

XOSE MARIA MELLA MARQUEZ (*)**Un análisis de componentes principales
y de contigüidad espacial para la determinación
de comarcas homogéneas:
una aplicación al caso de Galicia.**

1. INTRODUCCION

En este artículo se aborda un elemento conceptual relativamente novedoso en la delimitación de comarcas: la restricción de contigüidad espacial de las unidades territoriales. Tradicionalmente, tal restricción -debido fundamentalmente, sin duda, a razones de carácter procedimental e informático- se obviaba, a pesar de la falta de tratamiento del problema de auto-correlación espacial que ello comportaba. Este es el caso -por solo citar algunos ejemplos- de los trabajos de Berry (1), Pocock (2) y Reuben (3) en la literatura anglosajona, y de Aznar (4), Cabrer (5) y Prevasa (6) en la literatura española.

(*) Profesor de Teoría Económica de la Universidad Autónoma de Madrid.

(1) Berry, B.J.L., "A method for deriving multi-factor uniform regions". *Przegląd Geograficzny* (en inglés), t. XXXIII, 2, 1961, pp. 263-282.

(2) Pocock, D.C., Wishart, D., "Methods of deriving multifactor uniform regions". *Transactions of Institute of British Geographers*, 47, 1969, pp. 73-98.

(3) Reuben Gabriel, K., Et Al., "A multivariate statistical technique for regionalization". *Journal of Regional Science*, vol. 14, n° 1, 1974, pp. 89-106.

(4) Aznar, A., "Infraestructura y regionalización de las provincias españolas: una aplicación del análisis factorial." *Revista Española de Economía*, Mayo-Agosto, 1974, pp. 137-166.

(5) Cabrer, B., Ribera, M.J., *Los desequilibrios espaciales: una comarcalización del País Valenciano*. Ed. I.E.A.L. Valencia, 1982, pp. 133-ss.

(6) PREVASA, "Propuesta de comarcalización del País Valenciano desde el punto de vista de la acción regional". En Diputación Provincial de Valencia, *Taula redona sobre la comarcalización al País Valencià. I Treballs*. Valencia, 1980, pp. 93-120.

Pues bien, se trata de superar dicha insuficiencia, de modo que las comarcas homogéneas que se obtengan, no solo sean muy parecidas en los valores que observan las diferentes variables atribuidas a las unidades locacionales (en este caso, los municipios) que las constituyen, sino también vecinas y adyacentes en algún tramo de su perímetro municipal.

A tal efecto se ha establecido la delimitación comarcal como un problema taxonómico y, a continuación, se han propuesto las técnicas analíticas que permiten ceñir adecuadamente el concepto de comarca, siguiendo una rigurosa metodología: el análisis de componentes principales y el análisis de grupos con la restricción de contigüidad espacial.

2. LA COMARCALIZACION: UN PROBLEMA TAXONOMICO

En términos generales, se puede afirmar que un problema taxonómico consiste en reordenar los objetos dentro de un sistema de clases basándose en alguna valoración de los mismos. La finalidad de los procedimientos de clasificación es la de proporcionar un agrupamiento válido en función de un criterio previamente especificado por el investigador.

Grigg (7) considera que la comarcalización es un proceso similar al de clasificación. Si los objetos son unidades locacionales o espaciales, la clasificación produce "tipos comarcales". Y si la ordenación se hace teniendo en cuenta el principio de contigüidad de las áreas, entonces el resultado es una comarcalización. Así pues, de acuerdo con este principio, el proceso de comarcalización puede ser reducido a un problema taxonómico.

Por consiguiente, un problema taxonómico, desde el punto de vista de la delimitación de comarcas, puede ser concebido como aquel problema consistente en cómo agregar unidades locacionales para formar comarcas.

(7) Grigg, D.B., "The Logic of regional systems". *Annal of the association of American Geographers*, nº 55, 1965, pp. 465-491.

Dicha agregación puede ser representada, en un sentido formal, por un conjunto E (8), que puede definirse mediante una función:

$$E = f(S, R, A, F)$$

donde S es un conjunto finito de unidades espaciales exclusivas o, dicho de otro modo, que no se solapan; R es un conjunto de criterios; A es un procedimiento de agregación y F es un conjunto de diferentes agregaciones entre las que hay que elegir.

La taxonomía simétrica y cuantitativa -que sigue un itinerario de razonamiento que se atiene a una estricta normativa operativa- permite medir la afinidad o similaridad entre diferentes objetos o unidades locacionales y su ulterior clasificación. El resultado es la formación de ciertos grupos o "clusters".

Los principios operativos de una clasificación sistemática son (9):

1. Una clasificación es una función del objetivo perseguido.
2. Una clasificación está basada en características de diferente naturaleza.
3. Una clasificación es siempre relativa; es decir, es una función de la información disponible, de las hipótesis de partida y de los procedimientos utilizados.
4. Una clasificación se basa en las características de los objetos clasificados.
5. Una clasificación es exhaustiva, si no se omite ningún item.
6. Una clasificación es exclusiva, si los objetos o subconjuntos no se solapan.
7. Los principios de clasificación se mantienen iguales en todas las fases del análisis. Esto implica que las características de la clasificación son fijadas desde el principio, aunque lógicamente -como se verá más adelante- a diferentes niveles del proceso

(8) Keane, M.J., "A functional distance approach to regionalization". *Regional Studies*, vol. 12, 1978. pp. 379-386.

(9) Sokal, R.r., Sneath, P.H. *Principles of numerical taxonomy*. Freeman and Co. San Francisco. 1963.

de clasificación la utilización de un criterio u otro puede generar comarcalizaciones diferentes.

8. El principio de división o diferenciación elegido debe ser adecuado al objetivo perseguido, lo que resulta de algún modo arbitrario, aunque está justificado por el principio 1 (clasificación orientada a un objetivo). Conviene decir, además, que las características de reagrupamiento de los grandes grupos deben ser más importantes que las utilizadas para los pequeños grupos.
9. Las características cualitativas deben ser introducidas en el análisis, utilizando -póngase por caso- una escala ordinal.
10. La utilización de un principio de división o "clasificador" produce una jerarquía de clases. La consistencia lógica de la jerarquía vendrá garantizada por el cumplimiento de los principios cinco, seis, siete, ocho y nueve.

Por otra parte, cualquier análisis taxonómico comprende tres fases bien diferenciadas:

a) Sistematización o estudio científico de las clases, de la diversidad de objetos y de todas las relaciones entre ellos. Es la llamada fase de ordenación.

b) Clasificación u ordenación de los objetos dentro de los grupos, basándose a tal efecto en sus relaciones, es decir, en sus asociaciones por contigüidad.

c) La fase, propiamente taxonómica, que implica el estudio teórico de la clasificación, e incluye las bases, los principios y los procedimientos o sistemas de agrupamiento.

3. TAXONOMIA Y ANALISIS MULTIVARIANTE

Por lo tanto, el problema se reduce a un "análisis de grupos" ("cluster analysis", en la terminología anglosajona), cuyo objeto es la formación de grupos de objetos de tal forma que los elementos pertenecientes a un grupo tengan un alto grado de asociación natural entre ellos mismos, con la condición de que los grupos constituidos sean relativamente distintos unos de otros (10).

(10) Véase Anderberg, M.R., *Cluster Analysis for Applications*. Academic Press. New York, 1973.

Además, hay que tener en cuenta también que el número de formas en que se puede agrupar n objetos en m grupos es extraordinariamente elevado incluso para pequeños valores de n y m . Y es evidente que no interesan todas las formas en que se pueden agrupar los objetos, sino solo aquellas que permitan reproducir más fácilmente la estructura existente entre ellos. Sin embargo, el número de combinaciones posibles es relativamente grande y, por esa misma razón, es obligado recurrir a algunas técnicas de análisis multivariante.

Para resolver el referido problema, los geógrafos -hasta hace bien poco- empleaban un método esencialmente cartográfico estableciendo -en un conjunto E - subconjuntos $E_1, E_2, \dots, E_n$, de modo que las características $a, b, c, \dots, n$ variaran menos en el interior de E_i que cuando se pasase de E_i a cualquier otro subconjunto E_j .

En este terreno la geografía tradicional no tiene más posibilidades que confeccionar y comparar entre sí un gran número de mapas y describir sucesivamente una infinidad de modalidades de asociación en el espacio, eligiendo las que parezcan más idóneas. Al fin y al cabo, es un trabajo de "percepción visual del investigador" (11).

Pero, actualmente, se está ya en condiciones de afirmar que, tanto al nivel de la descripción objetiva de los hechos como de la asociación entre ellos, la comparación visual de los mapas -aunque se haga correctamente- no ofrece un medio seguro para determinar las correlaciones entre los fenómenos distribuidos espacialmente. Y, en consecuencia, -como dicen Racine y Raymond- "antes de llegar a la comparación de los mapas es preciso poner en evidencia dichas correlaciones y las tipografías de las estructuras" (12).

(11) Bertin, J., "Généralisation du traitement graphique de l'information". *Ann. Ec. Soc. Civ.*, n° 1, 1969, pp. 70-101.

(12) Racine, J.B., Raymond, H., *L'analyse quantitative en Géographie*, P.U.F., Paris, 1973, p. 130.

4. ANALISIS DE COMPONENTES PRINCIPALES Y ANALISIS DE GRUPOS

En efecto, precisamente para “poner en evidencia dichas correlaciones y las tipografías de las estructuras” este estudio se valdrá de dos importantes técnicas estadísticas: el análisis de componentes principales y el “cluster analysis” o “análisis de grupos”.

Se parte de una matriz de datos estadísticos, que sólo puede representarse en un espacio de dimensiones elevadas, lo que impide asimilar o “visualizar” la información que contiene. Se necesita, por tanto, representar estos datos en un espacio de pequeñas dimensiones, con la condición de que la pérdida de información sea mínima. Dos son las finalidades que se persiguen:

a) La determinación de la “estructura de las variables”; es decir, saber qué grupos de variables se forman y cuáles son las relaciones entre ellas. Así, si dos variables tienen altas ponderaciones en un mismo componente, se puede afirmar que tienen una pauta de comportamiento semejante y, por consiguiente, convendrá agruparlas.

b) La construcción de índices o indicadores para la formación de grupos de individuos sobre los que se han efectuado las observaciones de las variables. En este sentido, se utiliza la transformación de variables a componentes, si hay variables con pautas de conducta semejantes. Y se calcularán las ponderaciones que indican qué parte de la varianza de cada variable representa cada uno de los componentes. Si cada uno de ellos contribuye cada vez menos a tal representación, de acuerdo con el método de ejes principales, puede deducirse que los primeros componentes representarán una gran parte de la varianza de las variables originales. Y en lugar de usar éstas pueden utilizarse aquéllas que, por añadidura, gozan de una importante y útil propiedad que es la de ortogonalidad.

Pues bien, la primera de las finalidades mencionadas se alcanzará por medio de análisis de componentes principales y, la segunda, por medio del “análisis de grupos”.

4.1. Análisis de componentes principales.

La exposición de la fundamentación teórica del Análisis de componentes y regresión ortogonal: cuanto más próximo esté un enfoque. Aquí se seguirá básicamente el de Aznar (13), que a su vez parte del tratamiento de Malinvaud (14), sobre la relación entre componentes y regresión ortogonal: Cuanto más próximo esté un punto de una recta, mayor será su proyección, y sus coordenadas serán más o menos explicadas por la misma. Si se aplica esta idea a todos los puntos ("nube de puntos"), se podrá obtener la máxima información posible de las abscisas y ordenadas de los diversos puntos, es decir, de la varianza de las variables consideradas. Pero esto mismo es lo que se trata de hacer cuando se halla un componente, por eso se define a éste como la relación entre las variables que toma la máxima información de las mismas. Por consiguiente, se trata a los componentes como nuevas variables cuya varianza hay que maximizar, en función de las originales de las que son combinación lineal.

El punto de partida del análisis es una matriz de datos (X_{in}) con tantas filas o individuos, i , como unidades locacionales observadas y tantas columnas, n , como variables o atributos.

Sea, pues, la matriz original (X_{in}) (15), cuya matriz de varianzas y covarianzas es

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix} \quad (1)$$

(13) Aznar, A., *El Análisis Factorial en Econometría. Una aplicación a las provincias españolas*. Tesis Doctoral. Facultad de CC.EE. y EE, Universidad Complutense. Madrid, 1974.

(14) Malinvaud, E., *Métodos Estadísticos de la Econometría*. Ed. Ariel. Barcelona, 1967.

(15) Dados los atributos o variables, definitorias de las unidades locacionales, hay que hacerlas comparables. Esto se puede hacer -como es universalmente aceptado- estandarizando las variables. Por eso, la matriz original de datos (X_{in}) será transformada en una matriz estandarizada (Z_{in}). Como es sabido, la estandarización o

Sabemos, por otra parte, que la matriz que diagonaliza a otra, premultiplicando a ésta por su transpuesta y postmultiplicándola por ella misma, es la matriz formada por sus vectores característicos, es decir la matriz D . Como los vectores característicos de una matriz simétrica (Σ) son ortogonales, tendremos:

$$D' = D^{-1} \text{ o bien } D'D = I \quad (6)$$

Postmultiplicando los dos miembros de la expresión (5) por D'

$$D' \Sigma = \Lambda D' \quad (7)$$

De esta manera, llegamos al sistema que define los vectores y raíces características. Para que tenga solución, el determinante característico tendrá que ser igual a cero; o sea:

$$[\Sigma - \Lambda I] = \begin{bmatrix} \sigma_{11} - \lambda_1 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} - \lambda_2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{nn} & \sigma_{n2} & \dots & \sigma_{nn} - \lambda_n \end{bmatrix} \quad (8)$$

Así pues, ya podemos obtener las n raíces características y sus correspondientes vectores.

Por otra parte, se demuestra que la varianza de cada componente es igual a una raíz característica. Luego para obtener el primero se elegirá la mayor raíz y así, sucesivamente, de mayor a menor. Esto significa que al elegir la mayor raíz, se considera la cantidad máxima posible de la varianza de las variables. Lo que constituye precisamente uno de los objetivos del análisis.

Se podría también expresar los componentes en función de las variables o viceversa. No se efectuará, sin embargo. Baste con afirmar que los cuadrados de los coeficientes que corresponden a los componentes en la explicación de las variables indican la importancia relativa de cada componente en la explicación de cada variable. Cuanto mayor sea la proyección de una variable sobre un componente, mayor será el coeficiente respectivo. Asimismo se sabe que la suma de los cuadrados de aquellos coeficientes es igual a la raíz característica de cada componente. Luego, cuanto mayor es ésta, mayor es la importancia del componente.

Por último, todavía queda por abordar el problema de cuantos componentes seleccionar. Lógicamente se seleccionarán aquellos que provean el máximo de información y que reflejen la estructura subyacente de las variables. Según Lefebvre (16), se escogerán aquellos componentes que “absorban” o “den cuenta” de un porcentaje elevado de la varianza original, de modo que

$$\frac{1}{\text{Traza } \Sigma} = \text{Porcentaje de la varianza total explicado por el primer componente.}$$

teniendo en cuenta que Traza Σ es igual a la suma de las varianzas de las variables.

Por lo tanto, se retendrán los componentes que expliquen un porcentaje razonablemente alto, a partir del cual la explicación, por ser despreciable, carece ya de importancia. La práctica normal, que ofrece resultados satisfactorios en este tipo de problemas, es elegir aquellos componentes con raíces características mayores que uno. Se considera que las raíces menores que uno proveen componentes no significativos.

Ahora bien, en orden a una mejor identificación de las variables que están asociadas con cada componente, se acostumbra a efectuar una rotación de los mismos. El proceso de identificación puede ser mejorado si los ejes son representados de manera que los grupos de variables con características comunes puedan percibirse; en otras palabras, operando un cambio en el sistema de referencia, aunque manteniendo las propiedades de las variables.

El objetivo fundamental de la rotación es ajustar al máximo cada factor o componente a los diversos grupos de variables, eliminando todo posible efecto perturbador procedente de otras variables no pertenecientes a tales grupos. En concreto, este problema se resuelve aplicando el método “varimax”, “que permite hacer mayores las ponderaciones más elevadas y más pequeñas a las más reducidas” (17); es decir, maximizando la varianza de las ponderaciones de cada factor.

(16) Lefebvre, J., *Introduction aux analyses statistiques multidimensionnelles*. Masson. París, 1975.

(17) Ibid., p. 301.

4.2. Análisis de grupos.

Una vez satisfechas las condiciones de ortogonalidad y estandarización, se podrán comenzar a agrupar las observaciones más cercanas entre sí.

El procedimiento -llamado de "agrupamiento por etapas" o "stepwise grouping" (18)- es como sigue:

- Se establece una matriz de distancia D^2 , existente entre cada par de observaciones. Dicha matriz se forma previa elección de una medida de distancia.

- Se identifica la menor distancia y se agrupan las dos observaciones correspondientes.

- Se sustituye, en la matriz de distancias, el vector fila y el vector columna de esas dos observaciones por una sola fila y una sola columna que contienen la distancia del centroide del nuevo grupo de dos miembros a todas las demás observaciones.

- Se repite el proceso hasta que, finalmente, queda un solo grupo que contiene todas las observaciones.

Es el conjunto de estas secuencias lo que muestra el "árbol factorial de cadenas de reagrupamiento" (19) o "linkage tree", que define los subsistemas en que se divide el sistema espacial (sometido a un análisis "cluster") y que representa el peso obtenido por cada una de las unidades locacionales en cada componente.

En el límite -a menos que dos o varias observaciones tengan rigurosamente las mismas posiciones en el sistema de ejes- este árbol de relaciones tiene evidentemente tantas ramas como observaciones o $(n-1)$ observaciones, si consideramos que la primera etapa asocia obligatoriamente un primer par.

(18) Ward, J.H., "Hierarchical Grouping to optimize an objective function", *Journal of the American Statistical Association*, vol. 58, 1963, pp. 236-244.

(19) Racine, J.B. Raymond, H., op. cit., p. 211.

Berry (20) y King (21) reconocen que, desgraciadamente, no existe una solución analítica que permita decidir cuántos grupos deben ser identificados. No obstante, se puede determinar en qué etapa una nueva reducción del número de grupos aumenta de manera mínima la suma de los cuadrados de las distancias en el interior de un grupo, puesto que es justamente sobre este cálculo que se fundamenta -en cada etapa- la elección de los grupos que deben ser reunidos. Berry aconseja que esta decisión o elección se haga intuitivamente, teniendo en cuenta la naturaleza del problema estudiado. Y Ward (22) -que propone un algoritmo que lleva su mismo nombre- advierte que, conforme se tienen menos grupos, hay una mayor pérdida de información, y las unidades que los forman son menos homogéneas entre sí. Por esta razón, suele elegirse aquella agrupación (comarcalización) que corresponde al paso previo a aquél en que tiene lugar un fuerte incremento en la distancia dentro de los grupos.

5. METODOLOGIA DE DELIMITACION DE COMARCAS HOMOGENEAS

Recapitulando lo dicho hasta aquí, se puede establecer una metodología o un procedimiento analítico de delimitación de comarcas homogéneas como el que se muestra en la Figura 1.

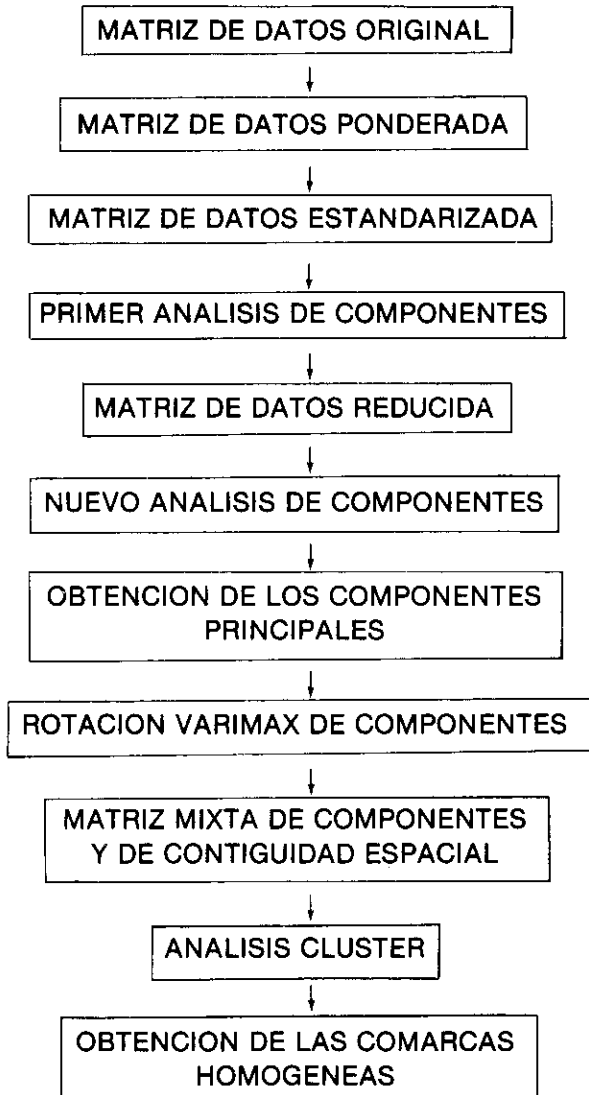
(20) Berry, B.J.L., "Grouping and Regionalising: an approach to the problem using multivariate analysis", en Garrison, W.L., Marble, D.F., *Quantitative Geography*. Northwestern University. Evanston, 1967, pp. 219-251.

(21) King, L.J., *Statistical Analysis in Geography*. Englewood Cliffs, Prentice Hall, 1969, p. 92.

(22) Ward, J.H. op. cit.

FIGURA 1

**DIAGRAMA DEL PROCEDIMIENTO ANALITICO DE LA
COMARCALIZACION HOMOGENEA DE GALICIA**



FUENTE: Elaboración propia.

El punto de partida del análisis es una matriz original de datos, que deben ser depurados y ponderados (matriz de datos ponderada) -normalmente, en este tipo de estudios, la ponderación se hace por el total de población de la unidad territorial- y, seguidamente, debido al hecho de que las variables vienen expresadas en diferentes magnitudes es preciso homogeneizarlas, por medio del conocido método de la tipificación o estandarización (matriz estandarizada). A partir de este momento, ya se puede proceder al primer análisis de componentes principales, que tiene por objeto hacer una primera selección de variables -descartando aquellas que sean redundantes o poco significativas- y obtener la matriz de datos reducida, a la que se le aplica de nuevo un análisis de componentes. Como consecuencia, se obtienen varios componentes principales y, en orden a su mejor identificación, se acostumbra a realizar una rotación "varimax" -en los términos en que se ha explicado más arriba- de los mismos. Finalmente, establecida una matriz mixta de componentes y de contigüidad espacial (representando cada observación por sus coordenadas, o por 1 ó 0 según que dos observaciones sean contiguas o no, respectivamente) se efectúa un análisis "cluster" o de agrupamiento, obteniendo de esta suerte las comarcas homogéneas.

En definitiva, se logran dos objetivos:

- a. Se identifica la estructura espacial de los fenómenos socio-económicos y su grado de homogeneidad versus heterogeneidad.
- b. Se asigna cada municipio a un ámbito espacial común del desarrollo (o subdesarrollo) económico y demográfico, que permita responder conjunta y unitariamente a los requerimientos propios de una planificación supramunicipal.

6. MATRIZ ORIGINAL DE DATOS

Como es sabido, un sistema regional viene definido por una serie de características asignables a un conjunto de unidades locacionales o espaciales. Un primer problema analítico aparece a la hora de realizar una selección adecuada de las variables definitorias de dicho sistema.

En efecto, sabido es que los sistemas regionales son extraordinariamente complejos y un gran número de propiedades de diversa naturaleza son necesarias para una comprensión cabal de su

problemática. Por lo tanto, se ha partido de una batería suficientemente amplia de variables que, en un principio, han parecido relevantes del fenómeno a estudiar, en sus diversas vertientes (23).

Las variables elegidas -para cada uno de los 312 municipios de Galicia se clasificaron según su pertenencia a los grandes perfiles de una región -y de acuerdo con las características peculiares de los asentamientos de la población gallega- y han sido las siguientes (24):

1. POBLACION

1. Población de derecho (25).
2. Índice de crecimiento de la población en el período 1965-1975.
3. Densidad de habitantes.
4. Densidad de población del municipio, deducida previamente la población del mayor núcleo.
5. Porcentaje de la población del mayor núcleo respecto de la población total del municipio.
6. Número de entidades de población por km².
7. Promedio de habitantes por entidad de población.

(23) Adviértase que ésta es una decisión subjetiva que, de alguna manera, condiciona el proceso analítico posterior. La elección es una y única, entre otras muchas posibles. La mayor o menor validez del análisis dependerá, en primer lugar, del grado de aprehensión de la realidad de las variables seleccionadas, lo que está en proporción directa del conocimiento previo que del problema, objeto de estudio, se demuestre poseer. Y, en segundo lugar, del grado de significatividad que muestren tales variables para alcanzar el objetivo propuesto.

(24) Que ofrece una buena medición del grado de polarización demográfica del municipio. La fuente de la que se ha extraído esta variable y -previo el tratamiento correspondiente- las variables de la 2 a la 27 ha sido I.N.E., *Explotación del Padrón Municipal de Habitantes de 1975* (inédita), Banesto, *Anuario del Mercado Español*, (varios años) e I.N.E. *Poblaciones de derecho y de hecho de los municipios españoles*. Madrid, 1977. Se entiende por población de derecho, los habitantes residentes, estén presentes o ausentes.

8. Promedio de habitantes de cada entidad de población deducida previamente la población del mayor núcleo.
9. Índice de urbanización (26).
10. Porcentaje de población hasta los 14 años de edad (27).
11. Porcentaje de población entre los 15 y 64 años de edad.
12. Porcentaje de población de 65 y más años de edad.
13. Tasa de residencia (28).
14. Tasa de actividad (29).

II. OCUPACION

15. Tasa de paro (30).
16. Porcentaje de activos en la agricultura (31).

(26) Dicho índice ha sido elaborado a partir de la ecuación $I_U = 2 \log P + \log T$, siendo I_U el índice de urbanización, P el promedio de habitantes por entidad de población y T el número de entidades de población por km^2 . Véase FUNDACION FOESSA. *Estudios sociológicos sobre la situación social de España, 1970*. Ed. Euramérica. Madrid, p. 1195.

(27) Definida, lógicamente, por el cociente Población de 0 a 14 años $\times$ 100/ Población total de derecho. El cálculo se ha hecho, de manera análoga, para las poblaciones de 15 a 64 y 65 y más años de edad.

(28) La tasa de residencia se define por el cociente población de derecho que ha vivido siempre en el municipio $\times$ 100 / población total de derecho.

(29) Definida por el cociente activos de 14 y más años $\times$ 100 / población de derecho de 15 y más años.

(30) Calculada por el cociente parados $\times$ 100 / total de activos. Es obvio que dicha variable -índice de paro registrado- no expresa claramente la realidad del desempleo en Galicia. Como es sabido, el concepto de parado -implícito en aquel índice- es el que corresponde a aquella persona que, estando en edad de trabajar, no ejerce una actividad remunerada durante una jornada laboral de 8 horas. En consecuencia, se trata básicamente de un concepto propio de una sociedad industrial, en donde predomina el trabajo por cuenta ajena, lo que evidentemente no se ajusta a las características de la economía gallega, caracterizada por altas tasas de paro encubierto. En cualquier caso, tal índice no se ha excluido apriorísticamente del análisis para que sean las propias cifras, y el procedimiento estadístico seguido, quienes se encarguen de hacerlo.

(31) Cociente de activos en agricultura $\times$ 100 / total de activos. De modo análogo, se han definido los índices de activos para los sectores industrial, construcción y de servicios.

- 17. Porcentaje de activos en la industria.
- 18. Porcentaje de activos en la construcción.
- 19. Porcentaje de activos en los servicios.

III. SITUACION PROFESIONAL

- 20. Porcentaje de empleadores (32).
- 21. Porcentaje de empresarios (33).
- 22. Porcentaje de asalariados (34).
- 23. Porcentaje de otros (35).

IV. NIVEL DE INSTRUCCION

- 24. Porcentaje de analfabetos (36).
- 25. Porcentaje de población con estudios primarios (37).
- 26. Porcentaje de población con estudios medios.
- 27. Porcentaje de población con estudios superiores.

(32) Empleadores $\times 100$ / activos totales. Por empleadores se entiende todo aquel patrono, empresario o profesional que emplea a personal.

(33) Empresarios sin asalariados y trabajadores independientes $\times 100$ / total activos. Empresarios son aquellos que no emplean personal, trabajadores por cuenta propia y miembros de cooperativas.

(34) Asalariados $\times 100$ / total activos. Asalariado es toda aquella persona que trabaja con carácter fijo o eventual por cualquier clase de remuneración, jornal o comisión.

(35) Los restantes activos (otros) $\times 100$ / total activos. "Otros" son las personas que trabajan sin remuneración reglamentaria en la empresa familiar y otros activos sin clasificar.

(36) Analfabetos $\times 100$ / población de derecho de 15 y más años de edad.

(37) Personas con estudios primarios $\times 100$ / población de derecho de 15 y más años de edad. Se ha hecho un cálculo análogo para el caso de las personas con estudios medios y superiores.

V. INFRAESTRUCTURA EDUCATIVA

28. N° de puestos escolares estatales de preescolar y E.G.B. por 1.000 habitantes (38).

VI. SANIDAD

29. N° de médicos por 1.000 habitantes (39).

VII. RELIEVE

30. Altitud media (40).

VIII. AGRICULTURA

31. N° de parcelas por explotación (41).
32. Porcentaje de la SAU (Superficie Agraria Util) sobre la superficie total (42).
33. Porcentaje total de tierras cultivadas.
34. Variación del número de explotaciones en la década 1962-72.
35. N° de habitantes por hectárea cultivada.

(38) Fuente: Delegación del I.N.E., en el M. De Educación y Ciencia. *Alumnos matriculados en E.G.B. Curso 1975-1976*.

(39) Colegio de Medicos de España. *Anuario Médico-Farmacéutico Nacional 1976*. Ed. Lectura y Estudio. Barcelona, 1976.

(40) I.N.E. *Reseñas Estadísticas Provinciales* (Diferentes años).

(41) Fuente: I.N.E. *Censo Agrario de España 1972*, a partir del cual se han obtenido las variables de las 31 a las 35.

(42) La superficie agrícola utilizada o útil comprende las tierras en alternativa o labradas, las de pastos y praderas permanentes, las de huerta no perenne y la correspondiente a las plantaciones no forestales. Véase Desclaude, F., Tondut, J., *La empresa agraria y su gestión*. Ed. Mundi-Prensa, 2ª Edición. Madrid, 1979, p. 269.

IX. TRANSPORTES Y COMUNICACIONES

- 36. N° de camiones por 1.000 habitantes (43).
- 37. N° de teléfonos por 1.000 habitantes.

X. FINANZAS Y COMERCIO

- 38. N° de bancos y cajas de ahorro por 1.000 habitantes.
- 39. N° de licencias comerciales por 1.000 habitantes.

XI. VIVIENDA

- 40. Porcentaje de viviendas agrarias sobre el total de viviendas.
- 41. Porcentaje de edificaciones de más de dos viviendas sobre el total de viviendas (44).

XII. SECTOR PUBLICO

- 42. Presupuesto municipal por habitante.

XIII. OCIO

- 43. Recaudación del impuesto de espectáculos por habitante.

XIV. ECONOMIA

- 44. Renta per capita en miles de pesetas.
- 45. Cuota de mercado.

(43) Banesto, op. cit., (Varios años). La misma fuente ha servido para la elaboración de las variables relativas al número de teléfono, número de bancos y cajas de ahorro, número de licencias comerciales, presupuesto municipal, recaudación del impuesto de espectáculos, renta per capita y cuota de mercado.

(44) I.N.E. *Censo de la Vivienda*.

XV. LOCALIZACION

46. Distancia a la capital de la provincia o a la ciudad de más de 50.000 habitantes más cercana.

XVI. FUNCIONALIDAD

47. Índice de centralidad (45).

Obviamente, el problema inmediato que se plantea puede formularse en los siguientes términos: ¿cómo abstraer y sistematizar la enorme masa de información disponible para el análisis de la estructura regional de Galicia?.

Normalmente, en economía regional se ha recurrido a distintas técnicas de análisis estadístico univariante, que naturalmente simplifican los estudios -y que resultan de una utilidad innegable- pero como contrapartida crean arduos problemas de errores de medición, significatividad, agregación y omisión de variables.

En consecuencia, ha resultado conveniente utilizar -en orden a la mejor adecuación posible del instrumental estadístico al problema planteado- métodos con capacidad para tratar numerosas variables, discontinuas o continuas, heterogéneas, etc. Tales métodos son los denominados de análisis multivariante, que serán los que se apliquen a la matriz original de datos de orden (312 x 47) (en donde las filas son las observaciones o municipios -unidades locacionales- y las columnas, las características o variables).

7. ANALISIS DE COMPONENTES PRINCIPALES ROTADOS

El método de componentes principales ha permitido responder a cuatro objetivos fundamentales del análisis:

(45) Dicho índice se ha definido utilizando como unidad de medida de interacción los flujos de "llamadas telefónicas". Así, un municipio tendrá un índice tanto más elevado cuanto mayor sea el porcentaje de comunicaciones telefónicas que centraliza. Se han tenido en cuenta, además, tanto las interacciones directas como las indirectas. Véase Mella Marquez, X.M., *Un modelo de simulación dinámica para la planificación regional. Una aplicación al caso de Galicia*. Tesis Doctoral. Universidad Autónoma de Madrid, 1982, pp. 582-ss. (Pendiente de publicación por el Instituto de Estudios de la Administración Local).

a) Eliminar la información redundante contenida en las variables.

b) Transformar el conjunto inicial de variables en otro, llamado de componentes, con la propiedad de que sean independientes entre sí dos a dos.

c) Sistematizar el análisis -mediante la selección de un número pequeño de componentes- de tal modo que sean capaces de reproducir un elevado porcentaje de información o varianza de las variables originales.

d) Determinar la estructura subyacente de relaciones entre las variables empleadas, de manera que a través de las nuevas variables o componentes principales podamos obtener las comarcas homogéneas.

Los diferentes análisis de componentes se han hecho a partir de las combinaciones de municipios y variables que se muestran en el Cuadro 1, porque, naturalmente, los resultados no son exactamente los mismos.

CUADRO 1

ANALISIS DE COMPONENTES PRINCIPALES

Nº DE MUNICIPIOS Nº VARIABLES	MUNICIPIOS				
	312	305	201	111	C > 196
47	*				
35	*	*	*	*	*

FUENTE: Elaboración propia.

La elección de tales matrices de municipios y variables viene justificada por la diferente naturaleza de aquéllos y de éstas, y, además -esto es fundamental- porque se ha tratado de buscar los grupos más representativos posibles, de modo que las relaciones entre las diferentes variables no se vean afectadas por efectos espúreos. Así, en primer lugar, se ha partido de una matriz de todos los municipios de Galicia (312) y 35 variables: descartándose, pues,

12 de ellas debido a que no alcanzaban un valor mínimo en la correlación con los componentes obtenidos.

A continuación, dado el enorme peso de las 7 principales ciudades sobre el sistema urbano gallego, se ha creído oportuno analizar los resultados de un análisis de componentes -de cuyas observaciones se excluyen a aquellas 7 ciudades- para despojar a los valores de los factores de los efectos perturbadores de las mismas.

Se ha realizado también un factorial con los municipios de crecimiento demográfico positivo (111) y otro con aquéllos con crecimiento negativo (201), con la intención de percibir los comportamientos específicos de cada uno de los grupos de municipios.

Pues bien, ya se puede proceder ahora a la obtención de aquellos componentes significativos, que permitan realizar la comarcalización homogénea de Galicia.

En primer lugar, se ha efectuado un análisis de componentes principales aplicado a la matriz original de datos (312 x 47) y ha podido comprobarse que 12 de dichas variables arrojaban unos "loading scores" o factores de carga no significativos: todos sus valores eran menores de 0.25 (46). Razón por la cual se han descartado y no se han utilizado en análisis posteriores. Dichas variables han sido: la tasa de paro, el porcentaje de empleadores, la tasa de analfabetos, el número de puestos escolares estatales de preescolar y E.G.B. por 1.000 habitantes, el número de médicos por 1.000 habitantes, el número de parcelas por explotación, el porcentaje de la SAU sobre la superficie total, el porcentaje total de tierras cultivadas, la variación del número de explotaciones en la década 1962-72, el número de bancos y cajas de ahorro por 1.000 habitantes, el porcentaje de edificaciones de más de dos viviendas sobre el total de viviendas y la cuota de mercado.

Seguidamente, a partir de la matriz de datos reducida (312x35) -mucho más manejable- se ha explicado nuevamente el método de componentes principales y se obtuvieron una serie de factores, que explicaban un alto porcentaje de la varianza original. Los resul-

(46) Este es el criterio recomendado por Jolliffe. Véase Jolliffe, I.J., "Discarding variables in a principal component analysis I: Artificial data". *Applied Statistics*, 21, 1972, pp. 160-173. También del mismo autor, "Discarding variables in a principal component analysis II: Real Data". *Applied Statistics*, 22, 1972, pp. 21-31.

tados obtenidos son los que se muestran -para los diversos conjuntos de municipios y sus respectivos componentes- en los Cuadros 2, 3, 4 y 5.

De la interpretación de dichos factores o componentes se han podido extraer las siguientes conclusiones:

a. El primer componente retenido (Cuadro 2) -considerando la totalidad de los municipios de Galicia (312)-, que explica el 44,9% de la varianza de las variables originales, exhibe fuertes ponderaciones factoriales en las variables relativas a la infraestructura comercial y de servicios -licencias comerciales, 0,800; número de teléfonos por 1.000 habitantes, 0,748; camiones por 1.000 habitantes, 0,696-; en la renta per capita (0,669), en el presupuesto municipal (0,722) y en la recaudación del impuesto de espectáculos (0,591). Dichas ponderaciones se mantienen igualmente elevadas en las tasas relativas a los niveles educativos primarios (0,549), medios (0,618) y superiores (0,673); en los porcentajes de activos industriales (0,570) y de servicios (0,518). Este primer componente está, asimismo, negativamente relacionado con el porcentaje de activos agrícolas (-0,381), con el porcentaje de empresarios (-0,269), con la tasa de residencia (-0,680) y con el porcentaje de viviendas agrarias (-0,471). Por lo tanto, parece razonable definir a este componente como una dimensión del "crecimiento económico y de la dotación de servicios".

Por otra parte, los primeros componentes de los restantes conjuntos de municipios no ofrecen ninguna novedad, salvo en el caso de los municipios regresivos (201) en el que, aparte de su menor poder explicativo (31,1%), los "scores" más significativos son los correspondientes al porcentaje de activos en servicios (0,707), a la tasa de asalarización (0,848), al porcentaje de activos en la construcción (0,655) y en la industria (0,541).

CUADRO 2**VALORES DEL PRIMER COMPONENTE PRINCIPAL ROTADO**

Número de Municipios	312	305	201	111
Variables				
Licencias comerciales	0.800	0.797		0.805
Presupuesto municipal	0.722	0.667	0.253	0.818
Teléfonos por 1000 habitantes	0.748	0.720	0.282	0.748
Camiones por 1000 habitantes	0.696	0.685		0.701
Tasa de escolarización superior	0.678	0.553		0.770
Renta per capita	0.669	0.664		0.689
% población del mayor núcleo	0.656	0.390		0.793
Tasa de escolarización media	0.618	0.538		0.879
Recaudación del impuesto de espectáculos	0.591	0.569	0.271	0.625
% de activos en servicios	0.518	0.384	0.707	0.625
Tasa de residencia	-0.680	-0.620	-0.307	-0.728
Tasa de escolarización primaria	0.549	0.318		0.811
% de empresarios	-0.269			-0.350
% de viviendas agrarias sobre el total de viviendas	-0.471	-0.491	-0.451	-0.501
% empleo agrícola	-0.381	-0.346	-0.869	-0.445
% de otros			-0.913	
Tasa de actividad			-0.721	
% de salarios			0.848	
% de activos en la construcción			0.655	
% de empleo en la industria	0.570		0.541	0.580
Densidad de población			0.256	
Índice de urbanización			0.289	
Densidad de población, deducida previamente la población del mayor núcleo			0.394	
% Varianza explicada	44.9	40.9	31.1	48.6

FUENTE: Elaboración propia.

También son significativos, pero negativamente, los “scores” del porcentaje de activos agrícolas (-0.869), del porcentaje de “otros” (-0.913), de la tasa de actividad (-0.721) y del porcentaje de viviendas agrarias (-0.451). Estos valores indican que también para esta clase de municipios, el primer componente está relacionado con el “crecimiento económico y la dotación de servicios”.

CUADRO 3

VALORES DEL SEGUNDO COMPONENTE PRINCIPAL ROTADO

Número de Municipios Variables	312	305	201	111
Población menor de 14 años	0.840	0.809		0.887
% de incremento demográfico	0.720	0.706		0.885
Índice de urbanización	0.605	0.529		0.618
% de asalariados	0.586	0.584		0.696
Densidad de población deducida previamente la población mayor núcleo	0.557	0.550		0.467
% de empleo industrial	0.471	0.450	0.468	0.539
% de población de 65 y más años de edad	-0.855	-0.796		-0.865
% de empresarios	-0.716	-0.679		-0.795
% de viviendas agrarias	-0.591	-0.500		-0.679
% de empleo agrícola	-0.504	-0.492	-0.290	-0.513
% de otros	-0.352	-0.316		-0.560
Densidad de población		0.768		
Renta per capita			0.809	0.373
Camiones por 1000 habitantes			0.776	0.569
Presupuestos municipales por habitante		0.684	0.288	
Licencias comerciales por 1000 habitantes			0.574	
Teléfonos por 1000 habitantes			0.572	0.387
Tasa de residencia			-0.341	
Altitud media				-0.779
Porcentaje de activos en servicios				0.417
% Varianza explicada	10.2	10.3	10.8	12.2

FUENTE: Elaboración propia.

b. El segundo componente (Cuadro 3), que explica el 10.2% de la varianza original, muestra sus factores de carga positivos más significativos en las variables relativas a la población menor de 14 años (0.840), en el porcentaje de incremento demográfico (0.720), en el índice de urbanización (0.605), en la tasa de asalarización (0.586), en la densidad del hinterland de la capital municipal (0.557) y en el porcentaje de activos industriales (0.471). Los factores de carga negativos son todos aquellos relacionados con variables de carácter agrario y de envejecimiento de la población, el porcentaje de empresarios (-0.716), el porcentaje de viviendas agrarias (-0.591), el porcentaje de activos agrícolas (-0.504) y el porcentaje de población mayor de 65 años de edad (-0.855). Así pues, no parece haber duda de que este factor da idea del dinamismo demográfico, del nivel de urbanización y del peso de la industria en el conjunto de las observaciones analizadas.

c. El tercer componente (Cuadro 4) da cuenta del 7.9% de la varianza original. Las ponderaciones mayores se encuentran en aquellas variables relacionadas con la ocupación y el empleo. Así, este factor está positivamente relacionado con la tasa de activos en la construcción (0.686), en la industria (0.529) y en los servicios (0.513), y negativamente relacionado con la tasa de activos agrícolas (-0.720), con la tasa de actividad (-0.787) y con el porcentaje de "otros" (-0.720). En consecuencia, este componente evidencia la fuerte oposición existente en Galicia entre la industria y la agricultura, y entre la ciudad y el campo: lo que, evidentemente, será muy útil para la agrupación, por una parte, de los municipios industriales y urbanos, y, por otra, de los municipios agrícolas y rurales.

CUADRO 4**VALORES DEL TERCER COMPONENTE PRINCIPAL ROTADO**

Número de municipios	312	305	201	111
Variables				
% de otros	-0.851	-0.841		-0.895
Tasa de actividad	-0.787	-0.759		-0.792
% empleo agrícola	-0.720	-0.702		-0.736
% empleo en la construcción	0.686	0.660		0.822
% de asalariados	0.683	0.616		0.770
% empleo industrial	0.529	0.468		0.679
% empleo en servicios	0.513	0.509	0.440	0.594
Recaudación de impuestos de espectáculos por habitante			0.678	
% población con estudios superiores			0.617	
Teléfonos por 1000 habitantes			0.561	
Licencias comerciales por 1000 habitantes			0.520	
Densidad de población			0.518	
% viviendas agrarias sobre el total de viviendas			-0.565	
Tasa de residencia			-0.373	-0.297
Altitud media				-0.335
Número de camiones por 1000 habitantes				0.327
% Varianza explicada	7.9	8.3	10.2	7.6

FUENTE: Elaboración propia.

d. El cuarto componente (Cuadro 5), que recoge el 6.4% de la varianza original, muestra un carácter mucho más concreto y localizado. Las variables con ponderaciones significativas son la densidad de población (0.908), el promedio de habitantes por entidad de población (0.863), la densidad de población deducida previamente la población del mayor núcleo (0.589) y el índice de urbanización (0.459). Es lógico que este componente representa a la concentración demográfica y al nivel de urbanización.

CUADRO 5

VALORES DEL CUARTO COMPONENTE PRINCIPAL ROTADO

Número de municipios Variables	312	305	201	111
Densidad de población	0.908	0.000	0.914	
Promedio de habitantes por entidad de población	0.863	0.796		0.888
Densidad de población, deducida previamente la población del mayor núcleo	0.589	0.000		0.659
Índice de urbanización	0.459	0.431		0.556
Recaudación del impuesto de espectáculos	-0.391	0.000		0.424
% de población del mayor núcleo respecto al total de la población	0.259	0.685		
Renta per capita	0.252	0.000		0.281
Número de teléfonos por 1000 habitantes	0.320	0.000		0.326
Médicos por 1000 habitantes			0.951	
Tasa de escolarización media			0.835	0.287
Tasa de escolarización primaria			0.812	
% activos en la construcción			0.544	
Tasa de actividad			-0.263	
% Varianza explicada	6,4	6,1	7,2	4,3

FUENTE: Elaboración propia.

e. Los componentes principales obtenidos con los restantes conjuntos de municipios -todos los municipios de Galicia, excluidos los correspondientes a las siete ciudades principales (305); los municipios regresivos, que han sufrido pérdidas de población (201), y los que han experimentado un incremento demográfico positivo (111)- no ofrecen, en general, resultados divergentes de los alcanzados para la matriz original de los 312 municipios. No obstante, pueden hacerse algunas matizaciones de relativo interés.

En efecto, el primer componente registra -en el conjunto de los 305 municipios y en el correspondiente a los 201- unas ponderaciones factoriales de las variables relativas a las dotaciones en servicios (licencias comerciales, número de teléfonos por 1.000

habitantes, número de camiones por 1.000 habitantes...) inferiores a las del caso en que se consideran incluidas las siete ciudades principales (Véase el Cuadro 2). Esto indica, sin duda, la fuerte polarización de los servicios en aquellas siete ciudades, que se ve reforzada, además, por un mayor nivel de la renta per capita y de los presupuestos municipales (47).

Si, asimismo, se observan los "scores loadings" del primer componente relativos al conjunto de aquellos municipios de mayor crecimiento demográfico, la afirmación anterior se ve, si cabe, todavía más enfatizada: el "crecimiento económico y las dotaciones en servicios" -que fue la interpretación que se ha hecho de este primer componente- es la clave explicativa del comportamiento de los municipios progresivos. Además, las ponderaciones relativas a las variables relacionadas con la agricultura y el mundo rural son también superiores -con signo negativo- a las obtenidas en los casos anteriores: lo que viene a indicar que el proceso de industrialización y de desruralización de estos 111 municipios es superior al existente en los demás. Los restantes componentes, aunque con menor capacidad explicativa, no contradicen, sin embargo, las observaciones anteriores.

f. Finalmente, resta todavía por analizar los componentes principales del conjunto de municipios de mayor centralidad ($C > 196$). En este caso, se han retenido cinco componentes, que explican, por orden decreciente, el 40.7%, el 16.3%, el 9.5%, el 8.1% y el 6.1% de la varianza de las variables originales (Véase el Cuadro 6).

(47) Ahora bien, una observación muy importante se impone: la disponibilidad tan sólo de datos municipales ha conducido inexcusablemente a la elaboración de indicadores igualmente municipales de los equipamientos sociales; pero el problema es que tales índices pueden originar -de hecho, originan- impresiones erróneas sobre la localización de dotaciones infraestructurales y generan resultados que ocultan las fuertes desigualdades existentes, en muchas ocasiones, a nivel intramunicipal. ¿Qué sucedería si se analizara la dispersión, de las mismas variables, por distritos urbanos o por parroquias en los municipios rurales?. Pues, sucedería, sin duda, que las heterogeneidades y las zonas intramunicipales desasistidas emergerían con toda su crudeza. Es lo que se conoce, entre los analistas de las desigualdades, con el desafortunado nombre de "falacia ecológica"; esto es, la atribución a cada individuo, o a pequeñas comunidades de individuos, de las observaciones correspondientes a agregados más amplios de los que forman parte. Véase -a modo de ejemplo- el estudio de Holtermann, S., "Areas of urban deprivation in Great Britain: an analysis of 1971 census data". *Social Trends*, nº 6, 1975.

El primer componente concentra los indicadores básicos de la estructura demográfica.

Así, pueden contemplarse las fuertes ponderaciones que registran la densidad de población (0.956), la densidad de población deducida previamente la población del mayor núcleo (0.695), el promedio de habitantes por entidad (0.890), el índice de urbanización (0.617) y la población de derecho del municipio (0.817). Este componente es un indicador de la importancia que, para el mantenimiento y desarrollo de la centralidad de un núcleo (0.818), tiene el tamaño urbano y la densidad de población del mismo; pero no sólo su propia densidad, sino de la densidad de su hinterland (0.695), al que provee de sus servicios más especializados y del que se sirve como fuente y reserva de mano de obra. Por lo demás, puede comprobarse como la distancia a la capital de la provincia es un factor de localización de la centralidad negativo (-0.273) y, a causa de la concentración demográfica, la ponderación de la presión de los habitantes sobre el suelo cultivado es muy alta (0.913).

El segundo componente relaciona el crecimiento demográfico (0.776) con la pirámide por edades de la población y la ocupación de la misma. En otras palabras, un mayor dinamismo demográfico implica el rejuvenecimiento de la población (porcentaje de menores de 14 años, 0.949; porcentaje de mayores de 65 años, -0.820), un menor peso de la agricultura en el conjunto de la economía (porcentaje de activos agrícolas, -0.465) y, como contrapartida, una mayor importancia de la industria (porcentaje de empleo industrial, 0.582).

El tercer componente no añade nada substancialmente diferente a los dos anteriores, mientras que el cuarto contrapone la tasa de actividad (-0.371) con el porcentaje de activos en la construcción (0.876); esto es, el crecimiento de este sector industrial provoca la disminución de las altas tasas de actividad de la agricultura. El quinto componente, en fin, no expresa otra cosa que la fuerte dispersión de la población (número de entidades por km², -0.874) afecta negativamente, como es lógico, al nivel de densidad del hinterland urbano (densidad de población, deducida previamente la población del mayor núcleo, 0.320).

CUADRO 6

VALORES DE LOS COMPONENTES PRINCIPALES ROTADOS.
MUNICIPIOS C > 196

Componentes Variables	I	II	III	IV	V
Indice de crecimiento demográfico		0.776			
Densidad de población	0.956		0.282		
Densidad de población, deducida previamente la población mayor núcleo	0.695	0.494			0.320
% de población del mayor núcleo respecto al total de la población	0.343		0.775		
Número de entidades por Km ² .					-0.874
Promedio de habitantes por entidad	0.890				
Promedio de habitantes de cada entidad, deducida la población del mayor núcleo	0.698	0.511			0.326
Indice de urbanización	0.617	0.580	0.395		
% habitantes < 14 años		0.949			
% 15-64 años		-0.731			
% habitantes < 65 años		-0.820	-0.288		
Tasa de actividad				-0.871	
% empleo agrícola		-0.465	-0.796		
% empleo industrial		0.582	0.469		
% empleo construcción				0.876	
% empleo servicios	0.284		0.803		
Habitantes por Ha. cultivada	0.913				
Distancia a la capital provincia	-0.273	-0.298			
Población de derecho	0.817		0.342		0.543
Indice de centralidad	0.818		0.346		
% Varianza explicada	40.7	16.3	9.5	8.1	6.1

FUENTE: Elaboración propia.

8. UNA COMARCALIZACION HOMOGENEA DE GALICIA: UN ANALISIS "CLUSTER" CON LA RESTRICCION DE CONTI- GUIDAD ESPACIAL

De los cuatro componentes principales -obtenidos en el epígrafe anterior- se han elegido los tres primeros, con los cuales se realizará una comarcalización homogénea de Galicia.

Las razones por las que se ha excluido el cuarto componente principal son varias. Por una parte, no admite una interpretación clara y diferenciada de los otros tres; y, por otra, su poder explicativo es relativamente bajo (6.4% de la varianza original). Además, existe todavía una razón adicional: los otros tres componentes concentran un alto porcentaje de la información original de nuestros datos (63%); lo que hace innecesaria la introducción de más información, que se presenta, por lo demás, como redundante.

En síntesis, los tres componentes principales retenidos han sido los siguientes:

- El crecimiento económico y la dotación de servicios.
- El dinamismo demográfico y el nivel de urbanización.
- La oposición campo-ciudad e industria-agricultura.

Pues bien -una vez seleccionados los tres componentes mencionados- uno se encuentra con un espacio tridimensional de carácter ortogonal, en el que pueden localizarse los municipios de acuerdo con los valores que observan en los diferentes componentes. Esta es precisamente la gran ventaja del análisis de componentes principales: en lugar de partir de las 35 variables iniciales se parte solamente de estos tres componentes, los cuales con una pérdida mínima de información- permiten agrupar adecuadamente las observaciones. Por consiguiente, en lo sucesivo, la matriz de datos será una matriz de orden (31x23).

El método de agrupación utilizado será el ya mencionado más arriba y que es conocido -en la literatura anglosajona- con el nombre de "cluster analysis", cuyo objetivo principal consiste -valga la insistencia- en obtener agrupaciones de individuos (municipios) a partir de una medida de similitud, de manera que quede garantizada la homogeneidad de los municipios que forman cada una de las agrupaciones y se asegure una clara diferenciación de

las mismas. La cuantificación de estos conceptos se basa en la definición de una medida de distancia entre pares de municipios, haciendo que cuanto menor sea esta distancia mayor será la similitud entre los mismos.

Dado que las dimensiones o componentes utilizados son ortogonales, puede utilizarse una medida de distancia euclídea, a partir de la expresión:

$$d_{tk} = \sum_{i=1}^n (F_{it} - F_{ik})^2{}^{1/2}$$

en donde, d_{tk} es la distancia existente entre los municipios t y k ; F_{it} y F_{ik} son los valores que los municipios t y k observan en el componente F_i , siendo $i = 1, 2, 3$.

A la matriz (312x312), cuyos elementos son las distancias euclídeas normalizadas entre municipios, suele llamársele -por ser ése su cometido fundamental- matriz de similitud.

Existen diferentes procedimientos para la realización del análisis "cluster" o Análisis de Grupos. En este caso, se ha elegido el procedimiento jerárquico aglomerativo, pues no se tenía un conocimiento apriorístico del número de grupos a formar. Los procedimientos jerárquicos proceden estableciendo una jerarquía entre los posibles grupos -en función del grado de homogeneidad de los mismos- y dan lugar a un árbol de relaciones; de tal forma que las sucesivas fusiones de grupos se realizan (considerando los n elementos de partida como n grupos) asegurándose que la distancia intragrupo sea mínima y la distancia intergrupos sea máxima.

Los algoritmos más socorridos para la aplicación del procedimiento anterior -y de uso bastante generalizado en las ciencias sociales- son los debidos a Berry (48) y Ward (49). Pero debe observarse que los resultados que se obtienen no son, en rigor, comar-

(48) Berry, B.J.L., Marble, D.F., *Spatial Analysis: a reader in statistical geography*, Prentice Hall, Englewood Cliffs, 1968, pp. 419-428.

(49) Ward, J.H., op. cit.

cas homogéneas; sino lo que Grigg (50) llama "tipos regionales/comarcales"; es decir, agrupaciones de municipios similares, pero no necesariamente contiguos (51).

Por consiguiente, -y esto es lo que nos parece más novedoso de nuestro estudio- para la obtención de comarcas homogéneas es necesario complementar la matriz de municipios y componentes con otra matriz que recoja la restricción de contigüidad espacial. Con tal finalidad se ha utilizado el algoritmo de Openshaw (52), que consiste, sencillamente, en lo siguiente:

— En primer lugar, se forma una matriz de municipios adyacentes o geográficamente contiguos ("matriz de contigüidades"). Dicha matriz puede construirse de dos formas distintas: asignando a cada municipio las coordenadas geográficas correspondientes o atribuyendo a los municipios (fila y columna de la matriz) contiguos un 1 y a los que no lo son un 0.

— Y, en segundo lugar, se adjunta a la matriz anterior la llamada matriz de similitud.

Como resultado se obtienen, para cada N-1 (siendo $N = 312, \dots, 2$) iteraciones, los grupos de municipios que no solo son similares, sino también contiguos; esto es, comarcas homogéneas.

Sin embargo, -por curiosidad analítica- se ha aplicado, en primer lugar, el algoritmo de Ward (y se han obtenido "tipos regio-

(50) Grigg, D.B., "Regions, models and classes", en Chorley, R.J., Haggett, P., *Models in Geography*. Ed. Methuen. London, 1967, pp. 461-509.

(51) Conviene insistir -ya se ha dicho al principio- que, en rigor, lo que han obtenido Aznar y Cabrer, al agrupar las provincias españolas, en el primer caso, y las comarcas del País Valenciano, en el segundo son "tipos regionales". Los procedimientos utilizados por ambos autores -algoritmo de Ward y algoritmo "centroid", respectivamente- no permiten la obtención de agrupaciones homogéneas que respeten la restricción de contigüidad; es decir, regiones o comarcas homogéneas. Véase Aznar, A., "Infraestructura y regionalización..." y Cabrer, B., Ribera, M.J., op. cit.

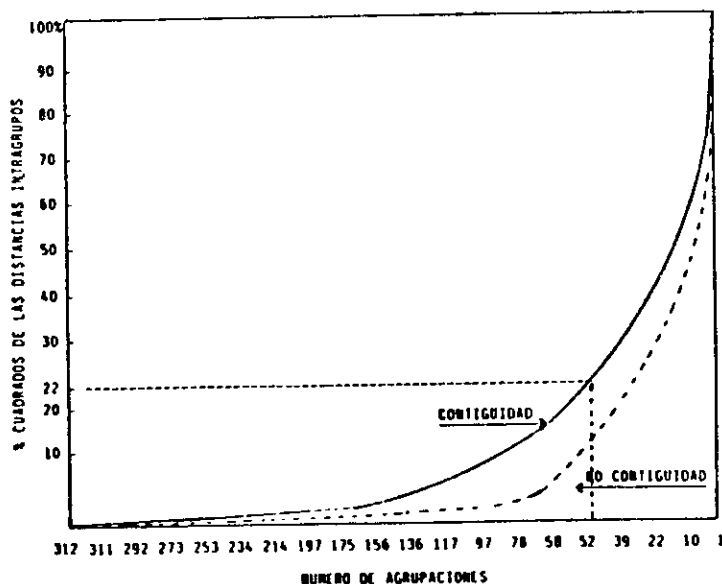
(52) Openshaw, S., "A regionalisation program for large data sets". *Computer Applications*, vol. 1, núms. 3-4, 1974, pp. 136-160. Este programa se ha puesto a punto en el Centro de Cálculo la Universidad de Santiago de Compostela. En este mismo Centro de Cálculo se ha procesado la matriz original de datos, a la que se le ha aplicado el programa de análisis de componentes principales de Dixon, W.J. (Ed.) *Biomedical Computer Programs. EMDP*. University of California Press. Berkeley. Los Angeles, 1975.

nales/comarcales") y, en segundo lugar, el algoritmo de Openshaw (y se ha obtenido una comarcalización homogénea de Galicia).

Comparando ambos "outputs", se puede advertir una mayor "eficiencia" en el caso de los "tipos regionales/comarcales" que en el relativo a las comarcas homogéneas. Por "eficiencia", se entiende aquí, la pérdida de información -de varianza- sufrida en cada itineración del proceso de agrupamiento.

La Figura 2 -en la que, en el eje vertical, se ha representado la suma de los cuadrados de las distancias intragrupos (como porcentaje de la suma de los cuadrados de todas las distancias de la matriz original) y, en el eje horizontal, se han expresado las sucesivas agrupaciones o "clusters" de municipios- permite distinguir la diferente "eficiencia" alcanzada.

FIGURA 2



A la izquierda de este diagrama se identifican los 312 municipios de Galicia totalmente separados unos de otros y, lógicamente, en ese punto, la distancia intra-grupos es cero. De izquierda a derecha, se van formando sucesivamente los diferentes grupos, obser-

vándose que, prácticamente, hasta la constitución de 78 grupos no se alcanza un porcentaje del 10%. Esto se explica fácilmente porque en el caso de los primeros grupos formados los municipios que los constituyen son muy parecidos. Pero también puede constatar que la comarcalización produce, a cada paso, una varianza intra-grupo superior al agrupamiento de "tipos regionales/comarcales" -no contigüedad-, hasta casi identificarse cuando las agrupaciones son muy pocas (10 o menos). En teoría, tal igualdad se obtiene en el momento en que se constituye una sola unidad: Galicia como única región.

CUADRO 7

COMARCAS HOMOGENEAS DE GALICIA

Comarca	Municipios
CEDEIRA	Cedeira, Cerdido, Mañón, Moeche y Ortigueira
AS PONTES DE G.R.	A Capela, As Pontes y Somozas
CURTIS	Aranga, Coirós, Curtis, Irixoa, Monfero, Oza dos Ríos, Sobrado, Vilarmajón y Vilasantar
CARRAL	Abegondo, Carral, Cereda y Cesuras
ARZUA-MELIDE	Arzúa, Bormorto, Melide, Santiso y Toques
ORDES	Frades, mesía, Ordes, Tordola y Trazo
NEGREIRA	A Baña, Negreira, Santa Comba y Val de Dubra
SANTIAGO	Ames, Boqueixón, Brión, O Pino, Oroso, Santiago, Touró y Vedra
PADRON	Dodro, Padrón, Pontecesures, Rois y Teo
CARBALLO	Cabana, Carballo, Coristanco, Laracha, Laxe, Maipica y Ponteceso
CEE	Camariñas, Cee, Corcubión, Dumbria, Fisterra, Muxia, Vimianzo y Zas
NOIA	Carnota, Lousane, Mazaricos, Muros, Noia, Outes y Porto do Son
RIBEIRA	Boré, A Pobra do C., Rianxo y Ribeira
O FERROL	Ares, Cabañas, Fene, O Ferrol, Miño, Mugardos, Narón, Neda, Pontedeume, San Saturniño y Valdoviño
A CORUÑA	Arteixo, Bergondo, Betanzos, Cambre, A Coruña, Culleredo, Oleiros, Paderna y Sada
VIVEIRO	Xove, Orol, Vicedo y Viveiro
CERVO-FOZ	Alfoz, Cervo, Foz y Valdeuro
RIBADEO	Barreiros, A Pontenova, Ribadeo y Trabada
VILLALBA	Abadín, Cospello, Muras, Villalba y Xerndade
MONDOÑEDO	Lourenzana, Mondoñedo, Pastoriza y Riotorto
LUGO	Begonte, Castro de Rei, Castroverde, Corgo, Friol, Guitinz, Guntín, Lugo, Meira, Outeiro de Rei, Palas de Rei, Pol, Portomarín y S. Vicente de Ribade
FONSAGRADA	Beleira, Fonsagrada, Nogueira de Muñiz y Ribeira de Piouira
SARRIA	Incio, Lancara, Paradela, Páramo, Camos y Sarria
BECCERREA	Baralla, Becerreá, Cervantes, Navia de Suarna, Os Negais, Padrafn e Triacastela
CHANTADA	Antas de Ulla, Carballedo, Chantada, Monterroso y Taboada
MONFORTE DE LEMOS	Bóveda, Monforte, Pantón, Pobra de Brión, Savinay y Sobor
QUIROGA	Folgoso de Caurel, Quiroga y Ribas de Sil
O CARBALLIÑO	Beazir, Boborás, O. Carballiño, Irixo, Maside, Piñor, Punxín, S. Cristobal de Cea y San Amaro
OURENSE	Amociro, Barbadáns, Coles, Nogueira de Rasuín, Ourense, Pereiro de Aguiar, A Peroxa, S. Ciprián de Viñas, Trón y Vilasamaria

FUENTE: Elaboración propia.

La comarcalización elegida ha sido aquella que forma 52 comarcas (Véase el Cuadro 7), porque es en ese momento cuando se produce una ruptura en la continuidad de la varianza analizada, produciéndose un salto del 22% al 26%.

CUADRO 7 (Continuación)
COMARCAS HOMOGENEAS DE GALICIA

Comarca	Municipios
ALLARIZ	Allariz, Taboadela y Xunqueira de Ambía.
MACEA	Baños de Molgas, Esgos, Maceda, Paderne de Allariz, Vilar de Barrio y Xunqueira de Espadanedo.
POBOA DE TRIVES	Chandrexa de Queixa, Manzaneda, Poba de Trives y Río.
CASTRO CALDELAS	Castro Caldelas, Montederramo, Parada do Sil y A Teixeira.
O BARCO DE V.	O Barco, O Bolo, Carballada, Larouco, Petín, A Rúa, Rubiana, Vilamartín de Valdeorras.
RIVADAVIA	Arnoia, Beade, Carballada de Avia, Castrelo, Cenlle, Leiro, Melón y Ribadavia.
CELANOVA	A Bola, Cartelle, Celanova, Cortegada, Gome sende, A Merca, Padrenda, Pontedeva, Quintela de Leirado, Ramirán y Vere a.
BANDE	Bande, Calvos de Randín, Entrín, Lobeira, Lovios y Muíños.
XINZO DA L.	Baltar, Blancos, Porqueira, Rairiz da Veiga, Sandiáns, Sarreaus, Trasmiras, Vilar de Santos y Xinzo de Limia.
VERÍN	Castrelo do Val, Cualedro, Laza, Monterrei, Oimbra, Riós y Verín.
VIANA DO BOLO	A Gudiña, A Mezquita, A Veiga, Viana, Vilarinho de Conso y Vilardevós.
VILAGARCIA DE A.	Cambados, Catoira, O Grove, Meis, Ribadumia, Vilagarcía y Vilanova de Arousa.
CALDAS DE REIS	Barro, Caldas de Reis, Campo Lameiro, Moraña, Portas y Valga.
PONTE CALDELAS	Cotobade, Fornelos de Montes, A Lana y Ponte Caldelas.
A ESTRADA	Cercedo, Cuntis, A Estrada y Forcarei.
LALÍN	Dozón, A Golada, Lalín, Rodeiro, Silleda y Vila de Cruces.
A GUARDIA	A Guardia, Oia, O Rosal y Tomiño.
PONTEAREAS	Mondariz, Mondariz Bañeairo, As Neves, Ponteareas, Salceda de Caselas y Salvaterra do Miño.
A CAÑIZA	Arbo, A Cañiza, Covelo y Crecente.
CANGAS	Bueu, Cangas, Marín, Moaña y Vilaboa.
VIGO	Balsosa, Gondomar, Mos, Migrán, Porriño y Vigo.
PONTEVEDRA	Meaño, Poio, Pontevedra y Sanxenxo.
REDONDELA	Pazos de Borbén, Redondela y Somaio.

FUENTE: Elaboración propia.

En general, se ha comprobado que la formación de las sucesivas comarcalizaciones homogéneas, obtenidas a partir de ese momento, se convierten cada vez en menos homogéneas -o más heterogéneas- en los casos en que las ligazones intra-grupos incorporan municipios caracterizados por su fuerte componente urbano e industrial. Este es el caso de la agregación de las comarcas de Ortigueira-As Pontes de García Rodríguez-O Ferrol, o el de las

comarcas de A Coruña-Curtis. Otro tanto puede observarse cuando se agrupan las comarcas de Vigo-Redondela, o Pontevedra-Ponte Caldelas.

Finalmente, es interesante observar que -más allá de la formación de las 52 comarcas homogéneas- las primeras agrupaciones o "clusters" que se forman son los relativos a la Galicia interior, por un lado, y a la Galicia litoral, por otro. En otras palabras, el proceso iterativo de comarcalización pone de manifiesto la existencia de las dos grandes "homogeneidades" de Galicia: la Galicia rural y la Galicia urbana. Es evidente que -a partir de determinado momento- especialmente cuando se unen las regiones de Vigo-Ponteareas-A Cañiza-Ourense, la varianza intragrupo experimenta una fuerte elevación, debido a la incorporación de varias dimensiones muy distintas: el campo y la ciudad, la industria y la agricultura, el crecimiento económico y la dotación de servicios, juntamente con el dinamismo demográfico y el mayor nivel de urbanización.

9. CONCLUSIONES

Muy resumidamente, se pueden extraer las siguientes:

1. Se ha obtenido una comarcalización homogénea de Galicia basada, no en un único índice de carácter natural-geográfico (comarcas naturales) o de tipo agrícola-ganadera (comarcas agrarias), sino en múltiples variables de muy diversa y variada naturaleza. Se debe añadir, además -a modo de comentario- que una comarcalización homogénea, obtenida de esta forma -esto es, con una amplia base de datos y con una metodología probada- asegura un uso eficiente de la asignación espacial de los recursos y ello, por dos razones:

— Por la similitud de problemática existente en varios municipios localizados contiguamente, que permite que éstos puedan ser conjuntamente planeados, administrados o financiados a fin de evitar una duplicación de esfuerzos o un despilfarro innecesario de recursos.

— Y por las economías de escala que conllevan los proyectos públicos realizados por unidades espaciales más amplias que los municipios.

2. Se ha podido comprobar la potencia de la metodología empleada para el tratamiento, clasificación y sistematización de un gran volumen de información estadística atribuible a un gran número de unidades locacionales (municipios).

3. Se ha constatado la utilidad empírica de la utilización conjunta de dos técnicas multivariantes: el análisis de componentes principales y el análisis de grupos.

Se ha podido superar los métodos tradicionales, de carácter esencialmente cartográfico, que se limitaban a un sencillo ejercicio de "percepción visual del investigador". Naturalmente, no se trata de sustituir tales métodos -por lo demás, siempre necesarios-, sino de corregir sus deficiencias y utilizarlos como procedimientos complementarios de metodologías -como las mencionadas- más rigurosas. Porque es evidente que la comparación visual de los mapas, aunque se haga correctamente, no ofrece un medio seguro para determinar las correlaciones entre los fenómenos distribuidos espacialmente.

5. Por último, se ha logrado obtener comarcas homogéneas y no solo unidades homogéneas -o "tipo comarcales"- combinando una matriz de similitud (derivada del análisis de grupos) con una matriz de contigüidad espacial. En este sentido, es conveniente afirmar que, desde el punto de vista informático, el algoritmo de Openshaw ha resultado muy eficaz.

RESUMEN

En este artículo, XOSE MARIA MELLA MARQUEZ realiza una comarcalización homogénea del espacio económico de Galicia. Introduce un elemento conceptual novedoso en este tipo de estudios: la restricción de contigüidad espacial de las unidades territoriales. Analiza, clasifica y sistematiza una amplia matriz de datos de carácter municipal. Constata la potencialidad de la metodología empleada y la utilidad empírica de la utilización conjunta de dos técnicas multivariantes, el análisis de componentes principales y el análisis de grupos. Y, finalmente, a la luz de los resultados obtenidos, plantea la necesidad de distinguir netamente -distinción que, por otra parte, no siempre se hace explícita en la literatura tradicional sobre el tema- dos conceptos bien diferentes: el concepto de comarca homogénea y el concepto de "tipos regionales/comarcales".

RESUME

Dans cet article, José Maria Mella Marquez, réalise un découpage homogène selon les cantons, de l'espace économique de la Galicie.

Il introduit un élément conceptuel nouveau pour ce genre d'étude: la restriction des contigüités spatiales des unités de territoire.

Il analyse, classifie et systématise une large matrice de données statistiques obtenues à l'échelle de la municipalité.

Il constate le potentiel que renferme la méthodologie employée et l'utilité empirique de l'utilisation conjointe des deux techniques multivariantes: l'analyse des principaux composants et l'analyse des groupes.

Et finalement à la lumière des résultats obtenus il pose la nécessité de distinguer nettement -distinction qui par ailleurs n'est pas toujours explicite dans la littérature traditionnelles sur ce sujet- entre deux concepts bien différents: le concept de cantons homogènes et le concept de types " régionaux / cantonnaux ".

SUMMARY

In this article, Xose Maria Mella Marquez carries out an homogeneous division into Counties of the Galician economic space. He introduces a new, in this type of studies, conceptual element: the restriction of spatial contiguity of territorial units. He analyses, classifies and systematises a large matrix of data of municipal nature. The validity of the methodology used and the empirical usefulness of the joined application of two multivariant techniques -analysis of principal components and analysis of groups- are highlighted. Finally, in the light of the results obtained, he poses the need to make a clear distinction between two well differentiated concepts: the concept of homogeneous County and the concept of "regional/County types" - a distinction which, on the other hand, is not always explicit in the traditional literature on this subject.

