

# ‘Clusters’ de Regiões na União Europeia

**Cristina del Campo**, Dpto. de Estadística e I.O. II, Universidad Complutense de Madrid, Campus de Somosaguas, 28223 – Madrid, Spain;  
E-mail: [campocc@ccee.ucm.es](mailto:campocc@ccee.ucm.es); Tel: +34 91 394 29 15

**Carlos M. F. Monteiro**, CEG-IST, Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Rovisco Pais, 1049-001 Lisbon, Portugal;  
E-mail: [cmont@mail.telepac.pt](mailto:cmont@mail.telepac.pt); Tel: +351 218417981

**João O. Soares**, CEG-IST, Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Rovisco Pais, 1049-001 Lisbon, Portugal;  
E-mail: [joaosouares@ist.utl.pt](mailto:joaosouares@ist.utl.pt); Tel: +351 218417207

## Resumo

Como é sabido, existem grandes diferenças de desenvolvimento entre as regiões que compõem a União Europeia. Essas diferenças têm aumentado com os sucessivos alargamentos, em particular com o recente alargamento a Leste. Atenta a estas diferenças, a Comissão Europeia estabeleceu um conjunto de prioridades para o período 2007-2013, que passam por apoiar diferenciadamente as regiões de acordo com a sua inserção em quatro grupos definidos com base no seu PIB *per capita*: *convergence regions*; *competitiveness and employment regions*; *phasing-in regions*; *phasing-out regions*. Este artigo faz uso de um conjunto mais rico de variáveis disponibilizadas pelo European Union Statistical Office (Eurostat) para agrupar as regiões da União Europeia. Mais concretamente, são utilizadas variáveis económicas, demográficas, de emprego e nível educativo, ao nível das regiões NUTS 2, as quais são sujeitas a análise estatística multivariada. A partir desta análise são identificados grupos de regiões com características semelhantes, sendo esta classificação confrontada com a referida classificação (univariada) da Comissão Europeia.

**Keywords:** desenvolvimento regional, indicadores regionais, estatística multivariada, política europeia.

**Classificação JEL:** R1; R58; O52; C1

## 1. Introdução

Foram múltiplos os passos dados até à integração europeia dos nossos dias: a criação da Comunidade Europeia do Carvão e do Aço (CECA) em 1951; a subsequente concretização em 1957 da Agência Europeia da Energia Atómica (EURATOM) e da Comunidade Económica Europeia (CEE); a entrada em vigor em 1993, após ratificação pelos países, do Tratado instituindo a União Europeia (UE); finalmente, o início da circulação de uma nova moeda, o Euro, em Janeiro de 2002.

Este processo de integração económica e política foi acompanhado pela progressiva adesão de novas nações, ampliando a diversidade cultural e linguística no espaço da União e, simultaneamente, ampliando desigualdades económicas e sociais no quadro das regiões europeias. Em particular, o último alargamento a vinte e cinco países trouxe uma descida do rendimento médio *per capita* em 12.5% com uma duplicação das disparidades económicas e sociais (Barnier, 2004).

De forma a combater essas assimetrias regionais, a Comissão Europeia anunciou já a intenção de continuar a alocar importantes recursos financeiros às políticas regionais (cerca de 1/3 do seu orçamento), tendo estabelecido as suas prioridades para o novo período 2007-2013 (veja-se European Commission, 2004). A primeira prioridade respeita à convergência das regiões e países menos desenvolvidos, englobando nomeadamente as regiões com um PIB *per capita* inferior a 75% da média comunitária – antigo Objectivo 1 (trinta e três *convergence regions* da Europa dos quinze – UE-15 – e trinta e sete *convergence regions* dos novos Estados membros – Figura 1). A segunda prioridade é a competitividade e o emprego (*competitiveness and employment regions*), concentrada nos problemas da sustentabilidade do crescimento das regiões já mais desenvolvidas (acima dos 75% da média comunitária), nomeadamente nas vertentes da Inovação e da Economia do Conhecimento. Para além dos dois grupos de regiões acima identificados com as respectivas prioridades, a Comissão destacou dois outros grupos que poderemos classificar como intermédios ou de transição. As *phasing-in regions* (onze regiões da UE-15 e uma região húngara – Figura 1) correspondem às regiões que só recentemente ultrapassaram o limiar do Objectivo 1, indo ter acesso a uma fatia própria do orçamento comunitário para assegurar a sua competitividade. O quarto e último grupo – *phasing-out regions* – compreende dezasseis regiões da UE-15, as quais, não fora o alargamento para vinte e cinco países com a consequente descida do PIB



comunitária. Ora, pode-se questionar o facto de esta visão da Comissão não levar em linha de conta outras dimensões socio-económicas e não permitir avaliar diferentes potenciais de desenvolvimento regional associados, por exemplo, à existência de populações mais jovens, com maior taxa de actividade; à maior densidade populacional com o seu efeito de atracção demográfica; ou um nível educacional superior. Essa visão não integra ainda diferentes realidades quanto ao desemprego. Acontece que há uma distinção clara entre regiões de convergência que, como algumas regiões de Leste, manifestam ainda sinais do seu atraso económico, mas contam com uma dinâmica demográfica e um elevado nível de formação escolar, e regiões desertificadas, deprimidas, dos membros mais antigos da União Europeia.

Veja-se a tabela 1, que apresenta as cinco regiões com maiores e menores valores de Produto, desemprego, densidade populacional e percentagem de activos com formação ao nível do secundário. Quanto ao Produto, as cinco regiões mais pobres são todas do leste, mais especificamente da Polónia; as mais ricas correspondem essencialmente a regiões urbanas da Europa do Norte. A taxa de desemprego aparece elevadíssima em regiões polacas, mas também em regiões ultramarinas francesas, enquanto é muito baixa numa região de convergência portuguesa – os Açores. A baixa densidade populacional é sobretudo marcante em regiões de países mais ricos da Europa – Finlândia, Suécia, Reino Unido e França. Por fim, quanto à percentagem de população com nível de formação educativa ao nível do secundário, factor marcante na qualificação e consequente produtividade da mão-de-obra, verificamos que os valores mais elevados são todos de regiões pobres (de convergência) de Leste, enquanto os valores mais baixos se referem a regiões também de convergência da EU15 – quatro de Portugal e uma de França (Corse).

Face a estes resultados, parece pois mais adequado proceder a uma segmentação das regiões europeias numa óptica multivariada, desafio a que se propõe este artigo. Concomitantemente analisar-se-ão as potenciais implicações de uma diferente classificação das regiões europeias tendo em vista a referida alocação de recursos financeiros no próximo horizonte orçamental de 2007-2013. Para esta análise multivariada serão utilizados dados socio-económicos publicados pelo *European Union Statistical Office* (Eurostat) ao nível das NUTS2, nível de desagregação do território

também usado no quadro da política regional europeia para a afectação dos recursos financeiros.

**Tabela 1 – Regiões com maiores e menores valores de Produto, desemprego, densidade populacional e percentagem de activos com formação ao nível do secundário**

<u>Pib per capita</u>		<u>Taxa de desemprego</u>	
pl31 Lubelskie (PL)	32.0	itd1 Provincia Autonoma Bolzano-Bozen (IT)	2.0
pl32 Podkarpackie (PL)	32.6	at32 Salzburg (AT)	2.3
pl62 Warminsko-Mazurskie (PL)	34.1	at33 Tirol (AT)	2.5
pl34 Podlaskie (PL)	35.1	nl34 Zeeland (NL)	2.5
pl33 Swietokrzyskie (PL)	35.7	pt2 Região Autón.dos Açores (PT)	2.5
...		...	
fr10 Île de France (FR)	176.0	pl43 Lubuskie (PL)	24.5
de60 Hamburg (DE)	187.8	pl42 Zachodniopomorskie (PL)	25.5
lu00 Luxembourg (Grand-Duché) (LU)	212.7	pl51 Dolnoslaskie (PL)	26.0
be10 Région de Bruxelles-Capitale (BE)	234.5	fr91 Guadeloupe (FR)	26.3
uki1 Inner London (UK)	315.4	fr94 Reunion (FR)	31.8
<u>Densidade populacional</u>		<u>Educação secundária</u>	
fr93 Guyane (FR)	2.10	fr83 Corse (FR)	7.13
se08 Övre Norrland (SE)	3.30	pt2 Região Autónoma dos Açores (PT)	8.90
fi1a Pohjois-Suomi (FI)	4.70	pt3 Região Autón.da Madeira (PT)	9.30
se07 Mellersta Norrland (SE)	5.23	pt11 Norte (PT)	9.72
ukm4 Highlands and Islands (UK)	9.30	pt16 Centro (PT)	10.62
...		...	
es63 Ciudad Autón. de Ceuta (ES)	3766.68	cz02 Strední Cechy (CZ)	80.18
de30 Berlin (DE)	3805.73	cz03 Jihozápad (CZ)	80.19
es64 Ciudad Autón. de Melilla (ES)	5545.92	sk04 Východné Slovensko (SK)	80.64
be10 Région de Bruxelles-Capitale (BE)	6146.47	cz07 Strední Morava (CZ)	81.24
uki1 Inner London (UK)	8946.45	cz05 Severovýchod (CZ)	81.45

Outros artigos aplicando a análise estatística multivariada a problemas sócio-económicos e, em particular, à classificação de diferentes tipos de regiões administrativas (autarquias ou regiões), podem ser encontrados na literatura: Cziráky *et al.* (2005), Aragon *et al.* (2003), González and Morini (2000), Soares *et al.* (2003), Peschel (1998), Pettersson (2001), Rován and Sambt (2003) ou ainda Rúa Vieites *et al.* (2003). Estes estudos abrangem, contudo, áreas restritas do espaço europeu: a Croácia, a região Midi-Pyrénées, a ilha de Tenerife, Portugal, os países do Mar Báltico, uma região da Suécia, a Eslovénia e a Galiza. Fora da Europa, refira-se Stimson *et al.* (2001), no caso norte-americano, e Hill *et al.* (1998) para a Austrália. Globalmente para a União Europeia há um estudo de Rúa Vieytes *et al.* (2000), mas refere-se ainda à fronteiras com 15 países e nele não constam dados para diversas regiões, para além de o seu

âmbito ser também diverso do deste trabalho. Por sua vez, é possível referir ainda outros estudos focados nas desigualdades do crescimento a nível regional, tendo por base distintas ferramentas analíticas e uma motivação associada à pesquisa das causas potenciais dessas desigualdades – Puga (2002) ou diversos artigos referidos em Rey e Janikas (2005, p.167, table 1).

O resto deste artigo encontra-se organizado da seguinte maneira: na secção 2 caracterizam-se os dados usados neste trabalho; a secção 3 é dedicada à busca de variáveis latentes capazes de representar o mais vasto conjunto de variáveis originais; a secção 4 debruça-se sobre a análise de *clusters*, procurando comparar o agrupamento de regiões similares obtido de acordo com a metodologia proposta com a classificação acima descrita correspondente à proposta da Comissão Europeia; finalmente, apresentam-se as principais conclusões do trabalho.

## **2. Dados estatísticos utilizados no trabalho**

As variáveis sócio-económicas consideradas no presente estudo (Tabela 2) foram seleccionadas na base de dados “Regio” do Eurostat. Correspondem a dezanove dos vinte e quatro Principais Indicadores Regionais (PIR) publicados no Terceiro Relatório Sobre a Coesão Económica e Social (*Third report on economic and social cohesion – European Commission, 2004*). As cinco variáveis que foram excluídas não satisfaziam os requisitos estabelecidos para integrarem a análise. A saber: serem dados actualizados, estarem disponíveis para as regiões dos vinte e cinco países, e serem expressos em forma de rácio, para evitar problemas de escala. Quanto aos dados relativos às regiões cujas fronteiras mudaram em Maio de 2003<sup>1</sup>, foram tomadas as seguintes decisões: quando as mudanças eram pequenas e existiam dados anteriores, estes dados eram retidos; quando tal não acontecia, essas regiões eram excluídas da análise (foram 13 em 254 as regiões NUTS2 excluídas da análise).

---

<sup>1</sup> Na Alemanha, Brandenburg foi dividida em duas regiões NUTS2. Em Espanha, Ceuta e Melilla também foi dividida em duas regiões. Em Itália, a região NUTS1 Nord-Ovest foi redefinida para incluir a Lombardia, anteriormente uma região NUTS1, Nord-Est passou a incluir a Emilia-Romagna, Centro a incluir a Lazio, e Sud a incluir Abruzzo-Molise e a Campania. Por sua vez, foi formada uma região NUTS1, Isole, abrangendo a Sardenha e a Sicília. Em Portugal, a região NUTS2 Lisboa e Vale do Tejo foi dividida, sendo o seu território espartilhado com o Centro e o Alentejo. Na Finlândia, quatro prévias NUTS2 incluídas na NUTS1 Manner-Suomi (todas menos Itä-Suomi) foram reclassificadas para formar três novas regiões NUTS2.

**Tabela 2 – Indicadores regionais europeus analisados**

<b>Código</b>	<b>Descrição</b>	<b>Ano</b>
<b>-Demografia-</b>		
ppdens	Densidade populacional (hab./km <sup>2</sup> )	2003
pop014	Percentagem da população com menos de 15 anos	2003
pop1564	Percentagem da população entre 15 e 64 anos	2003
pop65	Percentagem da população com 65 ou mais anos	2003
<b>-Economia-</b>		
gppps	PIB <i>per capita</i> (PPS), EU25=100	2002
Empagr	Emprego na agricultura (% do emprego total)	2002
empind	Emprego na indústria (% do total)	2002
empserv	Emprego nos serviços (% do total)	2002
patent	Pedidos de registo no EPO (European Patent Office) por milhões de habitantes, média anual	1999-2003
<b>-Emprego-</b>		
emptot	Taxa de desemprego total (desemprego nas idades 15-64 anos como % da população nessas idades)	2003
empf	Taxa de desemprego feminino (desemprego feminino nas idades 15-64 anos como % da população feminina nessas idades)	2003
empm	Taxa de desemprego masculino (desemprego masculino nas idades 15-64 anos como % da população masculina nessas idades)	2003
unemptot	Taxa de desemprego total (%)	2003
unemplt	Taxa de desemprego total de l/prazo (% do desemprego total)	2003
unempf	Taxa de desemprego feminino (%)	2003
unempy	Taxa de desemprego jovem (%)	2003
<b>-Educação-</b>		
lowedu	Percentagem da população activa com educação pré-primária e básica (níveis 0-2 ISCED 1997)	2003
mededu	Percentagem da população activa com educação secundária e pós-secundária (níveis 3-4 ISCED 1997)	2003
highedu	Percentagem da população activa com educação superior (níveis 5-6 ISCED 1997)	2003

As estatísticas descritivas apresentadas na Tabela 3 reflectem algumas diferenças muito acentuadas entre as regiões europeias. As mais significativas são a densidade populacional (um rácio de 1:4500 entre a menor e a maior densidade) e as patentes (0:745). Quanto à demografia, as diferenças na população potencialmente activa, i.e. dos 15 aos 64 anos, e na percentagem da população idosa, com 65 ou mais anos, são igualmente relevantes (amplitudes de 47.6%:72.6% e 8.5%:31%, respectivamente). Por sua vez, o PIB *per capita* mostra uma dispersão de 1:10, e o desemprego uma dispersão de 1:15, que se amplia para 1:20 no caso do desemprego de longo prazo. O valor de 1:10 é também o rácio encontrado quando se olha para a percentagem de população

com educação superior. Finalmente, refira-se que algumas das variáveis evidenciam excesso de kurtosis e enviesamento e, conseqüentemente, não seguem distribuições normais, um facto tido em linha de conta na escolha das técnicas utilizadas nas secções seguintes.

**Tabela 3 – Estatísticas descritivas dos indicadores regionais europeus**

	Min	Max	Mean	St. Dev.	Skew	Kurtosis
popdens	2.10	8946.45	385.78	926.37	5.73	39.62
pop014	10.14	23.33	16.68	2.47	-0.35	-0.06
pop1564	47.60	72.57	66.68	2.68	-1.90	12.59
pop65	8.59	31.10	16.64	3.13	0.70	2.38
gdppps	32.00	315.40	95.52	34.97	1.38	6.50
empagr	0.06	39.40	6.54	7.49	2.36	5.69
empserv	41.04	88.46	65.19	9.71	-0.14	-0.37
empind	11.43	46.33	28.27	7.38	0.04	-0.31
patent	0.00	745.88	106.35	130.31	2.17	5.85
emptot	40.10	78.60	63.35	8.20	-0.50	-0.12
empf	24.00	76.10	55.44	10.29	-0.64	0.18
empm	46.60	85.60	71.24	7.69	-0.77	0.31
unemptot	2.00	31.80	8.95	5.69	1.38	1.42
unemplt	4.09	80.44	38.45	15.80	0.23	-0.77
unempf	2.30	33.30	10.00	6.79	1.24	0.85
unempy	4.20	58.40	18.91	11.96	1.28	1.06
lowedu	3.30	86.30	25.97	17.88	1.12	0.76
mededu	7.13	81.45	49.38	16.36	-0.45	-0.15
highedu	4.80	48.24	23.35	8.20	0.29	-0.24

A correlação entre cada par de variáveis foi também calculada. (Tabela 4). Entre as correlações mais importantes, para além das óbvias associações entre variáveis da mesma categoria (demografia, economia, emprego e educação), dois aspectos merecem destaque. Um é a relevante correlação entre o PIB *per capita* e, respectivamente, o peso do emprego no sector de serviços (0.62), as patentes por milhão de habitantes (0.51), a taxa de desemprego total (0.50) e a percentagem de população activa com o ensino superior (0.49). O outro aspecto é a também significativa correlação entre a percentagem da população activa com o ensino superior e o peso do emprego nos serviços (0.58). Pelo contrário, as correlações entre activos com formação ao nível superior e os pesos dos sectores industrial e agrícola são ambas negativas (-0.31 e -0.44 respectivamente). Tal facto terá com certeza implicações na menor capacidade geradora de valor destes sectores face aos serviços e, conseqüentemente, regiões onde estes sectores estejam mais representados, terão tendência a apresentar valores inferiores do Produto *per capita*.

**Tabela 4 – Matriz das Correlações**

popdens	1																			
pop014	0.03	1																		
pop1564	0.18	-0.25	1																	
pop65	-0.18	<b>-0.56</b>	<b>-0.66</b>	1																
gdppps	<b>0.55</b>	0.00	-0.05	0.04	1															
empagr	-0.24	-0.06	0.07	-0.01	-0.52	1														
empserv	0.41	0.20	-0.22	0.04	<b>0.62</b>	<b>-0.67</b>	1													
empind	-0.30	-0.20	0.22	-0.04	-0.27	-0.17	<b>-0.62</b>	1												
patent	0.08	0.11	-0.07	-0.02	<b>0.51</b>	-0.41	0.28	0.06	1											
emptot	0.04	0.21	-0.27	0.07	<b>0.50</b>	-0.46	0.40	-0.05	0.47	1										
empf	0.07	0.26	-0.24	0.01	0.42	-0.46	0.40	-0.05	0.48	<b>0.93</b>	1									
empm	0.00	0.11	-0.26	0.13	<b>0.50</b>	-0.34	0.31	-0.05	0.35	<b>0.87</b>	<b>0.64</b>	1								
unemptot	0.00	-0.13	0.32	-0.18	-0.49	0.38	-0.32	0.02	-0.35	<b>-0.80</b>	<b>-0.63</b>	<b>-0.85</b>	1							
unemplt	-0.04	-0.33	0.28	0.01	-0.41	0.37	-0.40	0.15	-0.27	<b>-0.75</b>	<b>-0.66</b>	<b>-0.70</b>	<b>0.69</b>	1						
unempf	-0.08	-0.23	0.30	-0.08	-0.49	0.46	-0.37	0.01	-0.43	<b>-0.85</b>	<b>-0.78</b>	<b>-0.75</b>	<b>0.94</b>	<b>0.69</b>	1					
unempy	-0.04	-0.03	0.15	-0.11	-0.47	<b>0.51</b>	-0.29	-0.15	-0.47	<b>-0.84</b>	<b>-0.75</b>	<b>-0.77</b>	<b>0.84</b>	<b>0.62</b>	<b>0.88</b>	1				
lowedu	-0.12	-0.20	-0.06	0.20	-0.04	0.33	-0.16	-0.13	-0.30	-0.26	-0.47	0.08	-0.04	0.06	0.20	0.19	1			
mededu	-0.08	0.16	0.15	-0.25	-0.24	-0.10	-0.15	0.30	0.10	0.06	0.27	-0.24	0.16	0.12	-0.05	-0.03	<b>-0.86</b>	1		
highedu	0.34	0.09	-0.12	0.04	0.49	-0.44	<b>0.58</b>	-0.31	0.36	0.36	0.39	0.23	-0.17	-0.31	-0.24	-0.26	-0.34	-0.17	1	
		popdens	pop014	pop1564	pop65	gdppps	empagr	empserv	empind	patent	emptot	empf	empm	unemptot	Unemplt	unempf	unempy	lowedu	mededu	highedu

### 3. Indicadores regionais e dimensões sócio-económicas

As semelhanças socio-económicas entre as regiões NUTS 2 poderiam ser investigadas usando directamente os Principais Indicadores Regionais medidos pelo Eurostat. Contudo, em situações onde grupos de observações são formados com base nas variáveis originalmente medidas, o investigador tem que intervir a fim de escolher que variáveis originais usar no processo de formação de *clusters* (Hair *et al.*, 1998). Esta escolha é decisiva de modo a, por exemplo, evitar soluções que poderiam ser enviesadas para características que estejam representadas por mais variáveis originais do que outras - efeito “overweight”. Isto acontece com os dados actuais, onde um número diferente de indicadores regionais representa diferentes categorias de indicadores regionais - demografia, economia, emprego e educação (tabela 2).

Uma abordagem para lidar com este problema é usar um método da redução de dados tal como a análise factorial exploratória, que é capaz de identificar um conjunto mais pequeno de dimensões (variáveis latentes ou factores) não correlacionadas, cada uma associada a um conjunto de variáveis originais muito correlacionadas. Estes factores

podem subsequentemente ser usados como a base para a formação de *clusters*<sup>2</sup>, com a vantagem adicional de revelarem a estrutura dos dados. Neste estudo, a análise de componentes principais foi usada para a extracção dos factores. Este procedimento comum não faz nenhuma suposição relativamente à distribuição dos dados originais e permite também que um pequeno conjunto de componentes principais explique uma elevada proporção da variância total.

A primeira etapa para implementar esta metodologia passa pelo exame visual da matriz de correlações (tabela 4). Esta observação revela uma quantidade considerável de correlação nos dados, com todas as variáveis a terem pelo menos um coeficiente de correlação maior do que 0.5. Para além disso, o determinante da matriz de correlação é nulo, o que adicionalmente sustenta ser apropriado proceder com a análise factorial (Lattin, Carrol and Green, 2003, p.110). Por esta mesma razão, o teste de esfericidade de Bartlett (1950) não pôde ser calculado.

**Tabela 5 - Análise de Componentes Principais – Variância Explicada**

<b>Factor</b>	<b>Valor Próprio</b>	<b>% variância</b>	<b>% variância acumulada</b>
1	7.23	38.07	38.07
2	2.57	13.53	51.60
3	2.43	12.77	64.37
4	1.56	8.22	72.59
5	1.47	7.72	80.31
6	0.79	4.17	84.47
7	0.71	3.76	88.24
8	0.65	3.40	91.64
9	0.51	2.68	94.32
10	0.33	1.72	96.04
11	0.26	1.37	97.40
12	0.22	1.13	98.54
13	0.16	0.84	99.38
14	0.10	0.54	99.92
15	0.01	0.06	99.98
16	0.00	0.01	100.00
17	0.00	0.00	100.00
18	0.00	0.00	100.00
19	0.00	0.00	100.00

<sup>2</sup> Alguns autores sugerem que se multipliquem os vectores próprios pelas raízes quadradas dos valores próprios que lhes estão associados, de modo a que as variâncias das respectivas componentes principais igualem a variância explicada por aquelas componentes nos dados originais (Lattin et al., 2003, p. 274). Esta não foi no entanto a abordagem seguida uma vez que isto conduziria a dar pesos diferentes a cada categoria de Indicadores Regionais.

A etapa seguinte, a decisão do número de factores a reter, foi baseada no critério do valor próprio (Kaiser 1960). Consequentemente, os primeiros cinco factores com valores próprios superiores a 1 foram retidos (tabela 5). O critério de Ludlow (1999) aponta na mesma direcção uma vez que há uma clara diminuição da variância após o quinto factor. Além disso, a solução com cinco factores explica mais de 80% da variância total das variáveis originais, o que é um bom resultado de acordo Hair *et al.* (1998). A estrutura com cinco factores deu também a melhor solução do ponto de vista interpretativo quando comparada com as soluções com três, quatro e seis factores – depois de sujeitas a uma rotação Varimax. Este é um critério relevante uma vez que “na prática o investigador está interessado na interpretabilidade e no significado operacional das soluções factoriais” (Lattin, Carrol and Green, 2003).

A solução com cinco factores tem três méritos adicionais. Em primeiro lugar, quase todas as variáveis estão fortemente correlacionadas só com um factor. Em segundo lugar, todas as variáveis têm pelo menos um peso num factor com valor absoluto superior a 0.5, o que é considerado muito significativo (Hair *et al.*, 1998). Por último, a tabela 6 mostra que esta estrutura factorial explica entre 62% e 98% da variância de cada variável original, à excepção da variável “patent” para a qual explica somente 44% da respectiva variância. A solução com 5 factores obtida com rotação Varimax é mostrada na tabela 6, onde se omitem os pesos das variáveis originais nos factores cujo valor absoluto é menor do que 0.45.

No que respeita à interpretação dos factores, a tabela 6 mostra que os primeiros três factores estão relacionados essencialmente com três categorias de indicadores - emprego, economia e educação. O factor 1, **Factor Desemprego**, indica níveis elevados de desemprego, com correlações positivas fortes com indicadores de desemprego e correlações negativas com variáveis de emprego. Também se pode observar uma correlação negativa (relativamente baixa) com o número de patentes por milhão de habitantes, o que é um resultado esperado. O factor 2, **Factor Desenvolvimento Económico**, associado com níveis elevados do PIB *per capita* e um elevado número de empregos no sector de serviços, está também positivamente correlacionado com uma variável demográfica e uma variável de educação - respectivamente densidade populacional e percentagem da população activa com educação superior. Por isso, uma

região com um valor elevado neste factor é certamente rica, com uma grande oferta de serviços, e uma economia moderna e essencialmente urbana. O factor 3, **Factor Educação Média**, indica elevada percentagem da população activa com níveis de educação secundária e pós-secundária, e conseqüentemente baixa percentagem da população activa com níveis de educação pré-primária e básica.

**Tabela 6 – Solução com 5 factores obtida com rotação Varimax**

	F1	F2	F3	F4	F5	Comunalidades
popdens		0.71				0.62
pop014					-0.88	0.84
pop1564				0.89		0.88
pop65				-0.82	0.52	0.98
gdppps		0.74				0.81
empagr		-0.54				0.63
empind		-0.54			0.52	0.77
empserv		0.83				0.84
patent	-0.46					0.44
emptot	-0.92					0.92
empf	-0.78					0.85
empm	-0.90					0.86
unemptot	0.93					0.89
unemplt	0.77					0.72
unempf	0.92					0.88
unempy	0.91					0.88
lowedu			-0.95			0.95
mededu			0.91			0.89
highedu		0.74				0.62

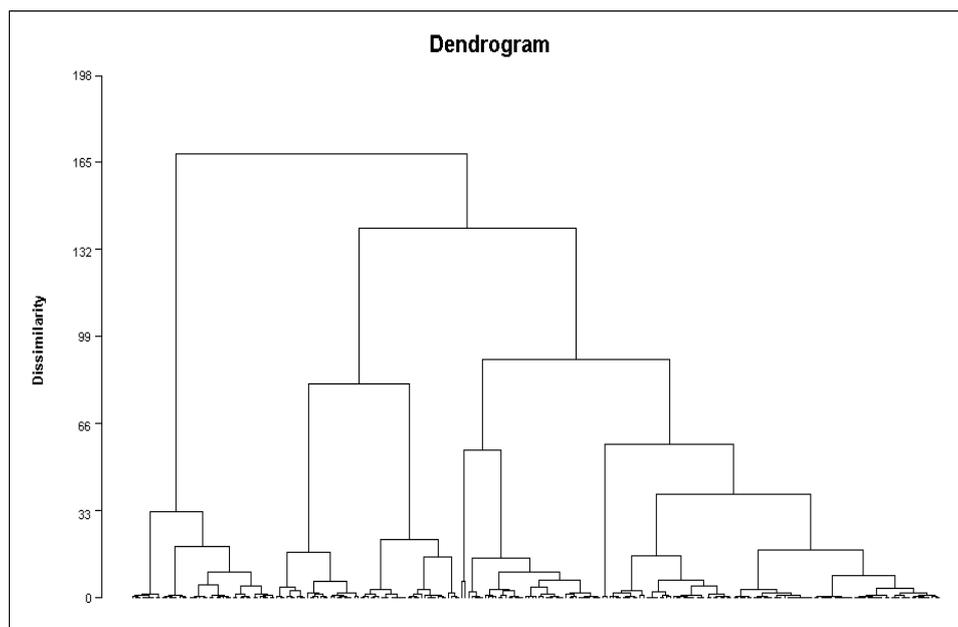
Quanto aos outros factores adicionais, o quarto factor está associado com uma elevada percentagem de adultos ativos (“pop1564”) e com uma percentagem reduzida de pessoas idosas (“pop65”), enquanto o quinto factor está muito negativamente correlacionado com a percentagem de crianças na população (“pop014”) e mostra uma razoável correlação positiva com a percentagem de pessoas idosas (“pop65”).

Estes dois factores estão relacionados com demografia – a quarta categoria de indicadores regionais - e assim, a preservação de ambos para a análise de *clusters* conduziria ao efeito “overweight” já mencionado. Uma vez que não parece desejável favorecer aspectos da demografia nas seguintes etapas do estudo e, ao mesmo tempo,

como a interpretabilidade da solução em que se forçou a extracção de 4 factores era menos óbvia, tomou-se a decisão de prosseguir com a análise de *clusters* mantendo apenas o quarto factor, o **Factor População Activa**, e deixar de fora o quinto factor.

#### 4. Multi-dimensionalidade do desenvolvimento económico e *clusters* regionais

Para procurar grupos de regiões NUTS2 empregaram-se inicialmente várias técnicas de agrupamento hierárquico usando os valores de cada região obtidos para cada um dos quatro factores referidos na Secção 3. O objectivo desta primeira etapa foi o de analisar os quadros de aglomeração e os dendrogramas a fim estabelecer o número de *clusters* a seleccionar. A melhor solução do ponto de vista interpretativo encontra-se ilustrada no dendrograma mostrado na Figura 2, correspondendo ao método de Ward em que se usou o quadrado da distância Euclideana (outros autores enfatizam o desempenho deste método - por exemplo Everitt, 1993, 2001; Punj e Stewart, 1983; Milligan, 1980).

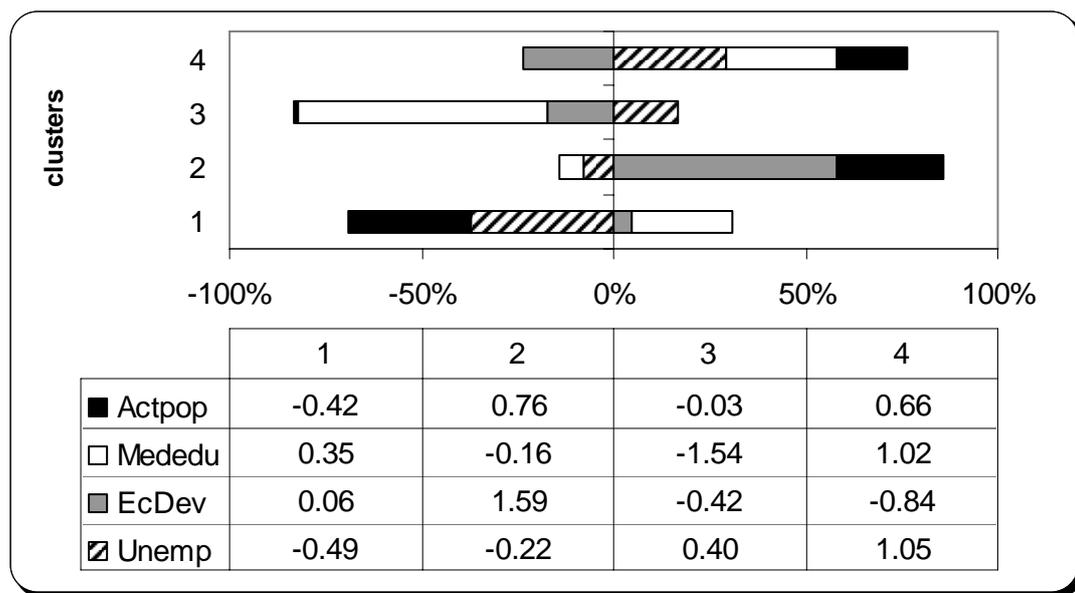


**Figura 2 – Dendrograma resultante do método de Ward**

Da análise deste dendrograma, nomeadamente a partir da análise dos sucessivos aumentos nas distâncias a que os *clusters* se juntam, pode-se concluir que uma escolha razoável para o número de grupos a ser seleccionado deve cair dentro do intervalo três a cinco *clusters*. Investigação adicional, partindo da solução cinco *clusters*, revelou que o cluster mais à esquerda na figura 2 corresponde aproximadamente às *convergence*

*regions* dos novos estados membros do Leste da Europa. Este é o último cluster a juntar-se, como resultado das enormes diferenças entre estes países e a grande maioria das regiões EU-15. Observando a figura 2 da esquerda para a direita, os segundo e terceiro *clusters* incluem um grande número de regiões do sul da UE. Estes *clusters* são os primeiros a juntarem-se quando se passa da solução cinco para quatro *clusters*. Finalmente, os dois últimos *clusters* correspondem aproximadamente às duas regiões mais ricas dentro da união. Estes dois *clusters* juntam-se quando se forma a solução três *clusters*.

A análise do parágrafo anterior constitui uma primeira abordagem. Contudo, os métodos hierárquicos impõem que um *cluster* não pode ser dividido depois de formado. Por outro lado, os métodos não-hierárquicos são mais flexíveis, permitindo que casos se separem dos *clusters* aos quais tinham sido previamente alocados. Consequentemente, e seguindo a metodologia sugerida por diversos autores (por ex. Lattin *et al*, 2003; Punj and Stewart, 1983), usou-se o método não-hierárquico k-means, tendo como sementes os centróides resultantes do método de Ward. Para além disso, de modo a comparar a solução que resulta desta metodologia com os quatro *clusters* propostos pela Comissão Europeia (figura 1), a atenção será centrada principalmente na solução quatro *clusters*.



Fig

Figura 3 – Centróides para os *clusters* (k-means)

Os resultados obtidos com o método k-means são, em grande parte, coincidentes com os resultados obtidos com o método de Ward. Assim, mais de 86% das regiões pertencem a *clusters* idênticos; um número que alcança os 100% para o *cluster* que integra quase todas as regiões dos novos estados membros (que corresponde ao *cluster* 4 da figura 3) e é mínimo (80%) para o maior *cluster* que integra a grande maioria dos países da UE-15 (*cluster* 1 da figura 3). Os valores médios dos quatro factores obtidos para cada um dos quatro *clusters* de regiões NUTS 2 são apresentados na figura 3. Diferenças significativas são mostradas nos perfis dos quatro *clusters*: os *clusters* 1 e 2 (particularmente este último) revelam bom desempenho económico; o *cluster* 3 identifica uma importante desvantagem no factor Educação Média; o *cluster* 4 exhibe valores positivos significativos para os factores Desemprego, Educação Média e População Activa.

Uma descrição mais detalhada dos quatro *clusters* é apresentada em seguida, e pode ser melhor compreendida olhando o mapa Europeu na figura 4:

*Cluster* 1 – Este é o maior *cluster* sendo formado por 117 regiões localizadas principalmente na **Europa do Norte e Central**. A população e a área são cerca de 43% do total da UE. O desemprego está abaixo da média assim como a percentagem de adultos activos relativamente à população idosa. Por outro lado, os valores dos factores Desenvolvimento Económico e Educação Média estão acima da média. Assim, o *cluster* 1 identifica regiões com um desenvolvimento económico razoável, um desemprego baixo, uma grande oferta de serviços e uma percentagem elevada da população activa com educação secundária e pós-secundária. Este *cluster* também integra duas regiões da antiga Europa Oriental: Közép-Magyarország, uma região da Hungria que foi classificada como uma região *phasing-in* pela CE, e, mais surpreendentemente, a Estónia, um país que foi classificado como uma região de convergência devido ao seu baixo PIB *per capita*. Neste caso, análise adicional revelou que a Estónia exhibe outras características em termos de desemprego, da percentagem da população activa, da percentagem do emprego nos serviços e da percentagem da população activa com educação superior que diferem dos valores médios dos outros países da Europa do Leste e, em particular, dos seus vizinhos Bálticos, Letónia e Lituânia. Assim, pode dizer-se que a Estónia exhibe sinais de ser um país mais rico do que, verdadeiramente, o seu PIB *per capita* sugere.

*Cluster 2* – Há 31 regiões neste conjunto e 13 delas são **regiões que incluem a capital** do respectivo país; por exemplo, Viena, Bruxelas, Berlim, Madrid, Île de France (Paris), Attiki (Atenas), Luxemburgo, Irlanda do sul e oriental (Dublin), Noord-Holland (Amsterdão), Estocolmo ou Inner London (Londres). Algumas são capitais dos novos países da UE, tais como Praga ou Bratislava. As restantes regiões que integram este *cluster* pertencem à Bélgica (4), à Alemanha (4), à Espanha (1), aos Países Baixos (5) ou ao Reino Unido (4). A população deste *cluster* é 16% do total enquanto que a área é somente 3% - é assim o *cluster* o mais denso. Relativamente aos valores dos quatro factores, o Desemprego está um pouco abaixo da média assim como a Educação Média. Em particular, este valor relativo à educação pode parecer surpreendente; no entanto, Rodríguez-Pose e Vilalta-Bufi (2005) concluíram que, em regiões europeias, outros factores como a satisfação no trabalho, o ajustamento entre as necessidades da oferta e a procura de emprego, e a imigração estão mais relacionados com o desempenho económico do que o nível de formação dos recursos humanos. Para além disso, a baixa percentagem da população activa com educação secundária e pós-secundária é inteiramente compensada pela percentagem da população que tem educação ao nível superior. Por outro lado, os valores dos factores Desenvolvimento Económico e População Activa estão consideravelmente acima da média. Assim, o *cluster 2* é formado por regiões muito ricas e densas.

*Cluster 3* – As 51 regiões neste *cluster* representam cerca de 21% da população e da área do total da UE. Geograficamente, todas as regiões estão no **Sul de Europa** sendo a única excepção a região ie01 (Border, Midlands and Western Ireland). O Desemprego está acima da média, enquanto que a percentagem de adultos activos relativamente à população idosa se encontra em torno da média. Por outro lado, os valores dos factores Desenvolvimento Económico e Educação Média estão abaixo da média, especialmente este último. A principal característica deste *cluster* é o baixo nível de educação acima do nível pré-primário e básico. Assim, o *cluster 3* identifica regiões desfavorecidas, com algum desemprego e baixos níveis de educação secundária e pós-secundária. Adicionalmente, muitas regiões do *cluster 3* são *convergence regions*.

*Cluster 4* – Há 43 regiões neste *cluster*, que representa cerca de 17% da população e da área total da UE. Todas as 43 regiões não pertencem a países da EU-15, à excepção de

seis regiões alemãs que pertencem à antiga República Democrática. As regiões do *cluster 4* têm os níveis mais elevados de desemprego assim como a percentagem mais elevada de educação secundária e pós-secundária. Além disso, os valores para o factor Desenvolvimento Económico estão bastante abaixo da média enquanto a percentagem de adultos activos relativamente à população idosa se encontra acima da média. Assim, o *cluster 4* além de identificar regiões do **Leste Europeu**, também identifica regiões com baixos rendimentos e com desemprego elevado. Todas as regiões do *cluster 4* são *convergence regions*.

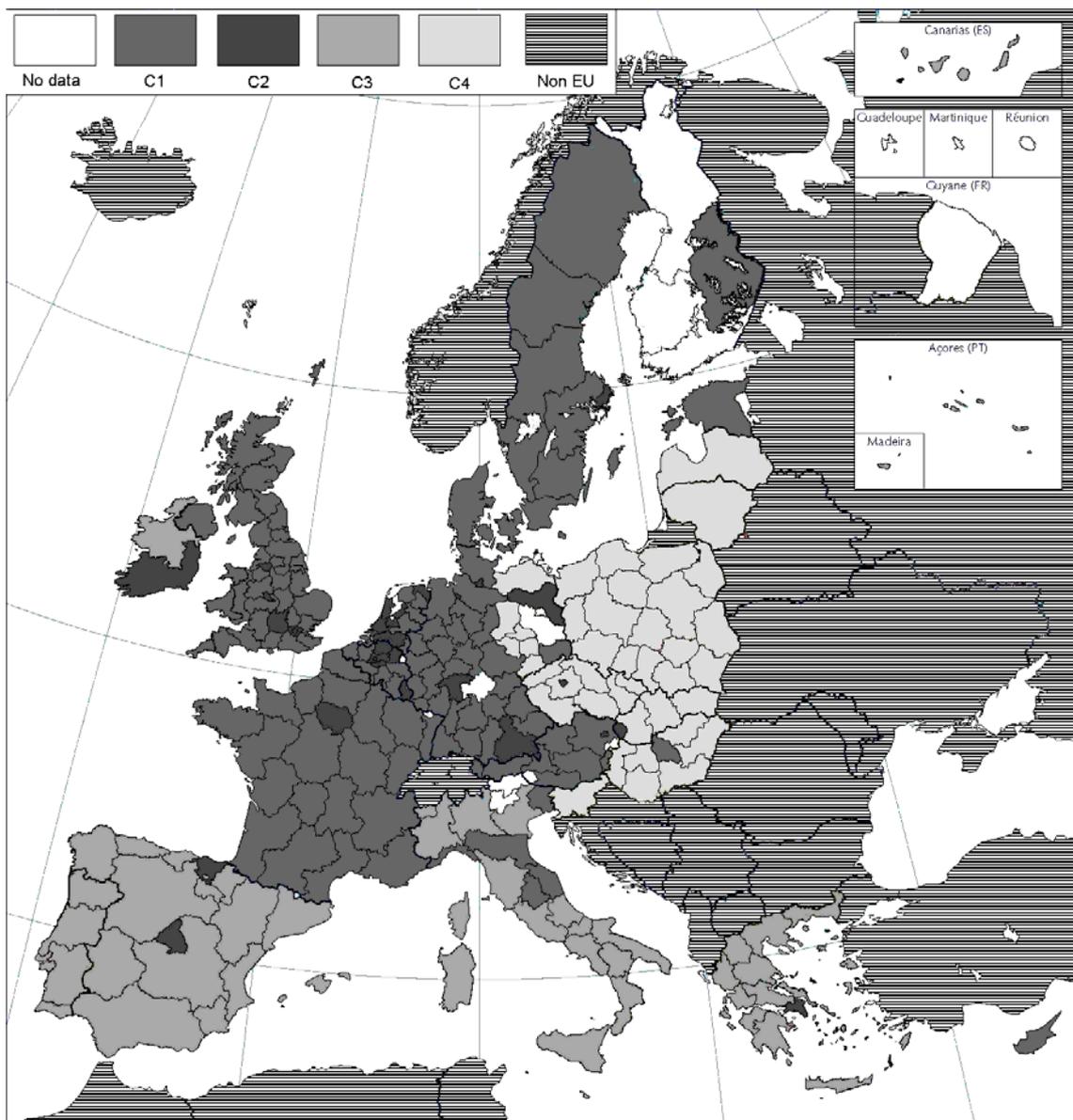


Figura 4 – EU-25: Quatro *Clusters* Sócio-Económicos de Regiões NUTS 2 (K-means)

O próximo assunto a ser analisado é o grau de concordância entre os *clusters* regionais descritos anteriormente e as quatro classes de regiões identificadas pela Comissão Europeia como a base para a política regional Europeia para o período 2007/2013 (estas classes foram descritas mais cedo na secção 1 e estão representadas na figura 1). A tabela 7 mostra que os *clusters* 1 e 2 são dominados pelas regiões *competitiveness and employment*; que o *cluster* 4 inclui quase exclusivamente *convergence regions*; e que o *cluster* 3, que inclui um grande número de regiões do sul da Europa, se encontra disperso por todas as classes da CE. Como conclusão, fica claro que a análise de *clusters* revelou dois grupos diferentes dentro das regiões *competitiveness and employment* e que o mesmo acontece para as *convergence regions*.

**Tabela 7 – Concordância entre as duas classificações**

		<i>Comp.Emp. Regions</i>	<i>Ph. Out Regions</i>	<i>Ph. In Regions</i>	<i>Conv. Regions</i>	<b>Total</b>
	<b>1</b>	105	4	4	4	117
<i>Clusters</i>	<b>2</b>	30	1	0	0	31
<b>K-means</b>	<b>3</b>	16	6	8	21	51
	<b>4</b>	0	2	0	40	42
<b>Total</b>		151	13	12	65	241

Com vista a avaliar o significado desta diferença e chegar a uma conclusão sobre a classificação mais conveniente para perseguir os objectivos de política regional Europeia, é útil voltar aos indicadores regionais. A tabela 8 mostra as médias dos diferentes indicadores calculados respectivamente para a classificação da CE e para a classificação 4-*clusters*. Uma análise comparativa revela que a solução 4-*clusters* permite uma distinção mais clara entre as diferentes regiões, o que é particularmente óbvio quando se observam os indicadores densidade populacional, PIB *per capita* e níveis de educação. Note-se que os dois principais grupos da classificação da CE - *convergence regions* e *competitiveness and employment regions* - incluem pelo menos dois grupos de regiões significativamente diferentes em termos destes indicadores. Os dois outros grupos - *phasing-in regions* e *phasing-out regions* – que se encontram dispersos pelos diferentes *clusters* na tabela 7, não revelam valores evidentemente diferentes em muitos indicadores. De facto, como o PIB *per capita*, a única base para a classificação da CE, varia entre 32% e 315.4% da média comunitária, a fixação exclusiva e arbitrária do valor 75%, como limiar separador de grupos de regiões, dificilmente poderia conduzir a uma distinção mais clara entre regiões.

**Tabela 8 – Médias dos indicadores regionais para as classificações CE e 4-clusters**

Indicadores Regionais	<i>Comp.Emp.</i> <i>Regions</i>	<i>Conv.</i> <i>Regions</i>	<i>Ph.In</i> <i>Regions</i>	<i>Ph.Out</i> <i>Regions</i>	<i>Cluster</i> <i>1</i>	<i>Cluster</i> <i>2</i>	<i>Cluster</i> <i>3</i>	<i>Cluster</i> <i>4</i>
popdens	481.6	130.6	307.1	190.2	226.3	1544.7	154.6	121.9
pop014	17.0	16.4	16.4	14.7	17.3	17.1	15.0	16.4
pop1564	66.0	68.0	67.3	66.2	65.1	68.3	66.9	69.3
pop65	17.0	15.6	16.3	19.0	17.5	14.5	18.1	14.3
gdppps	114.4	56.7	91.0	78.9	102.6	145.2	87.0	50.7
empagr	3.3	13.9	8.1	7.4	3.5	1.5	12.4	11.9
empserv	68.8	55.9	64.8	64.9	67.7	76.4	59.9	54.6
empind	28.0	30.2	27.2	27.7	28.8	22.0	27.6	33.5
patent	162.1	14.0	30.5	38.1	156.6	201.3	20.8	15.2
emptot	67.3	56.7	60.8	59.5	67.8	67.0	58.3	56.3
empf	60.0	48.3	50.0	49.8	61.3	60.0	44.7	50.8
empm	74.5	65.3	71.7	69.3	74.4	73.9	71.9	61.8
unemptot	6.3	13.5	8.9	11.5	6.2	6.9	9.8	15.3
unemplt	32.7	49.5	36.5	47.4	31.9	34.1	43.1	52.8
unempf	6.7	15.6	11.2	14.5	6.2	7.1	14.2	16.1
unempy	13.6	28.3	19.8	23.5	12.9	14.6	25.1	28.6
lowedu	23.5	26.9	41.0	32.9	19.7	21.3	54.2	11.5
mededu	48.9	55.3	36.4	42.6	53.3	44.4	27.3	70.5
highedu	25.8	17.5	21.8	23.4	24.9	32.1	18.5	17.6

## 5. Conclusões

A afectação de recursos financeiros no quadro da política regional europeia tem sido subordinada a regras que se podem considerar simples e objectivas, baseadas no essencial no desempenho das regiões em termos do seu PIB *per capita* e, quanto a este, adoptando um patamar correspondente a 75% da média comunitária para delimitar o acesso ao essencial dos fundos de coesão.

Tais simplicidade e objectividade, sendo de aplaudir, podem, contudo, esconder realidades mais complexas e conduzir a eventuais injustiças na afectação de fundos ou efeitos perversos. Foi por razões de acesso a fundos relativos ao Objectivo 1 que se deu em Portugal a redefinição da NUTS2 Lisboa e Vale do Tejo – a lógica financeira fez alterar uma divisão territorial existente. Por sua vez, é evidente que um limiar do Produto em 75% da média divide as regiões em dois grupos de amplitude bem desigual: um com o Produto variando entre 32% e 75% da média, e outro entre 75 e 315%. Ora, pese embora essa constatação, e com a ligeira nuance de diferenciar as regiões que recentemente ultrapassaram o limiar de 75% por razões de crescimento próprio ou estatístico (por efeito do alargamento), é este o critério que continuará a presidir à afectação de fundos para o novo período orçamental 2007-2013.

A vontade de procurar uma caracterização sócio-económica mais completa do que uma mera análise do indicador Produto, e a consequente curiosidade de confrontar os resultados dessa classificação com a classificação estabelecida pela Comissão de acordo com a lógica descrita, tais foram os grandes objectivos deste trabalho. Para tal, começou-se por procurar reduzir a informação contida nos Principais Indicadores Regionais publicados pelo Eurostat em quatro categorias – demografia, emprego, economia e educação. Os factores resultantes foram usados, com um peso igual entre si, para classificar as regiões europeias em quatro grupos alternativos aos da Comissão Europeia.

A comparação efectuada mostrou que os dois maiores grupos da classificação da Comissão – *convergence regions* and *competitiveness and employment regions* – englobam, cada um deles, dois ou mais grupos de regiões significativamente distintas. Por sua vez, os outros dois grupos – *phasing-in regions* e *phasing-out regions* –, não apresentam homogeneidade de acordo com a caracterização multivariada que foi empreendida, repartindo-se pelos diferentes grupos. A maior homogeneidade (semelhança) intra-grupal da caracterização sócio-económica efectuada ficou demonstrada claramente a partir da análise dos valores médios dos diferentes indicadores regionais para os diferentes grupos. Para mais, a análise multi-dimensional, não restrita à dimensão Produto, mostrou evidentes benefícios na caracterização das semelhanças e dissemelhanças entre regiões, permitindo o eventual desenho de medidas de política regional mais ajustadas às diferentes necessidades que se manifestam no território europeu. Como investigação a empreender em futuros trabalhos fica a análise comparativa de outras abordagens metodológicas ou de diferentes técnicas de classificação e da respectiva robustez.

#### **Referências bibliográficas:**

- Aragon, Y., Haughton, D., Haughton, J., Leconte, E., Malin, E., Ruiz-Gaen, A. and Thomas-Agnan, C. (2003) “Explaining the Pattern of Regional Unemployment: The Case of Midi-Pyrenees Region”. *Papers in Regional Science*, 82: 155-174.
- Barnier, M. (2004) Speech given to the European Parliament’s enlarged Conference of Presidents, European Parliament, Brussels, 18 February.

- Bartlett, M.S. (1950) "Tests of Significance of Factor Analysis", *British Journal of Psychology* (Statistical Section), 3: 77-85.
- Cziráky, D., Sambt, J., Rován, J. and Puljiz, J. (2005) "Regional development assessment: A structural equation approach", *European Journal of Operational Research*, in press.
- European Commission (2004) "A New Partnership for Cohesion – Convergence, Competitiveness, Cooperation. Third Report on Economic and Social Cohesion".
- Everitt, B. S. (1993) *Cluster Analysis*, 3rd Ed., London: Edward Arnold.
- Everitt, B. S.; Landau, S. and Leese, M. (2001) *Cluster Analysis*. London: Edward Arnold.
- González, J.I. and Morini, S. (2000): "Posicionamiento socioeconómico y empresarial de los municipios de la Isla de Tenerife" (Socio-economic and business affairs situation in Tenerife Island), Working Paper 2000-07, Universidad de La Laguna (in Spanish).
- Hair, J.F., Anderson, R.E., Tatham, R.L. and Black. (1998) *Multivariate Data Analysis*. Prentice Hall International.
- Hill, E.W., Brennan, J.F. and Wolman, H.L. (1998) "What is a Central City in the United States? Applying a Statistical Technique for Developing Taxonomies", *Urban Studies*, 35(1): 1935-1969.
- Kaiser, H. F. (1959) "The Application of Electronic Computers to Factor Analysis", in *Symposium on the Application of Computers to Psychological Problems*, American Psychological Association
- Lattin, J., Carrol, J.D. and Green, P. E., (2003) *Analysing Multivariate Data*, Duxbury: Thompson Learning.
- Ludlow, L.H. (1999) "The Structure of the Job Responsibilities Scale: A Multimethod Analysis", *Educational and Psychological Measurements*, 59(6): 962-975.
- Milligan, G. (1980) "An examination of the effect of six types of error perturbation on fifteen clustering algorithms", *Psychometrika*, 45 (September): 325-42.
- Peschel, K. (1998) "Perspectives of Regional Development around the Baltic Sea", *The Annals of Regional Science*, 32: 299-320.
- Pettersson, O. (2001) "Microregional Fragmentation in a Swedish County", *Papers in Regional Science*, 80: 389-409.
- Puga, D. (2002) "European regional policies in light of recent location theories", *Journal of Economic Geography*, 2: 373-406.

- Punj, G. and Stewart, D. (1983) "Cluster Analysis in Marketing Research; a Review and Suggestions for Application", *Journal of Marketing Research*, 20(May): 134-148.
- Rey, S.J. and Janikas, M.V. (2005) "Regional convergence, inequality and space", *Journal of Economic Geography*, 5: 155-176.
- Rodríguez-Pose, A. and Vilalta-Bufí, M. (2005) "Education, migration, and job satisfaction: the regional returns of human capital in the EU", *Journal of Economic Geography*, 5: 545-566.
- Rovan, J. and Sambt, J. (2003) "Socio-economic Differences among Slovenian Municipalities: A Cluster Analysis Approach" in *Development in Applied Statistics*, Ferligoj, A. and Mrvar A. (eds), pp. 265-278.
- Rúa Vieytes, A., Peralta Astudillo, M.J., Fernández Rodríguez, M.L. and Borrás Pala, F. (2000) *Tipología Socioeconómica de las Regiones Europeas. Comparativa Estadística* (Socio-economic Typology of the European Regions. Statistical Comparative), Madrid: Instituto de Estadística de la Comunidad de Madrid (in Spanish).
- Rúa Vieytes, A., Redondo Palomo, R. and del Campo, C. (2003) "Distribución municipal de la realidad socioeconómica gallega" (Galician Socio-economic Reality Municipal Distribution), *Revista Galega de Economía*, 12(2): 243-262 (in Spanish).
- Soares, J.O., Marquês, M.L. and Monteiro C.F. (2003) "A Multivariate Methodology to Uncover Regional Disparities: A Contribution to Improve European Union and Governmental Decisions". *European Journal of Operational Research*, 145: 121-135.
- Stimson R., Baum, S., Mullins, P. and O'Conor, K. (2001) "A typology of Community Opportunity and Vulnerability in Metropolitan Australia". *Papers in Regional Science*, 80: 45-66.