

Construcción de nivel socioeconómico predominante por AGEB mediante análisis multivariado

KAREN ARGÜELLES Y LUIS E. GARCÍA



Niveles socioeconómicos para AGEB's

Asociar niveles socioeconómicos con áreas pequeñas es una necesidad para muchos investigadores. Conocer *a priori* qué NSE es predominante en una AGEB puede ayudar a enfocar esfuerzos de campo, a estratificar marcos muestrales, a detectar áreas con un *target* más preciso (amas de casa de NSE ABC+, por ejemplo), etcétera. Esto por no mencionar todas las maravillas que se podrían hacer con *software* para manejo de información geo espacial.

Sabemos, claro, que en una misma AGEB pueden convivir varios NSE. Pero proponemos recurrir al concepto de “NSE predominante” o “NSE proxy”, es decir, asumir una cierta homogeneidad del nivel de bienestar de la mayoría de los hogares al interior de un área geográfica pequeña. Esperamos que una gran mayoría de hogares de un mismo edificio compartan un mismo NSE. Si bien las reglas de asignación de NSE aplican para hogares, no para AGEB's, INEGI publica un gran número de indicadores a nivel AGEB, de los cuales muchos están relacionados con el nivel de bienestar o desarrollo económico de esa área.

El maestro Francisco Sánchez Villarreal, en uno de sus trabajos de consultoría de muestreo, nos propuso un método para conseguir este resultado, que ahora compartimos con ustedes.

Variables de entrada

Para asignar cada AGEB a un NSE primero seleccionamos algunas variables reportadas por INEGI del levantamiento del Censo de Población y Vivienda 2010 al nivel de AGEB. INEGI reporta indicadores con información tanto a nivel de personas como a nivel de viviendas. Las variables consideradas para nuestro ejercicio se muestran en la tabla 1.

Las variables de la base de viviendas tienen ya de por sí una correlación directa con el nivel de bienestar económico de cada AGEB, es decir, si el indicador es más grande, el nivel socioeconómico de las personas en la AGEB debería ser más alto.

Creación de variables nuevas

Las variables anteriores no son directamente comparables entre AGEB's pues no toman en cuenta el tamaño de la AGEB. Además, las variables están en diferentes escalas, lo que tiene efectos no deseables en el cálculo de covarianzas. Por ello creamos tres nuevos indicadores a partir de las variables anteriores para los que el tamaño de la AGEB no influye y, además, tienen la misma escala (entre 0 y 1).

Tabla 1
Variables consideradas

	Variable	Descripción
Base de Personas	Habitantes	Total de habitantes
	Población>18 edad	Población mayores a 18 años de edad
	NIV_ESCO>Prepa	Población con un nivel de escolaridad mayor a la preparatoria
	AvgOfESC_ACUM	Promedio de la escolaridad acumulada de la población mayor a 18 años de edad
Base de Viviendas	Tohogviv	Total de hogares por vivienda
	Dis_comp	Distingue si la vivienda tiene o no computadora
	No disbie	Distingue a las viviendas que no tienen ninguno de los cuatro bienes captados

$$\text{ÍndiceCompu} = \frac{\text{Núm. viviendas en AGEB con computadora}}{\text{Núm. viviendas en AGEB con los 4 productos (compu, refri, lavadora y TV)}}$$

Estas tres variables además están fuertemente correlacionadas entre sí.

$$\text{ÍndiceEsc} = \frac{\text{Prom. años estudiados por persona en la AGEB}}{\text{Max. del promedio de años estudiados de todas las AGEBs}}$$

$$\text{ÍndiceEduSup} = \frac{\text{Núm. personas mayores a 18 años con educación media-superior}}{\text{Núm. personas mayores a 18 años en la AGEBs}}$$

Tabla 2
Matriz de correlaciones

	ÍndiceCompu	ÍndiceEduSup	ÍndiceEsc
ÍndiceCompu	1.00	0.87	0.84
ÍndiceEduSup	0.87	1.00	0.87
ÍndiceEsc	0.84	0.87	1.00

Análisis de componentes principales

Cada una de las tres variables anteriores podría ser usada por sí misma como una medida del nivel socioeconómico, pero ¿es posible obtener un solo indicador que ordene a las AGEB's por su nivel de bienestar considerando la información de las tres variables? La respuesta es sí. El análisis multivariado a través del análisis de componentes principales permite hacer precisamente eso.

Ya que las tres variables están en la misma escala y están fuertemente correlacionadas, no sorprende que la primera

componente explique el 89% de la varianza total de las tres. Es esta nueva variable, la primera componente, la que utilizamos para clasificar a las AGEB's.

Asignación de NSE a AGEB's

Ordenando todas las AGEB's descendientemente por esta primera componente, se determinaron los cortes para cada NSE aplicando la distribución acumulada publicada por la AMAI 2010 (8x7) a la proporción de la población acumulada de las AGEB's. En la tabla 4 se muestran los cortes de esa primera componente para cada ciudad/NSE.

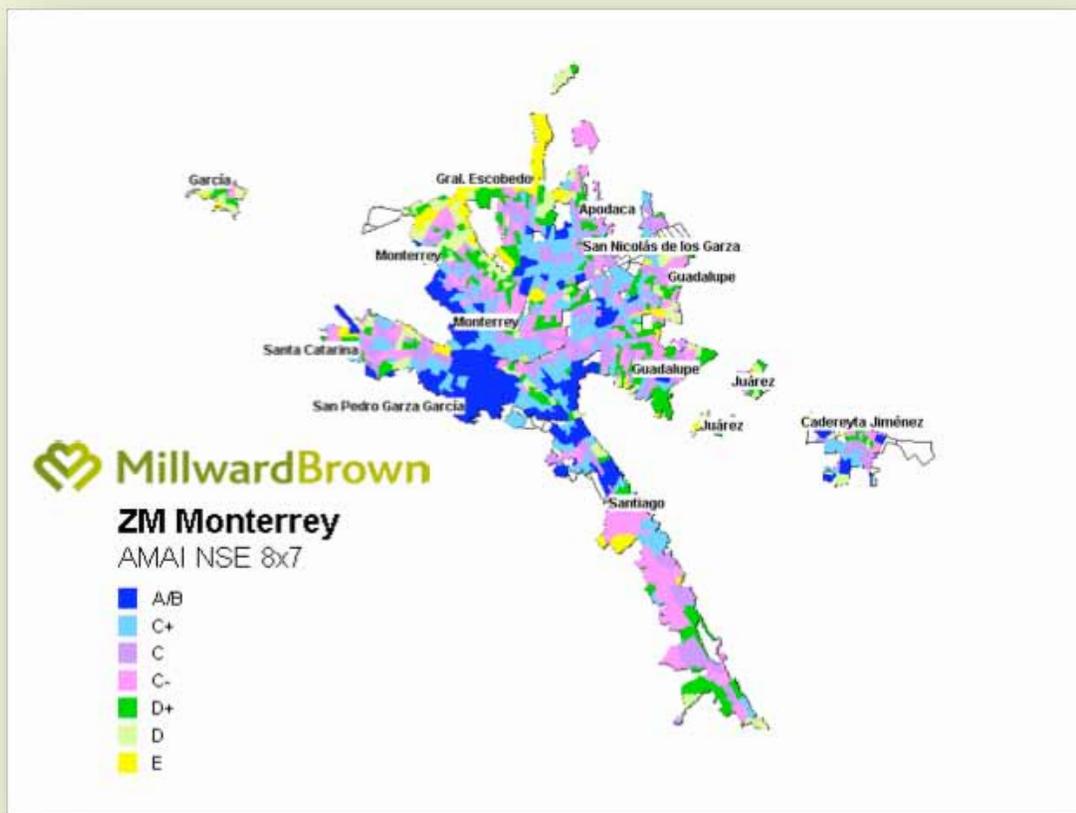
Tabla 3
Varianza total explicada

Componente	Valor característico	% de la varianza explicada	% acumulado	Total	% de la varianza
1	2.671	89.027	89.027	2.671	90.560
2	0.240	8.002	97.029		
3	0.089	2.971	100.00		

Método de extracción: Análisis de componentes principales.

Tabla 4
Cortes de primera componente para cada ciudad/NSE

NSE Proxy	Valle de México		Guadalajara		Monterrey	
	Límite superior de la variable proxy	% Poblac. Acumulado	Límite superior de la variable proxy	% Poblac. Acumulado	Límite superior de la variable proxy	% Poblac. Acumulado
A/B		6%		11%		8%
C+	2.41	18%	2.10	28%	2.75	25%
C	1.25	34%	0.91	47%	1.14	45%
C-	0.68	50%	0.25	65%	0.40	63%
D+	0.34	68%	-0.12	84%	0.00	82%
D	0.05	92%	-0.48	96%	-0.46	97%
E	-0.40	100%	-0.70	100%	-0.71	100%



Validación

Antes de dar por buena esta estimación, se realizaron dos validaciones independientes:

1. **Mediante las entrevistas de dos años de diversos estudios continuos levantados casa por casa pertenecientes al área metropolitana de la Ciudad de México.** De cada entrevista se tiene la clave completa de la AGEB a la que pertenece y se tiene también el NSE asignado por la regla AMAI 8x7. Con esta información se determinó un NSE predominante por AGEB, que fue igual al NSE que tuviera más entrevistas en la AGEB. Así clasificamos a 1,464 AGEB's metropolitanas y comparamos esta clasificación con la de la variable proxy:

- Más del 40% tuvo el mismo NSE predominante.
- Casi otro 40% tuvo una diferencia de un NSE, que puede deberse a la muestra efectiva lograda por campo en la AGEB, más que a una mala clasificación.
- Del 20% restante, 9% no tenía muestra suficiente para asignar un NSE.

2. **Mediante la ayuda de Google Maps,** se ubicaron las AGEB's de niveles completamente distintos que colindaran unas con otras. Los resultados se muestran en las páginas siguientes.

Comentarios finales

Insistimos en que **la regla AMAI aplica solamente para hogares**, no para AGEB's. Pero la clasificación que acabamos de describir es una propuesta metodológicamente sustentada para generar una herramienta de gran utilidad para optimizar el diseño y la implementación de proyectos de diversos ámbitos, como pueden ser muestreo, trabajo de campo, logística y georreferenciación. Además cuenta con la ventaja de utilizar información pública generada por el INEGI, aplicando un método estadístico ampliamente divulgado.



Zona Metropolitana:	Valle de México
NSE proxy que conlindan en AGEB's:	A/B y D
Delegación:	Cuajimalpa, D.F.
Mapinfo:	Google Maps



Santa Fe, Cuajimalpa.
NSE apreciativo A/B

La Granja, Cuajimalpa.
NSE apreciativo D/E



Municipio:	Monterrey, Nuevo León San Pedro Garza, Nuevo León
Mapinfo:	Google Maps



San Pedro Garza, Nuevo León.
NSE apreciativo A/B

Monterrey, Nuevo León.
NSE apreciativo E



Zona Metropolitana:

Guadalajara

NSE proxy que conlindan en AGEB's:

A/B y D

Municipio:

Tlaquepaque, Jalisco

El Salto, Jalisco

Mapinfo:

Google Maps

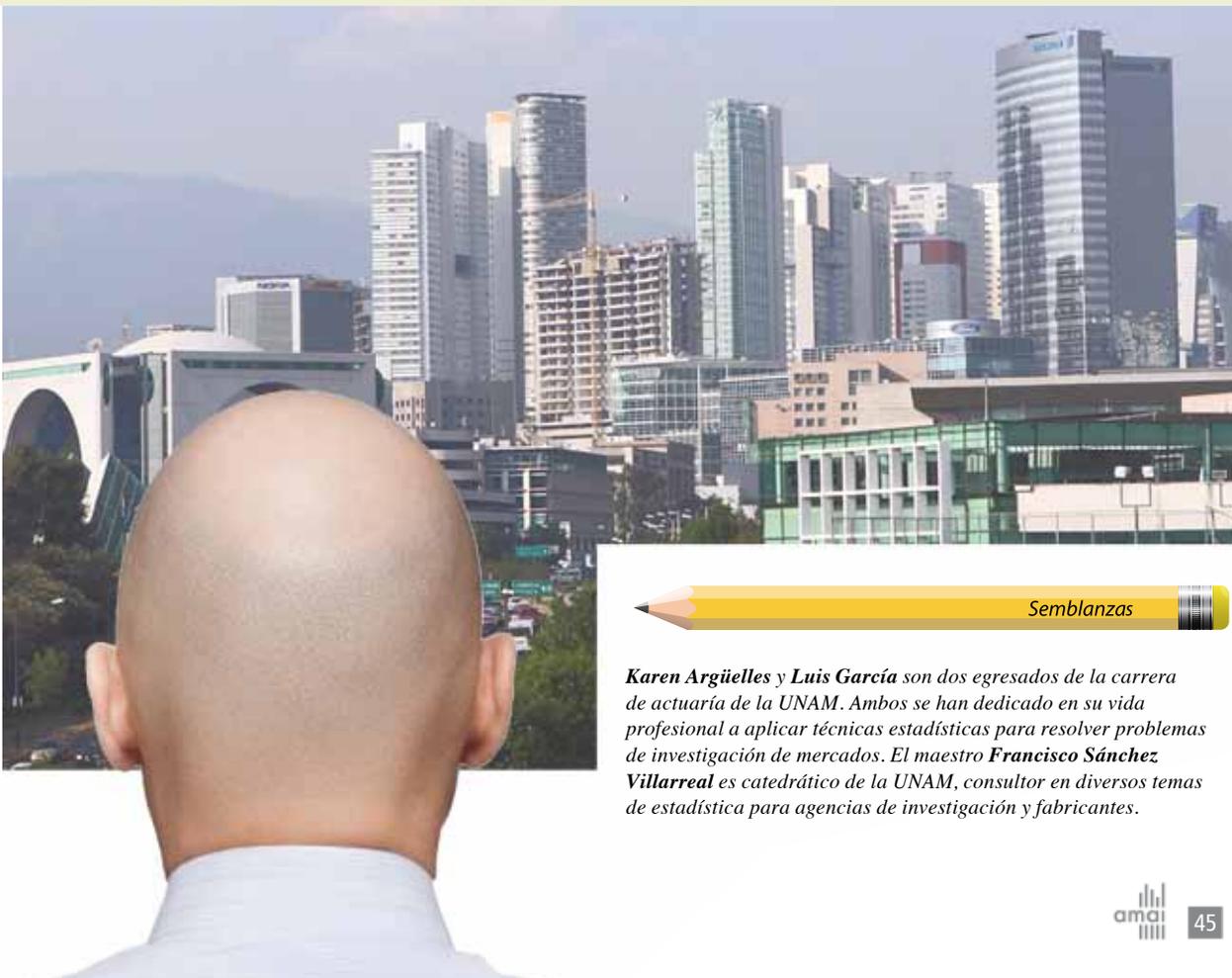


El Salto, Jalisco.

Tlaquepaque, Jalisco.

NSE apreciativo A/B

NSE apreciativo D



*Karen Argüelles y Luis García son dos egresados de la carrera de actuaría de la UNAM. Ambos se han dedicado en su vida profesional a aplicar técnicas estadísticas para resolver problemas de investigación de mercados. El maestro **Francisco Sánchez Villarreal** es catedrático de la UNAM, consultor en diversos temas de estadística para agencias de investigación y fabricantes.*