

Caracterização territorial do desemprego registado em Portugal

Aplicação da análise de clusters

ALCINA NUNES E ELISA BARROS

RESUMO: Apesar da discussão em torno da taxa de desemprego nacional, pouco se conhece quanto à distribuição territorial das características dos desempregados registados nos centros de emprego, mesmo que o conhecimento dessa distribuição possa ser de utilidade crucial no desenvolvimento de políticas públicas adequadas de combate ao desemprego. Este artigo procura ultrapassar tal lacuna, recorrendo à análise de clusters. A metodologia referida procura encontrar grupos de unidades territoriais (NUTII) que apresentem as mesmas características para os desempregados registados. Os resultados permitem concluir pelo agrupamento dos distritos portugueses em quatro grandes grupos, com uma distinção clara entre o Norte e o Sul e, ainda, o litoral e o interior de Portugal Continental, assim como uma alteração no perfil regional de desemprego registado.

Palavras-chave: Desemprego, Análise de Clusters, Distribuição Geográfica, Análise Regional

TITLE: Territorial characterization of the registered unemployment in Portugal: a cluster analysis application

ABSTRACT: The regional distribution of the unemployed individual characteristics is of core importance for the development of public policies that can fight the unemployment phenomenon, especially in times of crises. In this research work, the data mining cluster methodology allows finding groups of regional areas – considering for that the NUTII administrative regions – that share the same characteristics for the registered unemployed and, therefore, helps in a better understanding of the problem and possible solutions. Results for the Portuguese regions show a clear division of the territory among four regions – north and south and urban and rural areas of the country. The final results also show that the economic crisis imply changes in the geographical profile distribution of the registered unemployed.

Key words: Unemployment, Cluster Analysis, Geographical Distribution, Regional Analysis

ALCINA NUNES

alcina@ipb.pt

Doutora em Economia (Universidade de Coimbra), Professora Adjunta no Instituto Politécnico de Bragança, Portugal.

PhD in Economics (University of Coimbra), Assistant Professor at the Polytechnic Institute of Bragança, Portugal.

ELISA BARROS

ebarros@ipb.pt

Mestre em Tecnologia e Gestão (Universidade do Porto), Professora Adjunta no Instituto Politécnico de Bragança, Portugal.

Master in Technology and Management (Oporto University), Assistant Professor at the Polytechnic Institute of Bragança, Portugal.

INTRODUÇÃO

Um estudo das semelhanças/disparidades regionais, no mercado de trabalho português, não se pode limitar a uma mera análise descritiva dos fenómenos associados mas deve, antes, estabelecer padrões de comparação espacial entre os territórios alvo de análise, para que se desenvolvam políticas públicas, de âmbito tanto central como regional, adequadas ao combate dos problemas associados. A criação de políticas de emprego, que combatam o fenómeno persistente do desemprego, têm merecido particular atenção na economia portuguesa, no entanto, pouco se conhece relativamente ao perfil dos desempregados registados ao longo das regiões nacionais, para que se desenvolvam políticas especialmente direccionadas a perfis regionais de desemprego.

Em Portugal, são conhecidos alguns estudos que pretendem delimitar geográfica, económica e socialmente, as regiões nacionais em grupos mais ou menos homogéneos, de forma a promover um melhor conhecimento das estruturas socio-económicas, regionais e locais. Não são, no entanto, conhecidos estudos que forneçam uma análise particular do perfil regional do desemprego registado. Este tipo de análise começa a ser desenvolvido em outras economias, pois acredita-se serem de importância fulcral para a compreensão do fenómeno do desemprego a nível agregado. Veja-se o exemplo de estudos como o de López-Bazo *et al.* (2005) ou Fertig *et al.* (2006) que procuram compreender, a nível regional, padrões de desemprego anormalmente elevados ou o impacto de distintas políticas públicas de emprego, respectivamente. Por outro lado e de acordo com autores como Marelli (2007), economias que verificam alterações substanciais no seu Produto Interno Bruto (como, por exemplo, crescimentos rápidos) fazem-no, essencialmente, à custa de algumas regiões em detrimento de outras, tendendo a fomentar disparidades regionais que se observam, nomeadamente, através de diferentes perfis de desemprego regional. Os chamados países do alargamento puderam experimentar tal fenómeno, o que gerou a produção de estudos vários sobre o fenómeno regional de desemprego (Arandarenko, 2007; Babucea, 2009 e Nadiya, 2008).

Neste trabalho de investigação, vai procurar-se delimitar regiões do país – recorrendo à unidade territorial distrito – de acordo com um determinado perfil de desemprego registado. Assim, aplica-se uma análise classificatória onde as unidades territoriais se agrupam em classes, de acordo com as suas semelhanças, observáveis através do conjunto de variáveis explicativas apresentadas. Vai procurar-se detectar a presença de homogeneidade entre diversos distritos partindo de um método estatístico multivariado – a metodologia de *clusters*. Este método permite a obtenção de uma segmentação do território continental em regiões caracterizadas por um perfil que em «média» defina a população desempregada registada residente.

Tal segmentação será observada para os anos de 2008 e 2009, numa perspectiva comparativa *cross-section*. Vai procurar identificar-se padrões de continuidade ou de evolução nos perfis médios da população registada no conjunto dos 18 distritos que compõem o território de Portugal Continental, tendo em conta um conjunto de variáveis tornadas públicas pela instituição que gere o registo de desempregados em Portugal – o Instituto de Emprego e Formação Profissional (IEFP).

O artigo desenvolve-se da seguinte forma: na secção seguinte apresenta-se uma descrição do desemprego registado em Portugal Continental, tendo em conta um conjunto de características que descrevem os desempregados registados nos centros de emprego nacionais, e a sua evolução entre 2008 e 2009; posteriormente, desenvolve-se sucintamente a metodologia de análise seleccionada e que terá aplicação empírica na secção subsequente, na qual serão apresentados os resultados obtidos e, por último, as conclusões.

CARACTERIZAÇÃO NACIONAL E DISTRITAL DO DESEMPREGO REGISTRADO

Os desempregados registados nos Centros de Emprego nacionais apresentam um determinado número de características que os distinguem em aspectos tão básicos como o género, a idade, a escolaridade, a duração de desemprego ou a situação relativamente ao primeiro emprego. Tal caracterização está disponível publicamente ao nível concelhio e numa base mensal com o mês de Dezembro de cada ano a apresentar o *stock* de registos de desemprego final desse mesmo ano¹. Em termos globais, e considerando o território de Portugal Continental, a caracterização do desemprego registado, tanto em termos demográficos como laborais, apresenta-se na Tabela 1 (ver p. 12).

Os valores apresentados permitem visualizar as alterações na caracterização do desemprego registado entre 2008 e 2009, ano de crise financeira e económica mundial, em que se acentuou o registo de desempregados em Portugal. Esse incremento nos registos é notório em valor total como também ao nível de cada uma das características em análise, como se pode observar na Figura 1.

A distribuição das várias características altera-se entre características demográficas e laborais. Na Figura 1 (ver p. 13) é claro o crescimento do desemprego registado na população masculina, tornando a distribuição equitativa entre os géneros. É visível, ainda, o aumento da proporção de desemprego de menor duração e daqueles que procuram um novo emprego, assim como da proporção de desempregados de meia idade e com um nível de ensino mais elevado (nomeadamente com o ensino secundário).

TABELA 1

Caracterização global do desemprego registado em Portugal Continental

Características		2008	2009	Taxa de Crescimento (%)
Sexo	Homens	173.565	236.791	36,4
	Mulheres	228.980	267.984	17,0
Idade	<25 anos	53.732	64.116	19,3
	25-34 anos	93.046	119.441	28,4
	35-54 anos	176.083	229.054	30,1
	>=55 anos	79.684	92.164	15,7
Escolaridade	Menor 1º CB	21.728	27.408	26,1
	1º CB	119.557	142.665	19,3
	2º CB	74.864	96.529	28,9
	3º CB	78.734	99.976	27,0
	Secundário	70.486	94.442	34,0
	Superior	37.176	43.755	17,7
Situação face ao emprego	1º Emprego	32.262	37.556	16,4
	Novo Emprego	370.283	467.219	26,2
Duração de Desemprego	Menos 1 ano	259.288	329.358	27,0
	Mais 1 ano	143.257	175.417	22,4

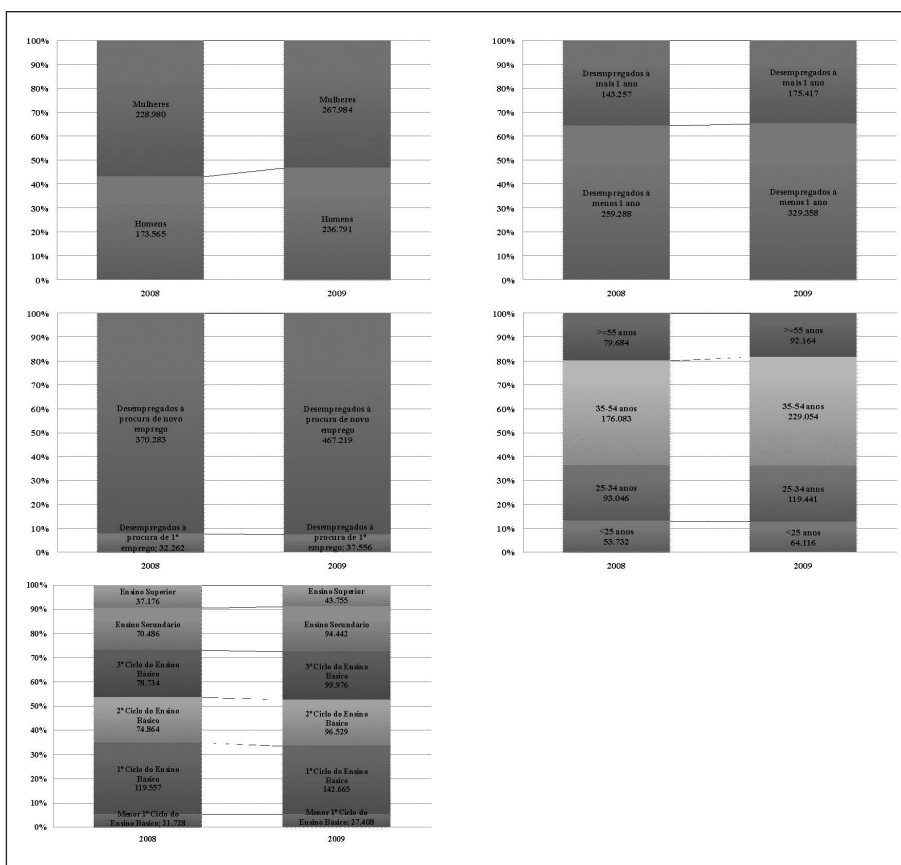
A distribuição das características do desemprego registado não é, no entanto, similar ao longo do território nacional. De facto, observam-se diferenças significativas relativamente a tal distribuição entre os vários distritos do país, tanto em termos demográficos como laborais. Vejam-se os exemplos apresentados na Figura 2 (ver p. 14).

Tendo em conta as características seleccionadas, pode observar-se, por exemplo, que os distritos de Lisboa, Setúbal e Santarém são os que apresentam uma menor disparidades entre géneros, em termos de desemprego registado, enquanto em distritos como Bragança, Guarda ou Porto, tal disparidade é significativa. No que se refere à duração do desemprego registado, Faro, Évora ou Portalegre constituem-se como os distritos com maior disparidade entre o período de duração dos seus desempregados, atenuando-se tal diferença em distritos como Porto, Braga ou Vila Real.

O exercício realizado na Figura 2 poderia repetir-se para outras características – a análise da distribuição dos registos de desemprego mostra um território com marcadas diferenças regionais na caracterização dos indivíduos inscritos nos Centros de

FIGURA 1

Evolução das características do desemprego registado entre 2008 e 2009



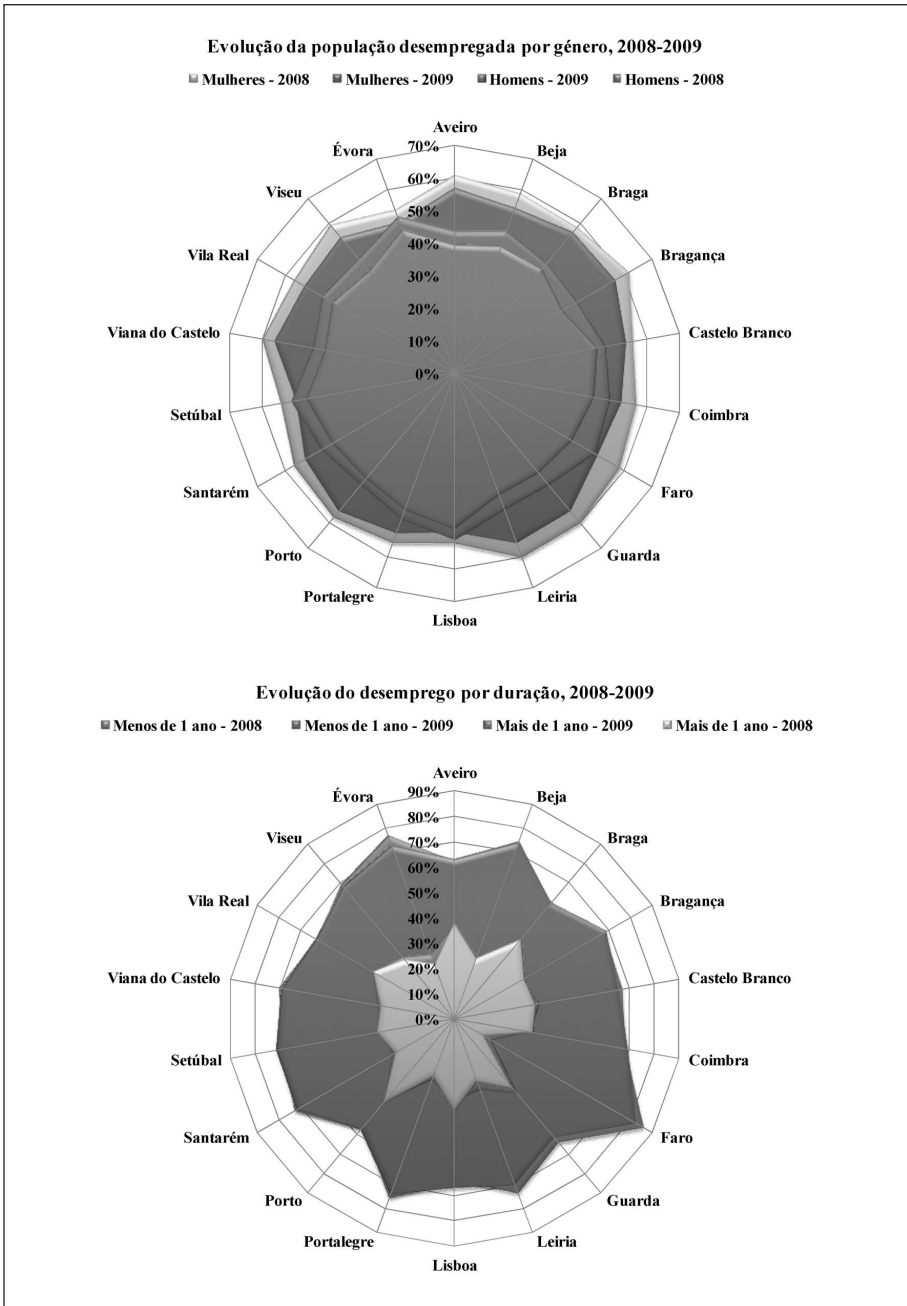
Emprego – diferenças estas que condicionam a política pública de emprego a implementar e os seus resultados.

BREVE INTRODUÇÃO À ANÁLISE DE CLUSTERS

Utilizada pela primeira vez por Tryon (1939), a análise de *clusters* consiste num conjunto de métodos da estatística multivariada que engloba diferentes algoritmos de classificação e optimização para organizar informações sobre variáveis e formar grupos homogéneos. Por outras palavras, a análise de *clusters* consiste numa análise exploratória de dados que pretende organizar diferentes variáveis para que o grau de associação entre duas variáveis seja máximo, se pertencem ao mesmo grupo, e mínimo, em caso contrário. Em resumo, o objectivo básico consiste em descobrir agrupamentos naturais de

FIGURA 2

Distribuição de características demográficas e laborais por distrito, 2008-2009



variáveis que apresentem, simultaneamente, coesão interna e separação externa. A análise de *clusters* não faz suposições em relação ao número de grupos ou às suas estruturas – o agrupamento é feito com base nas similaridades (ou dissimilaridades) entre os grupos caracterizadas pelas diferentes formas de cálculo das «distâncias».

Face ao exposto, como estamos a tratar informação associada com medidas de «similaridade» é necessário, numa primeira fase, organizar os dados observados em estruturas com significados físicos diferentes numa única medida padronizada de forma a que lhes seja garantida a mesma ponderação nos resultados finais.

Existem, depois, vários métodos usados para medir a distância ou similaridade entre os elementos de uma matriz de dados. Pretende-se inferir (i) sobre a distância entre observações do mesmo agrupamento e distância entre observações de agrupamentos diferentes, (ii) a dispersão das observações pertencentes a um agrupamento e (iii), a densidade das observações dentro e fora dos agrupamentos.

Sendo as variáveis a utilizar, no estudo, quantitativas, é aconselhada a utilização do método da *Distância Euclidiana* (Giudici e Figini, 2009) definido como a raiz quadrada do somatório dos quadrados das diferenças entre os valores de i e de j para todas as variáveis ($v=1,2,\dots,p$):

$$d_{ij} = \sqrt{\sum_{v=1}^p (X_{iv} - X_{jv})^2}$$

Para além de estabelecer qual a forma de cálculo da distância entre observações, é ainda necessário estabelecer a forma de cálculo da distância entre grupos. Tal passo é particularmente importante aquando da utilização de métodos hierárquicos de definição de *clusters*. Estes métodos permitem obter uma família de partições, associadas com subseqüentes níveis de agrupamento entre observações, que podem ser representadas graficamente por dendogramas, que não são mais que estruturas gráficas hierarquizadas em forma de árvore. De facto, representam partições que partem de divisões mais desagregadas de observações, para se irem aglomerando até todas estarem acondicionadas num único grupo.

Para a obtenção dos dendogramas podem aplicar-se diferentes técnicas, nenhuma melhor que a outra, *a priori*, pois não é possível seleccionar um método ideal para um determinado conjunto de informação. É necessário experimentar diferentes alternativas e comparar os respectivos resultados.

Neste trabalho, foram seleccionadas duas técnicas que se distinguem pelo facto de apenas exigirem o cálculo da matriz de distâncias entre observações, no caso do méto-

do de *complete-Linkage* (ou vizinho mais distante), e o cálculo desta matriz associada com a matriz original, como é o caso do método de Ward (Ward,1963).

No método de *complete linkage*, a distância entre dois grupos é definida como o máximo das distâncias entre cada observação de um grupo e cada observação de outro grupo:

$$d(\pi_1, \pi_2) = \max_{x \in \pi_1, y \in \pi_2} d(x, y)$$

No método de Ward, minimiza-se uma função objectivo, definida como a soma dos quadrados dos desvios das observações individuais relativamente à média do grupo em que são classificadas, cujo princípio visa criar grupos que possuam máxima coesão interna e máxima separação externa

$$d(\pi_1, \pi_2) = SP(\pi_1 \cup \pi_2) - (SP(\pi_1) + SP(\pi_2))$$

$$\text{Onde, } SP(\pi) = \sum_{x \in \pi} |x - \bar{x}_\pi|^2$$

Identificando o número de grupos a considerar, com a ajuda da leitura resultante dos dendogramas, está-se em condições de aplicar os Métodos de Optimização ou Métodos Não Hierárquicos. Estes baseiam-se directamente na escolha antecipada de um número de agrupamentos que conterão todos os casos. Procede-se, de seguida, a uma divisão de todos os casos pelos k grupos pré-estabelecidos e a melhor partição dos n casos será aquela que otimizar o critério escolhido. Um dos processos que se pode utilizar é o Método Partitivo Iterativo *k-means*. O método segue os seguintes passos: começa por fixar uma partição inicial dos indivíduos por um número de *clusters* pré-definido pelo analista; calcula, para cada *cluster*, o respectivo centróide; calcula as distâncias entre cada indivíduo e os centróides dos vários grupos; transfere cada indivíduo para o *cluster* relativamente ao qual se encontra a uma menor distância; recalcula os centróides de cada *cluster*; repete os passos anteriores até que todos os indivíduos se encontrem em *clusters* estabilizados e não seja possível efectuar mais transferências de indivíduos de um *cluster* para o outro. O método possui a desvantagem de limitar a procura óptima de partição dos indivíduos, já que se restringe ao número de *clusters* pré-definidos pelo utilizador. Tem, no entanto, a vantagem de definir as regiões pertencentes a cada *cluster*, bem como a distância de cada grupo.

APLICAÇÃO EMPÍRICA²

A aplicação empírica da análise de *clusters* que se apresenta neste trabalho segue os passos metodológicos sugeridos na literatura.

Começou por se seleccionar as observações em estudo. Face à disponibilidade concelhia do número de desempregados inscritos nos Centros de Emprego nacionais, caracterizados através de um conjunto distinto de variáveis, pode fazer-se uma agregação das observações concelhias por distrito. São, assim, os agrupamentos distritais, o objecto de estudo deste trabalho. Tal opção territorial explica-se por razões administrativas e de coesão política e social.

Escolhida a unidade territorial objecto de estudo foram, de seguida, seleccionadas, as variáveis a partir das quais foi obtida a informação necessária ao agrupamento dos distritos. As variáveis contemplam características individuais e laborais. Em termos de características individuais, está disponível informação relativa ao sexo, idade e escolaridade dos indivíduos registados como desempregados. Em termos de características laborais, é conhecida a duração do período de registo como desempregado, bem como a situação face ao mercado de trabalho. Para evitar que alguma medida de semelhança/distância reflectisse o peso das variáveis de maiores valores e dispersão, já que as variáveis se apresentam definidas em diferentes escalas de medida, procedeu-se à sua padronização prévia. Tal padronização consistiu no cálculo do peso relativo de cada variável no total de desempregados inscritos por distrito, ou seja, calculou-se o rácio de desempregados inscritos, por variável, face ao total de desempregados inscritos.

Face à natureza *cross-section* da análise que se apresenta, as etapas anteriores foram aplicadas ao ano de 2008 e 2009, respectivamente. Os valores registados, e aqui utilizados, recolheram-se em Dezembro de cada um dos anos e reflectem o *stock* de desempregados inscritos formalmente nos Centros de Emprego nacionais. Os resultados serão apresentados comparando os dois anos e as conclusões terão em conta as alterações ocorridas de um ano para o outro.

Aplicação dos métodos hierárquicos

Não se conhecendo, *a priori*, qual o método de classificação considerado ideal para aplicar aos dados em questão, foram, tal como explicitado na secção anterior, aplicados dois algoritmos distintos de classificação hierárquica dos distritos – o método de *complete linkage* e o método de Ward. Estes métodos permitiram a construção dos dendogramas apresentados nas Figuras 3 e 4 (p. 18) e, desta forma, uma primeira abordagem à definição do número de *clusters* a esperar na solução final de optimização. A visualização dos hipotéticos agrupamentos de distritos permite antecipar o número ideal dos mesmos e a sua composição.

A primeira conclusão a retirar é a de que os dois métodos apresentam resultados distintos, não só entre si, mas também entre os anos de 2008 e 2009. Repare-se, por

FIGURA 3
Dendrogramas obtidos pelo método do *complete linkage*, 2008-2009

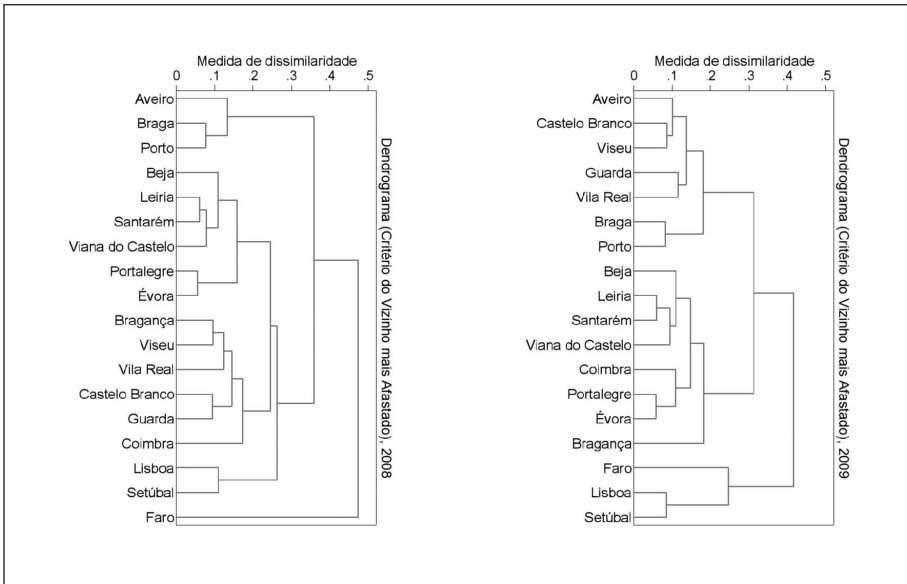
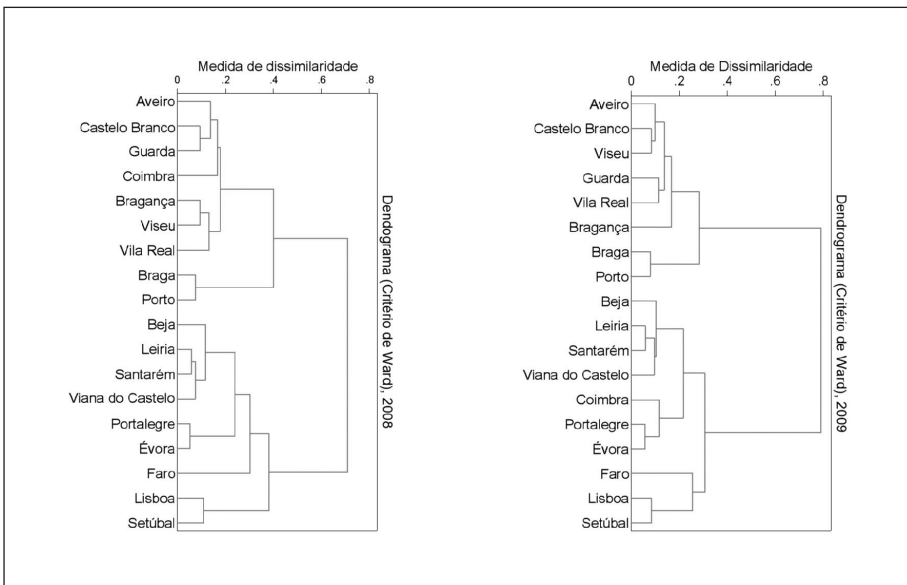


FIGURA 4
Dendrogramas obtidos pelo método de Ward, 2008-2009



exemplo, na distância máxima calculada por cada um dos métodos. Enquanto a distância máxima calculada no método de *complete linkage* não atinge os 0,5, no método de Ward tal distância sobe para cerca de 0,7. Por outro lado, enquanto no método de *complete linkage* as semelhanças parecem diminuir de 2008 para 2009, quando considerado o total de distritos, no método de Ward observa-se precisamente o contrário.

A análise dos agrupamentos apresentados em cada um dos métodos e para cada um dos períodos em análise permite ainda concluir pela semelhança óbvia entre alguns distritos – veja-se o caso dos distritos de Porto e Braga, dos distritos de Lisboa e Setúbal e ainda dos distritos de Portalegre, Évora, Leiria e Santarém. Estes conjuntos de distritos apresentam-se semelhantes entre si, mas distintos em relação uns aos outros. Em termos de disparidades saliente-se o distrito de Faro – cuja disparidade, em relação aos restantes, é notória em 2008 através do método *complete linkage*. Com o método de Ward e no ano de 2009, tal disparidade não é tão evidente, mas Faro mantém-se como um distrito que não é especialmente semelhante a nenhum outro.

Por fim, note-se que, para o ano de 2009, ambos os métodos utilizados, embora com as devidas diferenças de magnitude, parecem devolver os mesmos agrupamentos de distritos.

Aplicação da metodologia de optimização

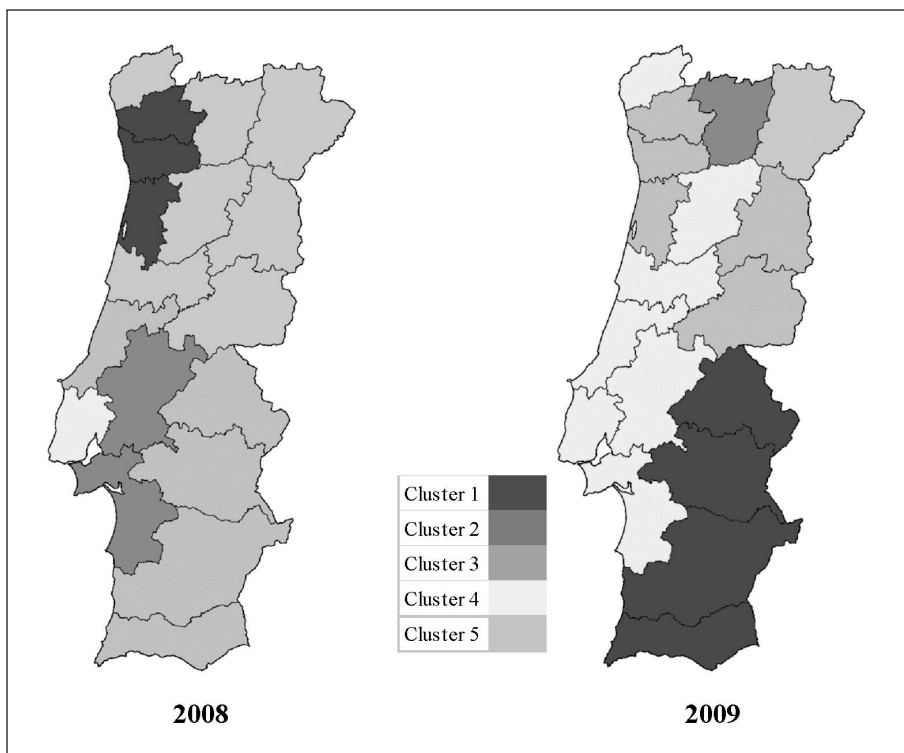
Os resultados obtidos através dos métodos hierárquicos podem ser, ou não, confirmados através dos resultados obtidos pelo método de optimização que seleccionámos para análise – o algoritmo *k-means*. Este algoritmo pressupõe que o número de agrupamentos seja conhecido. No entanto, esta hipótese é pouco realista na maioria dos problemas, uma vez que, normalmente, se explora um conjunto de dados com propriedades não conhecidas. Desta forma, um problema recorrente que surge não só quando se utiliza o algoritmo de *k-means*, mas também com outros métodos de geração de *clusters*, consiste em calcular o número de *clusters* k .

A abordagem, mais intuitiva e visual dos métodos hierárquicos, lançando pistas para a determinação do número de agrupamentos, não é, no entanto, conclusiva, relativamente ao número de agrupamentos de distritos a considerar. A observação visual parece indicar que os agrupamentos de distritos não devem ser inferiores a 5. Partindo deste valor intuitivo, aplicaram-se dois métodos que testam, no *software Stata*, o número ideal de *clusters* a considerar. Tanto o pseudo F-teste de *Calinski-Harabasz*, como o pseudo teste *Duda-Hart*, indicam que os distritos devem ser agrupados em cinco, sete ou mais *clusters*. Estes resultados repetem-se tanto em 2008

como em 2009. Seguindo uma regra de parcimónia de forma a evitar a perda excessiva de informação, optamos pela partição dos distritos em cinco *clusters*.

Os resultados da metodologia *5-means*, apresentam-se cartografados na Figura 5. Os perfis médios dos desempregados registados nos distritos que pertencem a cada *cluster* descrevem-se na Tabela 2 para o ano de 2008 e na Tabela 3 para o ano de 2009.

FIGURA 5
Distribuição espacial dos distritos por *cluster*, 2008-2009



Pela análise dos mapas de Portugal Continental, em 2008 e 2009, conclui-se que existe uma clara distinção entre distritos do norte e sul – para ambos os anos – e entre distritos do litoral e interior do país – o que é notório no ano de 2009. Conclui-se ainda que os resultados do método hierárquico se reforçam quando se analisa a proximidade entre os distritos do Porto e Braga surgindo, no entanto, com o método *k-means*, também uma proximidade entre estes distritos e o distrito de Aveiro que não se tinha visualizado antes e que persiste em ambos os períodos de análise. O distrito de Faro, que se havia apresentado como bastante distinto dos restantes, surge agora

próximo dos distritos sul interior do país, não se verificando uma alteração da classificação destes distritos entre 2008 e 2009.

Da análise visual da evolução da divisão geográfica do país entre 2008 e 2009, dois factos merecem destaque. Primeiro, Lisboa abandona o isolamento que apresentava em 2008, em termos de perfil dos seus desempregados registados, para se passar a assemelhar aos distritos do centro litoral do continente em 2009. Segundo, os distritos mais no norte interior do território – Vila Real e Bragança – surgem, em 2009, com perfis de desemprego registado muito distintos dos restantes distritos do país e, também, distintos dos perfis que apresentavam em 2008.

Os resultados visuais não explicitam, no entanto, qual o perfil do desempregado registado em cada um dos *clusters*. Para obter tais resultados, foi preciso calcular os valores médios das variáveis para cada um dos agrupamentos, para tentar perceber quais as variáveis que melhor os caracterizam. Tais resultados podem ser analisados nas tabelas seguintes (ver p. 22).

Em 2008, a divisão do território em cinco *clusters* resultou nos seguintes agrupamentos:

- *Cluster 1*: Aveiro, Braga e Porto. Este *cluster* é caracterizado por apresentar a maior percentagem de mulheres desempregadas e de trabalhadores com idade entre os 35 e os 54 anos, com o menor nível de escolaridade (inferior ou igual ao 1.º ciclo básico). É ainda o *cluster* que apresenta a maior proporção de indivíduos desempregados com uma duração do período de desemprego superior a um ano. Não se salientando, é dos *clusters* que apresenta mais desempregados à procura de novo emprego;

- *Cluster 2*: Santarém e Setúbal. Este é o segundo *cluster* com maior proporção de homens desempregados e aquele com maior proporção de trabalhadores com idade entre os 25 e os 54 anos, cuja escolaridade é essencialmente caracterizada pela frequência do 3.º ciclo básico. É ainda o *cluster* com menor proporção de desempregados que apresentam um grau de escolaridade superior. É o segundo maior *cluster* em termos de desempregados à procura de novo emprego e cuja duração de desemprego é inferior a 1 ano;

- *Cluster 3*: Bragança, Castelo Branco, Coimbra, Guarda, Viana do Castelo, Vila Real e Viseu. Sendo o *cluster* que apresenta o maior agrupamento de distritos, caracteriza-se essencialmente por ser o *cluster* que apresenta a maior proporção de indivíduos à procura de 1.º emprego (por oposição, é o que apresenta a menor proporção de indivíduos que já anteriormente haviam exercido uma actividade). Relativamente às restantes características analisadas, este *cluster* não se salienta especialmente. É o segundo com maior proporção de mulheres desempregadas, de trabalhadores jovens (com menos de 25 anos de idade) e com um nível de escolaridade superior.

TABELA 2
Proporção média de cada característica por *cluster*, 2008

Cluster		2008				
		1	2	3	4	5
Distritos		(3)	(2)	(7)	(1)	(5)
		Aveiro	Santarém	Bragança	Lisboa	Beja
		Braga	Setúbal	Castelo Branco		Évora
		Porto		Coimbra		Faro
				Guarda		Leiria
				Viana do Castelo		Portalegre
				Vila Real		
				Viseu		
Sexo	Homens	41,0	44,6	41,4	47,8	42,8
	Mulheres	59,0	55,4	58,6	52,	57,2
Idade	<25 anos	13,4	12,9	15,6	10,3	16,9
	25-34 anos	20,9	25,0	25,2	23,6	25,9
	35-54 anos	44,8	42,2	41,3	44,4	40,7
	>=55 anos	21,0	20,0	17,9	21,7	16,5
	Menor 1º CB	5,6	5,5	5,9	5,1	7,6
Escolaridade	1º CB	35,4	25,1	29,0	22,0	25,3
	2º CB	19,5	18,1	17,9	17,4	18,6
	3º CB	16,9	23,1	19,6	21,4	21,6
	Secundário	14,3	20,0	16,4	22,5	18,6
	Superior	8,3	8,1	11,1	11,6	8,3
Situação face ao emprego	1º Emprego	8,1	6,3	13,4	4,9	9,0
	Novo Emprego	91,9	93,7	86,6	95,1	91,0
Duração de Desemprego	Menos 1 ano	59,1	71,0	67,3	63,4	77,5
	Mais 1 ano	40,9	29,0	32,7	36,6	22,5

• *Cluster 4*: Lisboa. A unicidade deste agrupamento caracteriza-se pela maior proporção de homens desempregados, pela menor proporção de jovens menores de 25 anos e pela maior proporção de trabalhadores com mais de 55 anos. Caracteriza-se ainda pela maior disparidade entre níveis de escolaridade, sendo o *cluster* com maior proporção de desempregados registados sem qualquer nível de escolaridade e também o *cluster* com maior proporção de desempregados com o ensino superior. Por fim, é o *cluster* em que a proporção de desempregados à procura de novo emprego é maior;

• *Cluster 5*: Beja, Évora, Faro, Leiria e Portalegre. Este *cluster* caracteriza-se, essencialmente, pelo facto de apresentar a maior proporção de desempregados inscritos com idade inferior a 34 anos e, dentro desta faixa etária, daqueles com menos de 25 anos. Associado a esta constatação, este *cluster* é ainda caracterizado pelo facto de ser aquele que apresenta a maior proporção de desempregados de curta duração (inferior a

um ano). É também o *cluster* com piores resultados em termos de proporção de indivíduos sem qualquer escolaridade.

Em resumo, pode concluir-se que, em 2008, se destacam os factos que se seguem. Existe um *cluster* (*cluster 1*) caracterizado por desemprego essencialmente feminino, menos jovem e escolarizado e com períodos de duração dos registos de desemprego mais elevados. Existe ainda um *cluster* (*cluster 4*) que é, no conjunto dos agrupamentos do país, aquele em que a diferença entre desemprego feminino e masculino mais se esbate, ao contrário dos níveis de escolaridade muito díspares entre si. É também o *cluster* com maior proporção de desempregados à procura de novo emprego. Por fim, salienta-se o *cluster 5* caracterizado pela sua população desempregada jovem, à procura de primeiro emprego e com percursos de desemprego registado de curta duração (inferior a um ano).

Como já foi possível visualizar na Figura 5, com a evolução do desemprego agregado nacional, também ocorreram alterações em termos de perfis de desemprego regional. Vejam-se os resultados analíticos das alterações médias nos perfis de desemprego (ver Tabela 3, p. 24).

De seguida, é feita uma descrição das características médias que identificam o perfil de desempregado associado a cada *cluster*. Adicionalmente, são apresentadas as principais alterações que se verificaram em relação ao que ocorria em 2008.

- *Cluster 1*: Beja, Évora, Faro e Portalegre. Este *cluster* reúne, com a excepção do distrito de Leiria, os distritos que, em 2008, se reuniam sob a designação de *cluster 5*. Em 2009, este *cluster* caracteriza-se por ser aquele com maior proporção de homens desempregados e de jovens com idade inferior a 34 anos. Talvez devido à juventude dos seus desempregados seja o *cluster* que apresenta das maiores taxas de escolaridade, bem como a maior proporção de desempregados com períodos de duração de desemprego inferiores a um ano. Relativamente a 2009, este *cluster* manteve, essencialmente, as características que apresentava em 2008, com a excepção de ser o *cluster* com a maior proporção de desempregados à procura de 1.º emprego. De facto, o aumento dos níveis de desemprego em termos agregados, terá levado a que uma maior proporção de indivíduos perdesse o vínculo que tinha com o mercado de trabalho e se registasse, procurando um novo emprego;

- *Cluster 2*: Vila Real. Este *cluster* surge como uma das grandes constatações a efectuar em 2009 relativamente a 2008. Incluído, em 2008, num *cluster* que não apresentava características particularmente distintas dos restantes, surge agora desfasado. É o distrito com a maior proporção de desempregados com idade entre os 25 e os 34 anos, com a maior proporção de desempregados apenas com o 1.º ciclo do ensino

TABELA 3
Proporção média de cada característica por cluster, 2009

		2009				
<i>Cluster</i>		1	2	3	4	5
		(4)	(1)	(1)	(7)	(5)
Distritos		Beja Évora Faro Portalegre	Vila Real	Bragança	Coimbra Leiria Lisboa Santarém Setúbal Viana do Castelo Viseu	Aveiro Braga Castelo Branco Guarda Porto
Sexo	Homens	48,6	47,0	43,2	47,4	44,7
	Mulheres	51,4	53,0	56,8	52,6	55,3
Idade	<25 anos	16,4	15,7	15,2	13,1	13,4
	25-34 anos	26,3	23,5	24,5	25,3	21,6
	35-54 anos	41,8	43,9	41,9	43,7	45,2
	>= 55 anos	15,6	16,9	18,4	17,9	19,8
Escolaridade	Menor 1º CB	8,1	6,7	9,0	5,0	5,6
	1º CB	24,4	32,5	27,4	25,2	32,9
	2º CB	18,7	19,1	20,4	18,8	18,4
	3º CB	20,5	16,8	17,2	21,2	19,1
	Secundário	21,0	16,3	15,4	20,1	15,5
	Superior	7,3	8,7	10,6	9,6	8,5
Situação face ao emprego	1º Emprego	9,1	16,9	17,6	7,9	9,4
	Novo Emprego	90,9	83,1	82,4	92,1	90,6
Duração de Desemprego	Menos 1 ano	76,0	62,9	69,3	69,6	61,4
	Mais 1 ano	24,0	37,1	30,7	30,4	38,6

básico e dos distritos com maior proporção de indivíduos à procura de primeiro emprego, mas desempregados há mais de um ano;

- *Cluster 3*: Bragança. Tal como Vila Real, Bragança destaca-se em 2009 relativamente ao ano de 2008. O aumento do desemprego registado em Portugal acentuou a proporção de mulheres desempregadas e com idade superior a 35 anos, no distrito. É a região em que a proporção daqueles sem qualquer tipo de escolaridade é superior e, também, onde se encontra a maior proporção de indivíduos registados com níveis de ensino superior. Note-se ainda que, em 2009, se tornou o distrito com maior proporção de desempregados à procura de primeiro emprego, apresentando uma grande proporção de indivíduos com menos de um ano de desemprego;

- *Cluster 4*: Coimbra, Leiria, Lisboa, Santarém, Setúbal, Viana do Castelo e Viseu. Este agrupamento apresenta-se como aquele com a menor proporção de jovens desempregados com menos de 25 anos. Associado a este facto refira-se que é o *clus-*

ter com maior proporção de desempregados à procura de novo emprego. Este *cluster* agrupa os distritos do litoral centro, indiciando que as alterações no número de desempregados registados parece ter tido consequências semelhantes no perfil de desemprego contínuo desta região, geograficamente;

- *Cluster 5*: Aveiro, Braga, Castelo Branco, Guarda e Porto. Em relação a 2008, este *cluster* reúne os distritos que se incluíam no *cluster 1* e surge, em 2009, com uma composição deveras interessante. Aos três distritos do norte litoral acrescentam-se dois distritos do centro interior. Apesar da dispersão geográfica do *cluster*, os cinco distritos repartem entre si as características que se seguem. Em 2009, perdem para Bragança, a maior proporção de mulheres desempregadas, mas mantêm-se como um *cluster* em que a diferença entre homens e mulheres desempregados é maior. É o *cluster* com maior proporção de indivíduos registados com idade superior a 55 anos e, também, dos que apresenta menores níveis de escolaridade da sua população desempregada. Face à característica anterior, não será de estranhar que seja o agrupamento com maior proporção de desempregados de longa duração (duração superior a 1 ano).

CONCLUSÕES

Da análise de resultados, a principal conclusão a retirar é a de que existe uma clara distinção entre diferentes regiões do país, especialmente no que toca à zona litoral e interior e ao Norte e Sul. Podem determinar-se diferentes perfis de desemprego consoante a zona geográfica. Outros estudos, aplicando a mesma metodologia aqui adoptada, tinham chegado à conclusão de que a zona litoral apresenta características em termos de desenvolvimento distintas do interior, alertando para o facto de que qualquer política de desenvolvimento regional dever ter em conta esta realidade (Soares *et al.*, 2003). Estes autores chamam particularmente a atenção para o facto de qualquer política de desenvolvimento regional dever prevenir a migração de jovens para o litoral, através da criação de emprego nas áreas menos desenvolvidas do interior do País.

A segunda conclusão a retirar é a de que existe espaço para implementar diferentes políticas de emprego no território continental. Encontraram-se regiões onde se concentra o desemprego jovem, o desemprego de longa duração ou o desemprego feminino. Há ainda regiões caracterizadas por uma população registada com baixos níveis de escolaridade, sem qualquer ligação prévia ao mercado de trabalho ou, pelo contrário, com anteriores vínculos ao mesmo.

Por fim, com o acentuar dos registos de desemprego entre 2008 e 2009, nota-se também um acentuar da fragilidade estrutural nos perfis de desemprego de algumas regiões nacionais. Veja-se o caso de Vila Real e Bragança.

AGRADECIMENTOS

Agradecem-se os contributos dos participantes do 16.º Congresso da APDR, Funchal, onde uma versão deste artigo foi apresentada.

NOTAS

1. Ver o site do IEFEP para a recolha dos dados estatísticos disponíveis em: <http://www.iefp.pt/estatisticas/MercadoEmprego/ConcelhosEstatisticasMensais/Paginas/Home.aspx>.

2. Os resultados empíricos foram obtidos através do *software* econométrico Stata.

REFERÊNCIAS BIBLIOGRÁFICAS

ARANDARENKO, M. e JUVICIC, M. (2007), «Regional labour market differences in Serbia: assessment and policy recommendations». *The European Journal of Comparative Economics*, vol. 4(2), pp. 299-317.

BABUCEA, A.G. (2009), «Using cluster analysis for studying the proximity of registered unemployment at the level of counties at the beginning of the economic Crisis». *Annals of the Constantin Brâncuși*, University of Târgu Jiu, Economic Series, n.º 1/2009.

FERTIG, M.; SCHMIDT, C.M. e SCHNEIDER, H. (2006), «Active labour market policy in Germany – is there a successful policy strategy?», *Regional Science and Urban Economics*, vol. 36, pp. 399-430.

GIUDICI, P. e FIGINI, S. (2009), **Applied Data Mining for Business and Industry**. (2.ª ed.), John Wiley & Sons Ltd., Londres.

LÓPEZ-BAZO, E.; DEL BARRIO, T. e ARTÍS, M. (2005), «Geographical distribution of unemployment in Spain», *Regional Studies*, vol. 39(3), pp. 305-318.

MARELLI, E. (2007), «Specialization and convergence of European regions» *The European Journal of Comparative Economics*, vol. 2, pp. 149-178.

NADIYA, D. (2008), «Econometric and cluster analysis of potential and regional features of the labor market of Poland». *Ekonomia*, vol. 21, pp. 28-44.

SOARES, J.O.; MARQUES, M.M.L. e MONTEIRO, C.M.F. (2003), «A multivariate methodology to uncover regional disparities: a contribution to improve European Union and governmental decisions». *European Journal of Operational Research*, vol.145, pp. 121-135.

TRYON, R.C. (1939), **Cluster Analysis**. McGraw-Hill, Nova Iorque.

WARD, J. (1963), «Hierarchical grouping to optimize an objective function». *Journal of American Statistical Association*, vol. 58, pp. 236-244.