

Análisis de las Zonas de una Población basada en Datos Censales

¹M^a Beatriz Bernábe Loranca, ²Maria del Rocio Boone Rojas,

³Ramon Aguirre Vara, ⁴Ramiro López Sales

Aplicaciones Estadísticas.

^{1,2,3}*Facultad de Ciencias de la Computación, BUAP.*

Calle 14 sur y Avenida San Claudio, Colonia San Manuel, Puebla, México.

E-mail[bety, rboone, rav]@cs.buap.mx_ Fax 015 222 2295500 ext. 7225

⁴*Secretaría de Transportes y Vialidad del Distrito Federal*

Álvaro Obregón 269, 9º. Piso, Colonia Roma CP 06700. Tel (55) 5209 9913 ext. 1141

PlaneacionSTV@aol.com

RESUMEN

Con el fin de simplificar el análisis de las zonas de la población para la Zona Metropolitana del Valle de México (ZMVM), en este trabajo se parte de los resultados de clasificación obtenidos en [4] y lograr un número mas pequeño de clusters, para ello se ha planteado una función Z.

Dentro de las 4,925 áreas geoestadísticas básicas (AGEB) que componen la ZMVM,, los grupos de zonas que han sido identificadas por Z, ayudan a facilitar el análisis logrando así, una descripción aproximada de la relación que guardan los promedios de las variables dentro de cada categoría.

I INTRODUCCION

Cuando se tiene un gran conjunto de objetos a los que se desea agrupar en clases, sin una clasificación previa, nos enfrentamos a varios retos. Por ejemplo, primero se deben aplicar mecanismos de selección variables, reducir la dimensión del problema e interpretación de los resultados y posteriormente dar lugar a la categorización. Es justo lo que en este trabajo se persigue, más un número de clusters adecuado que nos permita avanzar hacia el objetivo inicial de este proyecto [1,9], el cual es obtener propiedades de grupos de zonas que puedan relacionarse con las características de la movilidad de los habitantes (representadas por los datos del censo más reciente) y así seguir acumulando resultados que ayuden a describir propuestas para resolver el problema de la distribución equilibrada del transporte.

II CLASIFICACIÓN

Después de la aplicación de técnicas multivariadas a los datos censales del año 2000 [3] para ZMVM, sabemos que existe una clara correlación entre el tipo de vivienda contra la relación de dependencia y grado de escolaridad, y, en combinación con la descripción de los 16 grupos de zonas en donde cada grupo se distingue por un color específico

(Fig. 1), se concluyó que a los clusters que contienen las zonas de la periferia de la ciudad, le corresponden las variables z109 (población empleada como trabajador inexperto), z136 (casas privadas con el alcantarillado conectado con la red pública) z141 (las casas con agua entubada en el predio) z142 (las casas privadas con agua entubada en el predio y conexión pública) y z161 (casas privadas con comodidades domésticas) [2,3,4].

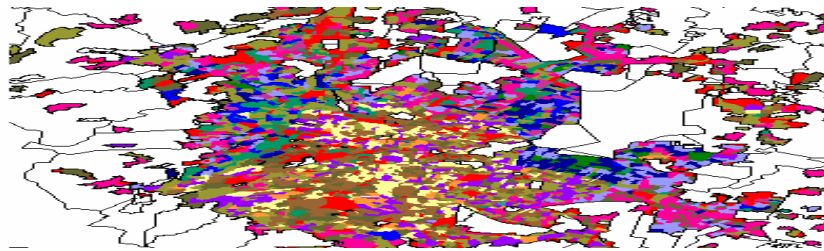


Figura 1. 16 Clusters para ZMVM

Dado que el número de grupos hasta este punto se eligió de forma experimental, fue necesario recurrir a un medio más confiable para obtener un número exacto de clusters. Así, se planteó una función que proporcionara el número más adecuado de conglomerados, menor a los números de conglomerados obtenidos en la fase experimental, para simplificar el análisis de las zonas de la población.

Para encontrar la función, primero se crearon clasificaciones con un número de conglomerados que variara entre 8 y 70. Los centroides iniciales en cada caso fueron elegidos aleatoriamente. La gráfica en la figura 8 muestra el promedio de la suma de los cuadrados, ponderado de acuerdo al número de elementos de cada conglomerado.

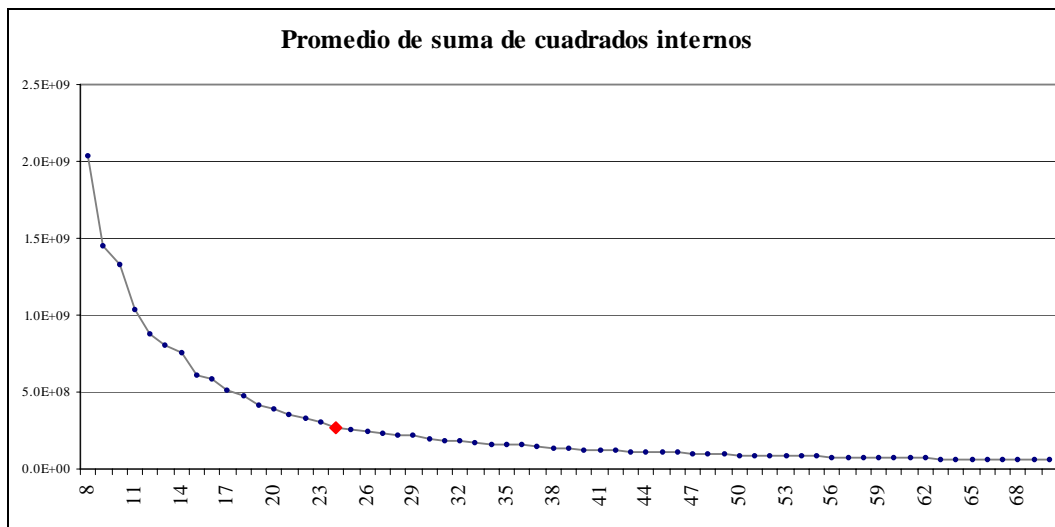


Figura 2 Promedio de suma de cuadrados intra-grupo para diferentes números de conglomerados.

Naturalmente, la suma de cuadrados dentro de un conglomerado disminuye conforme aumenta el número de conglomerados, ya que hay menos elementos dentro de cada grupo. Sin embargo, puede observarse que la *variación* del promedio de la suma de cuadrados entre tamaños sucesivos es mayor entre los tamaños pequeños.

Ya que el objetivo de este trabajo es obtener un número reducido de conglomerados que simplifiquen el análisis de las zonas poblacionales, se decidió adoptar la clasificación de 24 conglomerados ya que es la que más se aproxima a un punto de inflexión en la tabulación de la suma de los cuadrados internos.

La Tabla 2 muestra el número de zonas que corresponde a cada conglomerado.

Después de un análisis de esta tabla, concluimos que un rasgo común de todos los conglomerados es que aquellas variables que representan relaciones entre la población (z048, z049, z083, z163 y z164) tienen promedios muy cercanos al general en todos los conglomerados. Por esta razón, al hacer referencia a los promedios de las variables dentro de cada conglomerado se exceptúa a estas cinco.

11	Arriba	ninguno
3	Arriba	z128, z141, z142
6	Arriba	z136
8	Arriba	z136
12	Arriba	z128, z141, z142
9	Arriba	z109, z141, z142
2	Arriba	z136
4	Arriba	z136, z161
1	Divergente	z141, z161
15	Divergente	z109, z82, z118, z161
16	Arriba	z136, z161
13	Abajo	z136
14	Arriba	z160, z161
21	Arriba	z82, z118, z160, z161
7	Arriba	z109, z136, z142
5	Abajo	z142
22	Abajo	ninguna, z109, z141 muy abajo
18	Promedio	z109, z142, z136
19	Abajo	z109, z142
23	Abajo	z128, z136, z141

Tabla 2 Correspondencias entre Zonas y Conglomerados

Los mapas siguientes muestran los cuatro conglomerados con el mayor número de zonas.

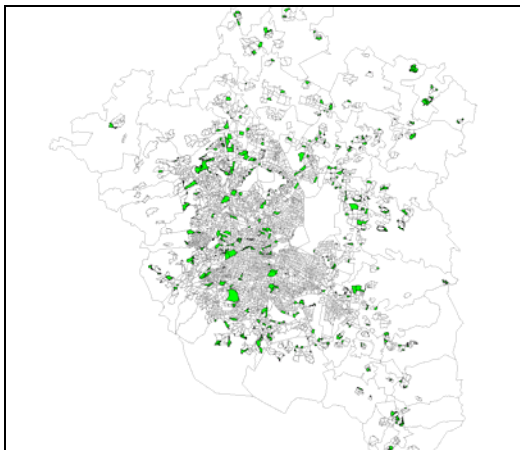


Figura 3 Cluster 10

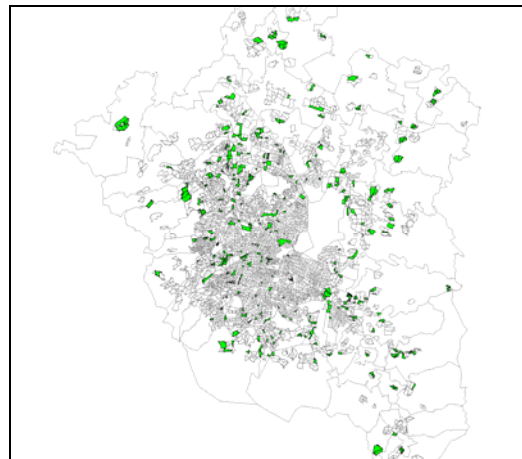


Figura 10 Cluster 20

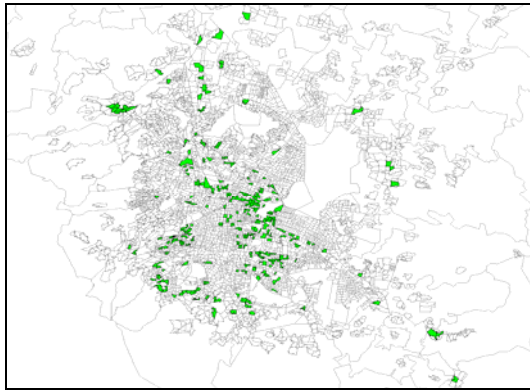


Figura 5 Cluster 17

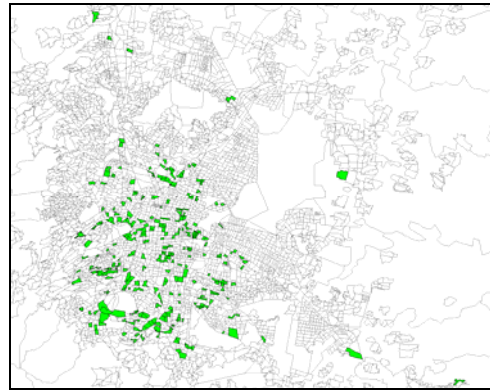


Figura 6 Cluster 24

Las zonas del cluster 10 (Fig. 3), se caracterizan por tener los promedios más bajos en sus variables. Este grupo se identifica claramente con las zonas deshabitadas o de muy escasa población. Como puede verse en el mapa, las zonas de este grupo se localizan por toda el área de estudio. Por el lado del cluster 20 (Fig. 4), las zonas de este grupo también tienen poca población, pero los promedios de sus variables están consistentemente arriba de los del grupo 10. La magnitud de cada promedio es como mínimo el doble del promedio del grupo 10, y en algunos mayor al triple. En este grupo se concentran zonas de escasa población.

En el cluster 17 que se muestra en la Fig. 5, en contraste presenta valores promedio muy cercanos al global, y se ubica principalmente en el centro de la ciudad. Los únicos promedios que se desvían de este comportamiento son los que corresponden a las variables z45, z54, z136 (por arriba) y z55, z109, z142 (por abajo), sin embargo en ningún caso la diferencia es significativa. Estas zonas presentan el perfil típico de la zona urbana consolidada.

Finalmente el Cluster 24 de la Fig. 6 presenta también una tendencia hacia el promedio global. Este cluster parece ser el más relevante de acuerdo a la figura 8 y se retomará el análisis que le corresponde aunado con los resultados de las secciones anteriores para presentar conclusiones finales.

4) CONCLUSIONES

El Cluster 24 corresponde al número mas adecuado de grupos, y se ha tomado como eje para presentar una descripción global.

En este Cluster pueden distinguirse tres tendencias principales:

Las variables están por debajo del promedio en su mayoría, mientras que las variables z109, z141 y z142 tienen promedios notoriamente inferiores, lo cual indica que en promedio estas zonas tienen un menor número de habitantes trabajando como jornalero o peón y que hay un menor número de casas con agua en el predio (en vez de la casa) o por acarreo de llave pública.

Las variables z82, z118, z136, z160 y z162 presentan promedios mayores que implican mayores niveles de ingreso y acceso a servicios.

De las tres características anteriores se deduce fácilmente que este grupo representa a zonas cuya población en conjunto presenta mayores niveles de ingresos y el resto de los grupos presentan el siguiente comportamiento:

Ocurre que la mayoría de los promedios están ya sea por arriba o por abajo del global, y unas cuantas variables resaltan en la dirección contraria, así los grupos presentan el siguiente comportamiento general (por orden creciente de tamaño): En el caso de que los promedios que no presenten una tendencia general (Divergente, Promedio) se listan los valores extremos en ambas direcciones. Los casos especiales son el grupo 22, que no presenta promedios en la “dirección opuesta”, que sin embargo tiene valores muy bajos para un par de variables, y el grupo 11 que no agrupa las zonas con más alto promedio en la mayoría de sus variables.

El análisis de promedios en los grupos obtenidos muestra que la mayoría de los grupos son claramente distinguibles, y además resalta un conjunto de variables que aparecen como rasgos relevantes en cuanto que sus promedios presentan un comportamiento opuesto a la mayoría de cada grupo.

Las variables que resaltan son z109 (Población ocupada como jornalero o peón) z136 (Viviendas particulares con drenaje conectado a la red pública) z141 (Viviendas particulares con agua entubada en el predio), z142 (Viviendas particulares con agua entubada por acarreo) y z161 (Viviendas particulares con todos los bienes). Estos resultados coinciden con la mayoría de los que se obtuvieron con el ACP de la sección 2.

Finalmente podemos decir que la aplicación del análisis de conglomerados en este caso permitió identificar importantes relaciones entre las variables utilizadas. También permite ubicar los grupos más representativos de las zonas de la ciudad.

Utilidad de los resultados

La clasificación obtenida permite distinguir claramente 24 categorías de zonas dentro de las 4,925 áreas geoestadísticas básicas (AGEB) que componen la ZMVM, basándose para ello en la relación que guardan los promedios de las variables dentro de cada categoría. Una consecuencia importante de obtener una clasificación que sea representativa de la interrelación de las variables es que pueden conducirse diversos análisis considerando a las zonas dentro de una categoría como una sola entidad; así, es más sencillo trabajar con grupos de zonas que con AGEBS individuales.

No perdiendo de vista el objetivo de este trabajo, los resultados aquí obtenidos no solo son aplicables a diversas áreas de estudios urbanos, también han aportado una descripción de zonas mas detallada que en [4]. Por otro lado, estos resultados sientan las bases hacia un análisis exhaustivo sobre los problemas de la distribución del transporte con respecto a las actividades de los habitantes. Finalmente, se desea proponer un modelo que permita explicar la generación y atracción de viajes los habitantes de la ZMVM, para ello, recurriendo a expertos en planeación y

transporte, se analizará la relación que guardan las zonas de cada categoría con el comportamiento de los viajes que se originan o se dirigen hacia ellas.

Referencias

- [1] Anderson, T.W, 1984, "An Introduction to Multivariate Statistical Analysis", 2nd Edition. Wiley.
- [2] Bernábe L.M.B, Aguirre V.R, & López S. R.; "Utilidad del Componente Principal para Caracterizar Rasgos Poblacionales" 2do.Congreso Latinoamericano de Matemáticos (UMALCA), Cancun, México, junio 2004.
- [3] Bernábe L.M.B, Aguirre V.R, & López S. R.; "Aplicación Del Análisis Factorial Sobre Datos Poblacionales" 2º Congreso Internacional sobre Innovación y Desarrollo Tecnológico (CIINDET 2004), Cuernavaca Morelos (noviembre 2004).
- [4] Bernábe L.M.B, Boone Rocio, Aguirre V.R,; "Análisis de Conglomerados en Problemas de Zonificación" 2º Congreso Nacional de Ciencias de la Computación, FCC, BUAP, Puebla, México, (noviembre 2004).
- [6]. Dixon, W.J, 1990, "BMDP Statistical Software Manual", Vol I, II. Dixon, W.J Eds, University of California Press, Berkeley, California.
- [7] Kachigan, S. K. "Multivariate Statistical Analysis. A Conceptual Introduction." Radius Press, 1991.
- [8] Loranca, M.B. & Olsina, L WSEAS International Conferences, ASCOM'04 Cancun, Mexico, May 12-15, 2004. (6th WSEAS International Conference on Algorithms, Scientific Computing, Modelling and Simulation). "Classification of Poputalion Zones Usin Multivariate Statitistical Techniques"; Cancun, México, junio 2004.
- [9]Loranca, M.B. & López, S.R; Aguirre V; "Application of Non-Supervised Classification to Population Data" ICEEE/CIE2004, International Conference on Electrical and Electronics Engineering, Acapulco Mexico, septiembre 2004
- [10] J. Macdonald and J. Braun (2003), "Data Analysis and Graphics Using R: An Example-Based Approach", Cambridge University Press, ISBN 0-521-81336-0
- [11]-<http://www.r-project.org> "Statistical Software Manual"
- [12]<http://www.cs.buap.mx/~bety/Investigacion.html>
- [13] XII Censo de Población y Vivienda. Resultados definitivos. INEGI. México 2001