

Design of Homogenous Territorial Units. A Methodological Proposal and Applications

Juan Carlos Duque Cardona

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tesisenxarxa.net) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tesisenred.net) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tesisenxarxa.net) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Departamento de Econometría Estadística y Economía Española
UNIVERSIDAD DE BARCELONA

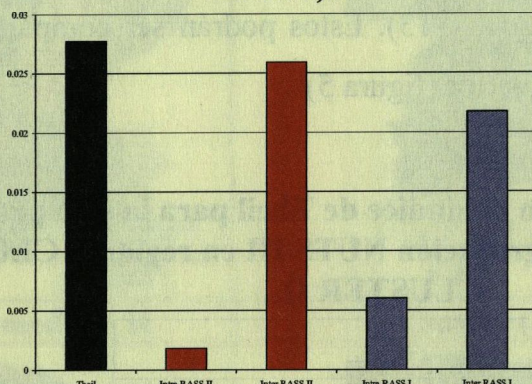
DESIGN OF HOMOGENOUS TERRITORIAL UNITS.
A Methodological Proposal and Applications.

Juan Carlos Duque Cardona

MAYO 2004

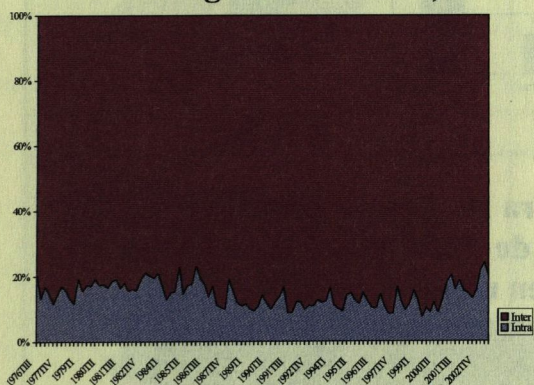
Tesis Doctoral para optar al título de Doctor en Estudios Empresariales
Director: Dr. Manuel Artís Ortuño
Doctorado en Estudios Empresariales
Programa: Técnicas y Análisis en la Empresa
Bienio 2000-2002

Figura 11. Descomposición del índice de Theil para la tasa promedio de desempleo 1976 – 2003. (agregación NUTS III en regiones RASS II y RASS I).



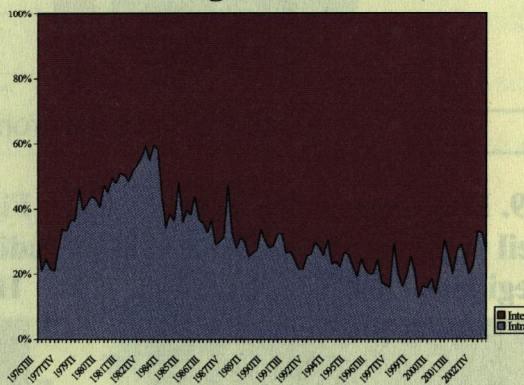
Fuente: Elaboración propia.

Figura 12. Descomposición del índice de Theil (agregación NUTS III en regiones RASS II).



Fuente: Elaboración propia.

Figura 13. Descomposición del índice de Theil (agregación NUTS III en regiones RASS I).



Fuente: Elaboración propia.

Los resultados muestran que el grado de desigualdad interna se reduce considerablemente con la utilización de los métodos analíticos, así, con el K-medias la proporción de desigualdad interna con respecto a la desigualdad entre grupos representa sólo el 4.68% y 11.98% para niveles de agregación CLUSTER II y CLUSTER I. Para el RASS, dichos porcentajes son del 6.54% y 21.64% para RASS I y RASS II respectivamente. Aunque para el RASS esta proporción es más elevada que para el caso del CLUSTER, se debe tener en cuenta que con éste último método, un importante número de regiones están compuestas por una provincia lo cual tiende a disminuir considerablemente el nivel de desigualdad interna.

Para ambos casos de agregación analítica cabe resaltar que el grado de desigualdad interna es superior a mayores niveles de agregación.

Por último, el cálculo del índice de Theil para cada uno de los trimestres disponibles muestra como la proporción de desigualdad intra y entre grupos presenta menores variaciones a lo largo del período analizado si se compara con los resultados obtenidos con las divisiones normativas NUTS. Esta homogeneidad temporal también se ve afectada al incrementar el nivel de agregación. Estos resultados muestran que con los métodos de agregación analíticos las regiones obtenidas ven menos afectadas por los posibles cambios en las estructuras de las regiones diseñadas, lo cual las hace más representativas a lo largo del período.

5.4. Comentarios finales

A continuación se presenta una comparación entre los resultados obtenidos con los métodos analíticos, K-medias y *RASS*, en función de diferentes características propias de los procesos de regionalización:

Homogeneidad: ambos métodos mejoran considerablemente el grado de homogeneidad a lo largo del período analizado. Para ambos métodos de agregación, se obtienen menores valores de desigualdad interna en las regiones, situación que el caso del K-medias se puede ver acentuado por el hecho de que 7 de las 15 regiones diseñadas (CLUSTER II) están conformadas por una sola provincia.

Forma de las regiones: con respecto a las formas que adoptan las regiones en los diferentes niveles de agregación, se puede observar como las formas conseguidas tras la aplicación del K-medias son mucho más irregulares que las obtenidas con el *RASS*. Con el K-medias a un nivel (CLUSTER II) hay regiones formadas por una provincia y otras por 9. A nivel CLUSTER I el rango varía entre 1 y 17 provincias. Para el caso

del RASS las regiones obtenidas son más equilibradas en términos del número de provincias por región, así, para el RASS II el número de provincias por región varía entre 2 y 4 y para RASS I el rango varía entre 5 y 11.

Control sobre el número de regiones: Una de las mayores desventajas vistas a la hora de aplicar el método cluster (K-medias) en dos etapas, es que no se tiene un control total sobre el número de regiones que se desean configurar, así, para el nivel de agregación CLUSTER I, fue imposible obtener 6 regiones, pues para un parámetro de $k=3$ se obtuvieron 7 regiones en la segunda etapa. Mientras que para un $k=2$ se obtuvieron 2 regiones en la segunda etapa. Por este motivo se decidió trabajar con 7 regiones a un nivel CLUSTER I. Dichos problemas no se presentan con el algoritmo RASS, pues el número de regiones a conformar esta determinado explícitamente por el investigador.

Así pues, tras los resultados obtenidos, es evidente que los investigadores no pueden estar ajenos a la sensibilidad de los resultados frente a diferentes métodos de regionalización utilizados y a las diferentes especificaciones del nivel de agregación.

6. Conclusiones y futuras líneas de investigación

En la presente tesis se han propuesto nuevas metodologías para realizar procesos de regionalización a partir de la agregación de unidades territoriales (áreas) teniendo en cuenta no sólo sus características, sino también, las relaciones que se establecen entre ellas. Para ello, se ha formulado un modelo de optimización lineal el cual, a partir de una matriz de continuidades y una matriz de relaciones, encuentra la agrupación óptima de las áreas en un número predeterminado de regiones. El criterio utilizado para definir la calidad de las agrupaciones es la medida de heterogeneidad dentro de cada una de las regiones configuradas, medida que debe ser minimizada con el fin de obtener regiones que contengan áreas muy similares. La posibilidad de plantear el problema de regionalización como un modelo lineal permite asegurar, por sus propiedades matemáticas, que la región factible es convexa y que, por tanto, es posible encontrar una solución óptima. Otra ventaja de este tipo de formulación es que su implementación es posible en una gran variedad de *software* comercial.

La metodología propuesta tiene, además, una gran capacidad para identificar configuraciones territoriales complejas en cuanto a su forma, así pues, se consigue formular un modelo que cumpla con la restricción de continuidad entre las áreas que pertenecen a una región, pero sin que la metodología utilizada para cumplir con esta restricción condicione de alguna manera la forma que pueden adoptar dichas regiones.

También se ha presentado un algoritmo, con el nombre de *RASS* (“*Regionalization Algorithm with Selective Search*”), que permite aumentar la capacidad de cómputo del modelo de optimización aprovechando los beneficios de aplicar la optimización directa en una porción del territorio que varía en cada ejecución gracias a una estrategia de búsqueda selectiva que define el esquema de entrada y salida de regiones. Estas características dotan al algoritmo de una buena capacidad de escapar de óptimos locales.

Por último, dos diferentes métodos para el diseño de regiones analíticas fueron aplicados, K-medias en dos etapas y *RASS*, utilizando las tasas de desempleo de las provincias españolas para el período 1976-2003. Los resultados muestran como las regiones analíticas tienen un mayor grado de homogeneidad y que dicha homogeneidad presenta un mayor grado de estabilidad a lo largo del período.

A continuación, se presentan las futuras líneas de investigación derivadas de la tesis que se han agrupado en dos apartados: teórico y empírico:

Teórico:

1. Formular funciones objetivo adicionales, apropiadas para otros contextos de regionalización.
2. Endogeneizar el número de regiones a diseñar, de tal forma que el modelo determine el número óptimo de regiones. Algunos indicadores han sido propuestos en el ámbito estadístico con el objetivo de determinar el número apropiado de grupos. Sin embargo, dichos estadísticos no se encuentran vinculados a ningún modelo de regionalización.
3. Explorar las implicaciones teóricas de los procesos de regionalización en el campo de la econometría espacial. ¿Existe algún vínculo entre los procesos de regionalización y la presencia de autocorrelación espacial?

Empírico:

1. Formular nuevos heurísticos que incorporen nuevas características útiles en otros contextos de regionalización geográfica.
2. Buscar nuevas aplicaciones de los modelos propuestos en análisis de tipo estructural, donde el objetivo final sea el diseño de regiones, o funcional, donde las regiones diseñadas representan un paso intermedio para un análisis econométrico aplicado.

7. Referencias

- Alonso, J. y Izquierdo, M. (1999): 'Disparidades regionales en el empleo y el desempleo', *Papeles de Economía Española*, 80, 79-99.
- Amrhein, C. G. y Flowerdew, R. (1992): 'The effect of data aggregation on a Poisson regression model of Canadian migration', *Environment and Planning A*, 24, 1381-91.
- Aurenhammer, F. (1991): 'Voronoi diagrams - a survey of a fundamental geometric data structure', *ACM Computing Surveys*, 23, 345-405.
- Dantzing, G. B. y Ramser, J. H. (1959): 'The truck dispatching problem', *Management Science*, 6, 80-91.
- Fisher, M. M. (1980): 'Regional taxonomy', *Regional Science and Urban Economics*, 10, 503-37.
- Fotheringham, A. S. y Wong, D. W. S. (1991): 'The modifiable areal unit problem in multivariate statistical analysis', *Environment and Planning A*, 23, 1025-44.
- Gordon, A. D. (1996): 'A survey of constrained classification', *Computational Statistics & Data Analysis*, 21, 17-29.
- Gordon, A. D. (1999): *Classification*. segunda edición, Boca Raton [etc.]: Chapman & Hall/CRC.
- Gower, J. C. y Legendre, P. (1986): 'Metric and euclidean properties of dissimilarity coefficients', *Journal of Classification*, 3, 5-48.
- Horn, M. E. T. (1995): 'Solution techniques for large regional partitioning problems', *Geographical Analysis*, 27, 230-48.
- Kohonen, T. (1984): *Self-Organisation and Associative Memory*. Berlin [etc.]: Springer.
- Laport, G. y Osman, I. H. (1995): 'Routing problems: A bibliography', *Annals of Operations Research*, 61, 227-62.
- López-Bazo, E., del Barrio, T. y Artís, M. (2002): 'The regional distribution of Spanish unemployment: A spatial analysis', *Papers in Regional Science*, 81, 365-89.
- Macmillan, B. y Pierce, T. (1994): 'Optimization modelling in GIS framework: the problem of political redistricting', en Fotheringham, S. y Rogerson, P. (eds.), *Spatial analysis and GIS*, London [etc.]: Taylor & Francis, pp 221-46.

- Maravalle, M. y Simeone, B. (1995): 'A spanning tree heuristic for regional clustering', *Communications in Statistics - Theory and Methods*, 24 (3): 625-39.
- Moran, P. (1948): 'The interpretation of statistical maps', *Journal of the Royal Statistical Society B*, 10, 243-251.
- Murtagh, F. (1985): 'A survey of algorithms for contiguity-constrained clustering and related problems', *The Computer Journal*, 28 (1): 82-8.
- Ohsumi, N. (1984): 'Practical techniques for areal clustering', en Diday, E., Jambu, M., Lebart, L., Pagès, J. y Tomassone, R. (eds.), *Data analysis and informatics*, Vol III, North-Holland, Amsterdam, pp 247-58.
- Openshaw, S. (1977): 'Algorithm3: a procedure to generate pseudo random aggregation of N zones into M zones where M is less than N', *Environment and Planning A*, 9, 1423-28.
- Openshaw, S. (1984): The modifiable areal unit problem, *Concepts and Techniques in Modern Geography*, 38 (GeoBooks, Norwich).
- Openshaw, S. (1992): 'Some suggestions concerning the development of artificial intelligence tools for spatial modelling and analysis in GIS', *The Annals of Regional Science*, 26, 35-51.
- Openshaw, S. y Taylor, P. J. (1981): 'The modifiable areal unit problem', in Wrigley, N. and Bennett, R. J. (eds.), *Quantitative Geography*, London, pp 60-70.
- Openshaw, S. y Wymer, C. (1995): 'Classifying and regionalizing census data', en Openshaw, S. (eds.), *Census Users Handbook*, Cambridge, UK: Geo Information International, pp 239-70.
- Perruchet, C. (1983): 'Constrained agglomerative hierarchical classification', *Pattern Recognition*, 16, 213-17.
- Theil, H. (1967). *Economics and Information Theory*, Chicago: Rand McNally and Company.
- Wise, S. M., Haining, R. P. y Ma, J. (1997): 'Regionalisation tools for exploratory spatial analysis of health data', en Fisher, M. M. y Gentis, A. (eds.), *Recent Developments in Spatial Analysis: Spatial statistics, behavioural modelling, and computational intelligence*, Berlin [etc.]: Springer, pp 83-100.

EXCIÒS DE
PRESTEC