

LAS REGIONES EUROPEAS: UNA TIPOLOGÍA BASADA EN LA APLICACIÓN DE TÉCNICAS MULTIVARIANTES

Bárbara Díaz Díez
Eugenia Cruces Pastor
Antonio Morillas Raya
Universidad de Málaga

En este trabajo se utiliza la información estadística referente a los indicadores regionales, suministrada por Eurostat, para elaborar una clasificación y caracterizar el espacio socioeconómico europeo en base a los mencionados indicadores, que, como se sabe, se refieren básicamente a datos demográficos, de mercado de trabajo y producto interior bruto. Para conseguir tal objetivo, se realiza una aproximación metodológica en dos fases: en primer lugar se utiliza un análisis *cluster*, con los indicadores regionales más relevantes, para evaluar las posibilidades de clasificación y número de grupos a formar. En una segunda etapa, se aplica un análisis discriminante con un doble objetivo: evaluar la bondad de las distintas clasificaciones derivadas del método anterior y encontrar los indicadores más significativos para diferenciar los diferentes espacios socioeconómicos obtenidos. Por último, se analizan las características socioeconómicas de las distintas agrupaciones regionales obtenidas.

Palabras clave: Estudios regionales, análisis cluster y análisis discriminante.

1. INTRODUCCIÓN

El presente trabajo trata de clasificar y caracterizar a las regiones europeas en grupos relativamente homogéneos, establecidos en base a una serie de indicadores suministrados por Eurostat, y señalar las variables más relevantes en tal agrupación.

Se analizan 11 de los 15 países que en la actualidad pertenecen a la Unión Europea. Se excluyen del estudio Grecia y el este de Alemania, por no disponer de información suficiente, y los estados de Austria, Suecia y Finlandia de

los que se carece de datos debido a su reciente integración. Dentro de las naciones consideradas no se han tenido en cuenta los indicadores relativos a Ceuta y Melilla (España), los Departamentos de Ultramar (Francia) y Azores y Madeira (Portugal), dada la peculiaridad de estos territorios.

Las unidades territoriales estadísticas (NUTS) sobre las que se realiza este estudio son las de nivel 1 y nivel 2, ya que parece ser que las diferencias regionales son menos obvias cuando se estudia la situación de las regiones de nivel 3. Excepto para España y Portugal, países en los que se ha utilizado el nivel 2, en el resto de los estados miembros se ha trabajado sobre el nivel 1, debido a la falta de información sobre ciertas variables en algunos de ellos.

2. EL ESPACIO SOCIOECONÓMICO EUROPEO

La imagen de la Unión Europea se caracteriza por las diferencias regionales, tanto dentro de cada país como entre los distintos Estados miembros. Bien es cierto que las disparidades entre las regiones son mayores que las que hay entre las naciones, por una simple razón: el mayor nivel de desagregación.

Las sucesivas ampliaciones que ha ido registrando la Unión Europea, han supuesto un aumento de estas disparidades intrarregionales. Ello ha sido debido, en gran parte, a que en los distintos Estados los problemas regionales existentes no eran lo mismo de relevantes. De los seis países fundadores, sólo Italia atribuía a la política regional un peso importante, pues era la única que presentaba serias diferencias entre sus regiones; casi todos los indicadores de desarrollo económico presentaban dos zonas bien diferenciadas: la del norte, desarrollada y dinámica, y la del sur, atrasada y periférica. La situación empeora con la entrada del Reino Unido, que también poseía una política regional fuerte, Irlanda, país tan pobre como el sur de Italia, y Dinamarca, la que menos problemas planteó. La posterior incorporación de Grecia agrava considerablemente la situación, ya que supuso que aumentase la ratio entre la región más rica y más pobre. La integración de Portugal y España, cuya política regional, con más o menos relevancia, ha estado siempre presente dentro de la política económica general, modificó substancialmente el mapa regional comunitario, en cuanto que supuso un aumento del peso de las regiones cuyo PIB/habitante estaba por debajo de la media comunitaria (Camagni, R.; Chesire, P. C.; Cuadrado, J. R. y Gaudemar, J. P.; 1991, pág. 27; y Cuadrado, J. R. y Mancha, J., 1995, pág. 48).

Para corregir los principales desequilibrios regionales, la Unión Europea potencia las medidas nacionales de política regional y trata de coordinarlas, como se sabe, con ayudas financieras adicionales con cargo a los fondos comunitarios. Estos fondos, como se sabe, centran su aplicación en cinco objetivos prioritarios que tratan de beneficiar a los países con mayores desequilibrios regionales. Los objetivos número 1, 2 y 5a son de carácter territorial, mientras que los restantes, número 3, 4 y 5b, afectan a todo el territorio comunitario. La adhesión a la Comunidad de Austria, Finlandia y Suecia ha motivado que se cree un nuevo objetivo, el número 6, con características similares a las del objetivo número 1.

Es bien conocido que las regiones incluidas en el objetivo número 1 son aquéllas con un PIB per cápita inferior al 75% de la media comunitaria. En

esta situación se encuentran, entre otras, todas las regiones de Grecia, Portugal e Irlanda, gran parte de las de España, y las del sur de Italia. Sólo dos países, Dinamarca y el Gran Ducado de Luxemburgo, no tienen ninguna región incluida en este objetivo. La población cubierta supone el 26,6% del total comunitario. Se caracterizan estas regiones por tener un gran sector agrícola con baja productividad y alto desempleo.

En el otro extremo, hay regiones con un PIB per cápita muy superior a la media comunitaria. Se incluyen Hamburgo, Darmstadt, Bremen, Oberbayen y Stuttgart en Alemania, Groningen en Holanda, Île de France en Francia, Greater London en el Reino Unido y Lombardy y Emilia Romagna en Italia. La mayoría de estas 10 regiones con economías metropolitanas, son clasificadas en un eje (llamado «banana azul»), que se extiende desde Greater London hasta Bélgica y Holanda a lo largo del Rin y hacia el interior del norte de Italia, llegando hasta Emilia Romagna. (Dunford, M., 1993, pág. 731).

El principal criterio para incluir a las regiones dentro del objetivo número 2, como se sabe, es que su índice de desempleo sea superior al de la media comunitaria, con lo cual la ayuda concedida con arreglo a este objetivo va encaminada a la reducción de las disparidades en las cifras de desempleo. La proporción de población cubierta es del 16,8%. Se encuentran regiones pertenecientes a Alemania (Berlín, Limburgo), Francia (Provenza, Ródano-Alpes), España (Cataluña, Madrid, País Vasco), Italia, Reino Unido.

El objetivo 5b, persigue fomentar el desarrollo de las zonas rurales. Se encuentran incluidas las regiones del sur de Italia, España y Portugal, East-Anglia y South-West en el Reino Unido, Lemosin y Champagne-Ardenne en Francia, Dinamarca y el Gran Ducado de Luxemburgo. La población a la que alcanza este objetivo es del 8% del total comunitario.

A pesar del aumento de la dotación financiera de los fondos con respecto a los primeros años y la concentración de sus intervenciones en las regiones menos desarrolladas de la Comunidad, su impacto ha sido muy limitado y, a veces, contraproducente para el desarrollo equilibrado de algunas regiones (Fontela, E. y Morilla, A., 1991), dada la escasa importancia relativa del presupuesto comunitario en comparación con el PIB de la Comunidad.

Los últimos datos reflejan, sin embargo, que la Unión Europea sigue estando muy lejos de ser una unidad integrada. Su diversidad interna es enorme. Las diferencias nacionales entre los países con niveles más bajos (Grecia, Portugal e Irlanda) y los que alcanzan los más altos (Luxemburgo —Gran Ducado— y Alemania occidental) plantean un reto a la nueva Europa que se quiere construir. El desequilibrio regional más señalado aparece en Italia donde sigue existiendo la histórica división norte-sur; sigue Francia donde se distingue entre París y el resto del país; España en la que se separa entre el noroeste y Madrid y el resto de las regiones, y el Reino Unido donde la mayor separación se traduce entre Greater London y el Home Counties y el resto del país. En contraposición las diferencias regionales son mínimas en Alemania occidental y Bélgica. Los hechos muestran que estas diferencias regionales son especialmente resistentes a su reducción (Dunford, M., 1993, págs. 733 y 734).

En diversos estudios sobre la convergencia histórica de las economías europeas, siempre se pone de manifiesto que dicha convergencia es más fuerte

para los países del centro que para los países de la periferia. La noción de situación periférica se utiliza, por lo general, para explicar el hecho de que algunas regiones fracasan repetidamente en sus intentos de alcanzar una situación similar a la de regiones con una ubicación más próxima al centro. Las regiones centrales parecen poseer una solidez natural mucho mayor que les facilita la adaptación a los cambios estructurales. Las regiones fronterizas de los Estados miembros suelen ser más pobres que el resto del país, aunque ello no es una tendencia uniforme. Los indicadores muestran que las regiones fronterizas densamente pobladas se encuentran en el centro de la Unión y que las escasamente pobladas se encuentran en la periferia. El PIB per cápita de las regiones fronterizas neerlandesas y belgas es mayor que el de las regiones alemanas vecinas, lo que contrasta con la situación a nivel nacional. La renta per cápita de las regiones fronterizas de España y Francia es muy similar, a pesar de que el PIB medio per cápita español es sólo dos tercios del francés (Comisión de las Comunidades Europeas, 1995, cap. 7).

El grado hacia la convergencia en términos reales que está experimentando las distintas regiones es lento y muy variable. Este acercamiento es menos dinámico en las regiones más atrasadas de la Unión, pertenecientes a España, Irlanda, Grecia y Portugal, que en las zonas más ricas, oeste y sur de Francia, sur de Alemania y países del Benelux. A pesar de todo ello, la evolución experimentada en algunas regiones sugiere que es posible, a largo plazo, conseguir una mayor integración (Comisión de las Comunidades Europeas, 1995, cap. 7).

Para valorar la intensidad de los problemas regionales, la mayoría de los estudios son realizados bajo una visión *unidimensional* de los indicadores. A lo sumo, como emplea la Comisión europea, utilizando un índice sintético que otorga arbitrariamente el mismo peso a las desviaciones del PIB y del desempleo con respecto a la media de la Unión (Camagni, R.; Chesire, P. C.; Cuadrado, J. R. y Gaudemar, J. P., 1991, pág. 25). El indicador más común de la desigualdad regional es el PIB por habitante. Son también utilizados el índice de desempleo, el tiempo medio de desplazamiento por carretera, ferrocarril o avión a cada uno de los grandes centros urbanos, etc.

En este trabajo, sin embargo, se pretende clasificar y caracterizar a las regiones en base a un estudio *multivariante*, usando simultáneamente un conjunto de indicadores demográficos, de mercado de trabajo y PIB por habitante, expresado en paridad de poder de compra, más adecuado para detectar proximidades y diferencias en bienestar.

3. INDICADORES UTILIZADOS Y METODOLOGÍA

En los últimos años, principalmente como consecuencia del desarrollo de la electrónica y la informática personal, se ha avanzado considerablemente en las aplicaciones prácticas de las técnicas multivariantes de análisis de datos. Dentro de esta rama de la Estadística, los trabajos iniciados por Fisher en 1936 acerca de la clasificación y discriminación han llegado a convertirse en auténticos modelos, que mejoran en gran manera a la capacidad humana para clasificar o discriminar.

No es nuestro propósito en este apartado realizar un análisis detallado de los distintos métodos utilizados, al tratarse de técnicas sobradamente utilizadas y conocidas, sino de dar una sencilla descripción del método utilizado en este trabajo, así como de detallar qué algoritmos concretos de resolución se han elegido.

Análisis «*cluster*», o de conglomerados, es un nombre genérico utilizado para identificar una gran variedad de procedimientos usados para establecer una clasificación. Estos métodos parten de un conjunto de datos con información relevante acerca de una serie de casos (en nuestro caso regiones) e intentan reorganizarlos en grupos homogéneos, en base a esa información. Así, el análisis cluster permite la utilización de diferentes algoritmos de clasificación. La mayoría de ellos son procedimientos relativamente simples y en muchas ocasiones no están basados en un cuerpo de razonamiento estadístico extenso. Un análisis cluster siempre situará a los elementos en distintos grupos, que además arrojarán resultados distintos dependiendo del método concreto de agrupación utilizado, por lo que la clave para aplicar con propiedad el análisis cluster es reconocer cuándo estos grupos son reales y no meras imposiciones del método a los datos. Es por esta razón que para la validación estadística de los resultados obtenidos, además de para evaluar la capacidad de discriminación entre los grupos que tienen las distintas variables, hayamos realizado un análisis discriminante.

La tarea de clasificar las regiones europeas supone la definición previa del conjunto de variables a utilizar. Como se ha dicho, estas variables se refieren a datos demográficos, de mercado de trabajo y producto interior bruto. En concreto, los indicadores considerados son: tasa de actividad total (TACTO), tasa de actividad de las mujeres (TMU), grado de dependencia o relación entre activos e inactivos (GRD), tasa de desempleo total (TDT), tasa de desempleo para las mujeres (TDM) y para los hombres (TDV), ocupación en agricultura (OAGR), ocupación en industria (OIND), ocupación en servicios (OSER), y PIB por habitante, en paridad de poder de compra (PIB/H P). Estos indicadores se encuentran en el anuario estadístico «Regiones» de Eurostat 1994, en el que se recopilan las estadísticas comparables más características de la situación social y económica de las regiones de la Unión Europea. Todas las variables son referentes al período 91, salvo la tasa de desempleo, que se encuentra disponible para 1993.

En la aplicación que presentamos, hemos utilizado un método de clasificación en dos fases: en primer lugar se ha utilizado una clasificación jerárquica (que no necesita de una definición previa del número de grupos) para determinar el número idóneo de grupos, y posteriormente se ha realizado una clasificación no jerárquica (en la que el investigador debe proponer el número de grupos), a partir de la información suministrada por el primer método.

Para el estudio de las similitudes o proximidades entre las distintas regiones se ha utilizado la distancia Euclídea. Como se sabe, esta distancia es sensible a la diferencia de escalas entre las variables, ya que en general aquéllas con mayor dispersión tendrán más impacto en el valor de similitud, por lo que se ha optado por la estandarización de las variables, de acuer-

do con la teoría al uso, para eliminar el sesgo introducido por las diferencias de escala¹.

El método jerárquico de clasificación comienza de manera que cada caso forma un grupo. En el primer paso, la distancia entre casos viene dada por la distancia Euclídea, pero cuando más adelante se trata de unir clusters hay que especificar qué regla de agregación se va a utilizar. En nuestro caso, utilizaremos el método de Ward. Este método utiliza una aproximación al análisis de la varianza para evaluar la distancia entre clusters, intentando minimizar la suma de los cuadrados de los residuos de cada dos hipotéticos clusters que pueden ser formados en cada paso.

La clasificación no jerárquica (K-means) se realiza a partir de la información proporcionada por el *dendograma* resultado del cluster jerárquico. Puede ser considerada como un análisis de la varianza «hacia atrás», en el sentido de que el test de significación para la hipótesis de que las medias de los grupos son diferentes en ANOVA evalúa la variabilidad entre grupos sobre la variabilidad intra grupo. En este método de clasificación, se intentará mover los elementos o casos de un grupo a otro para conseguir los resultados ANOVA más significativos.

Una vez obtenida la clasificación, se realiza un análisis discriminante con fines de evaluación y de explicación: evaluación de la bondad de la agrupación y explicación de las variables relevantes en la discriminación entre grupos. Se trata de determinar la contribución de cada variable clasificadora a la clasificación correcta de cada una de las regiones, así como determinar el grupo al que pertenecería un elemento (región) para el que se conocen los valores de las variables. Esto último permitirá que se pueda determinar qué regiones han sido mal clasificadas y evaluar la clasificación realizada por el análisis cluster. Se trata de una técnica parecida a la regresión, sólo que, en este caso, la variable dependiente es una variable categórica. El criterio para la obtención de la función discriminante es la maximización del ratio variabilidad entre grupos/variabilidad intra grupos.

4. RESULTADOS DEL ANÁLISIS

Se ha comentado que cuando se necesita clasificar una gran cantidad de información en grupos manejables y con significado, el análisis cluster es de gran utilidad. Dado que el objetivo es crear grupos relativamente homogéneos, en el análisis de un dendograma resultante de una agrupación jerárquica, donde se sugieren agrupaciones crecientes en número de casos, la cuestión es dónde «cortar» el gráfico de manera que se obtenga el número óptimo de grupos. Por otro lado, en un análisis cluster no jerárquico, se requiere al investigador

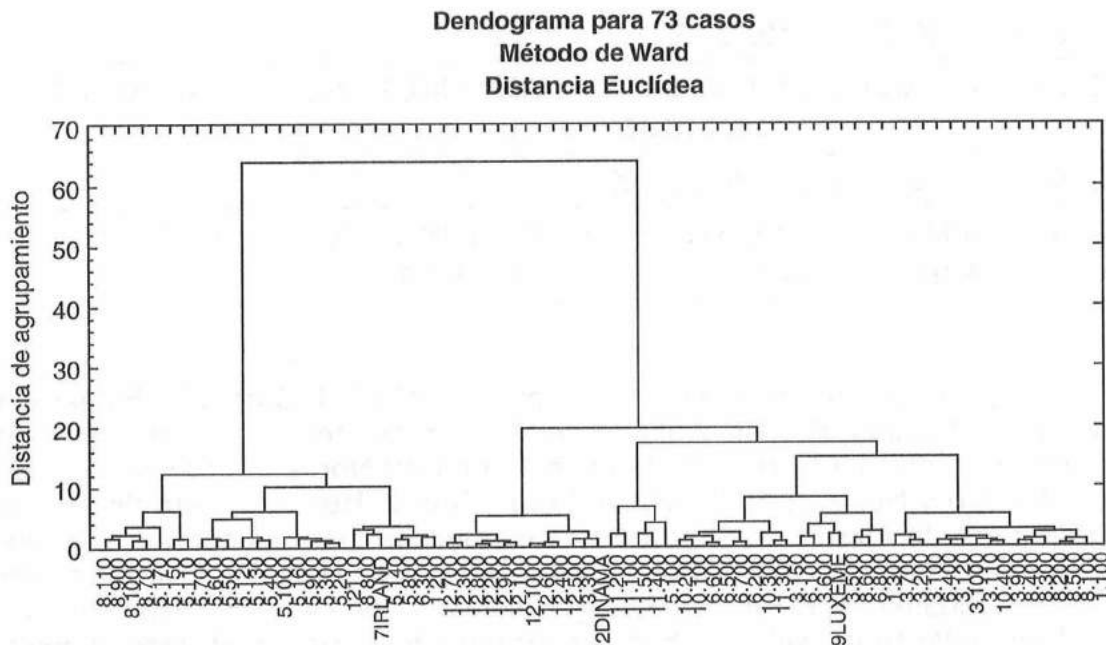
(1) Una medida alternativa a la distancia Euclídea, que incorpora directamente un procedimiento de estandarización, es la distancia de Mahalanobis, que, además, realiza un ajuste para tener en cuenta las correlaciones entre las variables, ya que un conjunto de variables altamente correlacionadas en el análisis cluster puede, implícitamente, sobreponderar ese conjunto de variables en el proceso de agrupamiento. El problema es que esta medida, que sería la más apropiada, no la incluye el software utilizado (STATISTICA), aunque sí otros paquetes como por ejemplo, el SAS.

especificar el número de grupos presentes en los datos, lo que requiere un procedimiento iterativo de prueba y error.

Gran parte de la dificultad para resolver este problema es la ausencia de una hipótesis nula, debido a la falta de una definición consistente de la estructura y contenido de un cluster. Una posible hipótesis nula sería: «no estructura» en un conjunto de datos, pero esta está lejos de ser una proposición clara y fácilmente manejable. En cualquier caso, los procedimientos heurísticos son los métodos más utilizados para decidir el nivel de «corte» en un dendograma, mediante la inspección subjetiva de los diferentes niveles del árbol.

El dendograma que resulta en la clasificación jerárquica de las regiones europeas en base a las variables especificadas, utilizando la distancia Euclídea y el método de Ward, se presenta en el gráfico 1.

Gráfico 1 RESULTADOS DEL CLUSTER JERÁRQUICO



Se ha trabajado con dos cortes en el dendograma: uno a la distancia de unión 12, que sugiere formar 6 clusters, y otro a una distancia inferior, con el que se tomarían 11 clusters. Por resultar mucho más manejable a la hora de interpretar los clusters, y, especialmente, por la buena respuesta al análisis discriminante que se realiza posteriormente, se trabajará con los resultados procedentes de la primera agrupación, aunque se comentarán algunos resultados obtenidos con 11 clusters.

En base a estos resultados, se ha realizado un análisis cluster no jerárquico, para 6 grupos. Las regiones que contiene cada cluster pueden verse en el cuadro 1. Dinamarca (2), Irlanda (7) y Luxemburgo (9) no se encuentran divididos en regiones. Las que comienzan por el número 1 pertenecen a Bélgi-

ca, las que comienzan por el número 3 pertenecen a Alemania, las que comienzan por el número 5 pertenecen a España, las regiones que comienzan por el número 6 pertenecen a Francia, por el 8 a Italia, por el 10 a Holanda, por el 11 a Portugal, y por el 12 al Reino Unido. Para más detalle, puede observarse en el anexo 1, la numeración correspondiente a cada región. Observando el cuadro 1 se ve como han quedado los grupos, que en una primera aproximación podemos describir como sigue.

Cuadro 1
RESULTADOS DEL ANÁLISIS CLUSTER.
REGIONES QUE INTEGRAN CADA GRUPO

Grupo 1	1.200, 5.500, 5.800, 5.120, 5.140, 6.200, 6.300, 6.500, 6.600, 6.800, 8.800, 10.100, 10.200, 12.110
Grupo 2	1.100, 3.100, 3.200, 3.700, 3.900, 3.1000, 3.110, 3.120, 6.400, 6.700, 8.100, 8.200, 8.300, 8400, 8.500, Luxemburgo, 10.400
Grupo 3	1.300, 3.500, 3.600, 6.100, 8.600, 10.300
Grupo 4	Dinamarca, 3.300, 3.150, 11.300, 12.100, 12.200, 12.300, 12.400, 12.500, 12.600, 12.700, 12.800, 12.900, 12.1000
Grupo 5	11.100, 11.200, 11.400, 11.500
Grupo 6	5.100, 5.200, 5.300, 5.400, 5.600, 5.700, 5.900, 5.1000, 5.110, 5.130, 5.150, 5.160, 5.170, Irlanda, 8.700, 8.900, 8.1000, 8.110

El grupo uno está formado, en esencia, por Madrid, Cataluña, Baleares y Navarra en España, Región Wallone de Bélgica, Oeste y parte mediterránea de Francia, Norte y Oeste de Holanda e Irlanda del Norte. El grupo dos lo forman: Centro y Norte de Italia, Luxemburgo, Sur de Holanda, Este de Francia y gran parte de Alemania. El grupo tres incluye las grandes capitales: Bruselas, Bremen, Hamburgo, Île de France, Lazio y Oeste de Holanda. El grupo cuatro lo forman Dinamarca, Berlín, Schleswig-Holstein, Lisboa y Reino Unido. El grupo 5 lo forman Portugal, salvo Lisboa; y el grupo 6 lo compone el resto de España, Irlanda y Sur de Italia. Para una mejor localización geográfica de las regiones que conforman cada grupo, puede consultarse el mapa presentado en las conclusiones.

Para la validación estadística de los resultados obtenidos se ha realizado un análisis discriminante, utilizando como valores de la variable categórica (variable dependiente) los números identificativos asignados a cada región, de acuerdo con los resultados del análisis cluster realizado anteriormente. La información estadística resultante puede contemplarse en el cuadro 2.

La *landa de Wilks* es un estadístico que expresa la significación estadística del poder de discriminación del modelo. Su valor está comprendido entre 0 y 1. Un valor de landa igual a cero significa perfecto poder de discriminación. En la cabecera y segunda columna del cuadro anterior, se encuentran,

Cuadro 2
RESUMEN DEL ANÁLISIS DISCRIMINANTE
 N.º DE VARIABLES EN EL MODELO: 5 (6 GRUPOS)
 LANDA DE WILKS: 0.00831 APROX. F (25.235) = 24.787 $p < 0.0000$

N = 73	Landa de Wilks	Landa parcial	F salida	Nivel de significación p	Tolerancia	1-Toler.
GRD	0,014	0,584	8,960	0,000	0,763	0,237
TDM	0,019	0,447	15,576	0,000	0,792	0,208
OAGR	0,016	0,509	12,140	0,000	0,805	0,195
OIND	0,013	0,633	7,320	0,000	0,875	0,125
PIB/HP	0,023	0,356	22,753	0,000	0,765	0,235

respectivamente, la landa de Wilks para el modelo final y la obtenida en el proceso iterativo, para después de que la respectiva variable (columna primera) ha entrado en el modelo. Como puede observarse, todas son relativamente cercana a cero, por lo que puede afirmarse que los resultados son plenamente satisfactorios desde el punto de vista estadístico. La *landa parcial de Wilks* (tercera columna) es la landa de Wilks para la contribución única de la respectiva variable a la discriminación entre grupos. Obviamente, cuanto más baja sea, mayor será esa contribución. La landa de Wilks puede convertirse en una F, por lo que se puede calcular el valor de significación de cada variable (columnas 4 y 5), pudiéndose comprobar la alta significación estadística de esa contribución en todos los casos.

Como puede observarse, no todas las variables seleccionadas para el análisis cluster permanecen finalmente en el modelo explicativo. Cuando el análisis se realiza paso a paso, se selecciona la variable con mayor contribución única a la discriminación entre grupos, es decir que tenga la mayor F. Por otro lado, en cada etapa del análisis se calcula la correlación múltiple (R^2) entre las variables que se han incluido en el modelo. El valor de tolerancia, que se calcula como $1 - R^2$, se considera como una medida de la redundancia de la nueva variable. Un valor de tolerancia muy bajo significa alta correlación, y por tanto redundancia de la variable que se incorpora. En nuestro caso, las variables que han quedado fuera del modelo final, presentaban una tolerancia inferior a 0.01 (más de un 99% de redundancia con las demás variables).

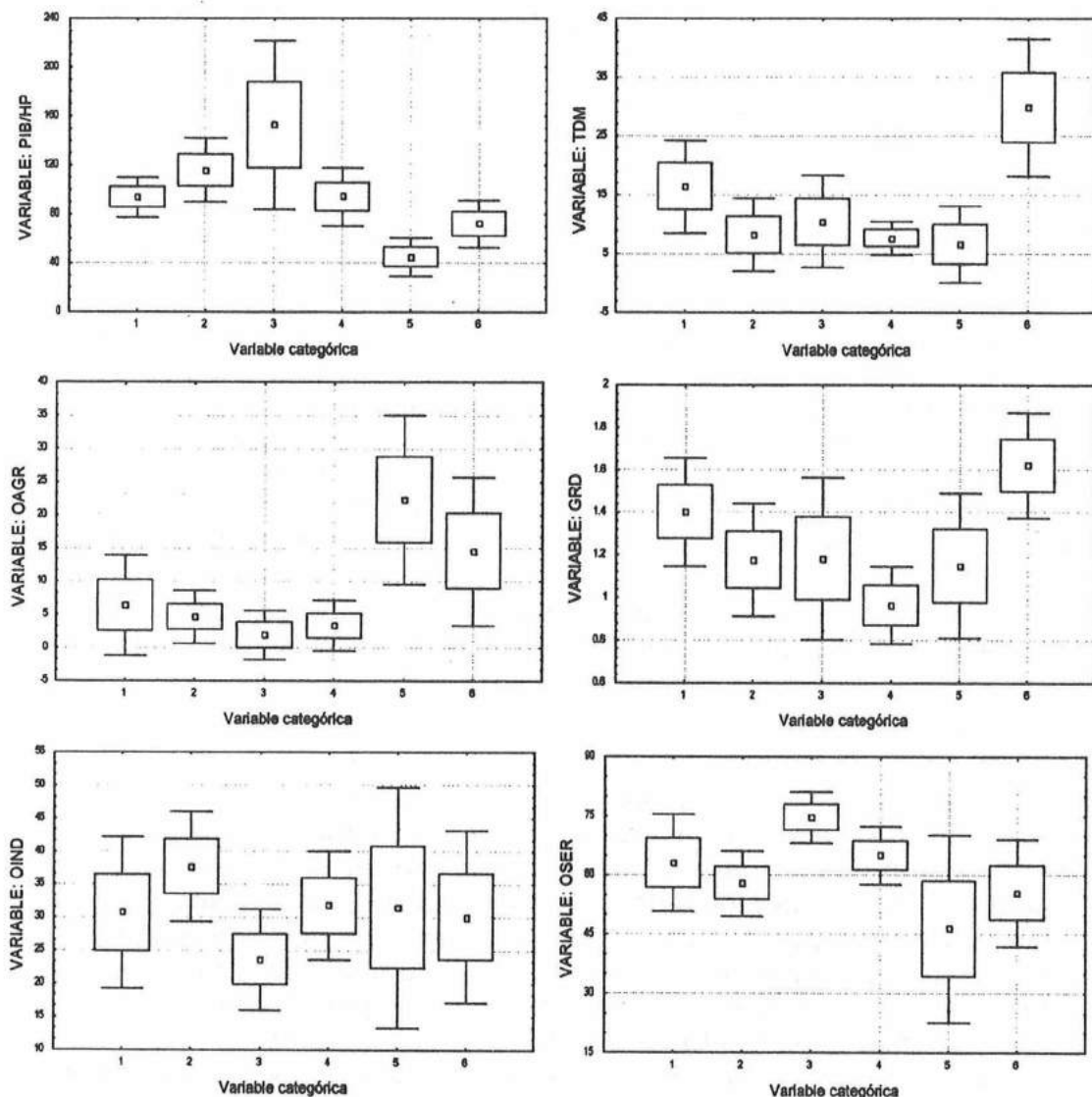
En resumen, las variables finalmente incorporadas en el modelo, ordenadas de mayor a menor contribución a la discriminación entre los seis grupos, son las siguientes:

1. PIB/habitante en paridad de poder de compra.
2. Tasa de desempleo de las mujeres
3. Ocupación agraria
4. Grado de dependencia
5. Ocupación industrial

Las características básicas de estas variables (media, dispersión y asimetría) en cada uno de los seis grupos, así como las de la variable ocupación

en servicios, que nos servirá posteriormente para interpretar algunas cuestiones, pueden contemplarse en el gráfico 2.

Gráfico 2
GRÁFICOS DE CAJA POR GRUPOS



Pero tan importante, al menos, como la significación estadística, comentada anteriormente, es la evaluación del agrupamiento realizado en base a la matriz de clasificación². Esta matriz, como se sabe, informa sobre el porcen-

(2) La finalidad de este estudio es descriptiva y no de predicción. En caso de haber querido utilizar el modelo para predecir, podría haberse optado por dos de estos procedimientos: dejar parte de la muestra para la validación del modelo, de manera que se predicen los grupos a los que pertenecerían estos elementos posteriormente y se ve el % de clasificados correctamente, o bien, se hacen n especificaciones del modelo, cada una de ellas con $n-1$ elementos, es decir, dejando fuera cada vez un elemento y prediciendo su valor, para posteriormente ver el % de clasificados correctamente.

taje de casos clasificados correctamente en cada grupo, haciendo uso de las funciones clasificadoras resultantes del análisis. Los resultados se presentan en el cuadro 3.

Cuadro 3
MATRIZ DE CLASIFICACIÓN.
FILAS: CLASIFICACIÓN OBSERVADA,
COLUMNAS: CLASIFICACIÓN PREDICHA

Grupos	% correctos	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6
Grupo 1	85,7142868	12	0	0	2	0	0
Grupo 2	94,1176453	0	16	0	1	0	0
Grupo 3	83,3333359	0	0	5	1	0	0
Grupo 4	100	0	0	0	14	0	0
Grupo 5	100	0	0	0	0	4	0
Grupo 6	100	0	0	0	0	0	18
TOTAL	94,520546	12	16	5	18	4	18

Puede verse que casi el 95% de las regiones están bien clasificadas. *Sólo hay cuatro regiones mal clasificadas*, todas ellas en el grupo 4. Del grupo 1 hay dos regiones: Irlanda del Norte y Este de Holanda; del grupo 2 se clasifica mal el sur de Holanda; y, finalmente, del grupo 3, el Oeste de Holanda. Ninguna región de los grupos 4, 5 y 6 aparece mal clasificada. El éxito del agrupamiento parece, pues, evidente.

Al realizar el análisis discriminante, se pueden obtener las distancias de Mahalanobis entre los distintos grupos (véase el cuadro 4). Obsérvese, que los grupos 5 y 6 son los que más se separan del resto, muy especialmente del grupo 3. Los grupos 1 y 2, son los más próximos entre sí. El grupo 1 está, por otro lado, más próximo al 4 y al 6 que al 3 o al 5. El grupo 2 se aproxima más al 3 y al 4 que al 5 ó 6. Finalmente, el grupo 5, está más próximo al 4 que el grupo 6 y, sin embargo, más alejado del 3 que éste. Toda esta información será de gran interés para caracterizar las macroregiones obtenidas, cuestión que se hará en las conclusiones de este trabajo.

Cuadro 4
DISTANCIAS DE MAHALANOBIS ENTRE GRUPOS

Grupos	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6
Grupo 1	0	6,878	24,911	17,689	36,630	20,712
Grupo 2	6,878	0	18,786	11,660	40,638	44,438
Grupo 3	24,911	18,786	0	35,521	92,641	69,321
Grupo 4	17,689	11,660	35,521	0	35,178	64,626
Grupo 5	36,630	40,638	92,641	35,178	0	53,604
Grupo 6	20,712	44,438	69,321	64,626	53,604	0

Como es bien sabido, para clasificar a los seis grupos son necesarias cinco funciones discriminantes. Todas ellas resultan ser altamente significativas (la más baja, al 0,2%). Se han obtenido los coeficientes estandarizados para cada una de estas funciones, con objeto de identificar las bases de su capacidad de discriminación. En la primera función tiene un mayor peso la tasa de desempleo de la mujer (véase el cuadro 5). Siguiendo con este razonamiento, puede verse que la segunda función discriminante está, básicamente, determinada por el PIB/habitante, en paridad de poder de compra. La tercera por la ocupación en agricultura, estando muy próximas la tasa de paro femenino y el grado de dependencia, la cuarta por la ocupación industrial y la quinta por el grado de dependencia.

Cuadro 5
DISTANCIAS DE MAHALANOBIS ENTRE GRUPOS

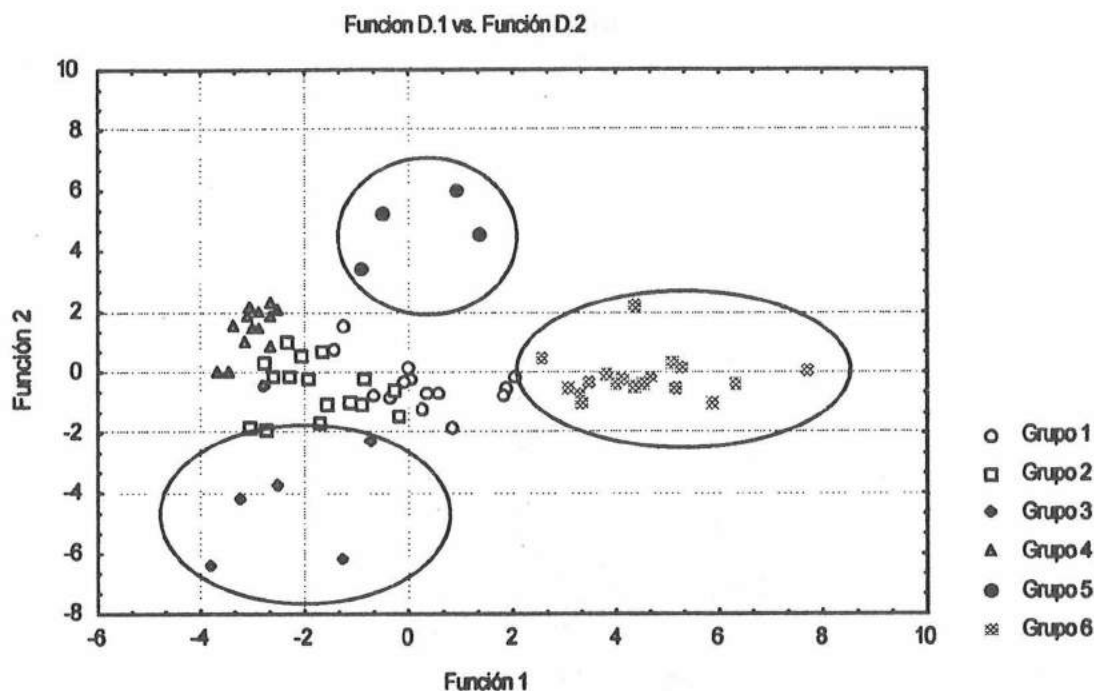
Variable	F.D. 1	F.D. 2	F.D. 3	F.D. 4	F.D. 5
GRD	0,398	-0,455	-0,611	-0,161	0,739
TDM	0,715	-0,209	0,607	-0,089	-0,575
OAGR	0,588	0,111	-0,665	0,483	-0,457
OIND	0,287	0,228	-0,455	-0,790	-0,421
PIB/HP	0,015	-0,986	-0,382	0,207	-0,381
Valor propio	8,563	3,275	0,749	0,448	0,162
% Acum. VARIANZA EXPLICADA	0,649	0,897	0,954	0,988	1,000

En el cuadro 5 también puede observarse los valores propios para cada función discriminante, así como la proporción acumulada de la varianza explicada para cada función. La primera función explica un 64,9% de la discriminación entre grupos, mientras que la segunda explica un 24,8%, la tercera el 5,7%, la cuarta el 3,4% y la quinta el 1,2%. Entre las dos primeras, explican el 89,7% y en ellas pesan fuertemente, como se ha visto, la tasa de desempleo de la mujer y el PIB/habitante, en paridad de poder de compra. Su representación gráfica puede verse en el gráfico 3. Para obviar dar la tabla de medidas de las variables canónicas, a la hora de explicar la discriminación entre grupos, haremos los comentarios de los respectivos papeles discriminantes sólo sobre estas dos funciones y sobre el propio gráfico.

Obsérvese que la función uno discrimina entre el grupo 6 y los demás, en especial del 3 y del 4, mientras que la función dos lo hace entre los grupos 5 y 3. En la tabla de medidas de las variables canónicas, no presentada aquí, las demás funciones presentan menores diferencias en las puntuaciones de los distintos grupos, pero puede decirse que, aunque la discriminación es menor, la función tres discrimina entre el grupo 4 y el 5, la función cuatro entre los grupos 3 y 5 y los demás, y la función cinco, aunque con mucha menos diferencia, el grupo 1 del resto.

Resumiendo, la discriminación más clara la hacen las funciones una y dos, donde se discrimina, básicamente en base a las variables tasa de desempleo

Gráfico 3 REPRESENTACIÓN DE LOS GRUPOS Y LA DISCRIMINACIÓN CON LAS FUNCIONES 1 Y 2



de la mujer y PIB/habitante, en paridad de poder de compra. Estas dos funciones contribuyen sobre todo a la separación de los grupos 3, 4, 5 y 6. Los grupos 1 y 2, sin embargo, presentan casos más o menos próximos a los demás, excepto al grupo 5, que está muy bien diferenciado. Veamos, seguidamente, cuáles son las conclusiones que pueden extraerse de todos estos resultados.

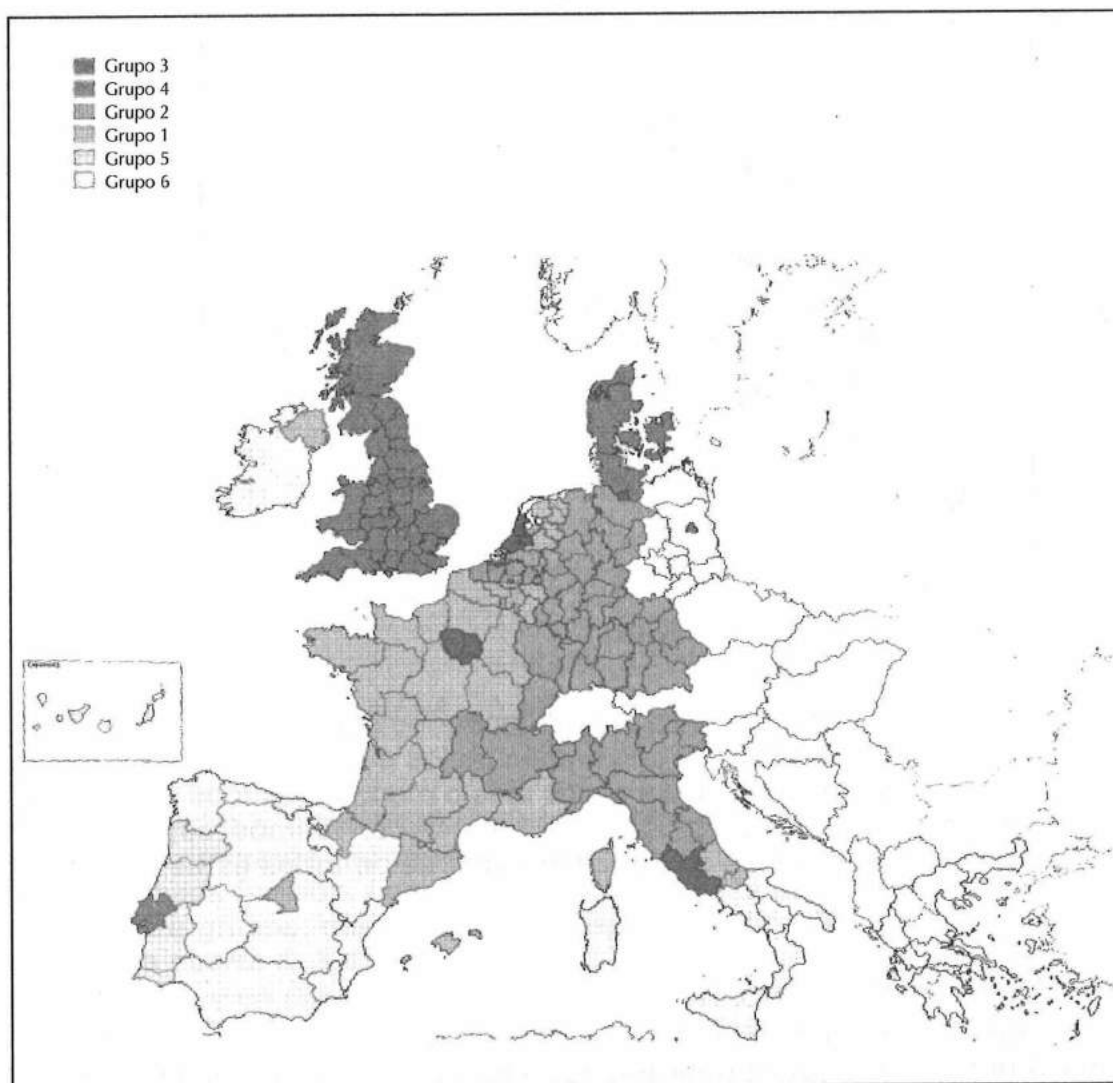
5. CONCLUSIONES

Parece evidente que la primera conclusión que hay que resaltar es el alto poder diferenciador que resulta tener la tasa de desempleo de la mujer. Así como el empleo de la capacidad adquisitiva (PIB/habitante) en la valoración de las desigualdades regionales es algo casi rutinario en los estudios relacionados con las políticas de cohesión social, no parece que este indicador, referente a la posición de la mujer en el mercado de trabajo y, por tanto, en la propia sociedad, haya merecido la atención que su capacidad discriminadora entre espacios comunitarios parece poner de relieve, según los resultados en este trabajo. Con esto, obviamente, no negamos la existencia de políticas específicas por parte de la Unión Europea y de los propios estados miembros, sino que queremos subrayar la necesidad de potenciarlas en aras a una mayor igualdad de oportunidades entre las mujeres de las distintas regiones europeas.

En segundo lugar, hay que señalar que, junto con las variables ya mencionadas, son la estructura de la producción y el grado de dependencia (tasa de actividad, en cierta forma) las que mejor explican las diferencias regionales en la Unión Europea. Tales diferencias dan lugar, según los resultados de este

trabajo, a seis grupos de regiones, cuyas características más relevantes son las siguientes (véase el mapa 1).

Mapa 1



Grupo 1: formado por un conjunto de regiones, situadas ligeramente por debajo de la media en poder adquisitivo, caracterizadas por su orientación básica a los servicios, poca importancia agraria y nivel aceptable en la industria. Su mercado de trabajo presenta una muy baja tasa de actividad, alta tasa de dependencia y una importante tasa de desempleo (entre el 10% y el 16%, en su mayoría), tanto global como femenino. Podrían caracterizarse como *regiones de desarrollo medio, con importantes problemas en el mercado de trabajo* y son, como ya se adelantó, las regiones más desarrolladas de España (Madrid, Cataluña, Baleares y Navarra), gran parte de Francia, excepto la región de París y algunas del Este, Irlanda del Norte y la región de Abruzy-Molise, en Italia.

Grupo 2: integrado por regiones con poder adquisitivo superior a la media comunitaria (segundas en este indicador), la mayor ocupación industrial, sin

mucho peso la de los servicios, y bajas tasas de desempleo (menos del 8%). Son las regiones, que podrían llamarse *industriales*, del centro económico e industrial europeo: la mayor parte de Alemania, Este de Francia y Norte de Italia.

Grupo 3: formado por las *grandes metrópolis*, con su actividad concentrada en los servicios (más del 75%, en general), la más baja ocupación industrial y prácticamente nula actividad agraria, pero con el mayor poder adquisitivo de toda la Unión Europea (entre 120 y 190). Su tasa de desempleo es ligeramente superior a la del grupo anterior (en torno al 10%). Se trata de regiones en las que hay, generalmente, una ciudad emblemática, en algún sentido, en el contexto europeo: París, Bruselas, Roma, Bremen, Hamburgo y Amsterdam.

Grupo 4: se trata de un conjunto de regiones realmente especial, por cuanto su rasgo común y, a su vez más diferenciador respecto al resto de grupos, es su alta tasa de actividad (menor grado de dependencia) global y, sobre todo, de la mujer, con una tasa de paro similar al grupo anterior, pero siendo inferior la de la mujer (menos del 8%) que la del hombre (más del 10%). Su actividad se reparte entre la industria y los servicios (segundo grupo en ambos indicadores), su poder adquisitivo está próximo a la media europea. Este grupo de regiones, que se caracterizan por su *buena situación en el mercado de trabajo y el equilibrio sectorial en la producción*, está formado por el Reino Unido (excepto Irlanda del Norte), Dinamarca, Berlín y Schleswing-Holstein en Alemania y la región de Lisboa, en Portugal, que ha tenido una evolución realmente favorable en los últimos años.

Grupo 5: está formado sólo por cuatro regiones cuyos rasgos diferenciados más importantes son el tener la más baja capacidad de compra (menos de 60) y la más alta ocupación en el sector agrario (entre el 18% y el 28%). No es de extrañar, por tanto, que sea el más distanciado del grupo 3. Sin embargo, su alta tasa de actividad, con una baja tasa de paro global y femenino, la menor de todas, le separa del grupo 6, al que podría aproximarse, en principio, por los dos primeros indicadores citados. Podrían caracterizarse como *regiones de base agraria, de pobreza relativa, pero sin tensiones en el mercado de trabajo*, y ocupan todo Portugal, excepto Lisboa.

Grupo 6: este último grupo, caracterizado por *sus graves problemas en el mercado de trabajo y su atraso económico*, está formado por Irlanda, el resto de España, la no presente en el grupo 1, y el sur de Italia. Se caracterizan estas regiones por tener, con mucha diferencia, la menor tasa de actividad (mayor grado de dependencia) y el mayor índice de paro, especialmente grave en la mujer (superior al 25%). Tienen una importante actividad agraria y, tanto los servicios como la industria, no alcanzan el nivel medio de desarrollo del resto de la Unión Europea. Su poder de compra está entre el 60% y el 80% (segundo más bajo, después del grupo 5). Su mayor distancia es al grupo 4, probablemente a causa de la baja tasa de actividad y el alto desempleo, especialmente entre la población femenina.

Hay que señalar, finalmente, que una mayor desagregación, en 11 grupos, que se ha ensayado, no arroja mucha más información que la diferenciación de Andalucía y Extremadura del resto de la España del grupo 6, del Algarve en Portugal, del sureste del Reino Unido (Greater London) y el arco mediterráneo en Francia.

ANEXO

NUMERACIÓN DE LAS DISTINTAS REGIONES

1	BELGIQUE	5,700	Aragón	8,900	Sud
1,100	Vlaams gewest	5,800	Madrid	8,1000	Sicilia
1,200	Région Wallone	5,900	Castilla-León	8,110	Sardegna
1,300	Bruxelles	5,1000	Castilla-La Mancha	9	LUXEMBOURG
2	DANMARK	5,110	Extremadura	10	NEDERLAND
3	BR DEUTSCHLAND	5,120	Cataluña	10,100	Noord-Nederland
3,100	Baden-Württemberg	5,130	Comunidad Valenciana	10,200	Oost-Nederland
3,200	Bayern	5,140	Baleares	10,300	West-Nederland
3,300	Berlín	5,150	Andalucía	10,400	Zuid-Nederland
3,400	Brandenburg	5,160	Murcia	11	PORTUGAL
3,500	Bremen	5,170	Canarias	11,100	Norte
3,600	Hamburg	6	FRANCE	11,200	Centro
3,700	Hessen	6,100	Île de France	11,300	Lisboa e Vale do Tejo
3,800	Mecklenburg	6,200	Bassin Parisien	11,400	Alentejo
3,900	Niedersachsen	6,300	Nord-Pas-de-Calais	11,500	Algarve
3,1000	Nordrhein-Westfalen	6,400	Est	12	UNITED KINGDOM
3,110	Rheinland-Pfalz	6,500	Ouest	12,100	North
3,120	Saarland	6,600	Suz-Ouest	12,200	Yorkshire and Humberside
3,130	Sachsen	6,700	Centre-Est	12,300	East Midlands
3,140	Sachsen-Anhalt	6,800	Méditerranée	12,400	East Anglia
3,150	Schleswig-Holstein	7	IRELAND	12,500	South East
3,160	Thüringen	8	ITALIA	12,600	South-West
5	ESPAÑA	8,100	Nord Ovest	12,700	West Midlands
5,100	Galicia	8,200	Lombardia	12,800	Nort-West
5,200	Asturias	8,300	Nord Est	12,900	Wales
5,300	Cantabria	8,400	Emilia-Romagna	12,1000	Scotland
5,400	País Vasco	8,500	Centro	12,110	Northern Ireland
5,500	Navarra	8,600	Lazio		
5,600	Rioja	8,700	Campania		
		8,800	Abruzzi-Molise		

REFERENCIAS BIBLIOGRÁFICAS

- Aldenderfer, M. S. y Blashfield, R. K. (1989): *Cluster analysis*, California, Sage University Paper.
- Camagni, R.; Chesire, P. C.; Cuadrado, J. R. y Gaudemar, J. P. (1991): «Las políticas regionales de la Comunidad Europea: pasado y futuro», *Estudios Regionales*, núm. 35.
- Comisión de las Comunidades Europeas (1995): *Quinto informe periódico de la Comisión sobre la situación y evolución socioeconómica de las regiones de la Comunidad*, Luxemburgo.
- Comission of the European Communities (1991): *Europe 2000: Outlook for the Development of the Community's Territory*, DGXVI, CEC, Brussels.
- Comission of the European Communities (1991): *The Regions in the 1990s. Fourth Periodic Report of the Socio-economic Situation and Development of the Regions of the Community*, DGXVI, CEC, Brussels.
- Cuadrado, J. R. (1991): «Las disparidades regionales en la Comunidad Europea y en España», *Economía Pública*, núm. 12.
- Cuadrado, J. R. y Mancha, T. (1995): «Política regional comunitaria: ventajas e implicaciones para España», *Economistas*, núm. 13.
- Cuadrado, J. R. y Suárez-Villa, L. (1992): «Integración económica y evolución de las disparidades regionales», *Papeles de Economía española*, núm. 51.
- Fontela, E. y Morillas, A. (1991): «El impacto de la política regional en Andalucía: una evaluación del Marco Comunitario de Apoyo», *Jornadas Escenarios europeos sobre la evolución tecnológica y la cohesión económica y social en la Comunidad Europea*.
- Hair, J. F. (1995): *Multivariate data analysis with readings*, New Jersey, Prentice Hall International.
- Iglesias, A.; Cantos, J. M. y García, A. (1995): «Las regiones frente a la integración europea. Realidad y coherencia de las políticas», *Ciudad y Territorio. Estudios Territoriales*, núm. 104.
- Lázaro, L. (1992): «La política regional comunitaria y los fondos estructurales ante el mercado único», *Estudios Territoriales*, núm. 38.
- Parlamento Europeo (1991): *Efectos regionales de las políticas comunitarias. CECA-CEE-CEEA*, Bruselas, Luxemburgo.

ABSTRACT

In this article is used the statistic information of regional indicators, given by Eurostat, in order to make a classification and to characterize the European socioeconomic space with those indicators that, as it is well-known, are basically from demography, labor market and GDP. To get that goal, we have tried a methodological approximation in two phases: first, he have applied an analysis cluster with the most relevant of the regional indicators, so as to evaluate the classification possibilities and to know how many groups could be recognized. Secondly, a discriminant analysis has been applied to both evaluate the suitability of the classification made by the first method, and find the most meaningful indicators in the task of differentiating the groups obtained. At last, we have studied the socioeconomic characteristics of those different regional clusters.

Key words: regional studies, cluster analysis and discriminant analysis.