

Nota metodológica - DOT DF

2014



Contenido

1. Metodología para estimar el nivel socioeconómico a nivel AGEB.....	2
2. Metodología para calcular el nivel socioeconómico promedio y el grado de marginación urbano promedio alrededor de las estaciones de transporte público.....	4
3. Metodología para estimar otras variables alrededor de las estaciones de transporte público.....	6
4. Acrónimos y abreviaturas.....	7
5. Bibliografía.....	7



1. Metodología para estimar el nivel socioeconómico a nivel AGEB

La siguiente metodología permite obtener una estimación del nivel socioeconómico¹ predominante a nivel de área geostadística básica (AGEB) para la Zona Metropolitana del Valle de México (ZMVM). Esta metodología fue adaptada a partir de una estimación utilizada por la Asociación Mexicana de Agencias de Inteligencia de Mercado y Opinión (AMAI) disponible en su página de internet.² En general, ésta consiste en realizar un análisis por componentes principales a partir de tres índices socioeconómicos distintos, con el fin de obtener un solo puntaje que permite ordenar a las AGEBS de acuerdo al nivel socioeconómico predominante. Posteriormente, se le asigna un valor a partir de la distribución de la población de la ZMVM que pertenece a cada nivel socioeconómico calculada previamente por AMAI (2010) utilizando la regla 8x7. Los datos utilizados provienen del censo de población de 2010 realizado por el INEGI.

En primer lugar, se definen los siguientes índices socioeconómicos a partir de los datos disponibles:

$$\text{Índice Compu.} = \frac{\text{Número de viviendas en la AGEB con comptuadora}}{\text{Número de viviendas con alguno de los 4 bienes (compu, refri, lavadora y TV)}}$$

$$\text{Índice Escolaridad} = \frac{\text{Promedio años estudiados en la AGEB}}{\text{Máximo del promedio de años estudiados de todas las AGEBS}}$$

$$\text{Índice Edu. Sup.} = \frac{\text{Número de individuos mayores de 25 en la AGEB con educación superior}}{\text{Número de individuos mayores a 25 años}}$$

Como podemos observar en el siguiente cuadro, estos tres índices están altamente correlacionados entre sí.

CUADRO 1: MATRIZ DE CORRELACIONES

	Índice Compu	Índice Escolaridad	Índice Edu. Sup.
Índice Compu	1	0.846	0.834
Índice Escolaridad	0.846	1	0.968
Índice Edu. Sup.	0.834	0.968	1

Fuente: elaboración de ITDP con datos de INEGI.

Para explotar la alta correlación que existe entre estas tres variables generadas y obtener una sola variable que explique la mayor parte de la varianza entre AGEBS hacemos un análisis por componentes principales. Los resultados de éste análisis se muestran en el siguiente cuadro:

¹ AMAI define el nivel socioeconómico como el bienestar económico y social del que disfruta un hogar, o que tan satisfechas están sus necesidades.

² <http://blog.amai.org/index.php/construccion-de-nivel-socioeconomico-predominante-por-ageb-mediante-analisis-multivariado/>



CUADRO 2: RESULTADOS DEL ANÁLISIS DE COMPONENTES PRINCIPALES

Componente	Valor característico	Porcentaje de la varianza explicada	Porcentaje acumulado
1	2.7674	0.9225	0.9225
2	0.2011	0.067	0.9895
3	0.0314	0.0105	1

Fuente: elaboración de ITDP con datos de INEGI.

Al utilizar éste método podemos reducir la información que aporta cada uno de estos índices a un solo puntaje que podemos utilizar para ordenar las AGEBS según su nivel promedio de bienestar. Es decir, este método reduce la correlación entre los índices a tres componentes principales, y como se puede observar en el cuadro anterior, al utilizar solamente el primer componente se captura el 92.25 por ciento de la varianza total, por lo que se puede utilizar solamente éste para aproximar el nivel socioeconómico relativo entre las AGEBS de la ZMVM.

Una vez realizado este análisis, ordenamos las AGEBS descendientemente según el puntaje obtenido a partir de este primer componente. Para asignar un nivel socioeconómico a cada AGEB ajustamos la proporción de la población que habita en ellas a la proporción de la población total de la ZMVM que pertenece a cada nivel socioeconómico. Para esto se utiliza la regla 8x7 desarrollada por AMAI (2010).

CUADRO 3: CORTES DEL PRIMER COMPONENTE PARA LA ZMVM (2010)

Nivel Socioeconómico	Límite superior del puntaje del primer componente	Número de AGEBS	Porcentaje	Porcentaje acumulado
A/B		459	8.31	8.31
C+	2.88	749	13.56	21.86
C	1.12	827	14.97	36.83
C-	0.2	763	13.81	50.64
D+	-0.36	808	14.62	65.27
D-	-0.87	1,219	22.06	87.33
E	-1.64	700	12.67	100

Fuente: elaboración de ITDP.



2. Metodología para calcular el nivel socioeconómico promedio y el grado de marginación urbano promedio alrededor de las estaciones de transporte público

Cuando se trata de obtener variables socioeconómicas de la población que habita alrededor de las estaciones de transporte público existe el problema de la información disponible se encuentra al nivel de Área Geoestadística Básica (AGEB).³ Esto provoca que al momento de definir el área de estudio, que en este caso es la población que reside dentro de un radio de 800, 1000 y 2000 metros alrededor de las estaciones de transporte masivo, encontremos con que las diferentes AGEBS que componen esta superficie tiene diferentes niveles de desarrollo social y, por lo tanto, diferentes puntajes en los índices de Nivel Socioeconómico (NSE) y Grado de Marginación Urbana (GMU). Para obtener un valor promedio de esta población de una manera congruente, se recurre al supuesto de que la población en cada AGEB se encuentra distribuida de manera uniforme. Utilizando este supuesto se puede, entonces, ponderar las variables socioeconómicas a nivel AGEB por la proporción de la población de éstas que se encuentra dentro del radio de 800, 1000 y 2000 metros de cada estación y utilizar estas variables ponderadas para calcular el NSE y el GMU.

Tanto el índice de NSE como el de GMU parten de un análisis por componentes principales que utiliza variables socioeconómicas del Censo de Población y Vivienda 2010. La metodología para calcular el NSE fue adaptada de una estimación utilizada por la Asociación Mexicana de Agencias de Inteligencia de Mercado y Opinión (AMAI),⁴ mientras que el GMU fue tomado de las estimaciones realizadas por el Consejo Nacional de Población (CONAPO).⁵ El análisis por componentes principales es un método que permite reducir el poder explicativo de varias variables socioeconómicas a una sola, es decir, permite extraer la información que tienen varias variables en un puntaje que después se puede utilizar para ordenar las AGEBS según su nivel de desarrollo social. Las categorías de cada índice se asignan asumiendo una distribución preexistente, en el caso del NSE una distribución de nivel socioeconómico realizada por AMAI para la ZMVM, y para el GMU una técnica de estratificación óptima desarrollada por Dalenius y Hodges para todas las AGEBS urbanas del país.

Utilizando los resultados a nivel AGEB de estos dos índices, la metodología propuesta consiste en los siguientes pasos:

1. Se obtienen los índices de NSE y GMU para las AGEBS que componen la Zona Metropolitana del Valle de México.

³ Las AGEBS son una subdivisión de los municipios o delegaciones que conforman el país, utilizada por el INEGI en los Censos de Población y Vivienda.

⁴ <http://blog.amai.org/index.php/construccion-de-nivel-socioeconomico-predominante-por-ageb-mediante-analisis-multivariado/>

⁵ http://www.conapo.gob.mx/es/CONAPO/Indice_de_marginacion_urbana_2010



2. Se definen los radios de 800 metros, 1000 metros, y 2000 metros alrededor de las estaciones de transporte y se identifican las AGEBs que componen esta superficie.
3. Suponiendo que la población se distribuye uniformemente dentro de cada AGEB, se ponderan las variables utilizadas para calcular cada uno de los índices por la proporción de la población total del radio que reside dentro de cada AGEB.⁶
4. Una vez obtenidas las variables ponderadas para estas superficies, se vuelve a realizar el análisis por componentes principales incluyéndolas en la muestra, es decir, se asume que son una especie de “AGEBs proxy”, y se calcula el puntaje que obtendrían dadas sus características socioeconómicas promedio.
5. Se utiliza el puntaje para ordenar a las superficies de los radios alrededor de las estaciones junto con las AGEBS.
6. Se le otorga a los radios de 800 metros, 1000 metros, y 2000 la categoría del índice de NSE y GMU que obtuvieron AGEBS con puntajes similares.

A pesar de las limitaciones teóricas de éste método, como es la duplicidad de áreas y el hecho de que la muestra ya no corresponde a aquella a la que se ajustó la distribución original, esto no es tan problemático, ya que lo que se desea hacer es simplemente comparar el puntaje relativo de estas superficies al de las AGEBS para las que sí se conoce el NSE y el GMU. Esta solución hace posible asignar una categoría del índice de NSE y GMU, a pesar de que la información geográfica no corresponde a nuestras áreas de estudio, si no que las áreas de estudio son compuestas por pedazos de otras áreas para las que sí existe información.

⁶ Realizar esto tiene sentido porque que las variables de entrada utilizadas en el cálculo de ambos índices son promedios a nivel AGEB. Por otro lado, el supuesto de distribución uniforme de la población de las AGEBS no es ideal, aunque las limitaciones en los datos disponibles nos obligan a recurrir a éste como mejor opción.



3. Metodología para estimar otras variables alrededor de las estaciones de transporte público

El cálculo de las variables alrededor de las estaciones de transporte público se dividió en 4 etapas:

Etapa 1.

Consistió en la elaboración de un mapa base de la Zona Metropolitana del Valle de México a nivel AGEB, con datos georreferenciados de INEGI. Para articular este mapa se utilizaron las bases de datos correspondientes al Distrito Federal, Estado de México e Hidalgo del Sistema para la Consulta de Información Censal (SCINCE) 2010; generándose como producto final una base cartográfica con su respectiva tabla de atributos. Tabla que contiene las variables población total, población con discapacidad, población económicamente activa ocupada y desocupada, viviendas totales, con disponibilidad de automóvil y viviendas desocupadas. La finalidad de tener este mapa base es poder extraer las variables requeridas por radio de influencia en estaciones de transporte público (500, 800, 1000, 2000 mts).

Etapa 2.

Para poder establecer los radios de influencia de la red de transporte público, primero se unieron las capas de estaciones para cada sistema de transporte público (sistema de transporte colectivo metro, metrobús, tren suburbano, mexibús y servicios de transportes eléctricos), donde cada punto representa una estación de transporte o bien la ubicación de un parabús. En la capa resultante se encuentran todas las estaciones disgregadas por sistema de transporte público y están representadas por puntos georreferenciados que cuentan con un número identificador para evitar traslapes, debido a que algunas estaciones de distintos sistemas tienen el mismo nombre. Los radios de influencia para cada una de las estaciones, se calcularon con el geoproceso *buffer* de ESRI⁷ a 500, 800, 1000 y 2000 metros; obteniendo como producto una capa formada por círculos cuyo centroide correspondía a cada identificador (ID).

Etapa 3.

En este proceso se intersectó cada capa de radios de influencia, con el mapa base de la Etapa 1, y para poder seleccionar las AGEBs que corresponden con cada círculo y extraer su información. Dado que los círculos que se generaron al calcular los radios de influencia tienen una circunferencia que segmentaba algunos AGEBs, se hizo un proceso para obtener la proporción de cada una dentro de los diferentes radios. El proceso consistió en calcular la superficie que se encontraba dentro de cada circunferencia y dividirla entre la superficie original de cada AGEB, obteniendo así una relación donde el número máximo era 1 y significaba que se debía tomar el total de cada variable. Para las que tenían resultado distinto a 1 se multiplicó este número por cada variable para tener una estimación de cada una.

⁷ Compañía estadounidense especializada en el desarrollo de software para sistemas de información geográfica



Etapa 4.

Para obtener los totales de las variable respectivas a cada radio de influencia, se hizo la sumatoria de los resultados de la etapa anterior, de acuerdo a cada uno de estos radios. Esto mediante el geoproceso *dissolve* de ESRI y en base al ID de cada estación; el cual se realizó para cada una de las capas de radios de influencia. Por último, para calcular los porcentajes de cada variable se dividió el resultado por variable de la Etapa 3, entre el total por unidad espacial, es decir por AGEB. Para las variables demográficas se dividió la población ocupada, desocupada y con discapacidad entre la población total. Para los indicadores de vivienda se dividieron las viviendas desocupadas y las viviendas con disponibilidad de automóvil entre el total de viviendas en cada AGEB.

La variable población por hectárea representa la densidad neta, es decir, la población total en la superficie habitable. Para obtener esta superficie se extrajeron del mapa base áreas verdes y las destinadas a grandes equipamientos como el Aeropuerto, el Bosque de Chapultepec y la Central de Abastos, y se calculó el área restante de cada AGEB en hectáreas. La población a dividir entre esta superficie fue la resultante del cálculo población total de la Etapa 3. Una vez obtenida la densidad por AGEB, se hizo un promedio de los resultados para cada radio de influencia.

4. Acrónimos y abreviaturas

AGEB: Área Geoestadística Básica.

AMAI: Asociación Mexicana de Agencias de Inteligencia de Mercado y Opinión.

CONAPO: Consejo Nacional de Población.

ESRI: Environmental Systems Research Institute.

GMU: Grado de Marginación Urbana.

INEGI: Instituto Nacional de Estadística y Geografía.

ITDP: Instituto de Políticas para el Transporte y el Desarrollo.

NSE: Nivel Socioeconómico.

SCINCE: Sistema para la Consulta de Información Censal.

ZMVM: Zona Metropolitana del Valle de México.

5. Bibliografía

- ARGÜELLES K, y GARCÍA L. HARO, Juan José.- 2013. Construcción de nivel socioeconómico predominante por AGEB mediante análisis multivariado. Disponible en: <http://blog.amai.org/index.php/construccion-de-nivel-socioeconomico-predominante-por-ageb-mediante-analisis-multivariado/> [Fecha de acceso: 6 de febrero, 2013].
- JOLLIFFE, I.T., 2002. Principal Component Analysis, second edition, New York: Springer-Verlag New York, Inc.

