

Especificidades Regionais do Emprego em Portugal: O Contributo da Análise Multivariada de Objectos Simbólicos

Resumo

Numa fase de profunda integração europeia, aumenta, paradoxalmente, a necessidade de conhecimento das especificidades regionais e locais, assistindo-se ao que diversos autores passaram a designar de processo de **glocalização** (aumento da importância do local num contexto de globalização).

O Mercado de Trabalho constitui uma das principais áreas de análise onde este processo é bem notório sendo indispensável a fundamentação das diversas *performances* do Mercado de Trabalho registadas inter-países e, mesmo, inter-regiões de um mesmo país.

Este estudo pretende contribuir para a compreensão das especificidades regionais do(s) mercado(s) de trabalho em Portugal, possibilitando assim a adopção de medidas de emprego que não se revelem demasiado ambíguas nas suas consequências.

Assim, com base no vasto conjunto de informação que consta do Inquérito ao Emprego, para o segundo trimestre de 1998, caracterizar-se-ão os mercados de trabalho ao nível das NUTS II (Nomenclatura de Unidades Territoriais para fins Estatísticos – nível II), dando ênfase à análise das características dos empregados.

Dada a elevada dimensão da informação em questão, recorre-se a um método relativamente inovador na área da análise de dados multivariados: a Análise de Dados de Objectos Simbólicos, agregando a informação de base com o intuito de extrair algumas das suas principais especificidades.

Ana Alexandrino da Silva
Direcção Regional de Lisboa
e Vale do Tejo/INE

Carlos Marcelo
Direcção Regional de Lisboa
e Vale do Tejo/INE

Duarte Rodrigues
Direcção Regional de Lisboa
e Vale do Tejo/INE

Introdução

Este estudo enquadra-se no objectivo de aprofundamento e divulgação do que podemos designar por uma nova filosofia de análise multivariada de dados: a criação e manipulação dos objectos simbólicos. Trata-se de uma metodologia inovadora, que está em contínuo desenvolvimento e aperfeiçoamento.

Assim, aplicaram-se os métodos associados à análise de objectos simbólicos aos dados do Inquérito ao Emprego, com o objectivo de aferir das principais especificidades regionais do conjunto dos empregados.

Num primeiro ponto faz-se uma breve apresentação dos objectos simbólicos e suas vantagens relativamente à análise clássica e ainda uma apresentação da informação que esteve na base deste estudo.

Posteriormente, apresentam-se os retratos territoriais de emprego ao nível de NUTS II, elaborados com base na aplicação de diversos métodos de análise multivariada aos objectos simbólicos construídos.

Por último, aborda-se ainda a importância de outras segmentações na discriminação das características dos empregados, nomeadamente, as segmentações por sexo e por escalões etários.

1. Apresentação dos Métodos e da Informação

1.1. Análise de Dados Simbólicos

Apesar das primeiras investigações no campo da análise de dados simbólicos se terem iniciado há cerca de doze anos com um *paper* de E.Diday (1988), a divulgação destas técnicas de análise beneficiou em grande medida da publicação recente do livro "Analysis of Symbolic Data: exploratory methods for extracting statistical information from complex data" (Bock et Diday; 2000). Este surge como o relatório científico final de um projecto europeu designado de SODAS (Symbolic Official Data Analysis System), financiado pela Comunidade Europeia através do EUROSTAT – programa DOSIS e desenvolvido de 1996 a 1999. O propósito deste projecto foi o desenvolvimento de um software para análise multivariada de dados, de fácil utilização, permitindo analisar conjuntos de informação com estruturas complexas. Neste projecto colaboraram 17 entidades europeias de 9 países, entre investigadores que definiram os métodos, programadores e equipas de teste. O INE colaborou neste projecto como equipa de teste.

1.1.1 O que são dados simbólicos?

Os dados simbólicos vêm dar resposta à necessidade de analisar informações de elevada dimensão e estrutura complexa, consonantes com as características das bases de dados dos nossos dias. São dados por tabelas, nas quais em cada célula não existe apenas um valor, mas vários, os quais podem ser ponderados, ligados por regras lógicas

(dependência hierárquica ou lógica das variáveis), ou ainda estruturados com base em taxonomias entre categorias. Cada célula pode ainda conter um intervalo ou uma distribuição de frequências, como casos particulares das possibilidades descritas.

Tendo como base esta nova geração de tabelas, às quais foi convencionado chamar objectos simbólicos (OS), foram desenvolvidos um conjunto de métodos que permitem visualizar, comparar, obter hierarquias e efectuar análises factoriais/discriminantes ou de clusters sobre estes dados agregados.

1.1.2 Que vantagens apresenta face à análise clássica?

Este método permite estender os tradicionais métodos de análise exploratória de informação às variáveis simbólicas que, tal como apresentado no ponto anterior, podem assumir valores como subconjuntos de categorias, extremos de intervalos ou distribuições de frequências, contrastando com a análise clássica, onde as variáveis apenas podem assumir um único número, categoria ou nível. Note-se que a análise clássica é um caso particular da análise de dados simbólicos.

A principal vantagem reside no facto de, com recurso à análise de objectos simbólicos, ser possível analisar as complexas bases de dados dos nossos dias, que raras vezes preenchem os requisitos necessários à análise clássica. Para além disso, a agregação dos dados iniciais sob a forma de objectos simbólicos, permite preservar a confidencialidade dos microdados.

Na medida em que ao longo deste estudo apenas se exploram parte das vantagens que a utilização de dados simbólicos permite, remete-se uma análise mais profunda desta questão para o livro supracitado (Bock et Diday; 2000), secção 1.7.

1.2. Informação utilizada

O Inquérito ao Emprego por NUTS II de Portugal (Norte, Centro, Lisboa e Vale do Tejo, Alentejo, Algarve, Açores e Madeira), para o 2º trimestre de 1998, constitui a fonte de informação deste estudo. Do vasto conjunto de informação passível de extrair deste inquérito, restringiu-se a análise às variáveis que permitem caracterizar os empregados no seu conjunto, focalizando-se então nos 21 197 indivíduos presentes na amostra deste inquérito que se encontravam empregados.

Por se tratar de um inquérito por amostragem, aos seus resultados está associado um erro, chamado erro de amostragem, o qual pode ser quantificado tomando como base o desenho da amostra utilizado. Assim, poderemos ter os resultados representados sobre a forma de intervalos de confiança (IC). Por questões de simplificação do cálculo exacto dos erros de amostragem neste exercício que se apresenta assume-se que a extracção da amostra foi efectuada utilizando uma amostragem aleatória simples. São calculados os IC a 95%, considerando as proporções das variáveis $\left(\left[p_i - 1,96 \sqrt{\frac{p_i(1-p_i)}{n_i}}, p_i + 1,96 \sqrt{\frac{p_i(1-p_i)}{n_i}} \right] \right)$ sendo (p_{ij}) a proporção dos indivíduos com a variável j no OS i e (n_i) o número de indivíduos no OS i cuja lista se segue:

Figura 1.1 – Variáveis e categorias utilizadas na análise

Variáveis	Categorias
Sexo	Feminino – <i>Fem</i> ; Masculino – <i>Masculino</i>
Escalão etário	Entre os 15 e os 24 anos – 15-24 anos ; Entre os 25 e os 34 anos – <i>25-34 anos</i> Entre os 35 e os 44 anos – <i>35-44 anos</i> ; Entre os 45 e os 54 anos – 45-54 anos Entre os 55 e os 64 anos – 55-64 anos
Situação Familiar	Solteiro; Casado ou junto; Divorciado ou separado; Viúvo
Trabalha em Portugal?	Sim; Não
Recebe alguma reforma ou pensão?	Sim; Não
Procurou ou pretende arranjar um emprego?	Procurou emprego; Pretende arranjar um emprego; Outros
Actividade principal	Agricultura e pesca – Primário Indústrias extractivas e transformadoras – Indústria Produção de electricidade, gás e água - <i>Elect_gás_água</i> Construção – Construção Comércio por grosso e a retalho – Comércio Alojamento e restauração – Aloj_Rest Transportes armazenagem e comunicações - <i>Transp_Comunic</i> Actividades financeiras, imobiliárias e serviços prestados às empresas - Act_Financeiras Administração pública – Admin_Pública Educação – Educação Saúde e acção social - <i>Saúde</i> Outras actividades de serviços colectivos, famílias com empregados domésticos, organismos internacionais - Out_Serv_Col
Profissão principal	Forças armadas – <i>Forças_Armadas</i> Quadros superiores da administração pública, dirigentes e quadros superiores de empresa - Quadros_superiores Especialistas das profissões intelectuais e científicas – Especialistas Técnicos e profissionais de nível intermédio – Técnicos Pessoal administrativo e similares – Administrativos Pessoal dos serviços e vendedores – Serviços Agricultores e trabalhadores qualificados da agricultura e pescas – Agricultores Operários, artífices e trabalhadores similares - Operários Operadores de instalações e máquinas e trabalhadores da montagem – Máquinas Trabalhadores não qualificados - Não_qualificados
Situação na profissão principal	Trabalha por conta própria – Conta_própria Trabalha por conta de outrem – <i>Conta_outrem</i> Outros casos – <i>Outros_casos</i>
Como encontrou a actual ocupação?	Por conta própria – Pl_conta_própria Por relações pessoais - Pl_relações_pessoais Por outros meios - <i>Pl_outros_meios</i>
Trabalha a tempo completo ou parcial?	A tempo completo - T/_completo ; A tempo parcial - <i>T/_parcial</i>
Está inscrito na segurança social?	Sim; Não

Variáveis	Categorias
Horário de trabalho efectivo	Menos de 24 horas – <i>Menos_de_24_horas</i> ; De 25 a 34 horas – <i>De_25_a_34_horas</i> De 35 a 39 horas – <i>De_35_a_39_horas</i> ; Mais de 40 horas – <i>Mais_de_40_horas</i>
Trabalha um n.º de horas diferente do normal para a sua profissão no seu local de trabalho?	Sim; Não
Gostaria de trabalhar um n.º diferente de horas das que faz actualmente?	Sim; Não
O actual emprego é o 1.º que tem ou já trabalhou antes?	Primeiro emprego; Já trabalhou antes
Para além da actividade principal exerce mais alguma actividade?	Sim; Não
Sabe ler ou escrever?	Sim; Não
Está actualmente a estudar ou a receber formação?	Sim; Não
Nível de instrução completo	Nenhum – <i>Nenhum</i> ; Obrigatório – <i>Obrigatório</i> ; Secundário – <i>Secundário</i> ; Superior – <i>Superior</i>
Condição perante o trabalho um ano antes	Estava empregado; Não estava empregado

NOTA: De todas as variáveis-categorias testadas optou-se por se analisar apenas as que se encontram a **bold** neste quadro. Para as variáveis analisadas apresenta-se a itálico o *label* que lhe está atribuído e que constará em algumas figuras ao longo do estudo.

As análises a efectuar terão como base diferentes tabelas de dados simbólicos, consoante o caso em estudo, em que o n.º de linhas é o n.º de objectos simbólicos (dependendo da análise em causa) e as colunas são as variáveis intervalares (intervalos de confiança a 95% para a proporção) associadas.

Por exemplo, para o caso em que o objecto simbólico é constituído apenas com base na região, temos o seguinte exemplo da descrição do objecto simbólico correspondente à região dos Açores (Figura 1.2) e de parte da tabela que contém todos os sete objectos simbólicos e sua caracterização (Figura 1.3).

Figura 1.2 – Exemplo da descrição de um objecto simbólico

Aço=	Act_Financeiras=[2.43:4.11]
And	Admin_Pública=[10.44:13.50]
And	Aloj_Rest=[2.63:4.37]
And	Comércio=[12.11:15.36]
And	Construção=[10.50:13.57]
And	Educação=[6.77:9.33]
And	Indústria=[9.40:12.34]
And	Out_Serv_Col=[8.39:11.20]
And	Primário=[16.14:19.77]
And	P/ conta própria=[24.37:28.53]
And	P/ relações pessoais=[24.72:28.90]
And	15 - 24 anos=[17.16:20.86]
And	45 - 54 anos=[14.95:18.47]
And	55 - 64 anos=[7.10:9.72]
And	De 35 a 39 horas=[20.10:24.02]
And	Mais de 40 horas=[65.39:69.81]
And	Menos de 24 horas=[5.14:7.43]
And	Nenhum=[7.50:10.18]
And	Obrigatório=[75.37:79.33]
And	Superior=[4.48:6.65]
And	Administrativos=[8.06:10.82]
And	Agricultores=[16.10:19.72]
And	Especialistas=[3.28:5.18]
And	Máquinas=[4.42:6.58]
And	Não qualificados=[15.24:18.79]
And	Operários=[19.94:23.85]
And	Quadros superiores=[2.62:4.36]
And	Técnicos=[6.55:9.09]

Por último de referir que os dados intervalares podem resultar¹:

- de imprecisão na avaliação de fenómenos associados a resultados extrapolados de inquéritos por amostragem, tal como os que se analisam neste estudo, sendo neste caso os limites dos intervalos dados pelo intervalo de confiança das estimativas;
- da variabilidade da informação, sendo que neste caso os limites do intervalo podem representar o mínimo e o máximo de uma dada distribuição “perfeitamente” conhecida.

Figura 1.3 – Exemplo de parte de uma tabela de objectos simbólicos e sua caracterização

	Out_Serv_Col	Primário	P/conta própria	P/relações pessoais	15 - 24 anos	45 - 54 anos	55 - 64 anos
AÇO	[8.39:11.20]	[16.14:19.77]	[24.37:28.53]	[24.72:28.90]	[17.16:20.86]	[14.95:18.47]	[7.10:9.72]
ALE	[4.66:6.99]	[10.12:13.32]	[18.03:22.01]	[33.24:38.01]	[10.00:13.18]	[19.14:23.21]	[13.25:16.81]
ALG	[4.41:6.37]	[9.22:11.87]	[23.51:27.27]	[26.79:30.70]	[10.83:13.67]	[20.50:24.09]	[12.00:14.95]
CEN	[3.39:4.93]	[16.68:19.66]	[31.45:35.09]	[25.32:28.74]	[12.19:14.83]	[18.60:21.69]	[14.34:17.15]
LVT	[7.62:9.21]	[3.77:4.94]	[18.15:20.41]	[41.84:44.68]	[11.71:13.62]	[21.57:23.98]	[11.20:13.07]
MAD	[5.31:7.59]	[11.95:15.11]	[16.40:19.97]	[29.76:34.07]	[15.37:18.85]	[14.96:18.41]	[10.25:13.23]
NOR	[4.96:6.04]	[10.83:12.35]	[20.76:22.72]	[38.58:40.90]	[18.21:20.08]	[16.71:18.52]	[9.07:10.48]

¹ Para mais detalhes ver (Bock e Diday; 2000).

2. Retratos Territoriais de Emprego

Com o intuito de caracterizar as principais especificidades do emprego a nível territorial, toda a análise elaborada neste ponto é feita com base em apenas sete objectos simbólicos, correspondentes às NUTS II de Portugal, ignorando outras características dos empregados que obviamente também contribuem decisivamente para a sua diferenciação. A incorporação de duas das principais características (sexo e escalão etário) será analisada no próximo ponto.

Assim, apresenta-se primeiro uma representação em “estrela” dos valores (intervalares) de cada categoria para cada objecto simbólico, para posteriormente se tentar, através de uma análise de componente principais, encontrar os principais vectores de caracterização do emprego, bem como o posicionamento relativo das diversas regiões nesses vectores. Por último, recorre-se ao método de *Divisive Clustering* para proceder a uma divisão das diversas regiões em grupos, consoante as suas similaridades/dissimilaridades, sendo ainda possível perceber qual a variável decisiva em cada partição de cada grupo.

2.1. Caracterização geral do emprego regional

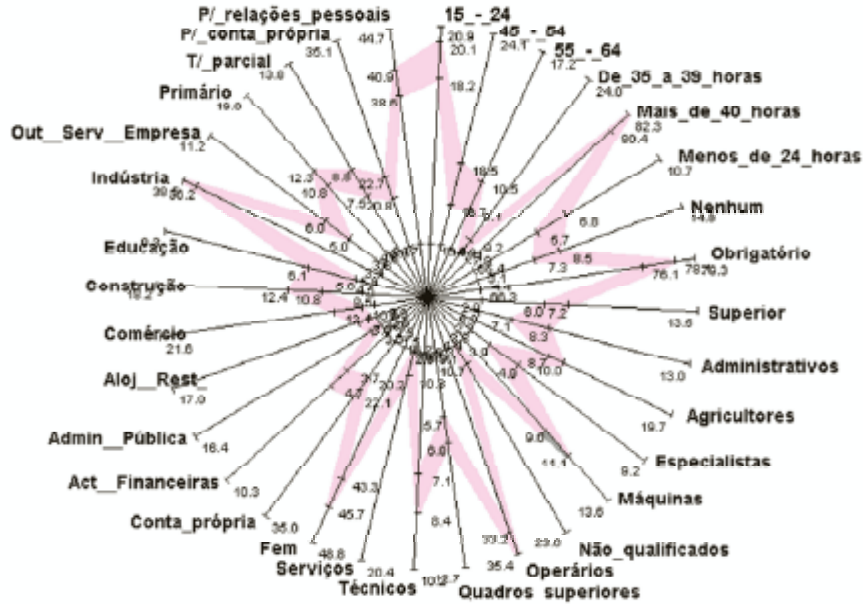
Para uma primeira caracterização do emprego regional com vista à apresentação das suas especificidades recorreu-se ao Editor de Objectos Simbólicos para a representação em “estrela” dos valores (intervalares) de cada categoria para cada objecto simbólico. Este tipo de gráficos permite evidenciar, em termos relativos, as características de um determinado objecto simbólico. Para uma análise correcta destes gráficos:

- os objectos simbólicos devem ser comparados entre si, em cada categoria;
- as categorias não devem ser comparadas entre si, em cada objecto simbólico, uma vez que apresentam escalas distintas.

Por exemplo, o facto de uma região ter um “pico” na categoria comércio significa apenas que comparativamente às restantes regiões esta assume o valor mais alto. Não estará correcto induzir que o comércio representa a característica com mais importância na região. Da mesma forma, o facto de a categoria comércio ter um “pico” maior que outra categoria, não significa necessariamente que apresente valores mais elevados². Por último, refira-se que a largura da “estrela” em cada categoria representa a variabilidade da mesma, neste caso expressa pela amplitude do intervalo de confiança associado.

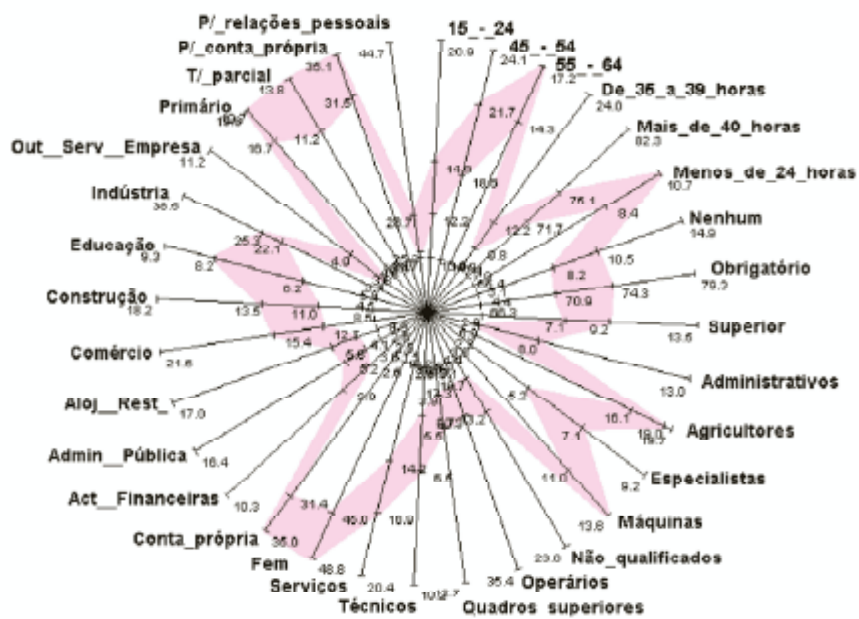
² Esta questão é bem visível se forem observados os valores das escalas.

Figura 2.1 – Região Norte



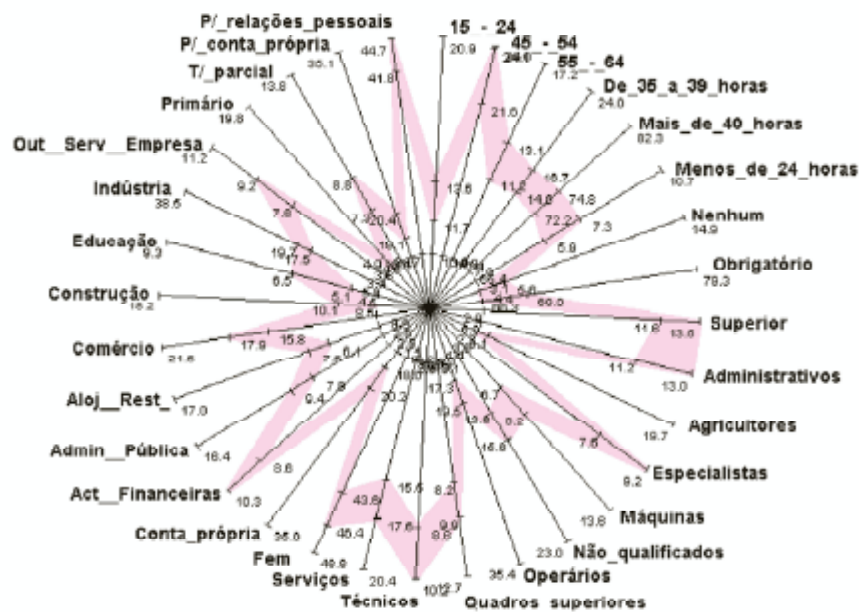
Comparativamente às outras regiões, são os empregados da região Norte que mais trabalham na *indústria* (cerca de 40%) e exercem a profissão de *operários*. O perfil que sobressai, em termos relativos, indicia uma mão-de-obra jovem (entre os 15 e os 24 anos), com o ensino *obrigatório*, onde o horário laboral excede as 40 horas semanais.

Figura 2.2 – Região Centro



Na região Centro, é visível a importância que o *primário* (cerca de 17%) tem no emprego da região, associado naturalmente a um peso elevado de indivíduos que exerce a profissão de *agricultor*. É também nesta região que a profissão de operador de *máquinas* assume um peso significativo, comparativamente às outras regiões portuguesas. As características que distinguem os empregados na região Centro das restantes regiões permitem traçar o seguinte perfil dominante: são do sexo *feminino*, têm entre *55 anos e 64 anos*, trabalham *menos de 24 horas* e a *tempo parcial*, por *conta própria* e conseguiram encontrar emprego também *por conta própria*.

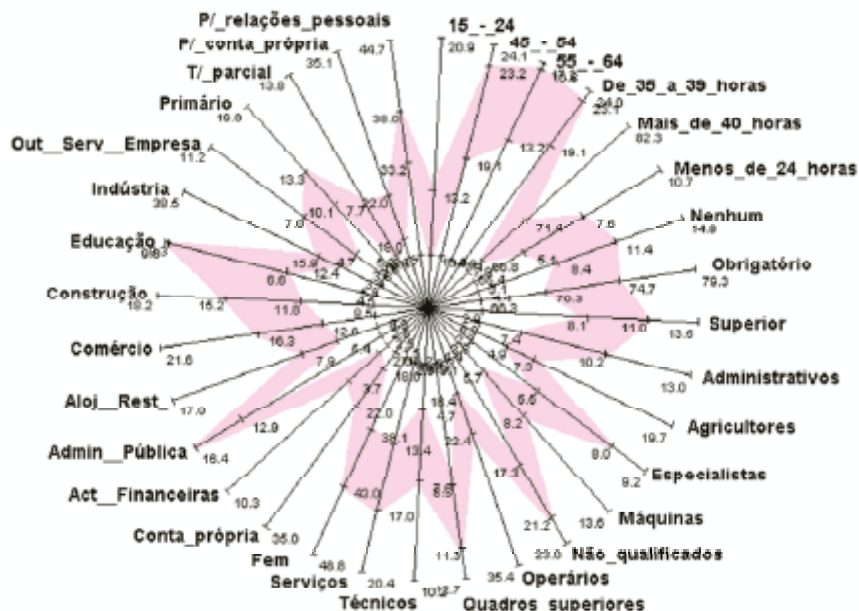
Figura 2.3 – Região Lisboa e Vale do Tejo



As actividades económicas em que os empregados na região de Lisboa e Vale do Tejo se distinguem das restantes regiões são *actividades financeiras* (cerca de 10%). É nesta região que os *especialistas*, os *técnicos* e os *administrativos* têm um peso mais expressivo, relativamente às restantes regiões. O empregado-tipo tem entre *45 e 54 anos*, o ensino *superior* e encontrou o presente trabalho através de *relações pessoais*.

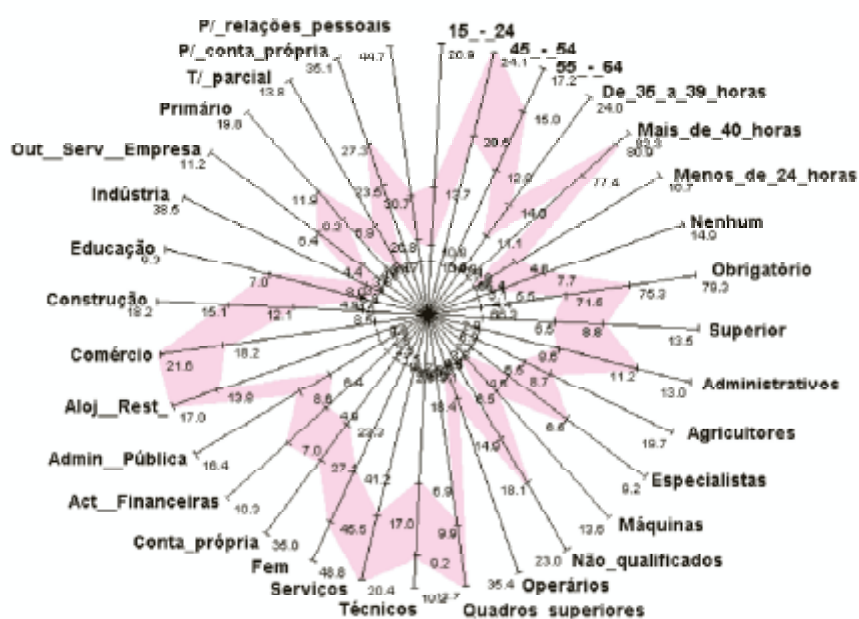
É no Alentejo que a *administração pública* (cerca de 16%) e a *educação* apresentam um peso superior relativamente às restantes regiões. Sobressai ainda uma maior frequência de pessoas entre os *45 e os 64 anos* que trabalham entre as *35 e as 39 horas*.

Figura 2.4- Região Alentejo



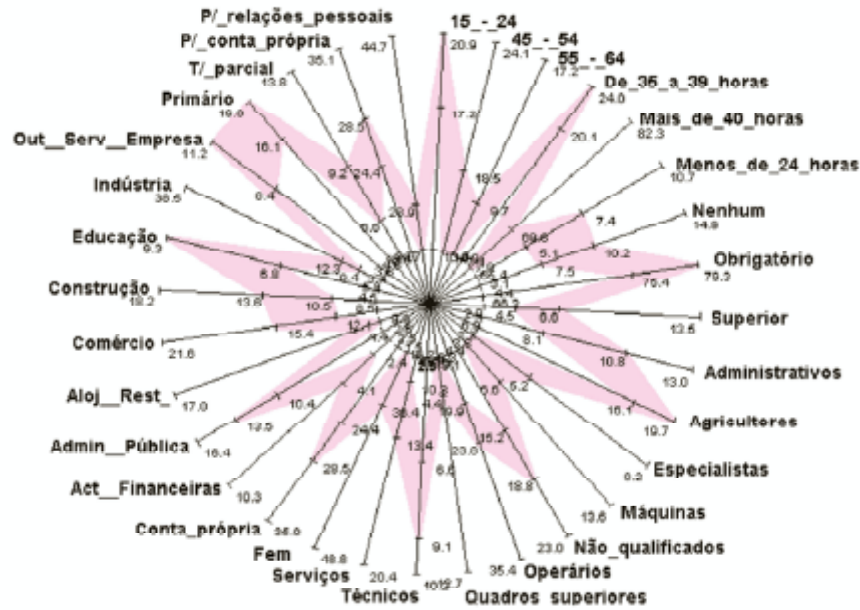
O alojamento e restauração assim como o comércio (cerca de 21%) surgem como as actividades económicas com mais expressão no Algarve relativamente às restantes regiões. As características que mais distinguem os empregados desta região são as seguintes: trabalham *mais de 40 horas*, são *quadros superiores* e pessoal dos *serviços* e têm entre *45 e 54 anos*.

Figura 2.5- Região Algarve



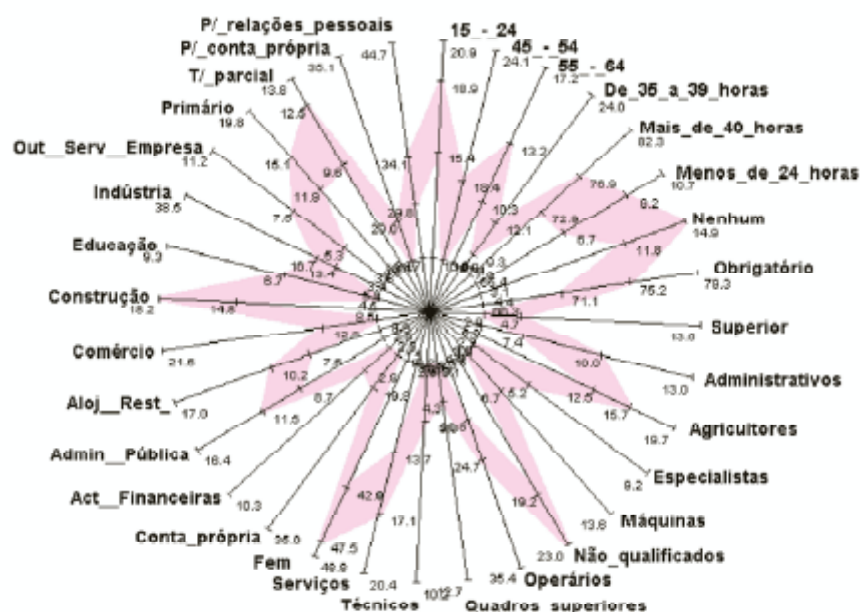
Nos Açores, o *primário* (cerca de 20%), os *outros serviços colectivos* e a *educação* são as actividades económicas que se diferenciam relativamente às outras regiões portuguesas. Os empregados desta região destacam-se das restantes pelas seguintes características: têm a profissão de *agricultores*, trabalham entre *35 e 39 horas*, têm entre *15 e 24 anos* e, no máximo, o ensino *obrigatório*.

Figura 2.6– Região Açores



A região da Madeira apenas sobressai das restantes pela actividade económica *construção* (cerca de 18%), a profissão *não qualificados*, o sexo *feminino* e *nenhum* nível de instrução.

Figura 2.7– Região Madeira



2.2. Principais vectores de caracterização do emprego

Através da análise factorial em componentes principais é possível reduzir a dimensão da informação e caracterizar os objectos simbólicos com base num número reduzido de novas variáveis (factores ou componentes principais) retendo grande parte da variância dessa informação. Neste caso, optou-se por analisar apenas

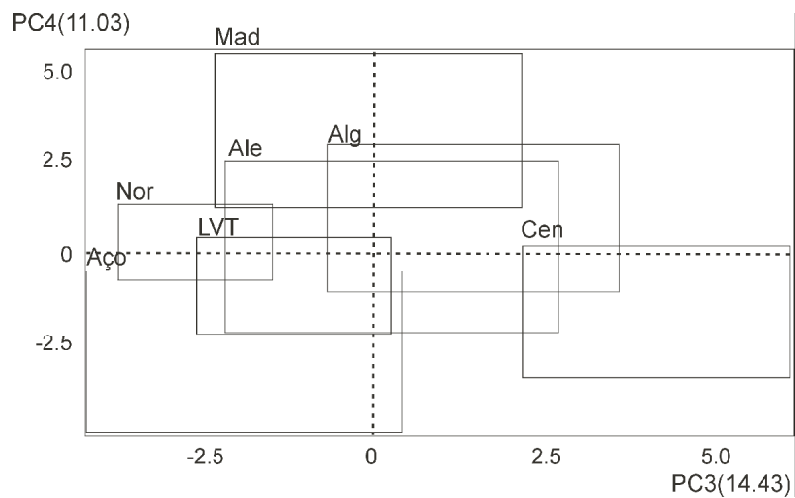
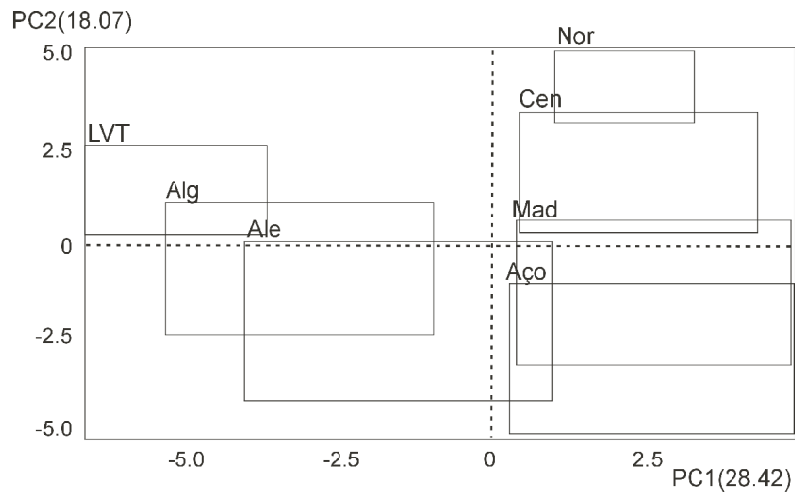
quatro factores que retêm em conjunto 72% da variância total da informação original (composta por 32 variáveis).

A análise dos planos factoriais, conjugada com as correlações entre as variáveis originais e as quatro componentes principais permite ter uma melhor percepção dos principais vectores que representam as especificidades regionais de emprego e que possibilitam discriminar as diversas regiões.

Figura 2.8– Matriz de Correlações Variáveis Iniciais / Componentes Principais (CP1-CP4)

	CP1	CP2	CP3	CP4
Actividades Financeiras	-0.787	0.312	-0.165	-0.183
Administração Pública	-0.183	-0.851	-0.153	0.018
Alojamento e Restauração	-0.498	-0.235	0.330	0.518
Comércio	-0.724	-0.158	0.291	-0.143
Construção	0.366	-0.393	0.216	0.646
Educação	0.151	-0.447	0.147	-0.509
Indústria	0.337	0.852	-0.219	-0.048
Outros serviços colectivos	-0.109	-0.421	-0.590	-0.340
Primário	0.806	-0.262	0.321	-0.267
Encontrou emprego por conta própria	0.314	0.107	0.609	-0.565
Encontrou emprego por relações pessoais	-0.472	0.475	-0.518	0.154
Idade entre os 15 e os 24 anos	0.678	0.089	-0.555	-0.031
Idade entre os 45 e os 54 anos	-0.765	0.108	0.339	-0.096
Idade entre os 55 e os 64 anos	-0.273	0.049	0.770	0.119
Trabalha entre 35 e 39 horas por semana	-0.203	-0.719	-0.206	-0.478
Trabalha mais de 40 horas por semana	-0.060	0.670	-0.040	0.461
Trabalha menos de 25 horas por semana	0.461	0.194	0.365	-0.095
Nenhum nível de instrução	0.643	-0.338	0.147	0.455
Ensino Obrigatório	0.645	-0.075	-0.285	-0.095
Ensino Superior	-0.762	0.237	0.094	-0.347
Administrativos	-0.659	-0.197	-0.303	-0.056
Agricultores	0.843	-0.205	0.258	-0.229
Especialistas	-0.693	0.131	0.211	-0.216
Máquinas	0.269	0.733	0.343	-0.291
Trabalhadores não Qualificados	-0.072	-0.804	-0.004	0.434
Operários	0.492	0.495	-0.585	0.134
Quadros Superiores	-0.826	0.044	0.085	0.097
Técnicos	-0.524	0.222	-0.378	-0.461
Serviços	-0.572	-0.250	0.574	0.245
Sexo Feminino	-0.134	0.640	0.413	0.431
Trabalhador por Conta Própria	0.306	0.077	0.620	-0.571
Tempo Parcial	0.537	0.247	0.406	0.035

Figura 2.9 – Planos Factoriais (CP1-CP4)³



No primeiro factor, que podemos designar como um **indicador de especialização no sector primário** ou o inverso de um indicador de terciarização, dado que apresenta fortes correlações positivas com as variáveis *primário* e *agricultores* e fortes

correlações negativas com as variáveis *comércio* e *actividades financeiras*. Apresenta ainda significativas correlações com características típicas das zonas agrícolas, como o baixo nível de instrução e um peso insignificante de *quadros superiores*. Por

³ associado a cada factor está, entre parênteses, a sua percentagem de cada variância explicada

último, de referir ainda, a correlação positiva com o trabalho a *tempo parcial* e com o escalão etário mais jovem (*15-24 anos*).

Da análise do plano factorial do primeiro factor é possível observar que se opõem as Regiões Autónomas, o Norte e o Centro às regiões de Lisboa e Vale do Tejo, Algarve e em menor escala ao Alentejo.

Também no segundo factor se apresenta como relevante a especialização sectorial, sendo que, neste caso, estamos perante um **indicador de industrialização** (correlações positivas com o sector da *indústria* e com a profissão de operador de *máquinas*), em oposição principalmente ao emprego público (forte correlação negativa com o emprego na *administração pública*). A duração semanal do horário de trabalho surge como outra importante característica deste factor, estando o horário mais longo (*+ de 40 horas*) associado ao emprego na *indústria* e o horário entre *35 e 39 horas* ao emprego na *administração pública*. Note-se que este emprego industrial está significativamente concentrado no sexo *feminino* e, segundo o plano factorial, localizado nas regiões Norte e Centro. Em oposição, as regiões dos Açores e do Alentejo apresentam um forte peso de emprego na *administração pública*.

No terceiro factor emerge a importância da **forma como as pessoas encontraram o emprego e do tipo de emprego** que têm (*conta própria ou conta de outrem*) como factor de caracterização. Assim, este factor é essencialmente um indicador do peso do emprego por *conta própria*, que normalmente está associado ao modo como foi encontrado o emprego (*por conta própria*) o que explica as fortes correlações com as variáveis

representativas da forma como encontra emprego. Outra característica importante deste factor é a forte presença dos escalões etários mais idosos, o que é consentâneo com a evolução mais natural de um trabalhador ao longo da sua vida profissional, ou seja, começa-se em jovem como trabalhador por *conta de outrem*, encontrando-se muitas vezes o emprego *por relações pessoais* e com o acumular de experiência ao longo da vida profissional, tende-se, em escalões etários mais avançados, a encontrar os empregos *por conta própria* ou mesmo a passar a trabalhar por *conta própria*. Existe ainda uma associação entre o tipo de emprego e a profissão que se exerce, sendo que os *operários* são trabalhadores por *conta de outrem*, enquanto que nos *serviços* a probabilidade de se ser trabalhador por *conta própria* se apresenta incomparavelmente mais elevada.

No plano factorial do terceiro factor é visível a oposição entre o Centro, como a região com maior peso do emprego por *conta própria* e as regiões de Açores, Norte e Lisboa e Vale do Tejo, como aquelas onde as relações pessoais apresentam uma importância decisiva na forma como se encontra emprego.

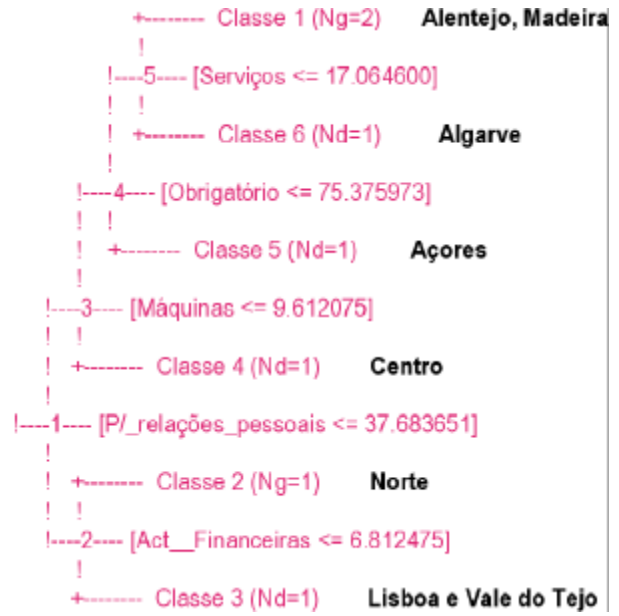
No quarto factor, a questão mais representativa é a **especialização sectorial na construção e no alojamento e restauração**, por sua vez associada a fraco peso de emprego por conta própria, bem como do modo de encontrar emprego pelos próprios meios. O facto de a Madeira apresentar o maior valor (destacado) no plano factorial poderá ser explicado pela especialização desta região no *turismo* e pela ampliação do seu aeroporto e das suas vias rodoviárias que terá implicado uma forte expansão da actividade da *construção*.

2.3. Similaridades/dissimilaridades regionais de emprego

Recorrendo ao método de *Divisive Clustering* é possível construir grupos de regiões de acordo com as suas semelhanças, identificando-se inclusivamente qual a variável que explica a separação de grupos. Este método permite uma separação iterativa do conjunto dos objectos simbólicos no número de *clusters* definido pelo utilizador, até ao limite em que se obtém um número de *clusters* igual a $(k-1)$, sendo k o nº de objectos simbólicos.

Assim, numa primeira iteração assiste-se a uma separação entre as regiões Norte e Lisboa e Vale do Tejo e as restantes, com base nos valores assumidos na variável encontrar emprego através de *relações pessoais*, sendo que esta forma é mais importante nestas duas grandes regiões, confirmando as conclusões que se retiraram da análise do terceiro factor, na secção anterior. Posteriormente apresenta-se a divisão do *cluster* formado pelo Norte e Lisboa e Vale do Tejo em dois novos *clusters*, com base no peso do emprego das *actividades financeiras*, muito superior na região de Lisboa e Vale do Tejo. Esta partição na terceira iteração significa que o *cluster* formado pelo Norte e Lisboa e Vale do Tejo é mais heterogéneo do que o formado pelas restantes cinco regiões. Numa terceira fase, a região Centro separa-se do *cluster* onde estava inserida, pelo enorme peso da profissão de operador de *máquinas* que caracteriza esta região.

Figura 2.10– Árvore de Classificação⁴



⁴ O nº inscrito em cada nó indica a ordem da divisão, sendo que, de forma genérica, a classe (Ng) equivale à verificação da condição de desigualdade para a variável que define a partição, enquanto a classe (Nd) corresponde à negação dessa condição.

Figura 2.1.1.- *Clusters* Formados nas Diversas Iterações

1ª iteração – Partição em 2 <i>clusters</i>					
Cluster 1					Cluster 2
"AÇO" "ALE" "ALG" "CEN" "MAD"					"LVT" "NOR"
2ª iteração – Partição em 3 <i>clusters</i>					
Cluster 1	Cluster 2		Cluster 3		
"AÇO" "ALE" "ALG" "CEN" "MAD"	"NOR"		"LVT"		
3ª iteração – Partição em 4 <i>clusters</i>					
Cluster 1	Cluster 2	Cluster 3		Cluster 4	
"AÇO" "ALE" "ALG" "MAD"	"NOR"	"LVT"		"CEN"	
4ª iteração – Partição em 5 <i>clusters</i>					
Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	
"ALE" "ALG" "MAD"	"NOR"	"LVT"	"CEN"	"AÇO"	
5ª iteração – Partição em 6 <i>clusters</i>					
Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
"ALE" "MAD"	"NOR"	"LVT"	"CEN"	"AÇO"	"ALG"

3. Outras especificidades de emprego: sexo e escalão etário

Uma vez conhecidas as principais especificidades regionais de emprego ir-se-á neste ponto aprofundar a análise, introduzindo na construção dos objectos simbólicos a segmentação por sexo e, posteriormente, por escalões etários. Assim para a secção 3.1. foram construídos 14 objectos simbólicos (7 regiões x 2 sexos), enquanto para a secção 3.2. foram construídos 35 objectos simbólicos (7 regiões x 5 escalões etários).

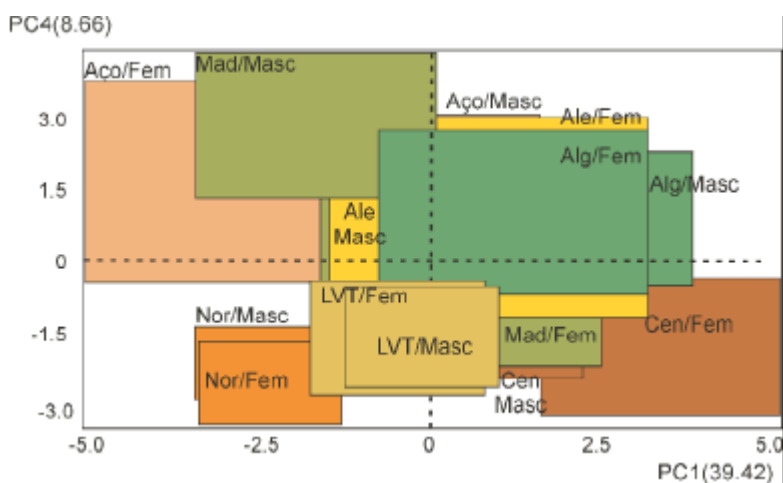
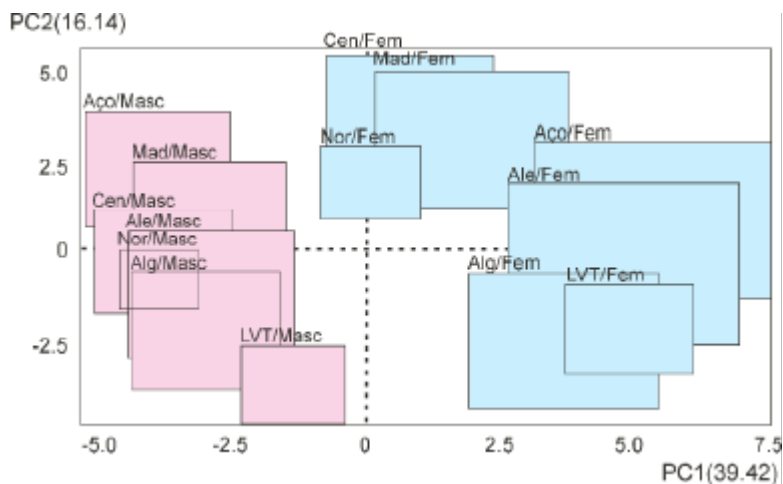
3.1. A segmentação região-sexo

Para analisar qual a contribuição da segmentação dos empregados por sexo para a compreensão das especificidades, optou-se por aplicar o método das componentes principais de forma análoga ao apresentado no ponto anterior, com a diferença de que neste caso estamos a trabalhar com 14 objectos simbólicos, que resultam do cruzamento das regiões com o sexo. A análise dos quatro factores equivale a uma retenção de 75% da variância total da informação original, composta agora por 31 variáveis (não se incorporou a variável sexo feminino, por esta ter estado na origem da construção dos objectos simbólicos).

Figura 3.1– Matriz de Correlações Variáveis Iniciais/Componentes Principais (CP1 - CP4)

	CP1	CP2	CP3	CP4
Actividades Financeiras	0.197	-0.681	0.025	-0.348
Administração Pública	-0.050	-0.166	0.025	0.738
Alojamento e Restauração	0.463	-0.286	0.398	0.258
Comércio	0.023	-0.707	0.348	0.193
Construção	-0.864	-0.253	-0.072	0.297
Educação	0.874	0.266	-0.019	0.015
Indústria	-0.260	0.210	-0.378	-0.787
Outros serviços colectivos	0.886	0.162	-0.251	0.044
Primário	-0.424	0.690	0.424	-0.008
Encontrou emprego por conta própria	-0.615	0.227	0.496	-0.157
Encontrou emprego por relações pessoais	0.126	-0.317	-0.243	-0.561
Idade entre os 15 e os 24 anos	-0.306	0.354	-0.664	0.097
Idade entre os 45 e os 54 anos	0.137	-0.518	0.521	-0.207
Idade entre os 55 e os 64 anos	-0.293	-0.131	0.713	-0.133
Trabalha entre 35 e 39 horas por semana	0.784	0.028	-0.111	0.318
Trabalha mais de 40 horas por semana	-0.895	-0.318	-0.044	-0.018
Trabalha menos de 25 horas por semana	0.745	0.497	0.127	-0.284
Nenhum nível de instrução	-0.103	0.691	0.236	0.180
Ensino Obrigatório	-0.806	-0.023	-0.281	0.225
Ensino Superior	0.800	-0.241	0.101	-0.259
Administrativos	0.878	-0.105	-0.068	0.027
Agricultores	-0.417	0.748	0.349	-0.006
Especialistas	0.759	-0.180	0.210	-0.239
Máquinas	-0.849	-0.108	0.053	-0.271
Trabalhadores não Qualificados	0.815	0.102	-0.127	0.418
Operários	-0.870	-0.094	-0.364	-0.118
Quadros Superiores	-0.178	-0.832	0.213	0.035
Serviços	0.827	0.036	0.349	0.047
Técnicos	0.550	-0.388	-0.295	-0.178
Trabalhador por Conta Própria	-0.609	0.224	0.510	-0.138
Tempo Parcial	0.712	0.525	0.143	-0.297

Figura 3.2- Planos Factoriais (CP1 - CP2 e CP3 - CP4)



Desta análise de componentes principais, a principal ideia a retirar é a importância do sexo na especificidade das características do emprego, na medida em que o primeiro factor, que neste caso apresenta um elevado poder explicativo (retém 40% da variância) permite identificar as especificidades do emprego feminino (ver Figu-

ra 3.1 e Figura 3.2 – factor 1), nomeadamente: forte peso nos sectores da *educação* e *outros serviços* colectivos e com um peso insignificante na *construção*; trabalhando menos de 40 horas, localizando-se essencialmente no escalão entre as 35 e as 39 horas; com níveis de instrução mais elevados (correlações positivas com o ensino *superior*

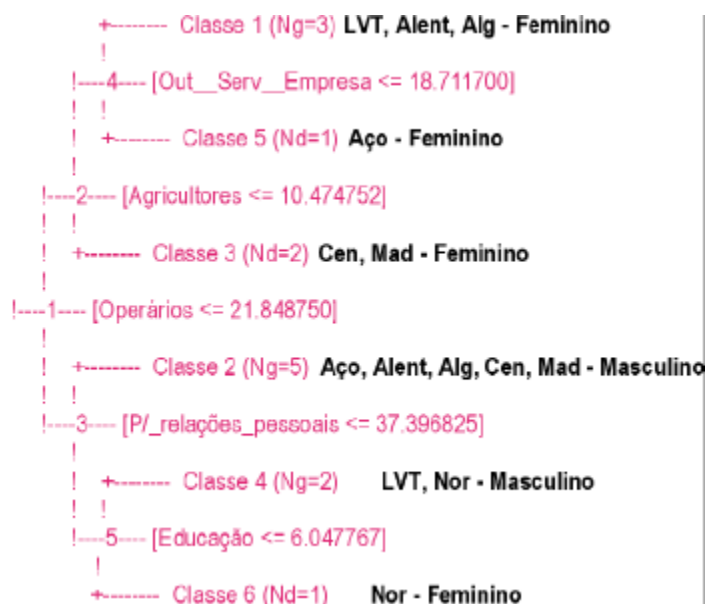
e negativas com *nenhum* grau de ensino); predomínio de *não qualificados* e *serviços* em detrimento de operadores de máquinas e operários.

Os restantes três factores transmitem ideias já presentes na análise de componentes principais elaborada no ponto anterior, havendo inclusivamente uma certa correspondência entre ambas. Assim, a fundamentação do segundo factor, da importância do primário, corresponde à que estava presente no primeiro factor da análise de componentes principais do ponto 2.2.; o indicador de envelhecimento dos empregados resultante deste terceiro factor, também já estava presente de alguma forma no anterior terceiro factor; por último, a oposição entre o peso do emprego da administração pública e o emprego industrial presente neste quarto factor, era também uma conclusão que se retirava do anterior segundo factor.

De realçar que, à excepção do primeiro factor, nos restantes três as oposições que se obtêm são quase sempre em termos regionais (visível nos planos factoriais), estando os dois sexos para uma mesma região posicionados sempre próximos nos planos factoriais. O único caso onde esta aproximação não é tão visível é no segundo factor que, como se analisou anteriormente, constitui um indicador do peso do sector primário, onde nas regiões do Norte, Centro e Madeira se denota um certo distanciamento no eixo factorial entre mulheres e homens, indicando uma maior expressão do emprego feminino no sector primário nestas regiões.

Recorrendo também neste caso à análise dos resultados do método de *Divisive Clustering* é possível confirmar a ideia anterior da importância do sexo nas especificidades do emprego, dado que a primeira divisão em dois *clusters* opõe os objectos simbólicos constituídos pelo sexo feminino de todas as regiões, com excepção da região Norte, aos restantes objectos simbólicos, com base no peso dos *operários*, muito superior no caso dos homens e das mulheres da região Norte. Da segunda partição resulta a formação de um *cluster* que engloba o sexo feminino das regiões Centro e Madeira associada ao elevado peso que o emprego feminino apresenta no grupo profissional dos *agricultores*. Destaque ainda para a terceira partição onde se forma um *cluster* composto pela região Norte (masculino e feminino) e pela região de Lisboa e Vale do Tejo (masculino) com base na maior importância das *relações pessoais* na forma como se encontram o emprego nestas regiões-sexo.

Figura 3.3 - Árvore de Classificação (Região - Sexo)



3.2. A segmentação região-escalão etário

Neste caso, devido a dificuldades inerentes à análise dos planos factoriais para os 35 objectos simbólicos e, conseqüentemente, à interpretação dos resultados da análise de componentes principais, optou-se por recorrer apenas ao método *Divisive Clustering* como forma de perceber as especificidades de emprego que derivam das diferenças de idade dos empregados.

Assim, na análise da árvore de classificação (Figura 3.4) é notória a clara distinção entre o último escalão etário (55-64 anos) e os restantes, na medida em que a primeira partição isola num só

grupo este escalão etário para todas as regiões e ainda o escalão imediatamente anterior (45-54 anos) para a região Centro. Esta primeira partição é feita com base no peso do sector primário, muito superior no escalão etário mais elevado.

Numa segunda partição, procede-se à separação do *cluster* que agrupava todos os escalões menos idosos em todas as regiões, isolando um novo grupo constituído, regra geral, pelos escalões entre os 15 e 24 anos e 25 e 34 anos⁵, que detêm um menor peso do emprego por conta própria.

⁵ Excepções constituídas pela região dos Açores e pelo escalão 25-34 anos na região Centro.

Figura 3.4 - Árvore de Classificação (Região - Escalão Etário)



4. Conclusão

Constituindo a identificação das especificidades regionais, ao nível de NUTS II de Portugal, o principal objectivo deste estudo e após as diversas caracterizações que foram sendo feitas ao longo do estudo com recurso a diversos métodos de análise estatística para objectos simbólicos (metodologia pioneira e inovadora), apresenta-se aqui um quadro síntese dessas principais especificidades⁶:

De forma complementar procedeu-se à análise da importância de outras segmentações (sexo e escalão etário) na compreensão das características do emprego, tendo-se verificado que ainda existe uma distinção significativa entre o emprego feminino e o masculino. O feminino caracteriza-se essencialmente por ser mais concentrado nas actividades *educação* e *outros serviços colectivos*, associado a valores mais elevados no pessoal dos *serviços* e *não qualificados*, deter horários semanais de trabalho mais reduzidos (predominância do escalão *35 a 39 horas*) e apresentar níveis de

⁶ Alerta-se no entanto para a necessária relativização das conclusões presentes neste quadro, uma vez que elas derivam directamente dos métodos de análise utilizados.

Figura 4.1 - Quadro Resumo das Principais Especificidades Regionais do Emprego

Regiões Variáveis	Norte	Centro	LVT	Alentejo	Algarve	Açores	Madeira
Actividades	Primário; Indústria	Primário; Indústria	Actividades Financeiras; Comércio	Adm. Pública; Educação	Aloj. e Restauração; Comércio	Primário; Adm. Pública; Out. Serviços colectivos; Educação	Primário; Construção; Aloj. e restauração
Profissão	Operários; Agricultores; Operadores de máquinas	Agricultores; Operadores de máquinas	Quadros superiores; Especialistas; Técnicos; Administrativos;	Não qualificados	Quadros superiores; Pessoal dos serviços	Agricultores Não qualificados	Agricultores Não qualificados
Idade	15 – 24 anos	15 - 24 anos; 55 – 64 anos	45 – 54 anos	45 – 54 anos; 55 – 64 anos	45 – 54 anos	15 – 24 anos	15 – 24 anos; 25 – 34 anos
Sexo	Feminino	Feminino				Masculino	Feminino
Nível de Instrução	Nenhum; Até ao obrigatório	Nenhum; Até ao obrigatório	Secundário; Superior		Superior	Nenhum; Até ao obrigatório	Nenhum; Até ao obrigatório
Situação na profissão		Conta Própria	Conta de Outrem	Conta de Outrem		Conta Própria	Conta de Outrem
Como encontrou emprego	Através de relações pessoais	Por conta própria	Através de relações pessoais			Por conta própria	
Horário	+ de 40 horas	- de 24 horas		35 a 39 horas	+ de 40 horas	35 a 39 horas	25 – 34 horas
Tempo completo ou parcial	Parcial	Parcial		Completo	Completo	Parcial	Parcial

instrução mais elevados. Na segmentação por escalões etários, o elevado peso do primário no escalão dos *55 aos 64 anos* permite uma clara distinção entre este escalão e os restantes.

Pretende-se no futuro um aprofundamento da utilização destes métodos de análise no estudo das características do(s) mercado(s) de trabalho regionais, que poderá passar, por exemplo, pela extensão da análise aos desempregados e também pela análise do Mercado de Trabalho da região de Lisboa e

Vale do Tejo, recorrendo ao Inquérito por NUTS III existente para essa região para o período 1997-1999.

Os resultados obtidos encorajam à utilização da análise de dados simbólicos a outras áreas. Esta metodologia pode também ser aplicada no controlo da qualidade do processo de Produção Estatística, associado ao tratamento dos dados dos inquéritos, no que respeita à análise dos erros de não amostragem, nomeadamente, nas fases de recolha, registo e codificação.

Referências Bibliográficas

- BOCK, H. and DIDAY, E. (ed) (2000); Analysis of Symbolic Data: exploratory methods for extracting statistical information from complex data; Spriger-Verlag; Berlin.
- INE (1993), Inquérito ao Emprego, Anos 90, Metodologia, Série Estudos N° 66.
- INE (1998), Boletim Trimestral do Inquérito ao Emprego.