

Perfil de pobreza de los habitantes bogotanos a través de métodos multivariados y modelos logísticos

Poverty Profile of Bogotá's Habitants through of Multivariate and Logistic Methods

JAIME FRANCISCO PANTOJA BENAVIDES

Licenciado en Matemáticas de la Universidad Distrital Francisco José de Caldas y Especialista en Estadística de la Universidad Nacional de Colombia. Docente de vinculación ocasional adscrito a la Facultad Tecnológica de la Universidad Distrital e investigador auxiliar del grupo de investigación en Robótica Móvil Autónoma (ROMA), de la misma Universidad.

jfpantojab@udistrital.edu.co

ÓSCAR ORLANDO MELO MARTÍNEZ

Magíster en Estadística. Profesor del Departamento de Estadística, Facultad de Ciencias de la Universidad Nacional de Colombia. Integrante del grupo de investigación en Estadística Aplicada a la Investigación Experimental, Industria y Biotecnología, clasificado en categoría A por Colciencias

oomelom@unal.edu.co

CARLOS EDUARDO MELO MARTÍNEZ

Ingeniero Catastral y Geodesta de la Universidad Distrital Francisco José de Caldas, Especialista en Estadística de la Universidad Nacional de Colombia y Magíster en Economía de la Pontificia Universidad Javeriana. Profesor investigador de la Facultad de Ingeniería de la Universidad Distrital, programa de Ingeniería Catastral y Geodesia. Integrante del grupo de investigación Núcleo de Investigación de Datos Espaciales (NIDE).

cmelo@udistrital.edu.co

Clasificación del artículo: reflexión

Fecha de recepción: 22 de mayo de 2006

Fecha de aceptación: 13 de octubre de 2006

Palabras clave: indicador de pobreza, vulnerabilidad, componentes principales, correspondencias múltiples, modelo Logit.

Key words: poverty indicators, vulnerability, main components, multiple correspondences, Logit model.

RESUMEN

La formulación de modelos para determinar indicadores de pobreza es fundamental en el contexto social actual de la ciudad de Bogotá; ellos permiten definir acciones concretas sobre los factores relevantes que influyen en los resultados de las mediciones de pobreza de los hogares de la ciudad. En este trabajo se analizan variables socioeconómicas que inciden en el desarrollo de la ciudad y de sus habitantes; el estudio se basa en la información disponible a partir de la Encuesta Nacional de Hogares de 2000, y consiste en la aplicación de métodos estadísticos de análisis en componentes principales y análisis de correspondencias múltiples; el resultado final es la proposición de un modelo logístico para la estimación de la pobreza.

ABSTRACT

The estimation of models to determine poverty indicators in Bogotá is fundamental in the current social context of the city; it allows the establishment of specific actions on the most relevant factors that influence on families from Bogotá, which are affected by those indicators. Socio-economical variables that have impact on the development of the city and their habitants are analysed in this work. This paper is based on the National Survey of Homes 2000 by mean of the application of statistical methods in Analysis of main components and multiple correspondences analysis. The final result is the formulation of a *Logit* model to estimate poverty.

* * *

1. Introducción

Si bien la pobreza no es un problema reciente, ella no puede ser admitida como una condición negativa con la que se deba convivir. Su permanencia y expansión concierne a todos, no sólo por razones éticas y de justicia social, sino también porque ella se constituye en una verdadera amenaza para la estabilidad social, económica y política de las comunidades.

En la mayoría de los casos la pobreza se concibe como la carencia de ciertos bienes materiales a los que se asignan previamente unos atributos; luego se procede a cuantificar tales carencias en forma directa (método de identificación de necesidades básicas insatisfechas [NBI]), o detectando su presencia a través de la insuficiencia de ingreso para su satisfacción (método de la línea de pobreza [LP]). En el primer caso el énfasis se realiza en el consumo y en el segundo en el ingreso [1].

En la actualidad, existen diferentes métodos para medir la pobreza; aunque pueda afirmarse que unos son mejores, en últimas el problema no es encontrar el método más adecuado para tal fin, sino también plantear estrategias efectivas para la solución del problema, y encontrar la forma para que los pocos

recursos destinados para el control de este fenómeno lleguen a quienes los necesitan y se distribuyan de la mejor forma.

El objetivo principal de este artículo es presentar un perfil de pobreza para los ciudadanos bogotanos, a través del análisis de componentes principales (ACP), correspondencias múltiples (ACM) y modelos logísticos, todo con base en la Encuesta Nacional de Hogares 2000¹. Entre los propósitos del modelamiento estadístico se encuentran la descripción de información, la predicción y el análisis de un problema; por otra parte, la finalidad analítica de la regresión *Logit* en este caso es la explicación de las causas de la pobreza. El estudio se centra en analizar las condiciones de mayor incidencia en los niveles de pobreza y determinar en forma predictiva el cambio de la variable dependiente, cuando se

¹ La Encuesta Nacional de Hogares es un sistema de muestras de propósitos múltiples, a través del que se observarían los hogares para obtener datos y estimaciones intercensales, capaces de producir estadísticas básicas relacionadas con la situación demográfica, social y económica de la población colombiana. Es realizada por el Departamento Administrativo Nacional de Estadística DANE. Disponible en: <http://suamox03.dane.gov.co:7778/imgcurso/mettran/documento2.html>

modifican las causas, pretendiendo mostrar la necesidad de aplicar diferentes programas orientados a la disminución de los indicadores de pobreza.

2. Antecedentes

Diferentes estudios acerca de la pobreza se han realizado en Colombia²; asimismo, se han desarrollado proyectos liderados por los últimos alcaldes de la ciudad. Forman parte de este último grupo el proyecto *Bogotá cómo vamos*³ y el Plan de Desarrollo *Bogotá sin indiferencia. Un compromiso social contra la pobreza y la exclusión*⁴, el que se plantea que en la última década se ha producido un avance notorio en diferentes características de la ciudad y en las condiciones de vida de sus habitantes. De acuerdo con el modelo de NBI y los índices de calidad de vida y de desarrollo humano, los indicadores de pobreza medidos muestran que se están resolviendo los problemas de acceso a servicios básicos, en gran medida como resultado de las acciones de las administraciones distritales [1].

De acuerdo con el modelo de NBI, aunque se hable acerca de un mejoramiento de la ciudad y de una reducción de los indicadores de pobreza, el desempleo y la pobreza medidos según ingresos han aumentado drásticamente desde fines de la década del noventa. Entre 1997 y 2002 el porcentaje de la población por debajo de la línea de pobreza pasó de 35,1% a 50%, y el porcentaje por debajo de la línea de indigencia de 6,8% a 17%. En la actualidad, la mitad de los habitantes bogotanos (6,8 millones, aproximadamente) no tiene ingresos suficientes

para adquirir su canasta básica de bienes y servicios, y cerca de 1,156 millones no tienen los ingresos requeridos para satisfacer sus necesidades básicas de alimentación. Este es el gran drama que enfrenta la ciudad: Bogotá, se ha modernizado pero no se ha hecho más humana.

Para aproximarse a la población con carencia de bienes y servicios básicos, vivienda digna, canasta básica de alimentos, ingreso mínimo para adquirirla y otras características descriptivas del capital social se ha hecho uso de diversos indicadores fundamentados en distintos criterios. A la ciudad de Bogotá se han sido aplicados el Índice de Calidad de Vida (ICV), el Índice de Necesidades Básicas Insatisfechas (INBI), las líneas de pobreza (LP) y de miseria (LM), el Índice de Desarrollo Humano (IDH) y el Índice de Concentración del Ingreso (ICI) o Coeficiente de *Gini*. Todos ellos permiten evaluar la situación de la ciudad frente al resto del país [3, 8].

2.1. Índice de Necesidades Básicas Insatisfechas (INBI)

Mide la población pobre sobre la base de un conjunto de criterios; tienen un peso fundamental las características de las viviendas como materiales utilizados en su construcción, disponibilidad de servicios básicos y nivel de hacinamiento; los demás se relacionan con el grado de dependencia económica y el ausentismo escolar. Entre 1993 y 2001 el porcentaje de población con necesidades básicas insatisfechas disminuyó en 3,8 puntos porcentuales, al pasar de 17,2% a 13,4% (Tabla 1).

% Población bogotana con necesidades básicas insatisfechas (NBI)			
Localidad	1993	1999	2001
Ciudad Bolívar	35,3	29,2	26,2
Usme	34,6	26,9	23,8
San Cristóbal	25,9	20,0	17,8
Bosa	24,2	19,3	17,1
Santa Fé	24,7	18,4	16,3
Rafael Uribe Uribe	31,0	16,0	14,3
Tunjuelito	18,6	14,2	12,8

(continúa)

²Es el caso del Proyecto de indicadores de seguimiento de la pobreza en Colombia (1985), auspiciado por la Presidencia de la República, el Programa de las Naciones Unidas para el Desarrollo (PNUD), el Fondo de las Naciones para la Infancia (UNICEF), el Departamento Nacional de Planeación (DNP) y el DANE.

³El proyecto perteneció al Plan de Desarrollo Bogotá para vivir todos del mismo lado, liderado por A. Mockus durante su periodo de alcaldía (2001-2004).

⁴Liderado por L. E. Garzón, alcalde actual de la ciudad de Bogotá (2004-2008).

% Población bogotana con necesidades básicas insatisfechas (NBI)			
Localidad	1993	1999	2001
Kennedy	14,3	13,3	12,6
La Candelaria	17,4	13,3	12,0
Fontibón	13,9	12,5	11,7
Usaquén	11,9	11,8	11,4
Suba	13,8	11,1	10,1
Engativá	11,0	8,8	8,1
Los Mártires	10,6	8,8	8,1
Antonio Nariño	10,0	8,2	7,6
Puente Aranda	9,1	7,7	7,2
Barrios Unidos	8,8	7,4	6,9
Chapinero	7,2	6,3	5,9
Teusaquillo	4,2	4,2	4,2
Total Cabecera	17,2	14,5	13,4

Tabla 1. Población bogotana con necesidades básicas insatisfechas

Fuente: DABS (1997). Proyecciones de necesidades básicas insatisfechas con base en DANE

Línea	1991	1993	1995	1997	1998	1999	2000	2003
Línea de pobreza	37,4	42,4	40,0	35,08	36,57	43,5	49,6	40,8
Línea de indigencia	n.d.	7,7	7,6	6,83	7,75	11,3	14,9	11,2

Tabla 2. Línea de pobreza e indigencia en Bogotá

Fuente: DNP-UDS-DIOGS y DANE (2000). Encuesta Nacional de Hogares

De acuerdo con el nivel de ingresos de las familias de la ciudad, la situación de pobreza aumentó en 2003; el 40,8% de la población recibía el ingreso suficiente para adquirir la canasta normativa, en comparación con el mayor porcentaje en el año 2000 (49,6%); asimismo, en 2003 el 11,2% podía adquirir sus alimentos y bienes esenciales, mientras en 2000 este porcentaje era del 14,9%. Así, a diferencia de otros indicadores que denotan una situación más alentadora, estos muestran el deterioro de la economía de los hogares bogotanos, en particular por el desempleo y la crisis económica de la década del noventa. En términos absolutos, las personas pobres de Bogotá en 2000 se estimaron en 3,19 millones y las personas en condiciones de indigencia en 0,96 millones [1].

Aunque se refleja una leve mejoría en la satisfacción de necesidades para el promedio de la ciudad, aún hay localidades que distan mucho de la media aritmética. Las localidades que se encuentran por encima del promedio de la ciudad son Ciudad Bolívar (que casi duplica el promedio de Bogotá) y Usme. Las localidades con mayor NBI tienen también las mayores tasas de crecimiento [2].

2.2. Líneas de pobreza e indigencia

Estos dos indicadores se fundamentan en criterios relativos a la suficiencia del ingreso del hogar, frente a una canasta normativa de bienes y alimentos (línea de pobreza) o frente a los requisitos nutricionales mínimos o gastos esenciales de supervivencia (línea de indigencia o nivel de miseria). Las últimas estadísticas bogotanas se muestran en la tabla 2.

3. Análisis de la situación de pobreza de los ciudadanos bogotanos

Los programas computacionales *Spad 4.5*® y *SAS*® serán empleados para analizar los resultados de la Encuesta Nacional de Hogares 2000, con respecto a las condiciones sociales de los niveles de pobreza en la ciudad. Primero, se realiza un análisis descriptivo de la información, luego su análisis por componentes principales y de correspondencias múltiples y, finalmente, se presenta el modelo de estimación de pobreza para los ciudadanos bogotanos.

3.1 Análisis por métodos multivariados

Desde el año 2000 el DANE realiza la Encuesta Nacional de Hogares con periodicidad anual. Para

el presente estudio se selecciona una muestra de 9.125 individuos tomados de la encuesta realizada ese año. Se pretende identificar los factores fundamentales que pueden determinar los indicadores de pobreza de la ciudad mediante el estudio de la relación existente entre diferentes condiciones socioeconómicas: actividad laboral, nivel educativo, afiliación al régimen de salud, población en edad de trabajar, total personas por hogar, entre otras⁵[5].

De las diecisiete variables estudiadas, siete son discretas, dos continuas y ocho nominales; estas se etiquetan así:

- *Variables discretas*

- NHTJ: Número de horas trabajadas por el jefe del hogar.
- NOCU: Número de ocupados en el hogar.
- NDES: Número de desocupados en el hogar.
- NPET: Número de personas en edad de trabajar.
- TPHO: Total de personas en el hogar.
- TRES: Tiempo de residencia en años
- INDB: Decil de ingresos hogares Bogotá.

- *Variables continuas*

- NBIE: Necesidades básicas insatisfechas por inasistencia escolar.
- NBIC: Necesidades básicas insatisfechas alta dependencia económica.

- *Variables nominales*

- AFILI (1 = Afiliado al ISS, 2 = EPS, 3 = Régimen excepción (Fuerzas Militares, Magisterio o Ecopetrol), 4 = Régimen subsidiado y 5 = No afiliado).
- AFI (1= Afiliado, 2= No Afiliado).
- NIVELEDU (1= Ninguno, 2= Preescolar, 3= Primaria, 4= Secundaria, 5= Superior (Universitario), 6= No informa, 7= menor de 5 años).
- GENERO (1= Hombre, 2= Mujer).
- GEDAD (1= Menor de 1 año, 2= De 1 a 4 años, 3= De 5 a 14 años 4= De 15 a 44 años, 5= De 45 a 59 años, 6= Mayor de 60 años).
- PET (1= Ocupado, 2= Desocupado, 3= Inactivo, 4= Menor de 12 años).
- IPLP (1= Considerado No pobre, 2= Considerado Pobre).
- PILI (1= No considerado indigente, 2= Considerado indigente).

3.2 Caracterización de la muestra

A continuación, se presenta el análisis de la información haciendo uso de los procesos estadísticos básicos; se comienza con la descripción de las variables objeto de estudio [9].

⁵ En este trabajo no se consideran variables no cuantificadas en términos estadísticos, por ejemplo los desplazamientos del campo a las grandes ciudades debidos a factores de violencia y seguridad; estos pueden influir en la construcción de los indicadores de pobreza de una ciudad.

Modalidades Ident libelle	Avant apurement Eff. Poids		Après apurement Eff. Poids		histogramme des poids relatifs
Genero					
Homb - hombre	4333	4333.00	4333	4333.00	*****
Muje - mujer	4792	4792.00	4792	4792.00	*****
Afli	1840	1840.00	1840	1840.00	*****
Iss - afiliado iss	2759	2759.00	2759	2759.00	*****
Eps - afiliado eps					
Rege - régimen excepción	713	713.00	713	713.00	****
Sisb - regsub	407	407.00	407	407.00	***
Naf - no afiliado	3406	3406.00	3406	3406.00	*****
Afi					
Afi - afiliado	5719	5719.00	5719	5719.00	*****
Noaf - no afiliado	3406	3406.00	3406	3406.00	*****
Gedad			==== Ventilee ====		
1 - < Un año	147	147.00			
2 - Entre 1 a 4	700	700.00	725	725.00	****
3 - Entre 5 a 14	1592	1592.00	1615	1615.00	*****
4 - Entre 15 a 44	4722	4722.00	4754	4754.00	*****
5 - Entre 45 a 59	1104	1104.00	1142	1142.00	*****
6 - > De 60	860	860.00	889	889.00	*****
Niveledu			==== Ventilee ====		
Ning - ninguno	171	171.00			
Ning - ninguno	200	200.00	241	241.00	**
Esco - preescolar	2357	2357.00	2400	2400.00	*****
P0rim - primaria	3873	3873.00	3918	3918.00	*****
0Secu - secundaria	1646	1646.00	1677	1677.00	*****
Univ - universitaria	31	31.00			
Ninf - no informa	847	847.00			
Men5 - < 5 años			889	889.00	*****
Pet					
Ocup - ocupado	3730	3730.00	3730	3730.00	*****
Deso - desocupado	909	909.00	909	909.00	*****
Inac - inactivo	2490	2490.00	2490	2490.00	*****
Me12 - menor de 12 años	1996	1996.00	1996	1996.00	*****
Iplpdane					
Npbr - no pobre	5493	5493.00	5493	5493.00	*****
Pbre - pobre	632	3632.00	3632	3632	*****
Iplidane					
Nidg - no indigencia	8154	8154.00			*****
Indg - indigencia	971	971.00			*** *****

Cuadro 1. Histograma de pesos relativos de variables analizadas

De los 9.125 individuos, 4.333 son hombres (el 47,5%) y 4.792 son mujeres (el 52,5%); 3.406 (el 37,4%) no se encuentran afiliados al sistema de salud y 5.719 (el 62,7%) sí. De estos últimos, 2.759 (el 30,2%) se encuentran afiliados a una EPS privada, 1.840 (el 20,2%) al Instituto de Seguros Sociales (ISS)⁶, 713 (7,8%) a un régimen denominado de excepción⁷ y 407 (4,5%) al régimen subsidiado. El mismo grupo de afiliados se encuentra constituido por 2.665 hombres (46,6%) y 3.054 (53,4%) mujeres (cuadro 1).

Con respecto a la edad, la muestra es relativamente joven: 4.722 individuos (el 52,1%) se encuentran entre los quince y cuarenta y cuatro años de edad. En relación con el nivel educativo, 2.353 individuos (el 26,3%) han cursado educación básica primaria y 3.873 (el 42,9%) educación básica secundaria. Con respecto a la ocupación, 3.730 individuos (el 40,9%) se mantienen ocupados y 5.395 (el 59,1%) desocupados, inactivos o no están en edad de trabajar (menores de doce años). Finalmente, 3.632 (el 39,8%) y 971 individuos (el 10,6%) están ubicados por encima de la línea de pobreza (LP) e indigencia (LI), respectivamente.

2.2. Análisis de componentes principales (ACP)

El ACP recurre a dos representaciones geométricas: una para comparar a los individuos (nube de individuos) y otra para estudiar las relaciones entre las variables (nube de variables). Estas representaciones implican transformaciones de la tabla de datos. La transformación más utilizada es la de la estandarización de los datos, es decir, res-

⁶ El ISS es la entidad estatal prestadora de servicios de salud.

⁷ De acuerdo con las leyes colombianas, hasta el Acto Legislativo 01 de 2005 pertenecían al régimen de excepción los servidores públicos de Ecopetrol, miembros de las Fuerzas Militares, Policía Nacional, Ministerio de Defensa, Hospital Militar y Hospital de la Policía. Luego de esta disposición mantienen el régimen el Presidente de la República, el Régimen de la Fuerza Pública, el Régimen del Magisterio, los cuales se mantendrán transitoriamente hasta el 31 de julio de 2010.

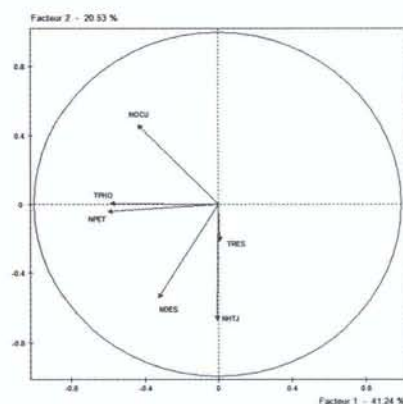
tar la media (centrado) y dividir por la desviación estándar (reducido), lo que da origen al análisis en componentes principales normado. No obstante, la comprensión del ACP es más sencilla siguiendo el proceso para el ACP no normado, que corresponde a una transformación de centrado solamente (resta de la media) [9].

Las variables continuas estudiadas se relacionan en la matriz de correlaciones (cuadro 2).

	NHTJ	NOCU	NDES	NPET	TPHO	TRES
NHTJ	1.00					
NOCU	-0.15	1.00				
NDES	0.13	-0.05	1.00			
NPET	0.05	0.59	0.43	1.00		
TPHO	0.01	0.53	0.41	0.83	1.00	
TRES	0.06	-0.01	0.00	0.09	-0.10	1.00

Cuadro 2. Matriz de correlaciones.

La matriz anterior brinda elementos de descripción de las asociaciones entre las variables activas. Las variables de mayor correlación son: a) TPHO y NPET (0,83), b) NOCU y NPET (0,59), c) TPHO y NOCU (0,53), d) NPET y NDES (0,43) y e) TPHO y NOCU (0,41), lo cual puede verificarse en la gráfica 1.



Gráfica 1. Primer plano factorial

Con respecto a las variables discretas se muestran la lista de valores propios y los porcentajes de varianza; la suma de los primeros es igual al número de variables (seis). Los dos primeros ejes representan cerca de la mitad de la inercia (61,78%), como se observa en el cuadro 3. Analizando el histograma se observa que existe una concentración de la variabilidad en un subespacio de dos dimensiones, es decir, en el primer plano factorial. Sin embargo, podría considerarse también el tercer eje factorial, con el que se acumularía una inercia total de 78,84%.

Para efectos de este estudio sólo se considera el primer plano factorial, debido a que en el tercero no se acumula gran cantidad de inercia (17,06%); además, para facilitar el análisis.

Aquí es importante resaltar que los valores propios indican las inercias o porcentajes de variabilidad en cada uno de los ejes (formando planos factoriales).

El primer plano factorial se construye con los dos primeros ejes, que son los correspondientes a los de mayor inercia (es decir, dos valores propios mayores). A partir de este plano se determinan las variables que están mejor representadas (un vector-variable cercano al borde de la circunferencia se encuentra bien representado y uno cercano al centro la variable mal representado). Por otra parte, la cercanía entre vectores indica el grado de correlación entre las variables. Finalmente, a la ubicación de los vectores en los ejes del plano factorial se pretende asociar un comportamiento; por ejemplo, si a la izquierda del primer eje se encuentran los valores altos de variables como número de personas con alto nivel de escolaridad y empleadas con altos ingresos, se podría pensar que a la izquierda están ubicadas las personas no pobres y a la derecha los pobres; en consecuencia éste sería un eje de pobreza.

HISTOGRAMME DES 6 PREMIERES VALEURS PROPRES

Numero	Valeur Propre	Pourcent.	Pourcent cumule
1	2.4746	41.24	41.24
2	1.2320	20.53	61.78
3	1.0237	17.06	78.84
4	0.8162	13.60	92.44
5	0.3125	5.21	97.65
6	0.1410	2.35	100.00

Cuadro 3. Valores propios

Las afirmaciones anteriores se pueden verificar a través de los resultados obtenidos en el cuadro 4, en el que se muestran las coordenadas de las variables sobre los cinco primeros ejes y las coordenadas de los extremos de los ejes unitarios. Puede apreciarse que NPET (-0,94), TPHO (-0,92) y NOCU (-0,69) tienen los valores más altos en las coordenadas y,

por tanto, una mejor representación en el primer eje factorial, mientras en el segundo eje están mejor representadas NDES (-0.60), NHTJ (-0,74) y NOCU (0,51). Este segundo factor se puede caracterizar como un eje de ocupación; en la parte inferior se encuentran los no ocupados y en el superior los ocupados.

Variables	Coordonnees Correlations variable-facteur					Anciens axes unitaires				
	1	2	3	4	5	1	2	3	4	5
IDEN - LIBELLE COURT										
NHTJ - NHTJEFE	-0.01	-0.74	-0.04	-0.66	-0.04	-0.01	-0.67	-0.04	-0.74	-0.08
NOCU - NOCUPADO	-0.69	0.51	-0.21	-0.30	-0.37	-0.44	0.46	-0.20	-0.33	-0.65
NDES - NDESOCU	-0.51	-0.60	0.25	0.49	-0.27	-0.32	-0.54	0.25	0.54	-0.49
NPET - NPET	-0.94	-0.05	-0.11	-0.01	0.14	-0.60	-0.04	-0.11	-0.01	0.24
TPHO - TPHOG	-0.92	0.01	0.11	-0.02	0.29	-0.59	0.01	0.11	0.03	0.51
TRES - TRESIDE	0.01	-0.23	-0.94	0.22	0.03	0.01	-0.21	-0.93	0.24	0.05

Cuadro 4. Coordenadas de las variables activas

3.3 Análisis de correspondencias múltiples (ACM)

Se utiliza en el análisis de tablas de individuos descritos por variables categóricas. La asociación

entre variables se presenta porque más o menos los mismos individuos asumen las mismas modalidades de diferentes variables [9].

	HOMB	MUJE	ISS	EPS	REGE	SISB	NAFI	AFI	NOAF	2	3	4	5	6
HOMB	4333	0												
MUJE	0	4792												
ISS	838	1002	1840	0	0	0	0							
EPS	1298	1461	2759	0	0	0								
REGE	345	368	0		713	0	0							
SISB	184	223	0	0	0	407	0							
NAFI	1668	1738	0	0	0	0	3406							
AFI	2665	3054	1840	2759	713	407	0	5719	0					
NOAF	1668	1738	0	0	0	0	3406	0	3406					
2	351	374	103	229	67	22	304	421	304	725	0	0	0	0
3	837	778	271	482	178	73	611	1004	611	0	1615	0	0	0
4	2217	2537	811	1501	345	168	1929	2825	1929	0	0	4754	0	0
5	532	610	329	337	67	74	335	807	335	0	0	0	1142	0
6	396	493	326	210	56	70	227	662	227	0	0	0	0	889

Cuadro 5. Tabla de Burt

En el cuadro 5 se observan los diferentes cruces de las modalidades de cada una de las variables continuas ilustrativas y las nominales activas, presentadas en la tabla de *Burt*; éste es un cuadro de contingencia múltiple. Se observa que del GEDAD, 1 (menor de un año); NIVELEDU, NINGUNO y NO INFORMA, no aparecen; esto sucede porque el programa SPAD 4.5 los elimina por tener una baja frecuencia [6][9].

4. Modelo Logit para la estimación de la pobreza de los ciudadanos bogotanos

Un modelo estadístico tiene como finalidad principal explicar el comportamiento (en términos de variabilidad) de las variables que hacen parte de él; está compuesto por la variable a explicar (dependiente o respuesta; para este caso “el individuo

es considerado pobre”) y las variables explicativas (independientes o regresoras) [7].

El modelo de estimación de pobreza de los bogotanos se formula con el objeto de mostrar que los indicadores de pobreza están asociados con variables socioeconómicas como número de horas trabajadas normalmente por el jefe del hogar, número de desocupados, número de personas en edad de trabajar, tipo de afiliación al régimen de salud y nivel educativo, entre otras.

Variabes económicas y demográficas como ingreso económico, cantidad de personas en el hogar, sexo (género), educación, empleo y desempleo se

constituyen en complementarias, para construir un modelo de probabilidad con la finalidad de clasificar a un individuo como pobre o no pobre, con base en la información reportada por la población a través de la encuesta ya citada.

A continuación, se presentan los resultados del modelo de pobreza implementado en el programa SAS®⁸, obtenidos mediante el procedimiento de regresión paso a paso. La variable dependiente en el presente estudio es $Y=1$, si “el individuo es considerado pobre”, o $Y=0$ si “el individuo es considerado no pobre”. Del total de la muestra, 3.632 individuos (el 39,8%) se clasifican como pobres y 5.493 (60,2%) como no pobres.

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3536.6927	17	<.0001
Score	2922.6313	17	<.0001
Wald	2037.0444	17	<.0001

Cuadro 8. Ajuste del modelo

La hipótesis nula en el modelo plantea que, $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ VRS. H_a : Al menos un $\beta_i \neq 0, i = 1, 2, K, k$. Al observar los valores p de los test *Likelihood Ra-*

tio, Score y Wald se concluye que al menos una de las variables explicativas a un nivel de significancia 5% se relaciona con la variable explicada.

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Standard Wald

Parameter	DF	Estimate	Error	Chi-Square	Pr > ChiSq
SEXO	1	-0.1346	0.0523	6.6213	0.0101
OCUPADO	1	0.9075	0.1318	47.4424	<.0001
DESOCUPADO	1	0.6973	0.1484	22.0938	<.0001
INACTIVO	1	0.3258	0.1184	7.5654	0.0059
TAMANO	1	0.4576	0.0173	698.9198	<.0001
NE_PREESCOLAR	1	-0.5993	0.2037	8.6544	0.0033
NE_PRIMARIA	1	-0.7811	0.1266	38.0542	<.0001

Continúa

⁸ SAS, Mediante el procedimiento LOGISTIC. 0.

The LOGISTIC Procedure
Analysis of Maximum Likelihood Estimates

Parameter	DF	Standard Wald		Chi-Square	Pr > ChiSq
		Estimate	Error		
NE_SECUNDARIA	1	-1.3764	0.1444	90.7895	<.0001
NE_UNIVERSITARIO	1	-2.7452	0.1613	289.7264	<.0001
REGCONT	1	-0.9445	0.0547	297.6277	<.0001
REGSUB	1	0.4795	0.1007	22.6705	<.0001
REGEXC	1	-1.7594	0.1475	142.3038	<.0001
EDAD2	1	0.5916	0.1105	28.6681	<.0001
EDAD3	1	1.0917	0.1213	80.9424	<.0001
EDAD4	1	0.6391	0.0879	52.8872	<.0001
EDAD6	1	-0.5195	0.1208	18.4820	<.0001
NRO_OCUPADOS	1	-1.0567	0.0349	917.0889	<.0001

Cuadro 7. Estimación de parámetros

En el cuadro 7 se presentan las estimaciones de los parámetros asociados con las variables explicativas y sus respectivos test de significancia; se observa que las diecisiete variables explicativas están asociadas con la variable explicada. Las variables con más alta asociación con la variable dependiente son NRO_OCUPADOS, TAMANO, REGCONT, NE_UNIVERSITARIO y REGEXC (Wald Chi-Square = 917,0889; 698,9198; 297,6277; 289,7264; 142,3038, respectivamente).

Del cuadro 8 es de resaltar el alto porcentaje de concordancia del modelo (82,7%), lo que significa que el valor obtenido por él es el mismo observado en este porcentaje de los casos. Asimismo, se aprecia la ventaja comparativa de quienes pertenecen al grupo EDAD3 (2.979 individuos); es decir, una persona entre los cinco y catorce años tiene aproximadamente tres veces más riesgo de ser pobre que una persona en EDAD1 (menores de un año de edad).

Effect	Odds Ratio Estimates		
	Estimate	Confidence	Limits
SEXO	0.874	0.789	0.968
OCUPADO	2.478	1.914	3.208
DESOCUPADO	2.008	1.502	2.686
INACTIVO	1.385	1.098	1.747
TAMANO	1.580	1.528	1.635
NE_PREECOLAR	0.549	0.368	0.819
NE_PRIMARIA	0.458	0.357	0.587
NE_SECUNDARIA	0.252	0.190	0.335

Continúa

Odds Ratio Estimates
Point 95% Wald

Effect	Estimate	Confidence	Limits
NE_UNIVERSITARIO	0.064	0.047	0.088
REGCONT 0.389 0.349 0.433 REGSUB	1.615	1.326	1.968
REGEXC	0.172	0.129	0.230
EDAD2	1.807	1.455	2.244
EDAD3 2.979 2.349 3.779			
EDAD4 1.895 1.595 2.251			
EDAD6 0.595 0.469 0.754			
NRO_OCUPADOS 0.348 0.325 0.372			

The LOGISTIC Procedure
Association of Predicted Probabilities and Observed Responses

Percent Concordant	82.7	Somers' D	0.657
Percent Discordant	17.0	Gamma	0.658
Percent Tied	0.2	Tau-a	0.315
Pairs	19950576	c	0.828

Cuadro 8. Estimación de las razones de riesgo

Con respecto a la población en edad de trabajar, los individuos ocupados y desocupados tienen aproximadamente dos veces más de riesgo de ser pobres, con respecto a quienes no lo están. En relación con la forma de afiliación a regímenes de salud, quienes pertenecen al régimen subsidiado tienen aproximadamente dos veces más riesgo de ser pobres que los no afiliados a ningún régimen de salud.

El modelo *Logit* para la clasificación de pobreza muestra que un individuo de género masculino tiene 12,6% menos de riesgo de ser pobre que uno de género femenino. Un individuo con nivel educativo preescolar tiene el 45,1% de probabilidad de ser clasificado como pobre, mientras que los preparados a nivel superior (universitarios) son menos propensos a la pobreza en un 93,6%. Así, a menor nivel educativo la posibilidad de ser pobre aumenta. Esta es una de las categorías elegidas

como base para la estimación del modelo, es decir, el nivel educativo marca diferencia en el indicador de pobreza, que se concentra en los grupos más vulnerables: los niños o las personas sin posibilidad de educación [7].

El modelo de regresión *Logit* para la estimación de pobreza establece que:

$$\log\left(\frac{\hat{P}(Y=1)}{1-\hat{P}(Y=1)}\right) = \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + L + \hat{\beta}_k X_k \quad (1)$$

Es decir,

$$\hat{P}(Y=1) = \left(\frac{1}{1 + \exp[-(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + L + \hat{\beta}_k X_k)]} \right) \quad (2)$$

Para $Y=1$, “el individuo es considerado pobre” se tiene el modelo siguiente:

$$\hat{P}(Y=1) = \left(\frac{1}{1 + \exp[-(-0.1346 SEXO + 0.9075 OCUPADO + L - 1.0567 NRO_OCUPADOS)]} \right) \quad (3)$$

re-creaciones

La probabilidad de ser pobre de una mujer inactiva que pertenece a un hogar de seis individuos, nivel educativo primaria, sin afiliación al régimen de

salud y mayor de sesenta años ($X_1=1, X_2=6, X_3=1, X_4=1$) se calcula reemplazando las estimaciones de los β 's y los valores de $X_1, X_2, X_3, X_4, X_5, X_6$ en (2) de la siguiente manera:

$$\hat{P}(Y=1) = \left(\frac{1}{1 + \exp[-(-0.1346(0) + 0.3258(1) + 0.4676(6) - 0.7811(1) - 0.5195(1))]} \right) = 0.86186$$

Este resultado significa que aproximadamente el 86,2% de los individuos que presentan estas características son clasificados como pobres. Si las condiciones del hogar cambiaran en la cantidad

de personas que lo conforman, por ejemplo sólo dos individuos por hogar, conservando las demás características descritas en el ejemplo anterior se tiene:

$$\hat{P}(Y=1) = \left(\frac{1}{1 + \exp[-(-0.1346(0) + 0.3258(1) + 0.4676(2) - 0.7811(1) - 0.5195(1))]} \right) = 0.4901$$

Con base en los anteriores resultados, el riesgo de pobreza disminuye en 37,2%; esto indica que la cantidad de personas que conviven en el mismo hogar es determinante entre los factores de pobreza.

Si se modifica el nivel educativo a universitario y la edad se establece entre quince y cuarenta y cuatro años, manteniendo constantes las demás características, esto es, ($X_1=1, X_2=6, X_3=1, X_4=1$) se tiene:

$$P(Y=1) = \left(\frac{1}{1 + \exp[-(-0.1346(0) + 0.3258(1) + 0.4576(6) - 2.7452(1) + 0.6391(1))]} \right) = 0.72418$$

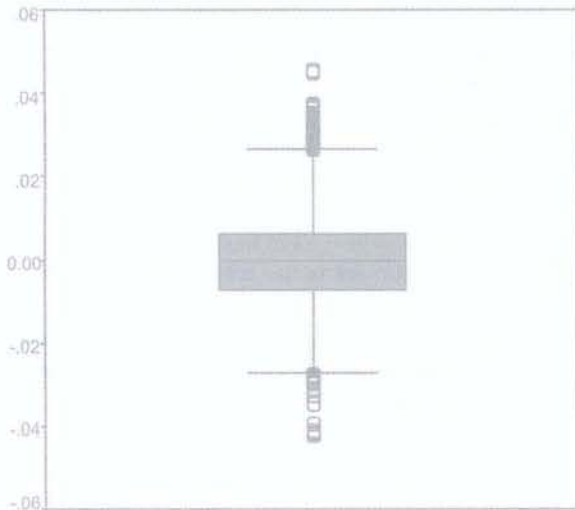
La probabilidad de pobreza de un individuo disminuye en el 13,8%, en comparación con las condiciones iniciales (primer ejemplo); esto permite concluir que cuando mejora el nivel educativo de la población y durante la edad productiva las condiciones de vida de los individuos tienden a mejorar, esto es, hay menos riesgo de ser considerado pobre.

La razón de *odds* para TAMANO es:

$\exp(\hat{\beta}_5) = \exp(0.4576) = 1.5803$, lo que significa que quienes conviven en hogares con mayor cantidad de individuos, manteniendo las demás variables constantes, tienen el 58% más de probabilidad de ser clasificadas como pobres que quienes conviven en hogares menos numerosos.

El intervalo de confianza se interpreta analizando si éste contiene o no el número 1. Si lo contiene puede concluirse que es igualmente riesgoso estar en un nivel de la variable que en otro; en otras palabras, que la variable no tiene efecto significativo sobre la respuesta. Si el intervalo no contiene a 1 se deberá observar si sus dos extremos son mayores o menores que 1; si son >1 se concluye que la probabilidad de ser clasificado "pobre" es mayor, por las características o condiciones del individuo. Por ejemplo, la variable EDAD3 con intervalo [2,349; 3,779] significa que las personas entre cinco y catorce años de edad tienen un alto riesgo (2,979, esto es, aproximadamente tres veces más) de ser clasificadas como pobres, manteniendo constantes

las otras variables. Si los extremos son menores que 1 se tiene la situación inversa.



Gráfica 2. Box-plot para residuales

Finalmente, el análisis de residuos y datos influyentes se puede interpretar a partir de la gráfica 2. El *box-plot* correspondiente muestra los residuos estudentizados. En verdad, no existen puntos muy influyentes, como se aprecia en la tabla 3, teniendo en cuenta que estos se comparan con una $\chi^2_{(17,0.05)} = 35.72$. En consecuencia, puede concluirse que el ajuste de los datos es correcto, es decir, no se observan casos que tengan un efecto significativo sobre las estimaciones de los parámetros del modelo.

5. Conclusiones

- El riesgo de ser clasificado como pobre aumenta o disminuye en la medida en que se modifiquen ciertas variables asociadas con la pobreza. Los factores aquí considerados se restringen a la información disponible, a partir de la aplicación de la Encuesta Nacional de Hogares realizada por el DANE.
- Estudios de esta naturaleza pueden servir de reflexión a las entidades encargadas del desarrollo social de la ciudad, para definir y/o fortalecer planes de acción especialmente encaminados al control de natalidad, la creación de mejores oportunidades laborales (empleos dignos y bien remunerados) y programas de formación académica acordes con las necesidades de la ciudad.
- Existen otros factores sociales que inciden directamente en los indicadores de pobreza. Es el caso de los desplazamientos masivos a la ciudad, la guerrilla, el paramilitarismo, la falta de oportunidades de empleo y estudio y la alta violencia intrafamiliar; dado que ellos no han sido eje del presente estudio, es parte del trabajo futuro la realización de estudios estadísticos centrados en cuantificar tales efectos.

DATO	DBET	DCHI
1	-0.04221	14.32826
2	-0.04181	12.72082
3	-0.04123	5.745475
4	-0.03975	8.621776
5	-0.03473	2.998135
6	-0.03447	19.81746
7	-0.03243	1.942488
8	-0.03243	1.942488
9	-0.03233	9.199762
10	-0.03137	1.896979
<hr/>		
9116	0.034865	25.31739
9117	0.035062	9.016276
9118	0.035089	5.896968
9119	0.036083	8.441767
9120	0.036435	73.51301
9121	0.036963	11.17301
9122	0.037656	11.05385
9123	0.037656	11.05385
9124	0.044761	8.199514
9125	0.04534	6.175394

Tabla 3. Datos extremos influyentes

Referencias bibliográficas

- [1] Casa Editorial El Tiempo, Fundación Corona y Cámara de Comercio de Bogotá D.C. (2003). *Bogotá cómo vamos. La pobreza en Bogotá*.
- [2] Arboleda, J. (2002). *Monografías Altos de Cazucá y El Úval. La voz de los pobres ¿Hay alguien que nos escuche?* Deepa Narayan, Banco Mundial, Bogotá D.C.
- [3] Bustelo, E. S. (1998). *Pobreza moral. Reflexiones sobre la política social amoral y la utopía posible*. Trabajo presentado al Seminario Internacional, 27 y 28 de octubre. Bogotá, Colombia.
- [4] Corredor, M. C. (1999, marzo) *Pobreza y desigualdad reflexiones conceptuales y de medición*. CINEP, Santafé de Bogotá, Colombia.
- [5] Díaz, L. G. (2002). *Estadística multivariada y métodos*. Universidad Nacional de Colombia, Bogotá.
- [6] Hair, J., Anderson, R., Tatham, R. y Black, W. (1999). *Análisis Multivariante*. 5ª ed., Prentice Hall.
- [7] Kleinbaum, D.G. (1994). *Logistic Regression: A Self-Learning Text. Statistics in the Health Sciences*. Springer-Verlag. Social Science Research Instituto, New York.
- [8] Muñoz, M: (1999). “Los indicadores de pobreza utilizados en Colombia: una crítica”. Artículo publicado en: Corredor, M.C. (1999, marzo).
- [9] Pardo, C. E. y Cabarcas, G. (2001). *Métodos estadísticos multivariados en investigación social*. Simposio de Estadística.