

Nota técnica del cálculo del Rezago Social en las AGEB urbanas de México

Introducción

Con el propósito de contar con la mayor información posible de indicadores de desarrollo social, en mayores niveles de desagregación territorial, el CONEVAL da a conocer la estratificación de las Áreas Geoestadísticas Básicas (AGEB) de las localidades urbanas del país en tres Grados de Rezago Social (GRS), de acuerdo con la información disponible del INEGI.

Según la definición de INEGI, una AGEB urbana es un área geográfica ocupada por un conjunto de manzanas perfectamente delimitadas por calles, avenidas, andadores o cualquier otro rasgo de fácil identificación en el terreno y cuyo uso del suelo es principalmente habitacional, industrial, de servicios, comercial, etcétera, y sólo son asignadas al interior de las localidades urbanas que son aquellas con población mayor o igual a 2,500 habitantes.

Para la estimación del número de GRS y la distribución de las AGEB en esos grados, se empleó la metodología estadística de Análisis de Clases Latentes (ACL), la cual permite estimar tanto el número óptimo de estratos como la distribución de las observaciones en esos estratos, empleando la información de las variables de cada observación en sus valores originales.

La metodología de ACL empleada a nivel AGEB es diferente a la usada en la clasificación de las entidades, municipios y localidades en los diferentes Grados de Rezago Social; sin embargo, se utilizan los mismos indicadores en todos los niveles de desagregación.

Los GRS estimados para las AGEB son Alto, Medio y Bajo, distribuyéndose entre esos grados a 51,034 AGEB de todo el país, de acuerdo a la información disponible del Censo de Población y Vivienda 2010.

1. Metodología

La estratificación a partir del ACL permite representar la heterogeneidad no observable en un conjunto de observaciones analizadas (Hagenaars y McCutcheon, 2002; Muthén y Muthén, 2001). A partir de una serie de características (variables latentes) agregadas a nivel de AGEB, el ACL asigna a cada AGEB una probabilidad de pertenecer a algún grupo (estrato) de AGEB similares entre sí. La cantidad de clases no está definida a priori, sino que se determina mediante un modelo estadístico que especifica la distribución conjunta de las variables latentes para establecer la cantidad de grupos más adecuada, lo cual sucede en el momento en que se minimizan las diferencias al interior de cada grupo y se maximizan entre ellos.

El número de variables utilizadas para el análisis fueron 14, de las cuales 11 son las mismas a las empleadas en la estimación del Índice de Rezago Social (IRS) en los niveles estatal, municipal y de localidades. Estas variables guardan relación con algunos de los indicadores marcados en la

Ley General de Desarrollo Social para la medición de la pobreza. Sin embargo, dado que no contempla a todos los indicadores y no se usa la metodología oficial, la agrupación de las AGEB en diferentes GRS no constituye una medición de pobreza, sino un análisis complementario de indicadores de carencias y rezago social en un nivel de desagregación territorial más pequeño que son las AGEB. La información para la construcción de estas variables se tomó de los resultados definitivos del Censo de Población y Vivienda 2010, proporcionados al CONEVAL por INEGI.

Las variables empleadas se muestran en cuadro I.

Cuadro I
Variables empleadas para la estimación del Rezago Social 2010, según nivel geográfico de desagregación

Indicador asociado de la Ley	Variables para la estimación del Rezago Social a nivel entidad, municipio y localidad, 2010*	Variables para la estimación del Rezago Social a nivel de AGEB, 2010*
	Total de variables: 11	Total de variables: 14
Rezago educativo	Población de 15 años o más analfabeta	Población de 15 años o más analfabeta
	Población de 6 a 14 años que no asiste a la escuela	Población de 6 a 14 años que no asiste a la escuela
	Población de 15 años y más con educación básica incompleta	Población de 15 años y más con educación básica incompleta
		Población de 15 a 24 años que no asiste a la escuela
Acceso a los servicios de salud	Población sin derechohabencia a servicios de salud	Población sin derechohabencia a servicios de salud
Calidad y espacios de la vivienda	Viviendas con piso de tierra	Viviendas con piso de tierra
		Personas que viven en hacinamiento
Servicios básicos en la vivienda	Viviendas que no disponen de excusado o sanitario	Viviendas que no disponen de excusado o sanitario
	Viviendas que no disponen de agua entubada de la red pública	Viviendas que no disponen de agua entubada de la red pública
	Viviendas que no disponen de drenaje	Viviendas que no disponen de drenaje
	Viviendas que no disponen de energía eléctrica	Viviendas que no disponen de energía eléctrica
Ingreso (bienes del hogar)	Viviendas que no disponen de lavadora	Viviendas que no disponen de lavadora
	Viviendas que no disponen de refrigerador	Viviendas que no disponen de refrigerador
		Viviendas que no disponen de teléfono fijo

*NOTA: todas las variables son en porcentaje.

Una vez que se construye la base de datos con estas variables, se realiza el ACL usando el programa Mplus¹. El análisis permitió establecer que el número óptimo de estratos es de tres, los cuales se obtienen aplicando los criterios siguientes:

- Menor índice BIC-Ajustado
- Menor valor de la Prueba de Lo-Mendel Rubin
- Mayor nivel de entropía
- Porcentajes de clasificación mayores al 5 por ciento en cada estrato.
- Probabilidades de pertenencia a cada estrato mayores del 90 por ciento.

¹ La sintaxis del programa se presenta en el anexo de este documento.

1.1. Tratamiento de variables concentradas en cero

Debido a que los valores de algunas de las variables usadas en el análisis se concentran en cero², se aplicó un tratamiento especial para esas variables conocido como Zero-Inflated Poisson (ZIP), que consiste en redondear los valores originales de las variables y optimizar una función de distribución Poisson como función objetivo en el ACL³.

2. Resultados generales

Como ya se mencionó anteriormente, el ACL dio como resultado que el número óptimo de estratos en los cuales deben dividirse las AGEB es de tres; cada estrato representa un GRS, los cuales se clasifican en GRS Alto, Medio y Bajo.

En la tabla I se presentan los resultados de las pruebas que permiten distinguir el número óptimo de estratos (GRS) a construir. Además, en la gráfica I se observa que los promedios de las variables en cada uno de los GRS son distinguibles; es decir, los promedios de las variables en el GRS Alto son mayores que en los otros estratos, y así sucesivamente para los grados Medio y Bajo.

Tabla I

Valores de las pruebas para determinar el número óptimo de estratos

Clases	Ajusted BIC	LO-MENDELL-RUBIN adjusted lrt value p-value		Entropía	Rango del % de AGEB por grupo	Rango de probabilidad de pertenencia a los grupos
1	7,438,448.9	-	-	-	100	1
2	5,960,381.9	1438596.23	0.0000	0.987	22.564 - 77.436	0.991 - 0.997
3	5,619,041.9	-	-	0.977	24.172 - 64.655	0.980 - 0.993
4	5,403,299.1	-	-	0.968	11.170 - 64.700	0.951 - 0.991
5	5,247,271.4	147929.048	0.0000	0.959	6.225 - 44.065	0.961 - 0.989

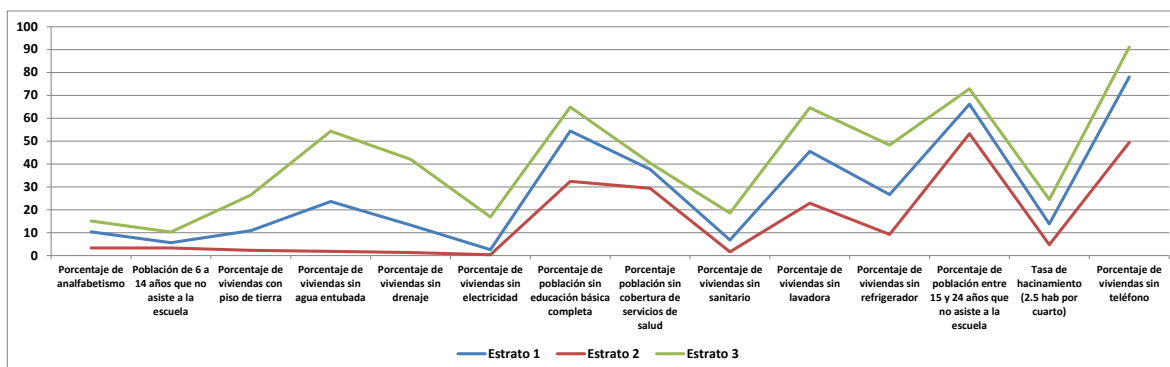
Fuente: Elaboración del CONEVAL con información del Censo de Población y Vivienda 2010

² Debido, por ejemplo, a que en zonas urbanas la cobertura en algunos servicios básicos en la vivienda y asistencia escolar son casi universales, por lo tanto la carencia en esos indicadores es cero para muchas AGEB.

³ Para una mayor explicación de los modelos ZIP, véase Broek (1995) *A Score Test for Zero Inflation in a Poisson Distribution*.

Gráfica I

Promedio de las variables en cada estrato (GRS)



Fuente: Elaboración del CONEVAL con información del Censo de Población y Vivienda 2010

En la tabla II se presenta la distribución de la población de las AGEB analizadas en cada uno de los GRS. De acuerdo a estos datos, se observa que en el GRS Alto aproximadamente se concentran el 10 por ciento de las AGEB analizadas; sin embargo, en estas AGEB se concentra poco menos del 2 por ciento de la población. Por otro lado, el GRS que concentra el mayor número de AGEB es el Bajo (alrededor del 65 por ciento de las AGEB) en donde habita el 83 por ciento de la población de estas AGEB urbanas.

Tabla II

Distribución de AGEB urbanas y población de acuerdo con su Grado de Rezago Social, México, 2010

Estrato	Número de personas	Porcentaje	Número de AGEB
Bajo	72,712,253	83.6	33,024
Medio	12,788,350	14.7	12,305
Alto	1,443,394	1.7	5,705
Total	86,943,997	100.0	51,034

Fuente: Elaboración del CONEVAL con información del Censo de Población y Vivienda 2010

De esta forma, quedan clasificadas las AGEB urbanas con información disponible en cada GRS, en donde habitan aproximadamente el 77.4 por ciento de la población total del país, de acuerdo a los datos del Censo de Población y Vivienda 2010.

Cabe recalcar que el ACL no implica la generación de algún índice en particular, ni tampoco el tratamiento o alguna transformación de los valores originales de las variables. Además, el mismo ACL permite determinar el número óptimo de estratos distinguibles entre si y homogéneos dentro de cada uno.

Referencias bibliográficas.

Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L. y Kirchner, U. (1999). *The Zero-Inflated Poisson Model and the Decayed, Missing and Filled Teeth Index in Dental Epidemiology*, Journal of the Royal Statistical Society, Series A, Vol. 162, No. 2, pp. 195-209.

CONEVAL (2010). *Metodología para la medición multidimensional de la pobreza en México*. México, CONEVAL.

Diario Oficial de la Federación (16 de junio de 2010). *Lineamientos y criterios generales para la definición, identificación y medición de la pobreza*, México, recuperado de:
http://dof.gob.mx/nota_detalle.php?codigo=5146940&fecha=16/06/2010 (2011, 22 de noviembre).

Hagenaars, Jacques y Allan McCutcheon (2002). *Applied Latent Class Analysis*, Reino Unido, Cambridge University Press.

Muthén, L. y Muthén B. (2010). *Mplus. Statistical Analysis with Latent Variables. User's Guide*, Estados Unidos, Muthén & Muthén.

Roeder, K., Lynch, K. G. y Nagin, D. S. (1999). *Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology*, Journal of the American Statistical Association, Vol. 94, No. 447, pp. 766-776.

Sepúlveda, Rosa A. (2003). *Análisis de Clases Latentes*, texto extraído de la tesis doctoral: *Contribuciones al Análisis de Clases Latentes en Presencia de Dependencia Local*, España, Departamento de Estadística, Universidad de Salamanca, mimeo.

Vargas, D., Méndez, I., Moreno, H., Nieto, L. E., Brunsdon, C., Tzavidis, N., Kwang-Kim, J. y Pérez, G. (2010). *Metodología de ajustes e imputación de indicadores de la pobreza por ingreso en áreas pequeñas*, México, El Colegio de México y CONEVAL, mimeo.

Van den Broek, Jan (1995). *A Score Test for Zero Inflation in a Poisson Distribution*, Biometrics, Vol. 51, No. 2, pp. 738-743.

Vargas, D., Pérez, G., De la Vega, J., Carmona, C. y Luján, C. (2011). *Calibración de indicadores socioeconómicos*, México, El Colegio de México y CONEVAL, mimeo.

Anexo

Sintaxis para obtener los GRS a partir del ACL

La sintaxis que se presenta a continuación debe correrse en el paquete estadístico Mplus6. Deberá también tener presente lo siguiente.

- La base de datos que contenga los valores de las variables debe estar en formato txt.
- El identificador de cada AGEB debe ser numérico, este identificador aparece en la sintaxis declarado con el nombre ID.
- El nombre de las variables no debe exceder 8 caracteres.
- Las variables deben estar en su valor original; en este caso en porcentaje.
- Para aquellas variables con alta concentración de ceros, sus valores deben ser redondeados al entero más cercano y deben ser declaradas como variables de conteo con la instrucción COUNT IS dentro de la sintaxis. En este caso, las variables siguientes fueron redondeadas: porcentaje de población de 15 años y más analfabeta, porcentaje de población de 6 a 14 años de edad que no asiste a la escuela, porcentaje de viviendas con piso de tierra, porcentaje de viviendas que no disponen de drenaje, porcentaje de viviendas que no disponen de electricidad, porcentaje de viviendas que no disponen de agua.
- Para obtener los valores de las pruebas, según el número de estratos, deberá correrse la sintaxis cambiando en cada caso el número de estratos en la instrucción CLASSES. En cada ocasión que se corra la sintaxis es recomendable guardar con nombres diferentes la sintaxis misma y la base de datos que se genera con las probabilidades de pertenencia a cada estrato.

A continuación se presenta la sintaxis para tres estratos.

!Nombre de la sintaxis.
TITLE: Instrucciones para la construcción de estratos mediante el análisis de clases latentes (ACL)

!Directorio de la base que se utiliza para la formación de las clases latentes.
DATA:
File is 'C:\Users\Desktop\rez_soc_ageb\base.txt';
FORMAT IS FREE;

!Nombre de las variables.
VARIABLE:
NAMES ARE
ID i_analfa i_asiesc i_tierra i_noagua i_nodren i_noelec i_ebasin i_ssalud
i_nosani i_nolava i_norefr i_asescj i_cuart2 i_sintel;

!Variables que se utilizan.
USEV ARE i_analfa i_asiesc i_tierra i_noagua i_nodren i_noelec i_ebasin
i_ssalud i_nosani i_nolava i_norefr i_asescj i_cuart2 i_sintel;

!Definición de variables count.
COUNT IS i_analfa i_asiesc i_tierra i_noagua i_nodren i_noelec (i);

!Número de clases.
CLASSES = C(3);

!Identificador de la variable.
IDVARIABLE IS ID;

ANALYSIS:
!Modelos mixtos.
TYPE = MIXTURE;

!Máximos locales.
STARTS 100 50;

!Resultados.
OUTPUT:

!SAMPSTAT: resultados de estadísticas descriptivas.
!TECH11: prueba de razón de verosimilitud del modelo ajustado.
!TECH7: estimaciones de las probabilidades para cada clases.
SAMPSTAT TECH11 TECH7;

!Base de datos final.
SAVEDATA:
FILE IS 'C:\Users\Desktop\rez_soc_ageb\clat_3.dat';

!Probabilidad de pertenecer a alguna clase.
SAVE=CPROB;
