

El análisis de conglomerados en la determinación de regiones homogéneas con base a la producción agrícola de maíz y sorgo en la región zuliana

Cluster analysis in the determination of homogeneous regions based on the production of maize (corn) and sorghum in the zuliana region

A. Apitz de Parra¹, M. J. Escalona F.², F. Taborda R.³

Resumen

En este estudio se presenta la delimitación de regiones geográficas usando la producción de rubros agrícolas. Para lograr esta clasificación se utilizó la técnica del análisis de conglomerados con el fin de obtener localizaciones geográficas homogéneas en la Región Zuliana. La regionalización obtenida permitió distinguir dos regiones, una eminentemente agrícola (municipio Baralt) y otra de escasa o mediana producción agrícola constituida por el resto de los municipios del estado-región. Las variables agrícolas usadas para conformar las regiones y la técnica resultaron operativas; además, permitieron determinar zonificaciones que deben ser consideradas periódicamente cuando se hacen estudios de planificación agrícola para necesidades de consumo alimentario de las poblaciones.

Palabras clave: Conglomerados, regiones, homogéneas, planificación, agrícola.

Abstract

This study presents the delimitation of geographic regions based on agricultural production. To achieve this classification, cluster analysis of conglomerates was employed in order to determine homogeneous geographic localities in the Zuliana region. The regionalization obtained allowed us to distinguish two regions, one eminently agricultural (Baralt Municipality) and another of little or medium agricultural production constituted by the rest of the municipalities in the state-region. The agricultural variables used to determine the regions, and the operational technique were functional; and furthermore permitted the determination of the zoning which should be periodically considered when making studies for agricultural planning and the consumption of agricultural produce by the population.

Key words: Clusters, regions, homogeneity, planning, agriculture.

Recibido el 06-06-1997 • Aceptado el 02-09-1997

1. Centro de Estudios Geográficos, Facultad de Humanidades y Educación. La Universidad del Zulia.

2. Centro de Estudios Matemáticos, Facultad de Humanidades y Educación. La Universidad del Zulia, Apartado 526, Maracaibo, Venezuela. E-mail: CEMAT@luz.ve

3. Facultad de Agronomía, La Universidad del Zulia.

Introducción

Tradicionalmente, el paisaje natural ha sido subdividido en unidades espaciales definidas por la homogeneidad de factores físicos, como la topografía, suelos y vegetación (2). Sin embargo, resulta de gran importancia dividir el espacio en unidades homogéneas con otro tipo de variables y técnicas de agrupamiento. La técnica de análisis de conglomerados puede ser de gran utilidad en estudios de diferentes escalas (local, regional, nacional) que consideren otros tipos de variables.

Bricker *et al.* (2) usaron las técnicas de análisis multivariado para resolver problemas de identificación y clasificación de comunicaciones en los Estados Unidos. Los citados autores trabajaron con formulaciones estadísticas y análisis asociados con el problema de identificación de personas utilizando como base la energía espectral representada por declaraciones acústicas. En ese trabajo se usaron dos cuerpos de datos, repartidos por repeticiones de declaraciones de palabras sueltas. El primer conjunto de datos estuvo formado por diez hablantes, cada uno de los cuales produjeron varias repeticiones de diez palabras corrientes usadas en conversaciones telefónicas. El segundo cuerpo de datos se obtuvo por acuerdo con el primer grupo de hablantes, éstos últimos constituyeron una población de 172 hablantes quienes repitieron los nombres de los cinco primeros dígitos (uno, dos, tres, cuatro y cinco) cinco veces cada uno.

Chen *et al.* (5) utilizaron las técnicas de conglomerados para

descubrir regiones homogéneas basándose en la variable cantidad y tipos de compañías industriales. Ese trabajo estuvo limitado a cuatro grandes grupos industriales - químicos, drogas, aceites domésticos y acero- especificado por sus estándares en los Estados Unidos. El agrupamiento se realizó basándose en la estructura y comportamiento de las firmas como miembros de grupos homogéneos para el desarrollo de ayudas para las mejores decisiones financieras concernientes a las operaciones de las firmas.

Lopes y Buarque de Lima (11) aplicaron el método de análisis de conglomerados usando la distancia (euclídea) para clasificar 50 grandes ciudades de Brasil, utilizando como variables el nivel de desarrollo y el tamaño funcional de las urbes.

Casp y Bernabeu (3) realizaron el análisis de las características sensoriales de 16 vinos tintos de la denominación de origen Valencia Uno, mediante el uso de técnicas estadísticas multivariadas.

No es la intención de este trabajo profundizar en la teoría que sustenta el método de los conglomerados, asunto que puede consultarse en trabajos especializados (6, 8, 10); sin embargo, se presentan las consideraciones básicas que permiten comprender sus técnicas al aplicarlo a un estudio de división del espacio geográfico en unidades homogéneas contiguas con base a rubros agrícolas. Este trabajo se realizó con el propósito de realizar una delimitación de regiones usando la producción de rubros agrícolas.

Materiales y métodos

En este trabajo se aplica el análisis de conglomerados para determinar regiones homogéneas basándose en la producción de maíz y sorgo, para el año 1990, en la región Zuliana. El estudio inicial contempló otras variables agrícolas, para el período 85-89. Sin embargo, para la elaboración de este trabajo se revisaron, nuevamente, las estadísticas en el Ministerio de Agricultura y Cría de la región zuliana; es por esto último que una limitación en las estadísticas agrícolas obtenidas para realizar este trabajo debe ser señalada. En efecto, las fuentes oficiales consultadas (12) sólo registran información detallada por rubros agrícolas por municipios hasta 1991; a partir de 1992 la información es presentada de manera generalizada, esto es, sin considerar la división político territorial de la región.

Los resultados obtenidos por cualquiera de los métodos de agrupamiento, sea por agrupamientos sucesivos o por series de divisiones, pueden ser presentados en forma de un diagrama bi-dimensional conocido como dendograma o diagrama de árbol (9).

Dentro de los métodos de agrupamiento se consideran en particular, los denominados métodos de enlace (*linkage methods*). Estos métodos de enlace son adecuados para agrupar tanto registros como variables, proceso que no se da con otros métodos de agrupamiento. Los criterios más conocidos de enlace son: el de vecino más cercano o mínima distancia (*single linkage*), el de vecino más

alejado o máxima distancia (*complete linkage*) y el de la distancia media (*average linkage*) (10).

Para eliminar la variabilidad en el conjunto de puntos u objetos iniciales se ponderan las distancias obtenidas. Estas distancias permiten agrupar los casos según criterios ya señalados de vecino más próximo, vecino más lejano, y distancia media.

En el primer paso del proceso de enlace se determina la matriz de las distancias y a partir de este punto se inicia un proceso que culmina con la ubicación de los distintos puntos en grupos homogéneos dentro de sí y heterogéneos entre sí. La secuencia de trabajo definido como un proceso iterativo consiste en ubicar la distancia según un criterio fijado inicialmente. Con la distancia se forma un primer grupo, el cual puede tener dos o más puntos o individuos o casos (6,9,10). La formalización matemática de este proceso se traduce en la sustitución de las filas y columnas respectivas por la fila y columna que pasan a representar al grupo, obteniéndose así una nueva matriz de distancias. Este último paso se repite hasta recomponer por similitudes todos los puntos de la matriz de datos originados (4, 6, 9, 10).

En este trabajo el criterio usado para determinar la cantidad de conglomerados, por escoger, fue el de la distancia mínima. La figura 1 contiene un diagrama de flujo de los procesos. Este algoritmo requiere los siguientes pasos:

Obtención de la matriz de datos. El caso en estudio son dos variables económicas del sector agrícola

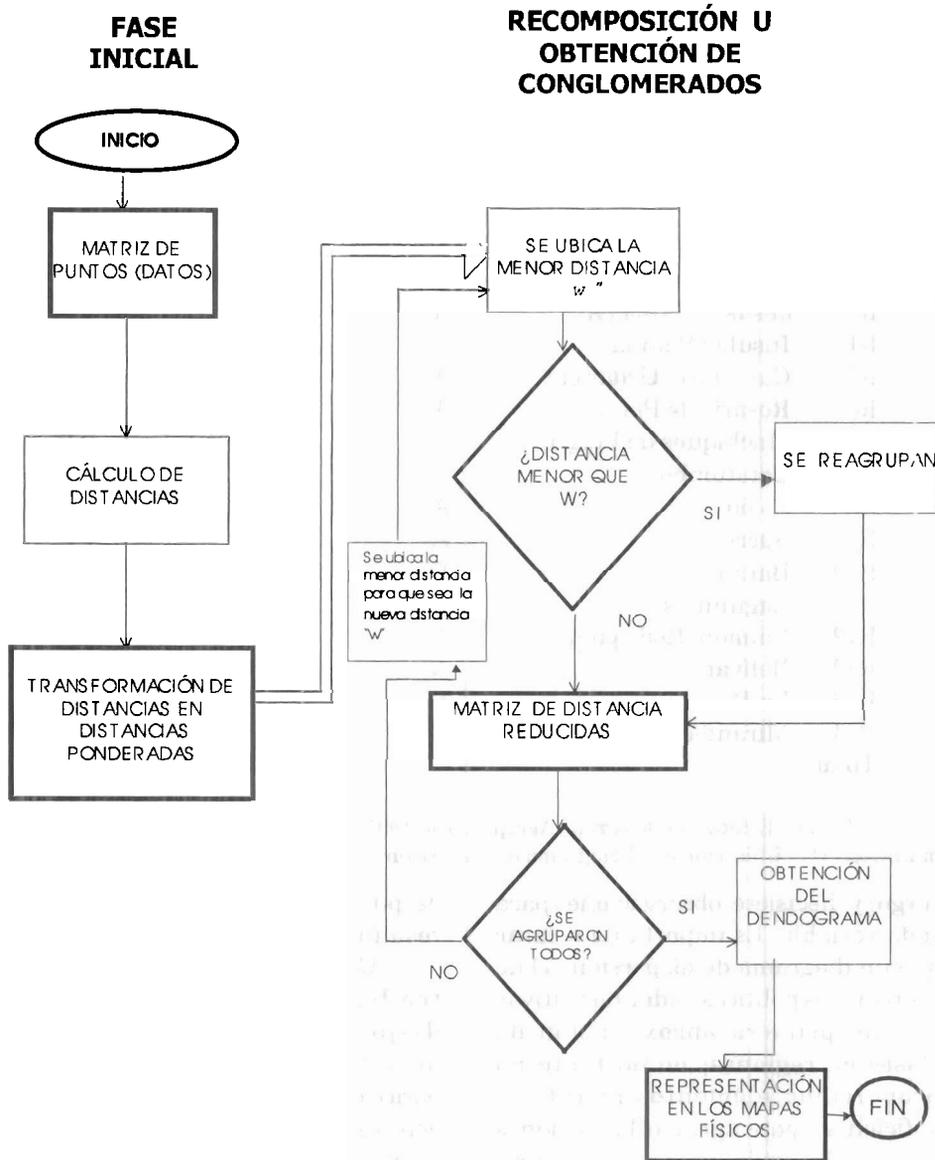


Figura 1. Diagrama de flujo de los procesos.

del estado Zulia. Se trabajó con 17 registros de estas variables; porque sólo se obtuvo información sobre maíz y sorgo por municipio, los cuales constituían el modelo de división

político - administrativa vigente para la época (cuadro 1). De ahí que, el arreglo de los datos es una matriz de 2x17, es decir, dos variables (producción de maíz, producción de

Cuadro 1. Producción de maíz y sorgo en el estado Zulia, por municipio 1990.

Municipio		Variables		
(+)		(*)	Maíz (t)	Sorgo (t)
R1	Páez	B	175,0	0,0
R2	Mara	A	0,0	0,0
R3	Maracaibo	D	0,0	973,72
R4	Jesús E, Lossada	A	0,0	0,0
R17	Insular Padilla	A	0,0	0,0
R5	Cañada de Urdaneta	A	0,0	0,0
R6	Rosario de Perija	A	0,0	0,0
R15	Machiques de Perija	B	165,00	0,0
R7	Catatumbo	A	0,0	0,0
R8	Colón	A	0,0	0,0
R9	Sucre	A	0,0	0,0
R10	Baralt	F	2163,4	1358,0
R11	Lagunillas	A	0,0	0,0
R12	Valmore Rodríguez	A	0,0	0,0
R13	Bolívar	A	0,0	0,0
R16	La Rita	E	281,6	941,6
R14	Miranda	C	645,0	0,0
Total			3430,0	3273,32

Fuente: Unidad Estatal de Desarrollo Agropecuario, 1991, MAC - Zulia. (+) Ubicación de cada municipio. (*) Ubicación en el diagrama de dispersión. (Figura 2).

sorgo) y diecisiete observaciones para cada variable. Es importante aclarar que un diagrama de dispersión (7) de las regiones políticas - administrativas da una primera aproximación de clases de regiones; no obstante no proporciona elementos analíticos suficientes para agrupar las regiones más próximas visualmente, con otras que se muestran en el gráfico relativamente distantes (figura 2). La base de datos inicial se transforma en una matriz de distancias donde cada componente de la matriz constituye la distancia entre dos puntos o regiones político - administrativa las cuales,

después de aplicar la técnica, pueden resultar separadas o agrupadas.

Cálculo de las distancias entre los puntos del diagrama de dispersión. Existen diversas fórmulas matemáticas y estadísticas para calcular estas distancias (4). Para este estudio el punto en el sentido geométrico - euclideo no diverge de los atributos de cada localización; porque aceptamos que ambas están contenidas en un solo concepto (11).

Recomposición de los puntos o construcción de conglomerados. En este proceso de recomposición se ubican los puntos que se encuentran a

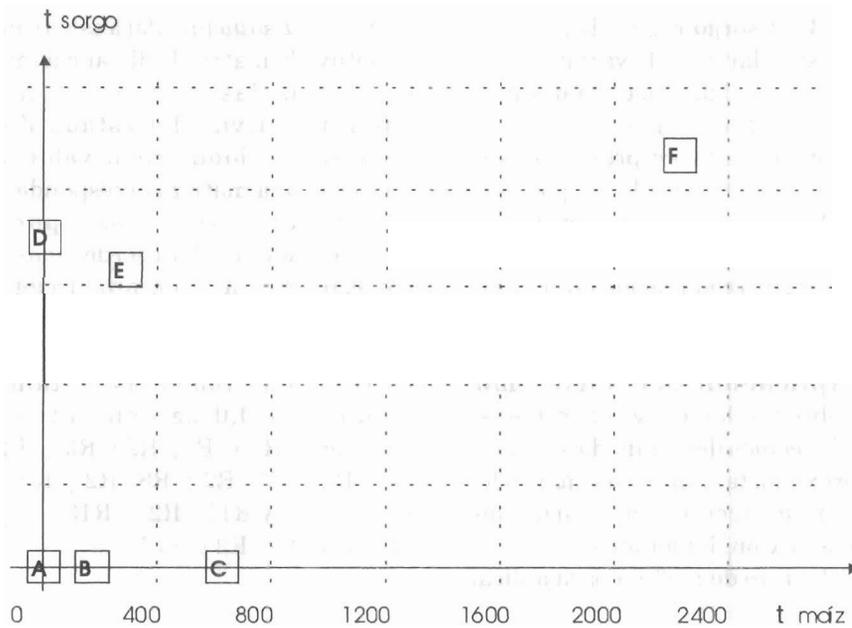


Figura 2. Diagrama de dispersión. Producción maíz y sorgo (Estado Zulia).

una distancia menor de la distancia mínima prefijada al nuevo grupo; enseguida se procede a fijar la fila y columna respectiva que reemplazará las existentes. En otras palabras, una región sustituye a las regiones que le son vecinas; porque aceptamos que la localización de un punto y sus atributos son lo mismo. El criterio para agrupar los vecinos, en este caso de estudio, se denomina el vecino más próximo (4, 6, 9, 10). La distancia entre un conglomerado C1 el cual agrupa tres objetos, esto es A (ABC) y un objeto W se calcula mediante $d_{C1W} = \min \{d_{AW}, d_{BW}, d_{CW}\}$, donde d_{AW} , d_{BW} y d_{CW} son distancias próximas de los objetos A y W, los agrupamientos de los objetos B y W, y los agrupamientos C y W,

respectivamente. Este paso se repite hasta que se agrupan los dos últimos conglomerados.

Construcción del dendograma. Cada conglomerado que se forma después de recomponer la matriz de distancia en cada iteración o repetición del paso anterior se representa por un enlace hasta esa distancia.

Como ilustración de este arreglo inicial se tiene el diagrama de puntos (figura 2), en el cual se aprecian los puntos entre los cuales se obtienen las distancias. Para este cálculo se utilizó la definición de distancia euclídea ponderada según: $d^2(A, B) = [(x_{A1} - x_{B1})/s_1]^2 + [(x_{A2} - x_{B2})/s_2]^2$ F1. donde: A es un punto con coordenadas $(x_{A1}, x_{A2}) = (t \text{ maíz región A; } t \text{ sorgo región A})$ y B tiene

coordenadas $(x_{B1}, x_{B2}) = (t \text{ maíz región B; } t \text{ sorgo región B})$; s_1 es el desvío estándar para la variable X_1 (t maíz) y s_2 es el desvío estándar para la variable X_2 (t sorgo).

Para los puntos representados en la figura 2, utilizando el paquete STATG (13), se producen las matrices de las distancias, las cuales fueron obtenidas para cada región en relación con las restantes usando la fórmula F_1 , (10).

Aplicación del algoritmo.

Para obtener los conglomerados se aplicó la técnica denominada el vecino más próximo, tal como se señaló en los aspectos básicos de la Teoría del Análisis de Conglomerados.

El punto de partida para aplicar

el algoritmo es la obtención de la *matriz de distancias*. Para este trabajo se obtuvo la matriz de distancia inicial con todas las regiones político administrativas del estado Zulia (cuadro 2), donde cada valor que aparece en la matriz corresponde a la distancia entre dos puntos representados por las coordenadas que corresponden a la de producción de sorgo y maíz de cada región. Observando la matriz de las distancias se tiene que todas las distancias menores a 0,0 se dan entre las regiones: R2 y R4; R2 y R5; R2 y R6; R2 y R7; R2 y R8; R2 y R9; R2 y R11; R2 y R12; R2 y R13, y por último entre R2 y R17.

Resultados y discusión

La primera matriz de reducción de distancias se obtiene de la sustitución o reemplazo de las regiones R4, R5, R6, R7, R8, R9, R11, R12, R13 y R17 por la región R2. El resultado es una nueva matriz de distancias formada por R1, C1, R3, R10, R14, R15, R16, y R17, donde C1 es la recomposición u obtención del conglomerado formado por los puntos R2, R4, R5, R6, R7, R8, R9, R11, R12, R13 y R17 (cuadro 3). En este cuadro, también, se muestra cual es el próximo agrupamiento, el mismo está indicado para los puntos o regiones R1 y R15. Esto se debe a que la distancia 0,0187 unidades es la menor entre todas las mostradas en el cuadro 3.

La figura 3 muestra la agrupación C1 para conformar un conglomerado. Este último es el inicio

de las iteraciones que permitirán al finalizar las recomposiciones tener la forma definitiva en el dendograma (9, 10).

De acuerdo con el trabajo de recomposición se formaron dos conglomerados a una distancia menor de 4,3590 unidades. La figura 4 muestra una síntesis gráfica de las nuevas agrupaciones ejecutadas.

La aplicación del método permite poner en evidencia con precisión estadística la existencia de dos regiones con características bien definidas, dedicadas a la producción de sorgo y maíz. Dadas las características estáticas de las variables consideradas, las inferencias obtenidas fueron enteramente descriptivas.

Según los resultados obtenidos aplicando la técnica de análisis de

Cuadro 2. Matriz de distancias con los puntos iniciales.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17
R1	0,0																
R2	0,328	0,0															
R3	2,255	2,23	0,0														
R4	0,328	0,0	2,231	0,0													
R5	0,328	0,0	2,231	0,0	0,0												
R6	0,328	0,0	2,231	0,0	0,0	0,0											
R7	0,328	0,0	2,231	0,0	0,0	0,0	0,0										
R8	0,328	0,0	2,231	0,0	0,0	0,0	0,0	0,0									
R9	0,328	0,0	2,231	0,0	0,0	0,0	0,0	0,0	0,0								
R10	4,859	5,12	4,155	5,115	5,115	5,115	5,115	5,115	5,115	0,0							
R11	0,328	0,0	2,231	0,0	0,0	0,0	0,0	0,0	0,0	5,115	0,0						
R12	0,328	0,0	2,231	0,0	0,0	0,0	0,0	0,0	0,0	5,115	0,0	0,0					
R13	0,328	0,0	2,231	0,0	0,0	0,0	0,0	0,0	0,0	5,115	0,0	0,0	0,0				
R14	0,882	1,21	2,538	1,211	1,21	1,211	1,211	1,211	1,211	4,219	1,211	1,21	1,21	0,0			
R15	0,018	0,31	2,252	0,309	0,31	0,309	0,309	0,309	0,309	4,873	0,309	0,31	0,31	0,901	0,0		
R16	2,165	2,22	0,534	2,219	2,22	2,219	2,219	2,219	2,219	3,658	2,219	2,22	2,22	2,26	2,16	0,0	
R17	0,328	0,0	2,231	0,0	0,0	0,0	0,0	0,0	0,0	5,115	0,0	0,0	0,0	1,211	3,09	2,22	0,0

Cuadro 3. Primera matriz de reducción de distancias.

	R1	C1	R3	R10	R14	R15	R16
R1	0,0						
C1	0,3285	0,0					
R3	2,2550	2,2310	0,0				
R10	4,8590	5,1156	4,1550	0,0			
R14	0,8823	1,2108	2,5384	4,2193	0,0		
R15	0,0187	0,3097	2,2524	4,8734	0,9011	0,0	
R16	2,1653	2,2192	0,5339	3,6589	2,2614	2,1671	0,0

conglomerados para determinar regiones homogéneas con base a la producción de maíz y sorgo en la región zuliana (1990), se formaron dos grupos de zonas si la distancia a considerar es superior a 4,3590 (figura 4) y tres grupos de regiones o zonas si la distancia es mayor que 2,2550 y menor que 4,3590 (figura 4).

Con respecto a la formación de tres grupos se obtuvo una estructura de nuevo ordenamiento, cuya característica en cada grupo es la siguiente:

1) El agrupamiento C5 unió las regiones con una producción muy baja en ambos rubros de menos de 300 t y un caso para el cual la producción de ambos rubros fue de menos de 900 t.

2) El agrupamiento C4 cuenta con una producción que podría definirse como intermedia entre ambos grupos; la producción es entre 301 t y 950 t; en tanto que en la producción registrada en sorgo está entre 901 t y 1000 t.

3) El agrupamiento formado por la Región R10, representada por el municipio Baralt, cuya producción de maíz fue de 2.163,4 t y de sorgo de 1358 t durante el año 1990.

Para el caso de la formación de dos grupos es posible advertir la siguiente estructura de reordenación:

1) Los municipios del estado Zulia en los cuales la producción de maíz fue menor a 2.163 t y con producción de sorgo de menos de 1358

▲ DISTANCIAS



Figura 3. Dendrograma para la primera matriz de reducción de distancias.

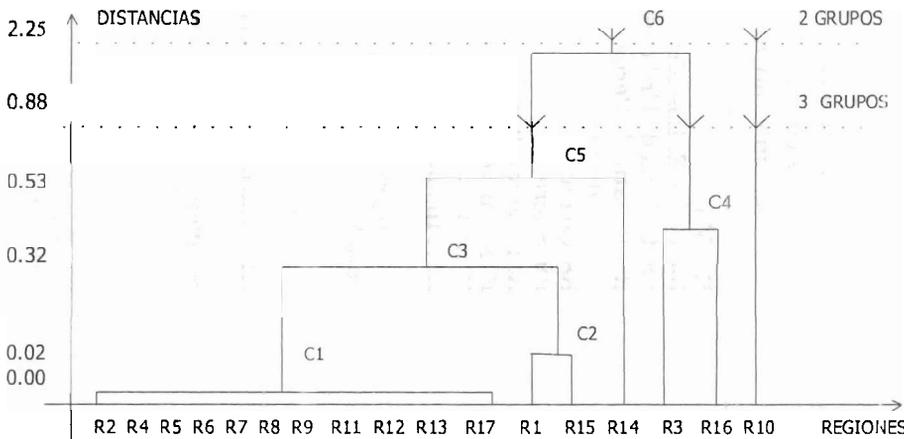


Figura 4. Dendrograma con la agrupación completa de las diecisiete regiones del Estado Zulia para los rubros agrícolas maíz y sorgo. Año 1990.

t conforman el conglomerado C6.

2) El municipio del estado Zulia con producción de maíz de 2163,4 t y producción de sorgo de 1358 t conformó el segundo grupo homogéneo o Región R10.

En general, éstos resultados muestran una regionalización homogénea de acuerdo a los niveles de producción de los rubros agrícolas considerados. El municipio Baralt presenta la mayor producción de maíz y sorgo con relación al conjunto formado por el resto de los municipios del estado Zulia. Los municipios La Rita y Maracaibo muestran una producción entre 900 t y 1000 t de sorgo. El resto de los municipios del estado Zulia registraron una producción muy baja del rubro maíz, entre 0,0 t y 650 t, en tanto que para el rubro sorgo, no hay registros de producción para el año seleccionado.

La figura 5 muestra una representación cartográfica en la cual

es posible observar las dos regiones homogéneas obtenidas de la aplicación de la técnica de análisis de conglomerados. La Región 1 representada por el municipio Baralt y la Región 2 representada por los 16 municipios restantes del estado Zulia, para el año 1990.

De la aplicación de la técnica de análisis de conglomerados se obtuvo no sólo la estructura de los grupos, sino que la misma corresponde a las características de los municipios eminentemente agrícolas como es el caso del municipio Baralt (R10), y de escasos o medianamente agrícolas, representado por el resto de los municipios.

El trabajo demuestra el poder analítico y de síntesis de la técnica de conglomerados que permitió clasificar el espacio geográfico en regiones homogéneas, con base a la producción agrícola de maíz y sorgo.

Una de las contribuciones del

Cuadro 4. Clasificación de municipio político-administrativos.

Grupo	Agrupados	observaciones del nuevo grupo (frecuencia)	Distancia mínima (similaridad)	Grado de generalización	Región	Municipio (división político-administrativa)
C1	11	11	0,00 (0,0)	0,0	1	R2: Mara R4: Jesús E. Lossada R5: Cañada de Urdaneta R6: Rosario de Perija R7: Catatumbo R8: Colón R9: Sucre R11: Lagunillas R12: Valmore Rodríguez R13: Bolívar R17: Insular Padilla
C2	2	2	0,0187 (0,01322)	0,828	1	Se integran once regiones para formar C1 R1: Páez
C3	2	13	0,3285 (0,2322)	14,55	1	R15: Machiques de Perija Se unen dos regiones para formar C2. C1 y C2 Se unen C1 con C2 para formar C3.

Cuadro 4. Continuación.

Grupo	agrupados	observaciones del nuevo grupo (frecuencia)	Distancia mínima (similaridad)	Grado de generalización	Región	Municipio (división político-administrativa)
C4	2	2	0,5339 (0,3775)	23,67	2	R3: Maracaibo R16: Cabimas
C5	2	14	0,8823 (0,2510)	15,73	1	Se forman dos grupos C3 y C4 (unión de las regiones R3 y R16) C3 R14: Miranda Se forma un grupo C5 (unión de C3 con R14) C4 y C5
C6	2	16	2,2556 (1,5949)	100	1	Se forma un grupo C6, con la unión de C4 y C5 En general, se forman dos grupos con C6 y R10
C7			4,1550 (2,988028)			Se agrupan todos con R10: BARALT

método es la obtención de regiones homogéneas del estado Zulia con base a la producción de las variables agrícolas estudiadas, la técnica aporta otras contribuciones. Así, a partir de la regionalización resultante de la aplicación de la técnica de análisis de conglomerados se pueden iniciar nuevas investigaciones. Usualmente, un trabajo de geografía concluye con una descripción del método por unidades espaciales homogéneas, en la cual la regionalización resultante se deriva de los criterios, generalmente

operativos, los cuales señalan objetivos de futuras actuaciones. De aquí, que cuando la regionalización homogénea refleja una determinada zonificación, los organismos de planificación en su búsqueda por ser cada vez más eficaces, deberán considerar la posibilidad de trabajar con regionalizaciones periódicas que permitan posteriormente hacer estudios para planificaciones agrícolas según las características de la región político-administrativa y las necesidades poblacionales.

Literatura citada

1. Bara, T., 1994. Gis-Based regionalization of natural landscape using derived landcover occurrence probabilities. p 34-43. In: Proceedings of the 1994 anual conference and exposition on GIS/LIS, Arizona, USA.
2. Bricker P.D., R. Gnanadesikan, M. V. Mathews, M. Pruzansky, P Tukey , K. Wachter and J. Warner. 1971. Statistical techniques for talk identification. J. Am. Telep. and Electr. 50 (4): 1427-1454.
3. Casp A. y A. Bernabeu. 1987. Caracterización sensorial de los vinos de la denominación de origen Valencia Uno. Vinos tintos. Rev. Agroquím. Tecnol. Aliment. 27(2): 237-245
4. Cormack A. D. 1971. A review of classification. J. Roy. Stat. Soc. Series A. 134: 321-367.
5. Chen H, R. Gnanadesikan and J. Kettenring. 1974. Statistical methods for grouping corporations. Sankhyá. The Indian Journal of Statistics 36 (serie B): 1-28.
6. Cuadras, C.M. 1991. Métodos de análisis multivariante, Editorial PPU, Barcelona
7. Fischer S, R. Dornbusch y R. Schmalensee. 1991. Economía. 2 edición. Editorial Mc Graw-hill, Madrid.
8. Gordon, A.D. 1987. A review of hierarchical classification. J. Roy. Stat. Soc. (A). 150. p:119-137.
9. Hartigan J.A. 1967. Representatior. of similarity matrices by trees. JASA. 62: 1140-1158.
10. Johnson R. and Wichern. 1982. Applied multivariate statistical analysis. Prentice - Hall, Englewood Cliff. New Jersey.
11. Lopes, M. y Buarque de Lima, O. 1978. Tendências atuais na geografia urbano/regional. Compilado por faissol S. Fundação instituto Brasileiro de geografia e estatística. IBGE. Rio de Janeiro p 113-124
12. Ministerio de Agricultura y Cría (MAC). 1991. Unidad estatal de desarrollo agropecuario (UEDA) y División de Planificación Agrícola, Región Zuliana. Estimaciones de producción agrícola
13. Statistical graphics corporation, 1992, *Statistical Graphics System*, versión 6, Manugistic, Cambridge, Estados Unidos (STATG).