ajustar modelos multinivel relativamente complejos con conocimiento estadístico relativamente exiguo.

En los últimos años el modelo de ecuaciones estructurales se ha constituido en un nuevo enfoque para analizar datos estadísticos. Se trata de una técnica de modelaje estadístico ampliamente usada en las ciencias sociales de los Estados Unidos y Europa. En general puede verse como una combinación del análisis factorial y el análisis de regresión o el análisis de senderos. El interés de esta aproximación a menudo se centra sobre los constructos teóricos que son representados como variables latentes. Las relaciones entre las variables latentes se identifican mediante coeficientes de regresión o por coeficientes “path” o de senderos (Hox y Berchger, 2001).

Los modelos de ecuaciones estructurales suelen representarse por gráficas del tipo siguiente:

Gráfica 1



En esta gráfica F_1 , F_2 y F_3 son variables latentes, y x_1, x_2, \dots, x_{14} son variables observadas o manifiestas, los términos $\epsilon_1, \dots, \epsilon_{14}$ corresponden al modelo de error de medición de cada variable latente; β_1 y β_2 son los coeficientes de sendero y δ es el error residual del modelo estructural. La estructura representada contiene el modelo de medición de cada variable latente (F_1 , F_2 y F_3), así como el modelo teórico, también denominado explicativo o estructural que relaciona a las variables latentes a

través de senderos. El primero, el de medición, establece que para hacer observables las variables latentes se emplean combinaciones lineales de las variables observadas (con pesos λ_{ij}), y, como en todo modelo de medición las variables observadas tienen una parte sistemática y otra de error, esta última se simboliza por ϵ_j . El modelo estructural supone que la variable latente F1, está relacionada con F2 mediante la covarianza ϕ_1 , F1 y F2 condicionan a F3 con coeficientes β_1 y β_2 , respectivamente.

Antes de seguir en esta línea es conveniente llamar la atención sobre tres temas metodológicos de la mayor importancia para la investigación aplicada.

El modelo supone que los conceptos teóricos representados por las variables latentes no son directamente observables. He aquí una primera ruptura con las técnicas de análisis de datos que habitualmente se emplean en las ciencias sociales en la Región ya que normalmente se aplican a las variables observadas como si representaran adecuadamente conceptos.

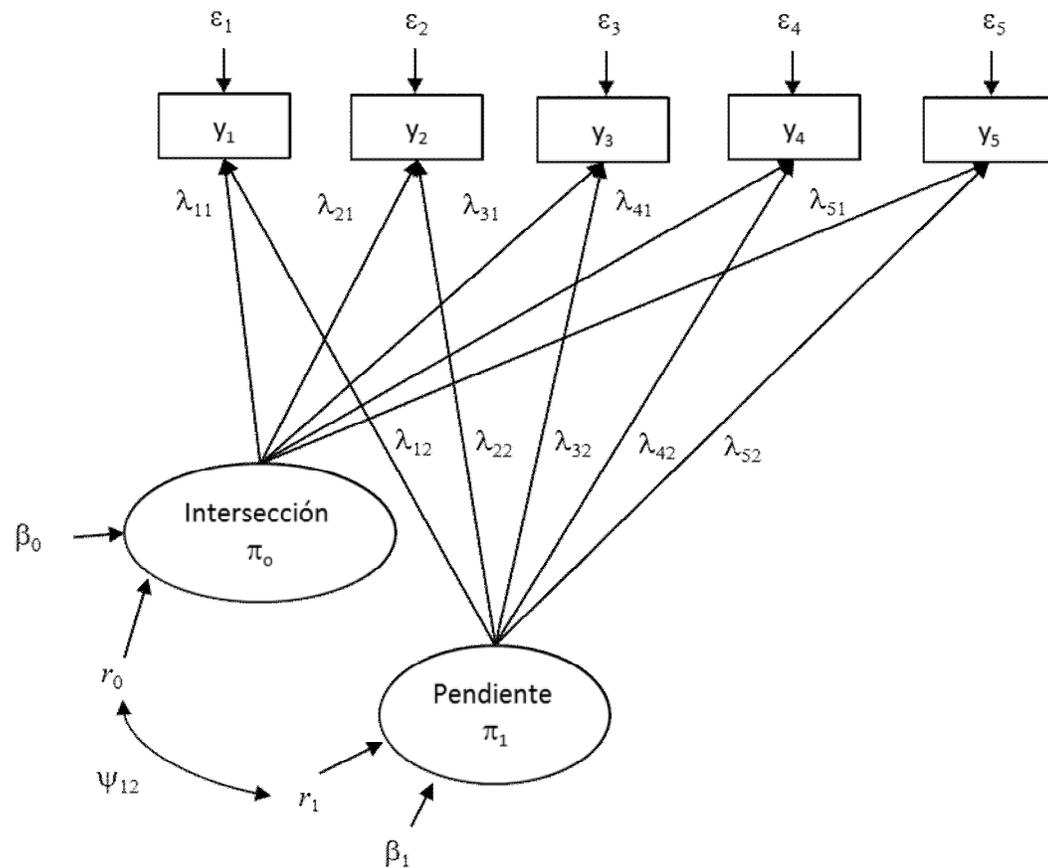
Este modelo incluye las imprecisiones de la medición al tomar en cuenta los errores de medición.

De acuerdo con los planteamientos de Cortés y Rubalcava (1993, 227 a 267) esta estructura no corresponde a ninguna de las técnicas habituales; la valoración de sus parámetros requiere ajustar de un modelo de ecuaciones estructurales.

Debe destacarse que este modelo apunta a temas metodológicos centrales para la investigación social empírica. Por una parte, entra en el plano de las teorías, las hipótesis y los conceptos, y por otra parte, hace explícita la operación de medición entendida como la relación entre los conceptos, su objetivación y los indicadores. En resumidas cuentas sintetiza el problema central de la investigación empírica: relacionar el producto del pensamiento (teoría, conceptos, hipótesis) con el material empírico que es el resultado de la construcción u organización de la información que se obtiene de la experiencia.

Si bien el modelo representado por la gráfica 1 es el más elemental, estas mismas ideas pueden representar modelos de crecimiento latente como sería el caso representado en la gráfica 2.

Gráfica 2



En este modelo la intersección y la pendiente de las variables a lo largo del tiempo (se suponen 5 mediciones en el tiempo) varían para cada observación de acuerdo con una parte sistemática representada por β_1 y β_0 , más una parte aleatoria simbolizada por r_1 y r_0 , que se correlacionan (es decir, covarían). Los valores estimados de las variables se modelan como la combinación de la intersección y la pendiente que no es la misma para cada observación pues ambas contienen un término de error. Los coeficientes λ representan los pesos con que concurren la intersección y la pendiente para formar el valor estimado de la variable. El modelo se puede hacer un poco más complejo (modelo condicional) si se considera que la pendiente y la intersección están condicionadas o dependen de las variables explicativas exógenas x_1, x_2, \dots, x_k .

El modelo de ecuaciones estructurales se ha desarrollado para realizar categorizaciones de variables latentes (análisis de clases latentes) y llevar a cabo análisis de transiciones entre las clases latentes, al estilo de la movilidad social. No viene al caso seguir destacando las múltiples aplicaciones de esta aproximación. Se trata de una estructura general que engloba una gran cantidad de técnicas estadísticas desde las más elementales hasta las muy desarrolladas, que posibilita aplicaciones muy variadas a diversos problemas y demanda el manejo de un solo paquete estadístico.

Este tipo de modelos se llaman de segunda generación, porque incorporan dos enfoques en la forma de analizar los datos, uno centrado en las variables y otro centrado en las personas. El primer enfoque es el que analiza las observaciones usando las variables y las agrupa en constructos latentes, a las cuales se les ha identificado como análisis factorial. El segundo enfoque se centra en las personas y ahora construye categorías de personas con características homogéneas dentro de grupos y heterogéneas entre grupos, que llamamos análisis de clases latentes. Ambos enfoques pertenecen a los modelos estructurales de segunda generación porque incorporan varios modelos tradicionales en uno sólo.

Las críticas a los modelos de ecuaciones estructurales se han concentrado sobre dos temas. Uno estadístico, que pone la atención en la importancia del supuesto de normalidad y los tamaños de muestra necesarios para tener resultados confiables, y otro relativo a la interpretación de los resultados.

En cuanto al número de observaciones la experiencia ha mostrado que el método de estimación máximo verosímil, aplicado bajo el supuesto que los datos siguen una distribución normal multidimensional, arroja estimaciones confiables con un número de observaciones “grande”²¹. Este mismo resultado se ha logrado con modelos de simulación.

Cuando falla el supuesto de distribución normal conjunta y se dispone de variables continuas o categóricas se pueden emplear métodos de estimación bayesianos o bien mínimos cuadrados parciales. Mediante técnicas de simulación de Monte Carlo se ha concluido que los métodos de estimación no paramétrico requiere

²¹ Por ejemplo, para el modelo de la gráfica 1 se requieren más de 200 observaciones para que la estimación máximo verosímil encuentre una solución satisfactoria.

contar con muestras de más de 1000 casos. Diversos ensayos numéricos han mostrado que aun desviándose de la normalidad el método de estimación máximo verosímil entrega, en la mayoría de los casos, buenas estimaciones si se cuenta con 400 o más casos. Otra alternativa consiste en corregir los valores de ji-cuadrada por el grado de no normalidad.

Si los datos son ordinales se puede suponer que son mediciones imprecisas de variable continuas. Con este supuesto se computan correlaciones policóricas y se procede a realizar una estimación no paramétrica, que requiere un número grande observaciones (más de 1000 casos). La experiencia ha mostrado que si se dispone de 5 o más categorías se puede hacer caso omiso de la ordinalidad de las variables, y se obtienen buenos resultados. Varios autores han propuesto métodos alternativos para datos no normales (Browne, 1984; Satorra & Bentler, 1994). Los métodos incluyen el uso de máxima verosimilitud robusta y también se conocen como estimadores tipo sándwich

El número de casos y el supuesto de normalidad se pueden enfrentar con los desarrollos que proporciona la misma disciplina Estadística. Para ello, se debe hacer énfasis en calcular el tamaño de muestra necesario o la potencia estadística asociada al modelo que permitan resolver este problema (Muthen & Muthen, 2002). Sin embargo, el tema que más debe preocupar es el relativo a la interpretación de los resultados. Es frecuente encontrar aplicaciones de esta técnica en que las interpretaciones son causales aunque los datos no sean experimentales (tal como suele acontecer con la interpretación de los modelos econométricos), dichas interpretaciones podrían ser correctas, sin embargo, su validez no proviene de la técnica estadística empleada, no hay nada en los modelos de ecuaciones estructurales que transforme las asociaciones estadísticas en relaciones causales.

V.- Conclusión

El centro de interés de este texto son los avatares que ha experimentado la Estadística aplicada a la investigación y su enseñanza, en el campo de la Sociología durante los últimos cuarenta años. Se ofrece un “modelo histórico”, que no

necesariamente es aplicable in toto a cada país iberoamericano, y que distingue tres etapas ordenadas en el tiempo: (i) Auge, (ii) Destierro y (iii) Resurgimiento.

En algunos países la Sociología Profesional surgió bajo el predominio de los temas metodológicos de la época del destierro y el papel de la estadística, tanto en la investigación como en la docencia, ha quedado circunscrito a la descriptiva, al análisis de asociación y a la correlación, sin considerar los enlaces con la inferencia estadística. Otros, han transitado de la segunda a la tercera etapa, sin pasar por el período de auge; la investigación que se realiza emplea las técnicas estadísticas más modernas y en sus aulas se las enseña utilizando grandes bases de datos y computadores personales poderosas, pero, hay que reconocer, su uso se limita a relativamente pocos investigadores y que la enseñanza se reduce a unas cuantos programas de postgrado en Sociología. Hay otros países en que la investigación y la docencia en Sociología transitaron de la época de auge a la del destierro y ahí han permanecido.

El análisis de lo acontecido con la Estadística en Ibero América muestra que hasta finales de la década de los años sesenta la estadística descriptiva, el muestreo y la inferencia estadística proporcionaban valiosos instrumentos de recopilación de información útiles para caracterizar poblaciones. De las técnicas para estudiar relaciones entre variables disponibles en esa época, la más usada en Sociología y Ciencia Política era el análisis de asociación y el de covarianzas de Lazarsfeld, este último especialmente válido para analizar las relaciones entre tres o más variables dicotómicas. El análisis de regresión era el instrumento estadístico más popular de los economistas y de los científicos políticos dedicados al análisis electoral. Se argumentaba que la diferencia entre los instrumentos de análisis de datos de la Estadística Social y de la Estadística Económica se originaba en la escala en que se medían las variables: predominante nominal y ordinal en el primer caso y de razón e intervalar en el segundo.

En los setenta y el primer quinquenio de los ochenta el conocimiento estadístico que había jugado un papel importante en la investigación social del período anterior ya no ayudaba a responder las preguntas que surgían del enfoque histórico estructural²²,

²² En esa época aún no se contaba con desarrollos estadísticos, disponibles hoy para analizar cambio estructural, ni tampoco con los conceptos elaborados por la teoría del caos ni con las leyes de evolución

centradas en la dinámica del cambio de las estructuras sociales. En las aulas de Ibero América esta disciplina fue reducida a su mínima expresión.

Sin embargo, continuó desarrollándose fuera de la Región y hubo importantes avances en la estadística teórica impulsados por las preguntas que surgían desde las ciencias sociales. En este período culmina el desarrollo de los modelos de regresión no lineal, en particular logit y probit y del análisis loglineal. La incorporación de variables no métricas en el lado derecho de la ecuación se complementó con la inclusión de variables dependientes dicotómicas en el lado izquierdo. Con estos avances empieza a desmoronarse el muro que separaba a la Estadística Social y a la Estadística Económica. A su vez los nuevos modelos estadísticos nacían articulados a la explicación social.

En los últimos años bajo la batuta del progreso tecnológico en la fabricación de computadoras personales cada vez más potentes, a precios sistemáticamente más bajos y un mercado en expansión que facilitó su acceso, combinado con amplia oferta de paquetes estadísticos, hizo posible “ajustar” modelos no lineales ya no sólo a variables dependientes dicotómicas sino también pluricotómicas y ordinales a bajo costo. Por otra parte, los vínculos estrechos entre los modelos teóricos dominantes en Economía, Sociología y Ciencia Política y los nuevos modelos estadísticos, vía la teoría de la elección racional o la explicación en la vertiente weberiana, llevaron a la proliferación de estudios sociales que utilizan los nuevos modelos. Todo esto aconteció en otras geografías, en las nuestras se registraron algunos estudios aislados aunque en los últimos años han sido cada vez más frecuentes.

Hay que destacar que en el período considerado ha variado el rol de la Estadística en el quehacer de los científicos sociales. En la Ibero América de los años sesenta dicha disciplina era parte constitutiva de la investigación social; estaba articulada a la teoría y metodología dominantes. En el segundo período dicha articulación se rompe cuando cambiaron radicalmente las preguntas de investigación, de cara a los acontecimientos sociales y políticos de la época; la teoría dominante; y la escasa adecuación del conocimiento estadístico del momento para ofrecer respuestas

de los sistemas complejos. Los estudios de la escuela piagetiana se asociaban más bien a la pedagogía y a la psicología, desconociéndose su veta epistemológica (García R, 2000: 11)

válidas a las nuevas preguntas. A partir del tercer período se advierte en Ibero América el inicio del reencuentro entre la teoría social, la metodología y la Estadística. Se vuelve así a la articulación inicial pero en un contexto en que su legitimidad es disputada por los métodos cualitativos.²³

En la medida que el modelo de regresión extiende sus dominios al incorporar variables dependientes no métricas termina por caer el muro entre la estadística aplicada a la Sociología y a la Economía; ya se puede usar para analizar tablas cruzadas con varios criterios de clasificación simultáneos.

El avance tecnológico hizo cada vez más fácil la aplicación del análisis estadístico. Para obtener resultados basta con tener un problema bien definido, disponer de información mínima respecto a los modelos estadísticos disponibles, los datos pertinentes, la máquina y los programas adecuados,. La interpretación es harina de otro costal pues requiere la concurrencia de conocimiento estadístico y de la disciplina en cuestión. Lo que sí es destacable es que esta labor se puede realizar ¡sin necesidad de gastar mucho tiempo en el estudio de la Estadística!

En cuanto a la docencia se abren dos caminos. Uno consiste en entregar los conocimientos estadísticos mínimos necesarios para aprender a interpretar las salidas de las computadoras. El otro, el tradicional, recorre la trayectoria que une a la estadística descriptiva con las técnicas modernas de análisis multivariado, pasando por el análisis de asociación, el muestreo, la inferencia estadística, análisis de varianza y regresión lineal.

La experiencia muestra que el primero de estos caminos tiene el inconveniente de que el investigador social, que sólo dispone de dicha formación estadística, suele sufrir serias limitaciones para incorporar los avances de la técnica, cuestión que no ocurre con quienes tienen una formación estadística más sólida. Pero hay claras diferencias en el tiempo que se debe invertir para proporcionar una u otra formación. Para enseñar a leer salidas de computadoras bastan dos o tres semestres mientras que una formación más estructurada, además de demandar mayor formación matemática, suele requerir cinco o seis semestres académicos.

²³ F. Cortés (2000: 103 a 132) hizo un análisis sistemático de las discusiones epistemológicas entre las investigaciones cualitativas y cuantitativas. El mismo autor estudia los procesos de generalización en las investigaciones estadísticas, experimentales y en los estudios cualitativos (2004).

No es fácil llegar a un balance entre tiempo y profundidad del conocimiento estadístico para científicos sociales. En el futuro próximo, en la medida que se intensifique la demanda estudiantil por este tipo de conocimiento, habrá que aprovechar experiencias desarrolladas en el extranjero y ensayar formas creativas de enseñanza que permitan dar a nuestros estudiantes buena formación estadística, en poco tiempo, y estrechamente ligada a las preocupaciones académicas que surgen de la reflexión problematizada de la evolución de nuestras realidades sociales.

Bibliografía

Aldrich John y Forrest Nelson (1984), *Linear Probability, Logit and Probit Models*, Sage University Paper, Series: Quantitative Applications in the Social Sciences, No. 45, Sage Publications, California.

Blalock Hubert (1961), *Causal Inference in non Experimental Research*, The University of North Carolina, Chapel Hill.

Borgatta Edgar, George W. Bohrnstedt (Eds) (1970), *Sociological Methodology*, Josse Bass, San Francisco.

Boudon Raymond (1970), *L'analyse mathématique des faits sociaux*, Plon, París.

Boudon Raymond (1974), "Propiedades individuales y propiedades colectivas: un problema de análisis ecológico", en *Metodología de las ciencias sociales II. Análisis empírico de la causalidad*, Laia, Barcelona.

Browne, M. W. (1984). Asymthotic distribution free methods in the analysis of covaraince structures. *British of Mathematics and Statistical Psychology*, 37, 62-83.

Bunge Mario (1979), *La investigación científica: Su estrategia y su filosofía*, Ariel Barcelona.

Carnap Rudolf (1959), "Psicología en lenguaje fisicalista", en Ayer A. J. (comp.) *El positivismo lógico*, Fondo de Cultura Económica, México.

Cortés Fernando (1967), "Algunos coeficientes de asociación en tablas de dos por dos". *Boletín de ELAS*, Año 1, Número 1.

Cortés Fernando (2000), "Algunos aspectos de la controversia entre investigación cualitativa y cuantitativa", *Argumentos: Estudios críticos de la sociedad*", División de Ciencias Sociales y Humanidades, Universidad Autónoma Metropolitana, Unidad Xoxchmilco.

Cortés Fernando (2004), "Selección no aleatoria y validez. A propósito de la evaluación cualitativa de Oportunidades", mimeo.

Cortés Fernando y Rosa María Rubalcava (1987), *Métodos estadísticos aplicados a la investigación social en ciencias sociales: análisis de asociación*, El Colegio de México, México.

Costner Herbert L. (ed) (1971), *Sociological Methodology*, Josse Bass, San Francisco.

Costner Herbert L. (ed) (1972), *Sociological Methodology*, Josse Bass, San Francisco.

García Rolando (2000), *El conocimiento en construcción: De las formulaciones de Jean Piaget a la teoría de los sistemas complejos*, Gedisa, Barcelona.

Gil Manuel (1997), *Conocimiento científico y acción social: Crítica epistemológica a la concepción de ciencia en Max Weber*, Gedisa, Barcelona.

Goldthorpe John (2000), *On Sociology: Numbers, Narratives and the Integration of Research and Theory*, Oxford University Press, Oxford.

Goodman Leo y William Kruskal (1954), "Measures of association for cross classifications" en *Journal of the American Statistical Association*, 49.

Goodman Leo y William Kruskal (1963), "Measures of association for cross classifications III: Approximate Sampling Theory" en *Journal of the American Statistical Association*.

Greene William (2003), *Econometrics Analysis*, Prentice Hall, New Jersey.

Hempel Carl G. (1988), *La explicación científica: Estudios sobre la filosofía de la ciencia*, Paidós, Barcelona.

Hernández Daniel, Mónica Orozco y Daniela Sotres (2000), "El impacto del Progreso en la inscripción a la secundaria: modelos multinivel para datos de matrícula escolar" en *Progreso, Más oportunidades para las familias pobres. Evaluación de resultados del Programa de Educación, Salud y Alimentación: Educación*. Secretaría de Desarrollo Social, México.

Hildebrand David, James Laing y Howard Rosenthal (1977), *Analysis of Ordinal Data*, Sage University Paper, Series: Quantitative Applications in the Social Sciences, No. 8, Sage Publications, California.

Johnston J. (1963) *Econometrics Methods*, McGraw Hill, Nueva York,.

Kendall Maurice y Alan Stuart (1960), *The Advanced Theory of Statistics Vol.II: Inference and Relationship*, Charles Griffin, London.

King Gary (1997), *A solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregated Data*, Princeton University Press, Princeton.

Lazarsfeld Paul (1974), "El álgebra de los sistemas dicotómicos", en *Metodología de las ciencias sociales II. Análisis empírico de la causalidad*, Laia, Barcelona.

Lazarsfeld Paul (1974), "La interpretación de las relaciones estadísticas como propiedad de investigación en Boudon Raymond, en *Metodología de las ciencias sociales II. Análisis empírico de la causalidad*, Laia, Barcelona.

Lazarsfeld Paul y Morris Rosenberg (coed) (1955). *The language of social research : a reader in the methodology of social research*, Free, Glencoe.

Luce R.D. y P.Supes (1965), "Preference, utility, and subjective probability", en Luce R.D. P.Supes E. Galanter (Eds.), *Handbook of Mthematical Psychology*", Vol. 3, John Wiley, New York.

MacKenzie Donald (1979), "Eugenics and the Rise of Mathematical Statistics in Britain", en Irvine John, Ian Miles y Jeff Evans, *Demystifying Social Statistics*, Pluto Press, Londres.

McFadden D. (1973), "Conditional logit analysis of qualitative choice behavior" en Zarembka P. (ed.) *Frontiers in Econometrics*, New York Academic Press.

Muthen, L., & Muthen, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 4, 599-620.

Neurath Otto (1959), "Proposiciones protocolares" en Ayer A. J. (comp.) *El positivismo lógico*, Fondo de Cultura Económica, México.

Popper Karl (1962), *La lógica de la investigación científica*, Tecnos, Madrid.

Przeworski Adam y Henry Teune (1972), *The Logic of Comparative Social Inquiry*, John Wiley, New York.

Raudenbusch Stephen y Anthony Bryk (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd edition*. Advanced Quatitative Techniques in the Social Sciences, Sage, California.

Robinson William (1950), "Ecological Correlation and the Behavior of Individuals", *American Sociological Review* 15.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. Clogg (Eds.), *Latent Variables Analysis, Applications to Development Research* (pp. 399-419). Thousand Oaks, CA: University of California, Los Angeles.

Siegel, Sydney (1956) *Nonparametric Statistics: for the behavioral sciences*, McGraw-Hill, New York.

Simon Herbert (1957), *Models of Man*, John Wiley, New Cork.

Williamson John (1990), "What Washington Means by Policy Reform", en *Latin American Adjustment. How much has happened?* Williamson John (ed.) Institute for International Economics, Washington D.C.

Williamson John (2003), "An agenda for Restarting Growth and Reform", en Kuczynski Pedro Pablo and John Williamson (eds.) *After the Washington Consensus: Restarting Growth and Reform in Latina America*, Institute for International Economics, Washington.

Wright Sewall (1934), "The Methods of Path Coefficients", *Annals of Mathematical Statistics*, V.

Yocelvezky Ricardo (1988), *La Democracia Cristiana Chilena y el gobierno de Eduardo Frei (1964-1970)*, Universidad Autónoma Metropolitana, Unidad Xochimilco.