

## REGRESIÓN LOGÍSTICA EN LA INVESTIGACIÓN SOCIAL:

# POTENCIALIDADES Y LIMITACIONES

FERNANDO CORTÉS -CES, COLMEX

## 1.- INTRODUCCIÓN

En este escrito intentamos dar una visión panorámica del uso de la regresión logística en las ciencias sociales, en general, y en la Ciencia Política, en particular. Para satisfacer este objetivo es necesario incluir algunos conceptos elementales del análisis de regresión. Especialmente si está dirigido a un público no especialista en métodos estadísticos. Varios de los conceptos fundamentales, en la versión estándar de ésta técnica estadística, tienen raíces en abordajes disímilos. Una mirada histórica ayuda a la mejor comprensión de los conceptos fundamentales en que descansa.

Con base en estas ideas, en la segunda sección entregamos las nociones elementales que, a nuestro juicio, allanan la comprensión de algunos conceptos básicos que se emplean en la regresión logística. También mostramos los puentes que unen a ésta técnica con el análisis de regresión lineal. En la tercera sección hay dos ejemplos en que se aplica regresión logística. El primero de ellos trata del análisis de votos tomando como unidad de análisis las divisiones político administrativas y tiene un carácter netamente teórico. El segundo ejemplo, muestra los resultados a los que se llegó, al estudiar las determinantes de la pobreza. En este caso la intención no es entrar en los vericuetos estadísticos sino limitarse al tipo de conocimiento que se genera al aplicar esta técnica de análisis de datos.

## 2.- DE LA REGRESIÓN LINEAL A LA REGRESIÓN LOGÍSTICA

Después de localizar las vertientes históricas de las que decantó el modelo de regresión, pasamos a exponer los lineamientos básicos que sustentan al análisis de regresión y terminamos con una discusión acerca de los procedimientos comúnmente utilizados para la estimación de los parámetros poblacionales.

### 2.1.- LOS ORÍGENES DEL ANÁLISIS DE REGRESIÓN

El análisis de regresión nace y se desarrolla en dos matrices culturales distintas: la francesa y la inglesa. En la primera vinculado a la astronomía y en la segunda a estudios eugenésicos.

El desarrollo del análisis de regresión, en Francia, 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 ellos envolvían, por una parte, observaciones astronómicas y por otra, la teoría de la gravitación (Stigler S. 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 que se tenía un número apreciable de observaciones y las segundas se reducían a unos pocos parámetros. Pareciera que en estos estudios no se dudaba acerca de la teoría, ésta era tomada por buena, las desviaciones entre los resultados que arroja el modelo y los datos observados, se suponía, tenían su origen en los errores de medición, los cuales, obviamente, eran considerados aleatorios.

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  $a$  y el coeficiente angular  $\beta$ . 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 S., 1986, p.12). La estimación mínimo cuadrática ordinaria así como sus variantes se ha utilizado profusamente en el ajuste de modelos de regresión, ya sea lineales o susceptibles de ser linealizados.

El desarrollo del análisis de regresión a mediados del siglo XIX en Inglaterra, en un ambiente cultural claramente distinto al francés; se enmarcó en el debate sobre las diferencias de clase, de raza y de inteligencia. No por casualidad fue Francis Galton (1822-1921) quién acuñó el término eugenesia. Su trabajo en estadística, genética y sicología de las diferencias individuales estuvo signado por su interés en mejorar la raza. La técnica de regresión le permitió predecir las características de los hijos a partir de los rasgos de los padres. El trabajo de Galton, lo continuó Karl Pearson y R. A. Fisher. El programa social y político que orientó su quehacer de investigación descansaba en las siguientes premisas: (I) suponía que la estructura de clases inglesa reflejaba las habilidades innatas de sus habitantes (II) que la cúspide de la estructura social estaba formada por la elite profesional y la base por los pobres, los desempleados y los criminales (III) que debían alentarse los nacimientos en los primeros y

limitarse en los segundos (MacKenzie D, 1979: 40). Aumentar la inteligencia y las habilidades de los nacionales de un país le permitiría a éste, por ser el país más apto, enfrentar con mayor probabilidad de éxito la lucha darwiniana entre las naciones (MacKenzie D., 1979: 42).

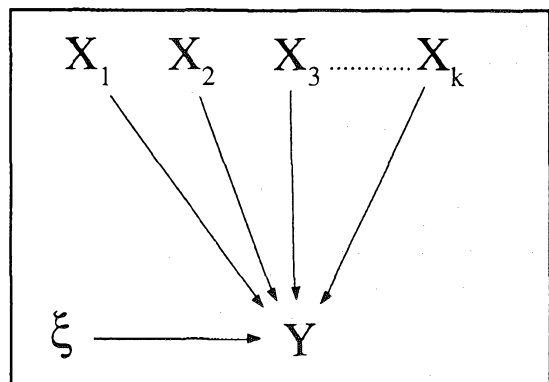
## 2.2.- EL MODELO DE REGRESIÓN: UNA MIRADA A VUELO DE PÁJARO

En los casos en que el marco conceptual lleva a la especificación de un modelo que postula la presencia de una variable explicada ( $Y$ ) por un conjunto de variables explicativas ( $X_1, X_2, \dots, X_k$ ), la técnica estadística apropiada es el análisis de regresión.

La proposición teórica que lleva a emplear esta herramienta supone que la variabilidad de la variable dependiente se puede explicar a través de las variables explicativas. El modelo estadístico correspondiente agrega que, además, opera como variable independiente un factor aleatorio no observable ( $\xi$ ), llamado error estocástico.

En esta gráfica tenemos  $k$  variables explicativas (desde  $X_1$  hasta  $X_k$ ) y la variable estocástica  $\xi$ . Nótese que no se han representado relaciones entre las variables explicativas ni entre éstas y el término de error. Las flechas que llegan a la variable explicada señalan que la teoría incluye la idea de dirección de la relación.

El ajuste del modelo estadístico especifica una serie de condiciones que debe satisfacer el término estocástico y que permiten estimar, no sólo si el modelo ajusta adecuadamente (para lo cual se tienen pruebas de hipótesis e índices de bondad de ajuste), sino también la significación y magnitud de los efectos de las variables explicativas.



La medición de los impactos de las variables explicativas requiere que no haya relación entre ellas o bien que sea tenue. En caso contrario se presentan dificultades para estimar la magnitud real de los efectos debido a que el modelo es incapaz de separar cuál corresponde a cada variable. Los procedimientos de estimación se complican si se detecta que no se cumplen los supuestos relativos al término de error. La construcción de pruebas estadísticas para decidir si se verifican empíricamente y el desarrollo de métodos especiales de ajuste han ocupado por años los esfuerzos de los econométricos.

El planteamiento clásico del modelo de regresión supone que todas las variables son métricas (medidas en escala de intervalo o de razón). Esta condición se erige en una fuerte barrera para aplicar ésta técnica estadística a los problemas típicos de las ciencias sociales. Sin embargo, esta limitación fue superada al introducir primero, *variables explicativas ficticias* (dicotómicas) que dieran cuenta de la presencia o ausencia de un evento particular (por ejemplo, el impacto sobre la función consumo de los años de paz o de guerra) (Johnston é., 1960: 221 a 228). Esta aproximación se generalizó a una o más variables pluricotómicas (Goldberger A., 1964: 218 a 226; Johnston é., 1984: 225 a 236). Es así como la regresión ganó en ductilidad, haciéndose más atractiva a los ojos de los científicos sociales. Sin embargo, el modelo, desarrollado hasta este punto, aún requería que la variable dependiente fuese métrica.

En efecto, el ajuste de un modelo de regresión lineal cuando la variable dependiente es dicotómica conlleva una serie de anomalías<sup>1</sup> en el modelo de regresión estándar. Estos problemas se superaron aplicando transformación logit a la variable dependiente (Theil H. 1971: 166 a 196; Aldrich é. y N. Forrest; 1984; Maddala G., 1983: 13 a 56; Agresti A., 1990: 79 a 119). La teoría estadística se ha generalizado para el caso en que la variable dependiente comporta más de dos categorías, pero su aplicación se ve restringida por la escasa disponibilidad de paquetes de cómputo que incluyan las rutinas de cálculo que permitan su empleo.

Como se puede apreciar, sucesivos avances liberaron al modelo de regresión de lo que alguna

vez se consideró su limitación más importante para aplicarlo al análisis de problemas sociales, haciéndolo cada vez más próximo al tan socorrido análisis de asociación.

### 2.3. ACERCA DE LOS MÉTODOS DE ESTIMACIÓN

En las ciencias sociales se recurre cada vez más a utilizar técnicas de análisis 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 le 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 sea 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  $\alpha$  y el coeficiente angular  $\beta$ .

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 explicativa están compuestos por una parte sistemática que varía con la variable explicativa y una parte aleatoria ( $\epsilon$ ), de modo que el sistema de ecuaciones responde a:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (i = 1, 2, \dots, n)$$

A las incógnitas  $\beta_0$  y  $\beta_1$  se agregan la media y la varianza de la variable aleatoria  $\epsilon$ . En la versión más simple de este modelo se supone que la esperanza matemática y la varianza de  $\epsilon$  son 0 y  $\sigma^2$  respectivamente. Nótese que el supuesto sostiene que todas las variables aleatorias  $\epsilon$ , una para cada ecuación

<sup>1</sup> Las estimaciones de la variable dependiente pueden escapar del intervalo  $[0, 1]$  y las varianzas de los estimadores no son mínimas.

ción, tienen la misma media y la misma varianza o, puesto en otros términos, que las medias y varianzas de las variables aleatorias son independientes de las observaciones.

Esta es una píldora sintética de los planteamientos estadísticos en que se basa el modelo de regresión y tiene por propósito disponer de un lenguaje básico para detenernos en lo que me parece más relevante: las ideas que llevan a distinguir la parte sistemática y aleatoria de la variable explicada y los argumentos que justifican la introducción del término  $\epsilon$ .

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, supongamos 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 este 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: 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, o bien los hijos...

Esta manera de conceptuar el 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. Dado un valor de  $X_i$ , es decir, de todas las unidades que tienen el mismo nivel de ingreso, puede asumir valores por arriba o por debajo del valor sistemático dependiendo del interjuego de 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 sólo los valores dados por la teoría y representados en la ecuación, sino también las perturbaciones en la medida: deficiencias en el tipo de instrumentos de captación, imprecisión de las respuestas, registros incompletos, etc. En nuestro ejemplo, esto quiere decir que los errores de medición en el

nivel de consumo de los hogares se consideran incorporados en  $\epsilon_i$ .

(III) En ocasiones se agrega como tercer factor la existencia de un elemento impredecible en la respuesta humana (Johnston J. 1984: Pág 14).

Me parece que detrás de la primera fuente que racionaliza la incorporación de un término estocástico en la ecuación de regresión subyace una imagen del mundo muy peculiar, según la cual para tener una explicación perfecta de la variable dependiente bastaría con incluir 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 si se quiere de observaciones) (Johnston J. 1984:14). Al estirar las consecuencias que entraña esta perspectiva se concluiría 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 de lo aleatorio se conceptúa 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.

Quiero llamar la atención sobre el hecho de que las ideas básicas que originaron la técnica mínimo cuadrática se encuentran presente 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. 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 magnitud tal que sería prácticamente imposible recabar los datos necesarios para tener más ecuaciones que incógnita. Además, habría que considerar que de todas maneras habría 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 aún 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  $\theta$  es  $\theta'$  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  $\theta'$ , dada las observaciones ( $Y$ ) y el modelo ( $M$ ). El método consiste en encontrar el valor de  $\theta'$  que maximiza la verosimilitud de que los datos fueron producidos por el modelo (King G.: 1989: 24). 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 atacar el problema.

Queremos destacar el hecho que el planteamiento del método de estimación máximo verosímil introduce el supuesto de que las observaciones se generan por un mecanismo aleatorio. Esta 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. La elección no es arbitraria sino que 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 el método de estimación mínimo cuadrático así como sus derivaciones al campo de la regresión. 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 diferentes atributos que les hacen aplicables a problemas de naturaleza distinta, sino que concep-

túan de manera distinta 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 que subyace a este procedimiento es que sería posible dar cuenta exacta de la variable en cuestión. Esta es una posibilidad teórica ya que aún cuando se consideraran todas las variables explicativas habría que generar suficientes casos como para tener más ecuaciones que incógnitas y estar en condiciones así 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 y a reemplazarla por una concepción que reconoce el carácter aleatorio de la misma y a reconocer el papel activo del sujeto en la construcción del objeto.

### 3.- REGRESIÓN LOGÍSTICA: DOS EJEMPLOS

En esta sección utilizaremos dos ejemplos para presentar las ideas, que a nuestro juicio, son las de mayor relevancia para comprender las potencialidades de la regresión logística aplicada a las ciencias sociales. El primero de ellos mostrará las potencialidades del lenguaje algebraico para llegar a conclusiones rigurosas, sin necesidad de recurrir a información empírica. El segundo, por el contrario, se incluye con la idea de delinear el uso de los resultados de un análisis concreto en la orientación de la política pública.

#### 3.1.- UN MODELO PARA ANALIZAR VOTOS

Es bastante frecuente que se disponga de datos sobre resultados de elecciones, referidos a alguna unidad administrativa electoral, como por ejemplo, municipios, distritos o estados; y que, además, se tenga algunas otras características asociadas a esas unidades administrativas. Casi siempre se trata de información censal, que para usarse adecuadamente

casi siempre requiere de algunos ajustes, ya sea en el tiempo o en el espacio<sup>2</sup>.

Supongamos, para simplificar, que por diversas razones llegamos a sostener que la proporción de votos que obtuvo un partido X en los estados de la República se puede explicar linealmente por la proporción de trabajadores que trabajan en el sector manufacturero. La ecuación que representaría esta idea sería:

$$V_i = \beta_0 + \beta_1 M_i$$

Donde  $i$  simboliza a las unidades de observación,  $V$  a la proporción de votos que obtuvo el partido y  $M$  la proporción de personas que trabajan en el sector manufacturero.

A partir de la simple observación de esta igualdad se podría conceputar a  $\beta_1$  como la propensión de los trabajadores de la manufactura a votar por el partido X. Por cada aumento unitario en la proporción de empleados en la manufactura el voto en favor de X variaría en  $\beta_1$ . Por simple extensión se podría concluir que  $\beta_0$  debería representar la propensión del resto de los trabajadores a votar en favor del partido X. Sin embargo, esta interpretación es incorrecta. En efecto, si existe un estado en que la proporción de personas que laboran en la manufactura es 1, es decir, no hay trabajadores en otras ramas, entonces la ecuación asume la forma:

$$V = \beta_0 + \beta_1$$

Ahora bien, este resultado dice que en estados en que sólo hay trabajadores manufactureros el voto en favor del partido X está formado por la propensión de éstos a votar X, y por la propensión de los trabajadores que laboran en las restantes ramas a votar X. Claramente esta conclusión contradice la premisa: sólo hay trabajadores en la manufactura.

Lo que ocurre es que el modelo está mal especificado. La forma correcta de escribirlo sería:

$$V_i = \beta_0(1 - M_i) + \beta_1 M_i$$

Donde  $(1 - M_i)$  representa a la proporción de trabajadores en las ramas no manufactureras. Una

vez que se disuelve el paréntesis y se factoriza, se llega a:

$$V_i = \beta_0 + (\beta_1 - \beta_0)M_i$$

Esta ecuación no presenta inconsistencias. Considérese que tenemos un estado en que  $M=1$ , es decir, está formado únicamente por trabajadores manufactureros. Entonces se llega a:

$$V = \beta_0 + (\beta_1 - \beta_0) = \beta_1$$

Los votos que obtiene el partido X en un estado totalmente manufacturero provienen sólo de los que laboran en esa rama. En el caso de un estado con las características exactamente opuestas, es decir, que nadie trabaja en la manufactura, la ecuación asume la forma:

$$V = \beta_0$$

Los votos para el partido X vienen de los trabajadores no manufactureros.

Además, hay que tomar en cuenta que si bien la ordenada al origen ( $\beta_0$ ) representa la propensión que tienen los trabajadores que laboran en las ramas no manufactureras a votar por el partido X, la pendiente, representa, en realidad, la diferencia entre la propensión de los trabajadores manufactureros y de los restantes a votar por él.

Una vez que se especifica correctamente el modelo y se interpretan adecuadamente los parámetros, habría que proceder a hacer el ajuste. Sin embargo, sabemos que la proporción de votos en favor de cualquier partido político no puede ser negativa, esto quiere decir que  $V_i$  debe cumplir siempre con la restricción de ser positiva o igual a cero para cualquier estado. El ajuste de una recta puede violar esta restricción, ya sea por el tamaño de los errores estándares asociados a los parámetros o porque el rango de variabilidad de la variable explicativa es estrecho. Con estas condiciones se puede llegar a obtener una ecuación estimada como la siguiente:

$$V_i = -0.1 + 0.7M_i$$

Este resultado rompe la restricción  $V > 0$  en todos los estados en que  $M < 0.1/0.7$ . En palabras, en los estados en que el porcentaje de empleados en la

<sup>2</sup> La exposición de este ejemplo se basa en el libro de Raymon Boudon (1968) citado en la bibliografía.

manufactura es inferior al 14% el modelo pronosticará que el porcentaje de votos es negativo. Este resultado absurdo se puede evitar aplicando una transformación logística a la variable proporción de votos en favor del partido X:

$$\frac{\text{Ln}V_i}{1-\text{Ln}V_i} = \gamma_0 + \gamma_1 I_i$$

El término de la izquierda, conocido con el nombre de logit o logito, representa una función que fluctúa entre menos y más infinito para todos los valores de  $V_i$  en el intervalo cerrado entre 0 y 1. Esta es la garantía que la proporción de votos por el partido X, calculada con esta ecuación, nunca arrojará un valor aberrante. Si el logaritmo Ln tiene base e, entonces la proporción de votos se calcula con la ecuación:

$$V_i = \frac{e^{\gamma_0 + \gamma_1 I_i}}{1 + e^{\gamma_0 + \gamma_1 I_i}}$$

### 3.2.- DETERMINANTES DE LA POBREZA EN MÉXICO. AJUSTE DE UNA REGRESIÓN LOGÍSTICA.

Al aplicar el método integrado para medir de la pobreza (Boltvinik *é.*, 1992) a los datos de hogares de la Encuesta Nacional de Ingresos y Gastos de los Hogares de 1992 (INEGI, 1992) se clasifica cada hogar de la muestra como pobre o no pobre. Interesa identificar los factores condicionantes de la pobreza. Dicho con mayor precisión, se trata de determinar los factores que tienen incidencia sobre la probabilidad de que un hogar sea pobre.

Notese que este planteamiento es coherente con el tipo de explicación propio de las ciencias sociales. Explicar un evento específico consiste en establecer las condiciones bajo las cuales siempre o usualmente acontece, esto es, usar enunciado generales a partir de los cuales se pueden inferir, con grados de certeza razonable, otros enunciados referidos a propiedades de eventos específicos. Para entender este tipo de eventos planteamos enunciados probabilísticos, que normalmente se traducen en que aseveraciones probabilísticas de que una unidad de-

terminada, bajo ciertas condiciones, se comporte de una manera determinada (Przeworski A., y Teune H., 1970:19).

Ahora bien, a continuación incluimos una serie de consideraciones respecto a las condiciones que estarían afectando o condicionando la probabilidad de que un hogar sea pobre.

Tanto la teoría del capital humano, el papel que adscriben los Estados a la educación, como el sentido común popular, sostienen que la instrucción es uno de los factores fundamentales que afecta la probabilidad de ser o no pobre. Más aún se considera que la educación formal no sólo es un instrumento privilegiado de movilidad social sino que, además, es uno de los elementos básicos para estimular la creatividad y la capacidad de trabajo autónomo que requieren las nuevas tecnologías. Estos argumentos permiten identificar a la educación como un factor que tendría incidencia sobre la probabilidad de ser pobre: a mayor educación menor probabilidad de ser pobre (World Bank, 1991: 52 a 69; Schultz T., 1993).

Ahora bien, el argumento maltusiano en que descansa la política demográfica de nuestros países (sin que México escape a ello) sostiene que la "familia pequeña vive mejor" o "vayámonos haciéndonos menos para vivir mejor" este *eslogan* se construye sobre la idea de que la probabilidad de ser pobre aumenta en la medida que mayor es el tamaño de la familia. Investigaciones de campo sostienen lo contrario, especialmente en épocas de crisis (González de la Rocha, 1988; Selby H., 1988; Tuirán R., 1993; Oliveira O., 1988; De Barbieri, 1989; Cortés F. y R. M. Rubalcava, 1991). Estos estudios han mostrado que ante los efectos de la política de estabilización y ajuste y de cambio estructural, los hogares pobres más numerosos han tenido la posibilidad de combatir los efectos perniciosos sobre las condiciones de vida de sus miembros aumentando la utilización de su fuerza de trabajo secundaria. Si se observa con detenimiento el segundo argumento debe notarse que descansa no en el tamaño del hogar sino en el número de perceptores, lo que lo hace mucho más atractivo para considerarse en un estudio de pobreza. En efecto, desde el punto de vista estrictamente lógico el tamaño del hogar no tendría por qué estar relacionado con la pobreza, ya que más bien depende de la cantidad de recursos con que se cuenta con relación al volumen de necesidades que hay que satisfacer, éstas sólo dependen en parte del

número de miembros del hogar. El tamaño presenta la peculiaridad de ser insensible al ciclo vital, en efecto, lo mismo daría que un hogar estuviese formado por dos adultos y tres hijos menores a los cinco años, que si contara con cinco adultos, todos trabajando. Indudablemente la probabilidad de ser pobre es mayor en el primer caso que en el segundo. En realidad, se debería esperar que la probabilidad de ser pobre estuviese más relacionada con el balance entre necesidades y recursos que con el tamaño del hogar (Chayanov A., 1974: 47 a 95; Cuéllar O., 1990: 19 a 53).

El ingreso es un factor que tiene una relación clara e indiscutible con la pobreza. En la medida que mayor es el ingreso con que cuenta el hogar menor será la probabilidad de que sea pobre.

También se argumenta que la pobreza tiende a ser mayor en las zonas rurales que en las urbanas, aunque en aquéllas es menos visible por su dispersión territorial. Sin embargo, si consideramos las fuertes corrientes migratorias rurales urbanas y su asentamiento en los márgenes de las ciudades, a lo largo de varias décadas, podría pensarse que el campo traslada a las urbes de América Latina los pobres que surgen en su seno como consecuencia de la producción y reproducción social. Si los que migran fuesen los pobres del campo y se moviesen hacia las zonas urbanas a una tasa superior a su reproducción, entonces se observaría más pobres en éstas que en aquél. Sin embargo, esta idea sería contradictoria con el hecho ampliamente documentado que los migrantes rurales no son los habitantes más pobres de las zonas rurales. Esto llevaría a concluir que la pobreza estaría más extendida en el campo que en las ciudades (González Soledad, 1994: 183 a 186; Boltvinik J., 1995: 79 a 91). Sea cual sea la explicación correcta cualquier estudio que pretenda identificar los factores de que depende la probabilidad de ser pobre no puede dejar fuera el contexto en que se asienta el hogar o la familia.

La literatura sobre género argumenta que a igualdad de trabajo y de preparación para desempeñarlo, la mujer recibe retribuciones menores. Aún más, las diversas corrientes del feminismo coinciden en señalar que son proporcionalmente más las mujeres que viven en condiciones de pobreza en comparación con los hombres (Salles V. y R. Tuirán, 1995: 68 a 71; Salles V., 1994: 62 a 69; Acosta F., 1994: 104 a 115). Tal vez la discriminación de que son objeto las mujeres en los diversos planos de la

vida social se sintetizan en sus condiciones de vida, de modo que la proporción de mujeres pobres es sustancialmente mayor que la de hombres pobres. Sin embargo, hay una corriente del feminismo que sostienen que el ingreso administrado por las mujeres se emplea con mayor eficiencia en el bienestar del hogar que cuando lo es por los hombres (Blumberg Rae, 1991: 91 a 107). De esto se colige, de que dos hogares iguales en todos los aspectos, excepto en que uno es encabezado por mujer y el otro por hombre, se podría esperar que el primero tuviese una probabilidad de no ser pobre mayor que la del segundo. En otros términos, debido a que la mujer tiende a asignar la totalidad o casi la totalidad de los recursos al hogar, a diferencia de los hombres que siempre distraen parte de ellos para otros propósitos, la pobreza tendería a extenderse más entre los hogares encabezados por hombres que entre los encabezados por mujeres.

Hay argumentos tanto a favor como en contra de que debería haber más mujeres pobres que hombres pobres. No hay mucha claridad sobre los alcances de esta discusión en la medida que es muy difícil desgajar a los hombres y las mujeres de sus hogares. En los hogares con jefe hombre hay tanto hombres como mujeres, así como también hay diversidad por sexo en los que tienen jefe mujer. Las discusiones en torno a este punto nos obligan a considerar al género como una de las determinantes de la pobreza de los hogares.

Por último, la ocupación es otro de los factores que está asociado a la pobreza. En efecto, no sólo posibilita tener mayor ingreso sino que también es una puerta importante para tener acceso a los sistemas de seguridad social. Los trabajadores de las empresas formales están cubiertos contra accidentes, tienen seguros de vejez, cobertura médica para ellos y su familia, atención hospitalaria, etc., es decir, una serie de beneficios atados al trabajo que hace más difícil que caigan en la pobreza. Algo similar sucede con los trabajadores al servicio del Estado. En definitiva se esperaría que hubiese una relación más o menos clara entre la ocupación y la probabilidad de ser pobre (Jusidman C. y M. Eternod, 1995: 105 a 110 y 151 a 159).

Estas son las razones que llevan a plantear el siguiente modelo de regresión:

$$\ln P/(1-P) = \beta_0 + \beta_1 E + \beta_2 TD + \beta_3 JM + \beta_4 C + \beta_5 SJ + \beta_6 SP + \epsilon, X$$



$\ln \text{Prob}[P/(1-P)]$  es el logaritmo natural de la razón de momios, es decir, el cociente entre la probabilidad de que el hogar sea pobre (P) y de que no lo sea (1-P).

EJ, simboliza la educación del jefe y está medida en años de instrucción formal.

TD, representa la tasa de dependencia, es decir, el tamaño del hogar dividido entre el número de perceptores.

IM, es el ingreso monetario medio por perceptor de los hogares, expresado en cientos de pesos de agosto de 1989.

C, es una variable dicotómica que representa el contexto y toma el valor 0 si el hogar está localizado en una zona urbana y 1 si lo está en el área rural.

Sé, es el sexo del jefe y asume el valor 1 si es hombre y 2 si es mujer.

SP, indica la composición por sexo de los perceptores. Toma el valor 0 si todos los perceptores son hombres o bien algunos son varones y otras mujeres. El valor 1 se asignó a los hogares que tienen perceptores exclusivamente femeninos, sin importar su número.

La sumatoria da cuenta de la posición en la ocupación. Esta es la única variable nominal no dicotómica. La manera de incorporarla al análisis es a través de variables ficticias (regresores) dicotómicas cuyo número es igual al total de categorías de la variable menos la unidad.

Recuérdese que el interés del estudio consiste en estimar la probabilidad de que el hogar sea pobre dado el conjunto de variables explicativas. Sin embargo, el modelo que se ha presentado no tiene explícitamente como variable explicada la probabilidad de ser pobre, sino una función de ella. Manipulaciones algebraicas simples permiten escribir esa misma igualdad de la manera siguiente:

$$\ln P = \frac{e^{\beta_0 + \beta_1 EJ + \beta_2 TD + \beta_3 IM + \beta_4 C + \beta_5 SJ + \beta_6 SP + \sum \beta_{7,i} X_i}}{1 + e^{\beta_0 + \beta_1 EJ + \beta_2 TD + \beta_3 IM + \beta_4 C + \beta_5 SJ + \beta_6 SP + \sum \beta_{7,i} X_i}}$$

O bien como:

$$\ln \frac{P}{1-P} = \beta_0 + \beta_1 EJ + \beta_2 TD + \beta_3 IM + \beta_4 C + \beta_5 SJ + \beta_6 SP + \sum \beta_{7,i} X_i$$

Con la ecuación de regresión lineal en el logaritmo de la razón de momios se estiman los valores de los parámetros  $\beta$ . Para calcular la probabilidad asociada a un hogar particular basta con sustituir en esta ecuación los valores de los estimadores y reemplazar el conjunto de variables explicativas que lo individualiza. Realizando las operaciones aritméticas involucradas en el cálculo de P se obtiene, por fin, el resultado.

Una vez ajustado el modelo y realizado el análisis se llegó a la conclusión de que en general, los factores explicativos de la probabilidad de que un hogar sea pobre se comportan de acuerdo con las hipótesis planteadas. La excepción son las variables asociadas al género. Para evaluar la tesis del "purse power" se construyó la variable dicotómica composición por sexo de los perceptores. La idea es que en los grupos domésticos en que sólo hay perceptoras el dinero se asignaría preferentemente a satisfacer las necesidades del hogar y de sus miembros de modo que, en igualdad de condiciones, tenderían a presentar probabilidades menores de ser pobres que los hogares en que sólo hay perceptores hombres o bien hombres y mujeres. Sin embargo, esta variable no tuvo ningún peso significativo en conformar la probabilidad; su valor no se altera sea cual sea la composición por sexo de los que generan el ingreso del hogar<sup>3</sup>. Este resultado podría llevar a concluir que la tesis del "poder de la cartera" es falsa, sin embargo, esta afirmación podría ser apresurada, tal vez habría que ser más cauto e investigar el punto en mayor profundidad. Existe la posibilidad de que el indicador utilizado no sea válido, es decir, que no represente el concepto a cabalidad.

Por otra parte, la jefatura del hogar sí juega un papel en la determinación de la probabilidad de estar o caer en la condición de pobreza. El único detalle es que el efecto fue inverso al esperado. Son los hogares de jefe mujer los que tienen una probabilidad más baja de ser pobres. Este resultado podría interpretarse en favor de la tesis de "purse power" y en contra del argumento de la explotación de género. Sin embargo, se debe evitar sacar conclusiones apresuradas ya que en este tema no hay un

<sup>3</sup> A continuación se revisarán brevemente las hipótesis básicas, para evitar repeticiones innecesarias toda vez que se afirme que una variable específica está relacionada con la probabilidad de ser pobre debe entenderse que las restantes variables se consideran fijas.

conjunto inequívoco de indicadores válidos que permitan extraer conclusiones de ese calibre con probabilidades bajas de equivocarse.

En todo caso el impacto del sexo del jefe sobre la probabilidad de que el hogar sea pobre fue pequeño (y bastante más débil que el que tuvo el contexto). Si el jefe es mujer y el hogar "medio" la probabilidad de ser pobre cae en aproximadamente 7% en tanto que el pasaje de zona urbana a rural aumenta la probabilidad en 17%.

La mayor probabilidad de ser pobre la tienen los peones y jornaleros (casi tres cuartos) en hogares medios. Cualquier otra posición en la ocupación la reduce sustancialmente, con efecto bastante mayor que el que presentó el contexto o la jefatura.

A mayor número de años de instrucción formal del jefe menor probabilidad de que el hogar sea pobre. Sin embargo, el efecto es despreciable en los casos extremos de los hogares "ricos" y de los "pobres". Si el jefe es un varón que se desempeña en calidad de peón o jornalero, su hogar está situado en una zona rural, presenta una alta tasa de dependencia y el ingreso por perceptor es pequeño; no importa cuántos años de instrucción tenga el jefe, el hogar seguirá sumido en la pobreza, su probabilidad de ser pobre se mantiene en la unidad. El efecto negativo de la instrucción no puede remontar el peso de las otras condicionantes de la probabilidad. Básicamente pasa lo mismo con los hogares "ricos". Su condición no se modifica con los años de instrucción del jefe. Sin embargo, esta variable sí reduce la probabilidad de pobreza en los "hogares" medios. Si las otras variables se acercan al promedio o a la moda, según sea el caso, y el jefe tiene cuatro o más años de instrucción, el hogar tiene una probabilidad menor a 0.5 de ser pobre.

Tal como se esperaba a mayor tasa de dependencia mayor probabilidad de que el hogar sea pobre. Aunque esta relación es directa la sensibilidad de los hogares "ricos" y "pobres" al cambio en la tasa de dependencia es similar al de la instrucción del jefe, por lo que no se abundará sobre este punto. El resultado del ajuste permite afirmar que un hogar "promedio" que tenga seis o más personas por perceptor tiene una probabilidad mayor a 0.5 de ser pobre. Nótese que el tamaño del hogar está considerado sólo indirectamente en esta variable.

La tasa de dependencia tiene mayor efecto sobre la probabilidad de estar en la condición de pobreza que la instrucción formal del jefe. Los co-

eficientes exponenciales muestran que el impacto de un año más de educación baja el momio en 13% mientras que el crecimiento de una unidad en la tasa de dependencia aumenta la probabilidad en casi un 40%. Este mismo hecho se refleja en el rango de variación de las curvas de probabilidad: la instrucción del jefe la hace disminuir de 0.6 a 0.1, mientras que la tasa de dependencia parte en 0.15 y llega hasta 1.

A mayor ingreso por perceptor menor probabilidad de que el hogar sea pobre. Esta es la única variable métrica que afecta la probabilidad de los hogares "pobres", aun cuando éstos deberían percibir la friolera de 13 salarios mínimos por mes para que su destino ineluctable no sea la pobreza. Dicho de otra manera, dado que estos hogares se caracterizan por poseer una serie de atributos que los sumen en la pobreza (tienen una probabilidad igual a uno), si se deseara disminuir la probabilidad, no se sacaría absolutamente nada si se les entregara un subsidio menor a 13 salarios mínimos. Se necesitarían por lo menos 20 salarios mínimos para que la probabilidad caiga a 0.5 o menos. Obviamente el impacto es mucho más fuerte en los hogares "medios" en que son necesarios un poco más de 5.5 salarios mínimos por perceptor para sacar a estos hogares de las garras de la pobreza. En los "ricos" el ingreso tiene un impacto sobre la probabilidad pero no la altera sustancialmente.

La conclusión que se obtiene al analizar el efecto de las variables métricas sobre la probabilidad de ser pobre es obvia: el ingreso por perceptor es el que mayor incidencia tiene. Su impacto es mucho más marcado que el de la educación formal del jefe o la tasa de dependencia. Esto quiere decir, que si el objetivo de política es atacar frontalmente el flagelo de la pobreza, el golpe más fuerte lo recibiría de las retribuciones que reciben los factores de la producción, es decir, sueldos, salarios, jornales, ganancias de los trabajadores por cuenta propia y de los pequeños empresarios, que se traducen en mayor cantidad de recursos económicos a disposición de los hogares.

La probabilidad de que un hogar sea pobre depende de la instrucción del jefe, de su posición en la ocupación y de su sexo. También está afectada por la tasa de dependencia del hogar, del ingreso medio de los perceptores y de su ubicación geográfica ya sea que esté sito en zonas rurales o urbanas. En general, este resultado coincide con lo que se

sabe sobre el fenómeno y con las hipótesis en boga (tal vez con la excepción de las relativas al género). El estudio realizado permitió considerar y dar cuenta empírica de la naturaleza multidimensional de la pobreza además de jerarquizar los factores que tienen incidencia sobre la probabilidad de que un hogar caiga en esa condición.

El modelo desarrollado permite ir un poco más allá, en la medida en que permite jerarquizar el efecto de los diferentes factores condicionantes de la pobreza de los hogares. Resulta claro que son las variables económicas las que tienen mayor incidencia sobre la probabilidad, tanto el ingreso por perceptor, como la posición en la ocupación. La otra condicionante que juega un papel importante es la tasa de dependencia que resulta ser una combinación económico-demográfica. Los restantes factores que tienen incidencia sobre la probabilidad de ser pobre tienen efectos significativos, con impactos bastante menores que los señalados en primer término.

Los datos, los análisis estadísticos y los argumentos presentados permiten afirmar que en el caso teórico extremo (hogares "pobres") las medidas de política que sólo atacan uno de los factores no tienen efecto significativo. No sería esa la manera de luchar contra un fenómeno multifacético. Esto se pudo observar aun en el caso de que se recurriera al expediente de transferir masivamente recursos a los perceptores de los hogares "pobres". El modelo enseña que la erradicación de la pobreza requiere de esfuerzos combinados que la ataquen tanto desde el ángulo económico, como social y cultural. Esfuerzos aislados estarían destinados al fracaso.

#### 4.- CONCLUSIÓN

En el análisis de problemas sociales la regresión logística es una técnica estadística que presenta varias virtudes:

(i) Puede conceptuarse como una extensión del análisis de asociación en dos sentidos. Por una parte, permite trabajar con un gran número de variables. Por otra parte, es fácil considerar, simultáneamente, cualquier número de categorías por variable. La única limitante es el número de casos.

(II) Supera al análisis de asociación, no sólo en su ductilidad para tratar simultáneamente un número grande de variables no métricas, con varias categorías, sino que, además, proporciona medicio-

nes del impacto que tiene cada variable explicativa, o sus categorías sobre la variable dependiente. A partir de los coeficientes de regresión se tienen estimaciones de los efectos, tanto en sentido como en magnitud, de las variables explicativas sobre la explicada. A partir de ellos se pueden establecer estrategias para tomar medidas de política, como las sugeridas en el análisis de las determinantes de la pobreza.

(III) La regresión logística permite expresar formal o matemática teorías o conjunto de hipótesis relativamente complejas. Sin embargo, su aplicación se limita a los casos en que el fenómeno que se quiere explicar corresponda a una estructura conceptual caracterizada por la inexistencia de relaciones entre los factores explicativos y que las variables independientes no presenten correlaciones estrechas (Cortés F. y Rubalcava. R.M., 1993).

(IV) Finalmente si entendemos que explicar, en ciencias sociales, consiste en asignar probabilidades a que una unidad se comporte de una manera determinada (bajo ciertas condiciones), entonces la regresión logística es un instrumento estadístico diseñado para explicar: responde a la pregunta ¿de qué depende la probabilidad asociada a un determinado evento? —por ejemplo de ser pobre—.

#### REFERENCIAS BIBLIOGRÁFICAS

- ACOSTA Félix, "Los estudios sobre jefatura de hogar femenina y pobreza en México y América Latina", en Grupo Interdisciplinario sobre Mujer, Trabajo y Pobreza, *Cuánto cuesta la pobreza de las mujeres: una perspectiva de América Latina y el Caribe*, UNIFEM, México, 1995.
- AGRESTI Alan, *Categorical Data Analysis*, John Wiley, New York, 1990.
- ALDRICH John, y FORREST Nelson, *Linear Probability, Logit, and Probit Models*, Sage University Paper, California, 1984.
- BLUMBERG Rae Lesser, "Income Under Female Versus Male Control: Hypotheses from a Theory of Gender Stratification and Data from the Third World", en Blumberg Rae Lesser (ed) *Gender, Family and Economy: The Triple Overlap*, Sage Publications, Newbury Park, 1991.
- BOLTVINIK Julio, "El método de medición integrada de la pobreza. Una propuesta para su desarrollo", *Comercio Exterior*, Banco Nacional de Comercio Exterior, México, Vol. 42, No. 4, 1992.

- BOLTVINIK Julio, "La pobreza en México. I. Metodologías y su evolución", *Salud Pública de México*, Vol. 37, No. 4, 1995.
- BOLTVINIK Julio, "La pobreza en México. II. Magnitud", *Salud Pública de México*, Vol. 37, No. 4, 1995.
- BOLTVINIK Julio, *Pobreza y Estratificación Social en México*, Instituto Nacional de Estadística Geografía e Informática, Aguascalientes, 1995.
- BOUDON Raymon, *L'analyse mathématique des fait sociaux*, Plon, París, 1968.
- DE BARBIERI Teresita, "La Mujer", *Demos 2, carta demográfica sobre México*, México, 1989.
- CORTÉS Fernando Y Oscar CUÉLLAR, *Crisis y reproducción social: los comerciantes del sector informal*, Flacso/Miguel Angel Porrúa, México, 1990.
- CORTÉS Fernando Y Rosa María RUBALCAVA, *Autoexplotación forzada y equidad por empobrecimiento*, Jornadas 120, El Colegio de México, México, 1991.
- CORTÉS Fernando y Rosa María RUBALCAVA, "Consideraciones sobre el uso de la estadística en las Ciencias Sociales", en Ignacio Méndez y Pablo González Casanova, *Matemáticas y Ciencias Sociales*, Centro de Investigaciones Interdisciplinarias y Miguel Angel Porrúa, México, mayo de 1993.
- CHAYANOV A. V., *La organización de la unidad económica campesina*, Ediciones Nueva Visión, Buenos Aires, 1974.
- GOLDBERGER Arthur, *Econometric Theory*, John Wiley, New York, 1964.
- GONZÁLEZ Soledad, "Mujeres, trabajo y pobreza en el campo mexicano: una revisión crítica de la bibliografía reciente", en Grupo Interdisciplinario sobre Mujer, Trabajo y Pobreza, *Cuánto cuesta la pobreza de las mujeres: una perspectiva de América Latina y el Caribe*, UNIFEM, México, 1995.
- GONZÁLEZ DE LA ROCHA Mercedes, «Economic crisis, domestic reorganization and women's work in Guadalajara», UCSD La Jolla/Ciesas Occidente, 1988.
- JOHNSTON J. *Econometric Methods*, International Student Edition, Singapore, 1984.
- JUSIDMAN Clara Y Marcela ETERNOD, *La participación de la población en la actividad económica en México*, 1995. Instituto Nacional de Estadística Geografía e Informática, Aguascalientes, 1995.
- MADDALA G. S. *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, New York, 1983.
- MCKENZIE Donald, «Eugenics and the Rise of Mathematical Statistics in Britain», en John Irvine, Ian Miles and Jedd Evans (eds.) *Desmistifying Social Statistics*, Pluto Press, Londres 1979.
- OLIVEIRA Orlandina DE, «El empleo femenino en tiempos de recesión económica: tendencias recientes», ponencia presentado al coloquio sobre fuerza de trabajo femenina urbana, México, UNAM, 1988.
- PRZEWORSKI Adam y Henry TEUNE, *The Logic of Comparative Social Inquiry*, John Wiley, New York, 1970.
- SALLES Vania, Pobreza, pobreza y más pobreza, en *Las mujeres en la pobreza*, GIMTRAP/El Colegio de México, México, 1994.
- SALLES Vania y Rodolfo TUIRÁN, "Mujeres y Hogares en México: Cada vez más pobres?", en Grupo Interdisciplinario sobre Mujer, Trabajo y Pobreza, *Cuánto cuesta la pobreza de las mujeres: una perspectiva de América Latina y el Caribe*, UNIFEM, México, 1995.
- SCHULTZ Theodore W., *The Economics of Being Poor*, Blackwell, Oxford, 1993.
- Selby Henry, Arthur D. Murphy, Stephen A. Lorenzen, *The Mexican Urban Household: Organizing for self defense*, University of Texas Press, Texas, 1990.
- THEIL Henry, *Principles of Econometrics*, John Wiley, New York, 1971.
- TUIRÁN Rodolfo, "Las respuestas de los hogares de sectores populares urbanos frente a la crisis: el caso de la Ciudad de México", en Béjar Navarro Raúl y Héctor Hernández Bringas, coordinadores, de *Población y Desigualdad Social en México*, CRIM-UNAM, México, 1993.
- STIGLER Stephen M., *The History of Statistics: The measurement of Uncertainty before 1990*, Belknap, Cambridge, Massashussets, 1986.
- WORLD BANK, *World Development Report 1991: The Challenge of Development*, Oxford University Press, Oxford, 1991.