

MANUEL FERNANDO PEREIRA ROCHA

**ANÁLISE DO COMPORTAMENTO DAS
ACTIVIDADES ECONÓMICAS POR GRUPO
SEGUNDO ÍNDICES DE VOLUME DE NEGÓCIOS,
EMPREGO, REMUNERAÇÕES E HORAS
TRABALHADAS NOS SERVIÇOS ENTRE
2009 e 2010**

Orientadora: Professora Doutora Ana Lorga da Silva

Co-Orientador: Professor Doutor António Costa

Universidade Lusófona de Humanidades e Tecnologias

Faculdade de Economia e Gestão

Lisboa

2012

MANUEL FERNANDO PEREIRA ROCHA

**ANÁLISE DO COMPORTAMENTO DAS
ACTIVIDADES ECONÓMICAS POR GRUPO
SEGUNDO ÍNDICES DE VOLUME DE NEGÓCIOS,
EMPREGO, REMUNERAÇÕES E HORAS
TRABALHADAS NOS SERVIÇOS ENTRE
2009 e 2010**

Dissertação apresentada para a obtenção do grau de Mestre em Gestão de Empresas, no Curso de Mestrado em Gestão de Empresas, conferido pela Universidade Lusófona de Humanidades e Tecnologias - Faculdade de Economia e Gestão

Orientadora: Professora Doutora Ana Lorga da Silva

Co-Orientador: Professor Doutor António Costa

**Universidade Lusófona de Humanidades e Tecnologias
Faculdade de Economia e Gestão**

Lisboa

2012

Agradecimentos

À Professora Doutora Ana Lorga, minha orientadora, um agradecimento muito especial por todo o tempo despendido, pelas sugestões, pelas muitas horas de trabalho disponibilizadas, pelo companheirismo e pela sua amizade, um muito obrigado.

Ao Professor Doutor António Costa, meu co-orientador, um muito obrigado pelas sugestões enriquecedoras.

À ULHT, em particular direcção de Gestão, quero agradecer pelas condições que me facultaram para que este trabalho pudesse ser realizado.

Ao Professor Doutor Ahmed Albatineh, professor de Bioestatística da Universidade Internacional de Florida. Pela disponibilidade enriquecedora através de troca de ideias via e-mail e pelos «papers» que disponibilizou, bem como ao Professor Doutor Douglas Steinley da Universidade de Missouri, Columbia.

Ao Professor Doutor Gilbert Saporta pela leitura e apreciação desta tese.

Ao Doutor Pedro Varela, Licenciado em Línguas e Literaturas Modernas - Variante de Estudos Portugueses da FCSH-UNL. Amigo de longa data. Um obrigado pela leitura e pelos conselhos gramaticais.

À minha família, particularmente aos meus pais, agradeço todo o apoio e incentivo que tornaram possível ultrapassar mais uma etapa da minha vida.

Por fim um agradecimento à minha companheira Cátia, que compreendeu toda a falta de tempo e apesar da reduzida disponibilidade sempre me apoiou.

Resumo

Os indicadores económicos, evoluem de acordo com as actividades que se desenvolvem no país. Assim progridem ou regridem em função do desempenho dos sectores de actividade económica. Estes indicadores agrupam-se em secções diversas. Deste modo interessa, verificar de que forma essas actividades, que se agrupam em situações distintas, se assemelham ou não entre si.

A classificação de variáveis e objectos semelhantes em grupos é uma importante actividade humana. Em todas as fases da vida, a classificação faz parte do processo de aprendizagem, principalmente, quando esta tem um peso enorme tal como quando uma criança aprende a distinguir o preto do branco, o pai da mãe. A classificação torna-se num processo do quotidiano onde no subconsciente estão a ser aplicadas ordenações consoante as situações com que nos deparamos constantemente.

Em temas actuais, como a economia, gestão e marketing a classificação surge para a investigação de relações entre conjuntos de objectos ou variáveis. Isto poderá resultar, ou não, num resumo de dados sob a forma de grupo, de objectos ou variáveis semelhantes entre si (dentro do mesmo grupo) e dissemelhantes em relação a grupos distintos.

Este trabalho tem como objectivo verificar de que forma as actividades económicas portuguesas são agrupadas entre si, mediante o estado da evolução económica nacional entre os anos de 2009 e 2010.

Nesta análise são trabalhadas vinte e sete actividades económicas, agrupadas em sete secções ou ramos diferentes de forma a conseguir verificar como se comportam áreas que contêm produtos ou serviços totalmente distintos.

Neste trabalho é utilizada a análise classificatória. Através deste método é estudado o nível de semelhança das actividades sob a representação de grupos. Nas diferentes áreas de actividades são estudados quatro índices: Volume de Negócios; Emprego; Remunerações e Horas trabalhadas.

Palavra Chave: Actividades Económicas, Gestão, Análise Classificatória, Coeficiente de correlação de Spearman e Índice de Rand

Abstract

The economic indicators evolve according to economic activities which develop in the country. Therefore progress or regress depends on the performance of activities of economic sector. It may be of interest to verify how these activities are grouped in different situations, by their level of similarity.

Classification of similar variables and groups is an important human activity. In all phases of life classification is part of the learning process, especially when learning carries an enormous weight, such as when a child learns to distinguish «black from white», or mother from father, making it an everyday process in the subconscious where they are applied; classification depends on the situations faced constantly.

In current topics such as economics, management and marketing, research emerges to investigate the relation between objects or variables that may or may not result in a summary of data through the form of groups of similar objects or variables.

This work is intended to verify how the Portuguese economic activities are grouped among themselves through economic evolution in 2009 and 2010.

This analysis uses twenty seven economic activities in seven distinct sections or branches, in order to be able to check whether there are areas that contain completely different products of services.

In this work the classification cluster analysis is used. Through this method, studies will be carried out to determine how similar activities are, under the representation of groups. Four indices are studied in different areas of activity Business: Volume; Employment; Salary and number of hours worked.

Keywords: Economic Activities, Management, Cluster Analysis, Rand Index, Spearman's correlation Coefficient.

Siglas e Abreviaturas

ARIMA – «Auto Regressive Integrated Moving Average»

CAE – Classificação das Actividades Económicas

CAE-Rev.3 – Classificação Portuguesa das Actividades Económicas

CEE – Comunidade Económica Europeia

ACHA – Análise Classificatória Hierárquica Ascendente

CITA – Classificação Internacional Tipo, por Indústria, de todos os Ramos de Actividade Económica

CNE – Centro Nacional de Estatística

CNBS – Classificação Nacional de Bens e Serviços

EUA – Estados Unidos da América

GPEARI – Gabinete de Planeamento, Estratégia, Avaliação e Relações Internacionais

INE – Instituto Nacional de Estatística

INEI – Índice de Novas Encomendas na Industria

IPC - Índice de Preços no Consumidor

IPCOP - Indicadores de Curto Prazo na área da Construção e Obras Públicas

ISM – «Insitute Supply Management»

IVNES – Indicadores de Curto Prazo na área dos Serviços

IVNEI – Indicadores de Curto Prazo na área da Industria

IVNECR – Indicadores de Curto Prazo na área de Comércio e Retalho

NACE – Nomenclatura Geral das Actividades Económicas das Comunidades Europeias

ONU – Organização das Nações Unidas

PIB – Produto Interno Bruto

ERRATA

Local:	Onde se lê:	Deve ler-se:
l.4, para.6, pág Resumo	Horas trabalhadas.	Horas trabalhadas nos serviços
l.4, para.1, pág 7	«preto do branco»	preto do branco
l.3, para.2, pág 7	resumo_de	resumo de
l.4, para.2, pág 7	é o do Segmentação	é o da Segmentação
l.1, para.3, pág 7	exemplo do Segmentação	exemplo da Segmentação
l.4, para.3, pág 7	Horas trabalhadas	Horas trabalhadas nos serviços
l.3, para.6, pág 8	Industria	Indústria
l.4, para.2, pág 9	Rand_	Rand
l.4, para.3, pág 9	apêndice 9 e 10	apêndices 9 e 10
l.1, para.1, pág 10	Normas elaboração	Normas de elaboração
l.1, para.2, pág 35	segmento de mercado	em segmentação de mercado
Eq (4.7), pág 44	$\sum_{r \in C_2} d_{rs}$	$\sum_{s \in C_2} d_{rs}$
l.2, para.2, pág 46	{4,5}	{1,2}
l.4, para.2, pág 46	{1,3}	{3,4}
l.3, para.4, pág 49	observa-se por esta amostra	pode inferir-se que
l.1, para.2, pág 51	A seguir a citação de Steinley, 2004	No supra citado artigo as simulações foram efectuadas para 2 a 8 grupos, considerando 50, 100, 200 e 300 objectos. Neste trabalho consideraram-se 7 grupos e 27 variáveis, admitindo que os resultados obtidos por Steinley poderão ser generalizados.
l.1, para.5, pág 54	Todos os índices estudados neste trabalho excepto PE e	Todos os índices estudados neste trabalho, excepto Peirce e

	k exibem uma distribuição assimétrica	Cohen, apresentam uma distribuição assimétrica como resultado de simulações de Monte Carlo - os dados são gerados aleatoriamente, a partir de uma distribuição binomial de acordo com o apresentado no ponto 5 em Albatineh (2010).
l.1, para.3, pág 55	Arabic	Arabie
l.1, para.2, pág 60	(Singletão)	Conjunto formado por um único elemento, conjunto singular também designado por alguns autores por Singletão
l.1, para.2, pág 67	Stanley	Steinley
l.2, tabela 5.6, pág 67	Recuperação de estrutura pobre	Estrutura pouco estável
l.3, tabela 5.6, pág 67	Recuperação de estrutura moderada	Estrutura moderadamente estável
l.4, tabela 5.6, pág 67	Recuperação de estrutura pobre	Estrutura pouco estável
l.5, tabela 5.6, pág 67	Recuperação de estrutura pobre	Estrutura pouco estável
l.1, para.3, pág 67	Stenley	Steinley
l.7,para.4, pág 67	manter-se nos mesmo	se manter nos mesmo

Índice

1	Introdução	7
1.1	Objectivos	8
1.2	Metodologia	8
1.3	Software Matlab.....	9
1.4	Estrutura do Trabalho	10
2	Conjuntura Nacional e Internacional	12
2.1	Introdução	12
2.2	Economia Internacional	12
2.3	Conjuntura Nacional	17
3	Apresentação e Condições de Recolha dos Dados	23
3.1	Introdução	23
3.2	Sazonalidade dos dados	23
3.3	Índices de Volume de Negócios, Emprego, Remunerações e Horas Trabalhadas nos Serviços	24
3.3.1	Introdução	24
3.3.2	Objectivos.....	24
3.3.3	Descrição	25
3.4	Secções e Actividades Económicas	26
3.4.1	Introdução	26
3.4.2	Descrição das Secções Económicas	26
3.4.3	Descrição das Actividades Económicas	27
3.4.4	Classificação Portuguesa das Actividades Económicas CAE.Rev.3.....	28
3.4.5	Objectivos CAE-Rev.3	29
3.4.6	Evolução História CAE-Rev.3	30
3.4.7	Como utilizar melhor esta Classificação	32

3.4.8	Sistemas de Codificação	33
4	Métodos de Classificação	34
4.1	Introdução	34
4.2	Classificação Hierárquica	36
4.2.1	Classificação Não Hierárquica	37
4.2.2	Classificação Hierárquica Ascendente	37
4.3	Medidas Proximidade e de Distância.....	39
4.3.1	Coeficiente de Semelhança «Correlation-coefficient»	39
4.4	Coeficientes de comparação entre elementos de em conjunto de dados	40
4.4.1	Semelhança.....	41
4.4.2	Dissemelhança	42
4.5	Alguns métodos de classificação hierárquica ascendente.....	43
4.5.1	Vizinho mais próximo	43
4.5.2	Média da distância entre os grupos.....	44
4.5.3	Vizinho mais afastado	44
4.5.4	Matriz Cofenética e Matriz Ultramétrica.....	45
4.6	Comparação de Estruturas Hierárquicas e de Partições.....	47
4.6.1	Coeficiente de correlação de Spearman e comparação de Estruturas Hierárquicas.....	47
4.6.2	Índice de Rand e outros Índices de Semelhança entre Partições.....	49
4.7	Metodologias de determinação do número de grupos	56
5	Estudo Empírico e Resultados	59
5.1	Introdução	59
5.2	Escolha do Método de Classificação Hierárquica a Utilizar	60
5.3	Método de Particionamento	62
5.4	Coeficiente de Correlação de Spearman aplicado à comparação das Estruturas Hierárquicas.....	63

5.5	Índice de Rand e outros Índices de Comparação Aplicados à Comparação de Partições	65
5.6	Demonstração dos Resultados	68
5.6.1	Volume Negócios	69
5.6.2	Emprego	73
5.6.3	Remunerações.....	77
5.6.4	Horas trabalhadas nos serviços.....	81
	Conclusão e Projecto Futuro.....	85
	Bibliografia.....	88
	Apêndices	91
	Anexos.....	105

Índice Figuras

Figura 2.1 Consumo Privado e Vendas a Retalho dos EUA (VH, em %)	12
Figura 2.2 PIB e Índices de Confiança na Indústria e nos Serviços dos EUA	13
Figura 2.3 PIB e Taxa de Inflação da China	14
Figura 2.4 Taxa de Juro a 3 Meses do Mercado Monetário (média do período, %)	15
Figura 2.5 Preço Médio Spot do Petróleo Brent	16
Figura 2.6 PIB e Indicadores de Síntese Económica	17
Figura 2.7 Fluxo de Comércio Internacional	18
Figura 2.8 Desemprego	19
Figura 2.9 Taxa de Variação do IPC	20
Figura 3.1 Sistema de Codificação	33
Figura 4.1 Exemplo de um Dendrograma com Cinco Variáveis	45
Figura 4.2 Gráfico de Frequências Relativas Acumuladas – Índice Horas com o Método Vizinho Mais Afastado no Ano 2009	58
Figura 5.1 Dendrograma de Horas com o Método do Vizinho Mais Próximo no Ano 2009	60
Figura 5.2 Dendrograma de Horas com o Método da Distância Entre os Grupos no Ano 2009	61
Figura 5.3 Dendrograma de Horas com o Método do Vizinho Mais Próximo no Ano 2009	61
Figura 5.4 Dendrograma de Volume Negócios com o Método do Vizinho Mais Afastado no Ano 2009	69
Figura 5.5 Dendrograma de Volume Negócios com o Método do Vizinho Mais Afastado no Ano 2010	69
Figura 5.6 Volume de Negócios – Índice Total Variações Homólogas, %	72
Figura 5.7 Dendrograma de Emprego com Método do Vizinho Mais Afastado no Ano 2009	73
Figura 5.8 Dendrograma de Emprego com Método do Vizinho Mais Afastado no Ano 2010	73
Figura 5.9 Emprego – Índice Total Variações Homólogas, %	76
Figura 5.10 Dendrograma de Remunerações com o Método do Vizinho Mais Afastado no Ano 2009	77

Figura 5.11 Dendrograma de Remunerações com o Método do Vizinho Mais Afastado no Ano 2010	77
Figura 5.12 Remunerações – Índice Total e Secções Variações Homólogas, %	80
Figura 5.13 Dendrograma de Horas com o Método do Vizinho Mais Afastado no Ano 2009	81
Figura 5.14 Dendrograma de Horas com o Método do Vizinho Mais Afastado no Ano 2010	81
Figura 5.15 Horas – Índice Total e Secções Variações Homólogas, %	84

Índice de Tabelas

Tabela 2.1 Principais Indicadores da Conjuntura Nacional	21
Tabela 2.2 Variação dos Índices de Volume de Negócios, Emprego, Remunerações, e Horas trabalhadas nos serviços entre 2009 e 2010.....	22
Tabela 3.1 Representação das Actividades Económicas	27
Tabela 4.1 Matriz Cofenética	46
Tabela 4.2 Matriz Ultramétrica	46
Tabela 4.3 Média de Classificações Arredondadas	48
Tabela 4.4 Tabela de Contingência	50
Tabela 4.5 «Quality of Cluster Recovery»	55
Tabela 5.1 Coeficiente de Correlação de Spearman – Resultados	63
Tabela 5.2 Comparação de Estruturas Hierárquicas 2009/2011 Utilizando o Método do Vizinho Mais Afastado para um Nível de Significância de 5%	64
Tabela 5.3 Comparação de Partições Através do Índice de Rand	65
Tabela 5.4 Comparação de Partições Através do Índice PE.....	66
Tabela 5.5:Comparação de Partições Através do Índice k	66
Tabela 5.6 Comparação de Partições Através do Índice de Rand Ajustado por Hubert e Arabie	67
Tabela 5.7 Grupo por Volume de Negócios	70
Tabela 5.8 Grupo por Emprego	74
Tabela 5.9 Grupo por Remunerações	78
Tabela 5.10 Grupo por Horas trabalhadas nos serviços	82

1 Introdução

A Classificação de variáveis e objectos semelhantes em grupos é uma importante actividade humana. Em todos os processos da vida a classificação faz parte do processo de aprendizagem principalmente quando esta tem um peso enorme, tal como quando uma criança aprende a distinguir o «preto do branco» ou o pai da mãe. A classificação torna-se, então, num processo do quotidiano, onde, no subconsciente, estão a ser aplicadas classificações consoante as situações com que nos deparamos constantemente.

Em temas actuais, como a economia, gestão e marketing a classificação surge para investigação de relações entre conjuntos de objectos ou variáveis. Tal poderá resultar, ou não, num resumo de dados, sob a forma de grupo, de objectos ou variáveis semelhantes. Um dos exemplos mais conhecidos é o do Segmentação de Mercado, sendo uma das mais fundamentais estratégias no conceito de mercado. Também aqui pode ser utilizada a análise classificatória para formar agrupamentos de pessoas de acordo com a sua semelhança pelo produto em questão.

Analisada a importância através do exemplo do Segmentação de Mercado, é com maior abertura que se verifica o enriquecimento deste tipo de estudo. Este trabalho trata do agrupamento de actividades económicas segundo quatro índices: Volume de Negócios, Emprego, Remunerações e Horas trabalhadas nos anos de 2009 e 2010. Deste modo interessa verificar de que forma essas actividades, que se agrupam em situações distintas, se assemelham entre si.

1.1 Objectivos

Este trabalho tem como objectivo verificar de que forma as actividades económicas portuguesas são agrupadas entre si mediante o estado da evolução económica nacional entre os anos de 2009 e 2010.

São analisados quatro índices: Volume de Negócios; Emprego; Remunerações e Horas trabalhadas nos serviços. Para cada um deles é apresentado o comportamento mensal de vinte sete actividades económicas que estão contidas em sete secções ou ramos distintos. Desta forma pretende-se verificar como se comportam áreas que contêm produtos ou serviços totalmente diferentes. Separado por índices, o resultado surgirá mediante a formação de actividades económicas em grupos que se assemelham entre si.

Como metodologia vão ser utilizados métodos de análise classificatória onde é estudado o nível de semelhança das actividades e a sua representação em grupos.

Uma pequena comparação será também tida em conta mediante o estudo de estatística descritiva apresentado mensalmente através do Boletim Mensal da Economia Portuguesa. As conjunturas nacional e internacional justificarão as variações dos resultados entre 2009 e 2010.

1.2 Metodologia

Para este trabalho é utilizada a análise classificatória, em particular a Análise Classificatória Hierárquica Ascendente (ACHA). Dos vários métodos existentes nesta metodologia, foram testados: Vizinho mais próximo, Média da distância (ligação) entre os grupos e Vizinho mais afastado, de forma a verificar com qual destes métodos se consegue obter os melhores resultados sobre as variáveis em estudo (vinte e sete actividades económicas).

A operação estatística que dá origem aos índices de Volume de Negócios, Emprego Remunerações e Horas trabalhadas nos serviços, resulta da fusão de cinco operações estatísticas relativas aos Indicadores de Curto Prazo na área dos Serviços (IVNES), Indústria (IVNEI e INEI), Comércio a Retalho (IVNECR) e Construção e Obras Públicas (IPCOP). Estes índices permitem conhecer a evolução, no curto prazo nestes sectores e respondem,

igualmente, ao Regulamento CE nº 1165/98 de 10 de Maio de 1998, alterado pelo regulamento de 1158/05 de 6 de Julho de 2005.¹

O agrupamento das actividades é mostrado sob a representação de estruturas hierárquicas e partições. A comparação das estruturas é efectuada através do coeficiente de correlação de Spearman e a concordância das partições por meio de índices tais como por exemplo o índice de Rand_entre outros.

1.3 Software Matlab

Para realização deste trabalho recorreu-se ao software Matlab² R2007a, versão 7.4.0. Este software foi escolhido devido à possibilidade de realizar as tarefas com as funções que se pretendia, tal como à sua rapidez de funcionamento. Este software é baseado em linguagem C e C++³. Todo o código apresentado no apêndice 9 e 10 foi construído pelo autor e aplicado a cada um dos quatro índices. Para além das funcionalidades incorporadas no software Matlab foi necessário construir uma função de raiz. Esta função serviu para construir as matrizes ultramétricas e cofenéticas como se pode observar no apêndice 9.

Pretende-se posteriormente enviar o algoritmo construído para determinar matrizes ultramétricas e cofenéticas (apêndice 9) para a equipa que gere o Matlab, de forma a sugerir a inserção desta função na biblioteca dos algoritmos que contemplam o Matlab. Pois, de acordo com diversas pesquisas e contactos efectuados, não foi encontrada nenhuma forma de determinação das matrizes ultramétricas por intermédio deste software.

Dendrograma

A representação gráfica, visto estar a tratar-se de classificação hierárquica ascendente, vai ser realizada através de dendrograma, que pode ser descrito como um tipo específico de diagrama que organiza determinadas variáveis ou indivíduos. Resulta de uma análise estatística de determinados dados, em que se emprega um método quantitativo que leva a agrupamentos e à sua ordenação hierárquica ascendente (neste trabalho), ou seja, em termos gráficos semelhante a ramos de uma árvore que se vão agrupando noutros sucessivamente.

¹ INE. Índices de Volume de Negócios e Emprego. código:136, versão 1.0

² MATLAB[®] is a high-level language and interactive environment that enable to perform computationally intensive tasks faster than with traditional programming languages such as C, C++, and Fortran. <http://www.mathworks.com>

³ Linguagem de programação

Normas elaboração e apresentação da tese

As normas utilizadas na elaboração desta tese são relativas ao Despacho nº 101/2009 - Actualização das normas para elaboração e apresentação de teses e dissertações na ULHT. Conforme regras apresentadas pelo Departamento de Gestão de Empresas da ULHT.

1.4 Estrutura do Trabalho

Este trabalho está organizado em cinco capítulos.

Este capítulo um é dedicado a uma pequena introdução, na qual se referem os objectivos do trabalho, a metodologia e a estrutura.

No capítulo dois são descritas as conjunturas nacional e internacional do ponto de vista económico. Nele são descritos os principais indicadores económicos que justificam o estado das economia nacional e internacional e consequente comportamento das actividades económicas para os anos de 2009 e 2010.

O capítulo três é dedicado à demonstração das condições de Recolha de Dados. São apresentados os «Índices de Volume de Negócios, Emprego, Remunerações e Horas Trabalhadas nos Serviços» que primam pela actualização uma vez que estes dados são actualizados mensalmente. É descrito de uma forma mais detalhada o procedimento efectuado no estudo estatístico para obtenção dos dados.

O capítulo quatro completa a demonstração da recolha de dados. Neste capítulo é explicada de uma forma mais completa a constituição das actividades económicas, a sua história, objectivos e a respectiva constituição por meio da Classificação Portuguesa das Actividades Económicas.

No capítulo cinco é mostrada a metodologia utilizada neste trabalho, ou seja, Classificação Hierárquica Ascendente, ou análise classificatória e os resultados empíricos. Numa fase introdutória é mostrada a sua utilidade através de exemplos do quotidiano e depois com exemplos ligados à gestão e ao marketing.

Os métodos inerentes à Classificação Hierárquica Ascendente são também aqui apresentados e estudados de forma a que seja possível concluir qual o método que vai permitir a obtenção dos melhores resultados.

Esses métodos são: Vizinho mais próximo, Média da distância entre os grupos e Vizinho mais afastado. Os resultados são então demonstrados sob forma de dendrogramas (diagrama em árvore) que permitem a demonstração dos resultados de uma forma única e de partições apresentadas sobre a forma de tabelas que, contendo mesma informação que os dendrogramas, permitem uma melhor visualização das actividades e respectivas secções.

Para comparar os resultados foram utilizados o coeficiente de correlação de Spearman para as estruturas hierárquicas e o índice de Rand, entre outros, para as partições.

Neste capítulo apresentam-se e interpretam-se também os resultados empíricos. São analisados dendrogramas, partições e representações em forma de tabelas por forma a mostrar a variação de grupo para grupo das actividades entre os anos 2009 e 2010 em cada um dos índices. Tal interpretação é comparada com o estudo estatístico descritivo elaborado pelo INE⁴ de forma a adicionar informação às conclusões a que se chegou. Os indicadores económicos apresentados inicialmente podem então ser novamente verificados e comparados com os resultados obtidos de forma a permitir tirar conclusões das estruturas obtidas e de que forma as actividades estão representadas no ano 2009 e no ano de 2010.

Por fim termina-se esta dissertação com uma síntese detalhada de todo o trabalho, o modo como se conseguiu chegar aos objectivos propostos e de que forma foi enriquecedor o seu resultado. Também são feitas algumas considerações finais e propostas de desenvolvimento de trabalho futuro. Este relatório finda com a Bibliografia, Apêndices e Anexos.

⁴ Através do Boletim Mensal da Economia Portuguesa

2 Conjuntura Nacional e Internacional

2.1 Introdução

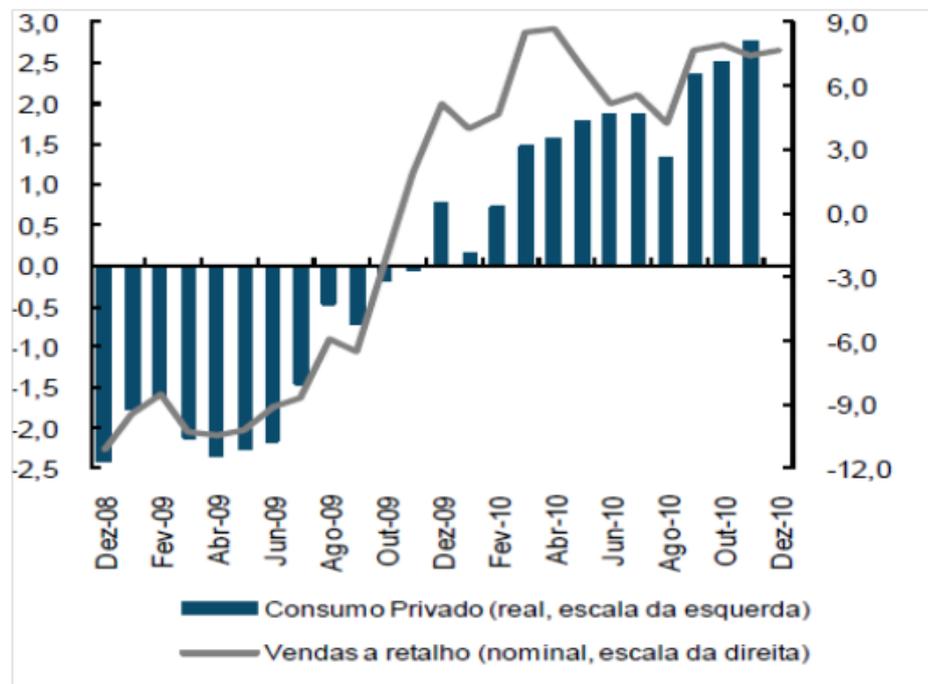
Neste capítulo serão apresentadas as variações dos indicadores económicos, tanto a nível nacional como internacional. Desta forma será possível analisar qual o sentimento económico para os anos de 2009 e de 2010, de forma a verificar-se quais as alterações entre esses dois anos. O comportamento das actividades económicas entre 2009 e 2010 será, também, justificado pelas variações destes indicadores económicos.

2.2 Economia Internacional

Consumo Privado e Vendas a Retalho

Do ponto de vista internacional (especificamente EUA⁵) ao longo do ano de 2009 existiu de uma forma generalizada uma pequena evolução favorável principalmente devido às vendas do comércio a retalho e produção industrial como se pode ver na figura 2.1. Na mesma figura pode-se verificar que, no ano seguinte (2010), o consumo privado apresentou um forte crescimento, prosseguindo a recuperação da economia.

Figura 2.1 Consumo Privado e Vendas a Retalho dos EUA (VH, em %)



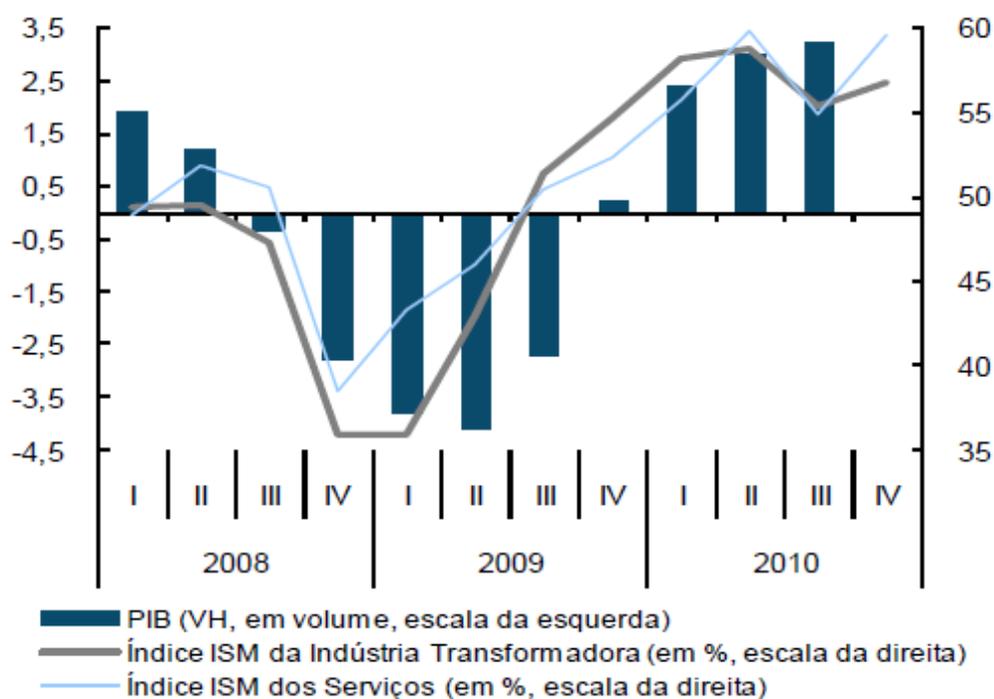
Fonte: Boletim Mensal Economia Portuguesa do GPEARI, Janeiro 2011

⁵ Estado Unidos da América

PIB e Índices de Confiança nos EUA

O principal factor que influenciou o forte crescimento das vendas e consumo privado, foi o fortalecimento da confiança dos empresários e dos consumidores nos EUA, conforme se pode verificar de seguida pela variação do PIB e pelo índice de medição de confiança ISM⁶, tanto a indústria como dos serviços.

Figura 2.2 PIB e Índices de Confiança na Indústria e nos Serviços dos EUA



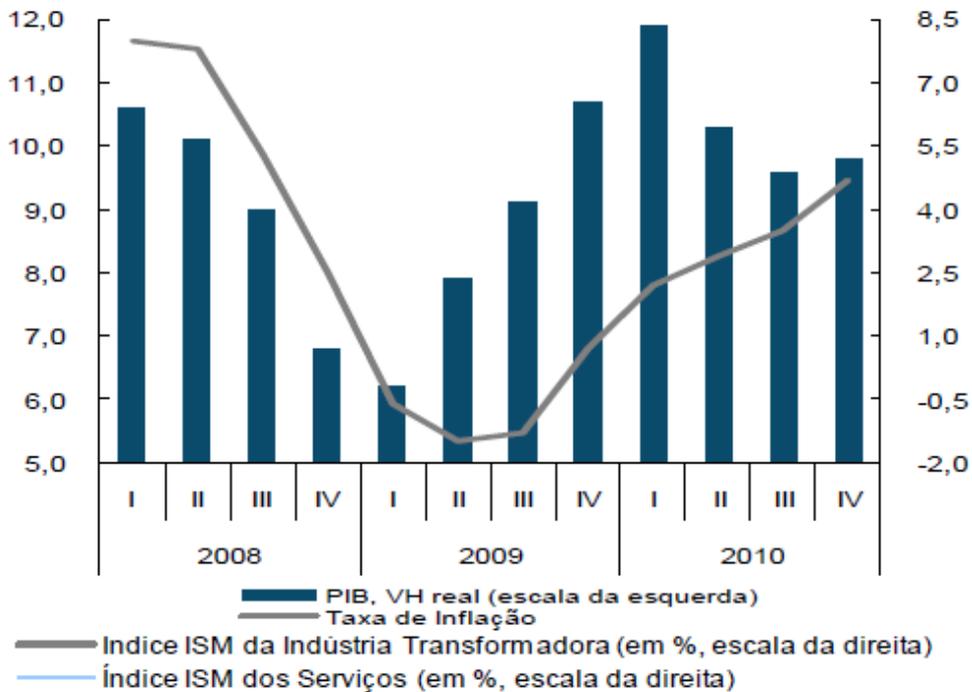
Fonte: Boletim Mensal Economia Portuguesa do GPEARI, Janeiro 2011

⁶ Índice mensal divulgado pelo Instituto de Gestão de Fornecimento, que controla a quantidade de actividade de produção que ocorreu no mês anterior

PIB e Taxa Inflação da China

Quanto a outra potência internacional, a China, verifica-se pela figura seguinte que o PIB teve um aumento significativo entre 2009 e 2010, devido sobretudo à continuação de um forte crescimento da procura interna. Ainda na figura seguinte se pode verificar que a taxa de inflação homóloga subiu. Esta aceleração na China resultou na implementação de algumas medidas restritivas no domínio monetário, nomeadamente na subida de taxas de juro de referência, de crédito e de débito, segundo o Boletim Mensal da Economia Portuguesa de Janeiro de 2011.

Figura 2.3 PIB e Taxa de Inflação da China

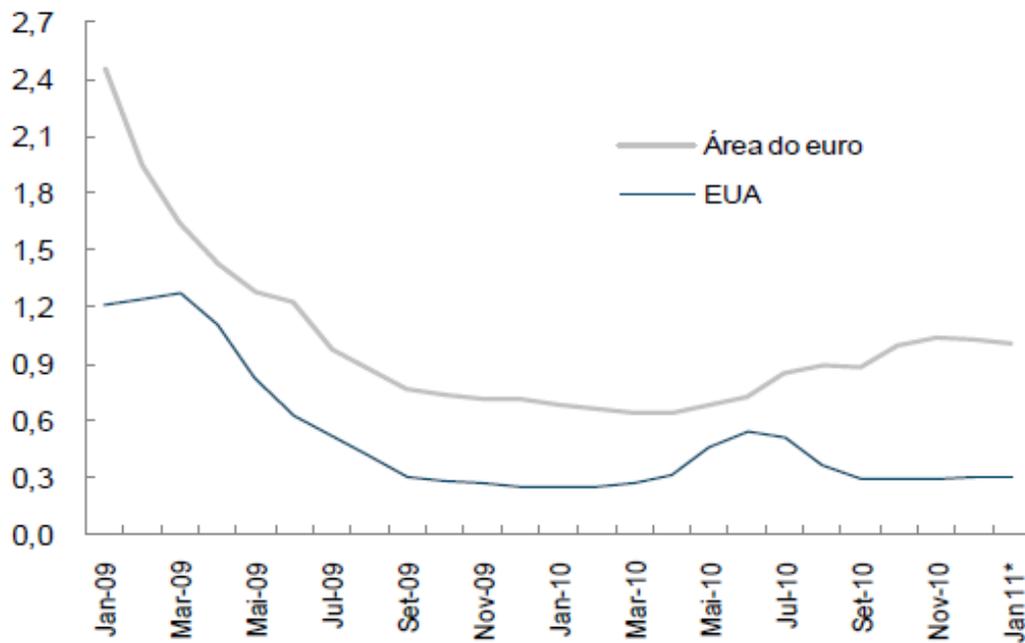


Fonte: Boletim Mensal Economia Portuguesa do GPEARI, Janeiro 2011

Taxa de Juro nos EUA

Relativamente a outro dado preponderante, isto é, a taxa de juro neste caso indicativo de 3 meses, nos EUA teve um ligeiro aumento durante 2010. Porém manteve-se abaixo do valor homólogo em 2009 que decresceu significativamente no 1º semestre. Em comparação com a zona euro tem-se uma pequena variação positiva entre Março e Julho de 2010, finalizando o ano com valores iguais aos homólogos. Verifica-se, da mesma forma, que nos EUA existe uma descida significativa no 1º semestre de 2009.

Figura 2.4 Taxa de Juro a 3 Meses do Mercado Monetário (média do período, %)

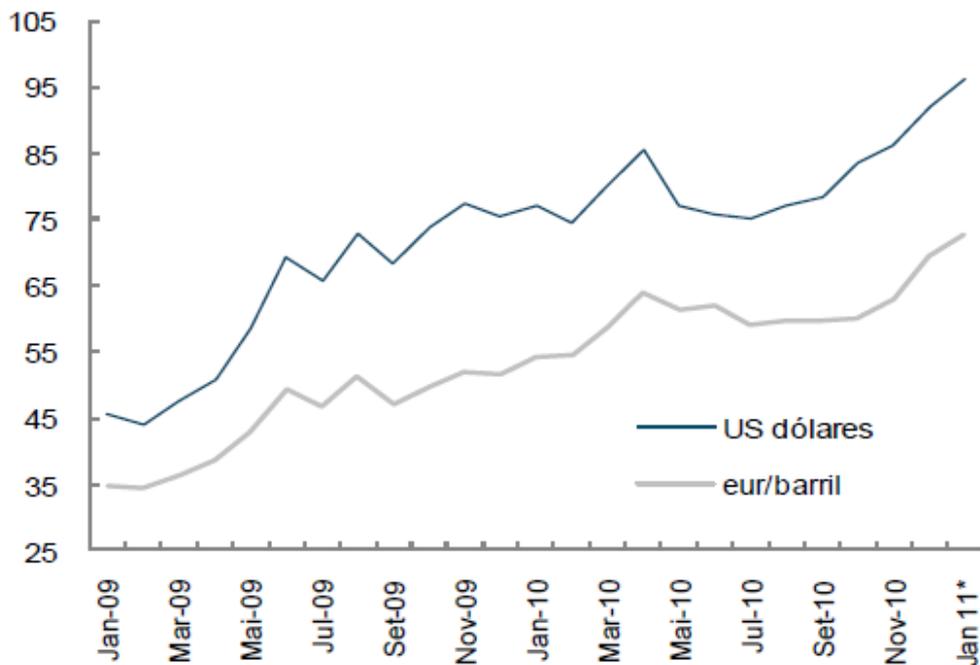


Fonte: Boletim Mensal Economia Portuguesa do GPEARI, Janeiro 2011

Preço Petróleo em dólares Americanos (US) e Euros

O preço do petróleo é um dos principais indicadores que influencia de uma forma geral todo o movimento proporcionado pelas actividades económicas. Constata-se pela figura abaixo que o crescimento do preço do petróleo aumentou significativamente de 2009 para 2010. Apesar de todas as actividades necessitarem deste produto, tem-se na apresentação dos resultados que os grupos irão variar consoante as actividades que mais necessitam deste produto. Este dado é apresentado comparativamente em dólares e euros/barril.

Figura 2.5 Preço Médio Spot do Petróleo Brent



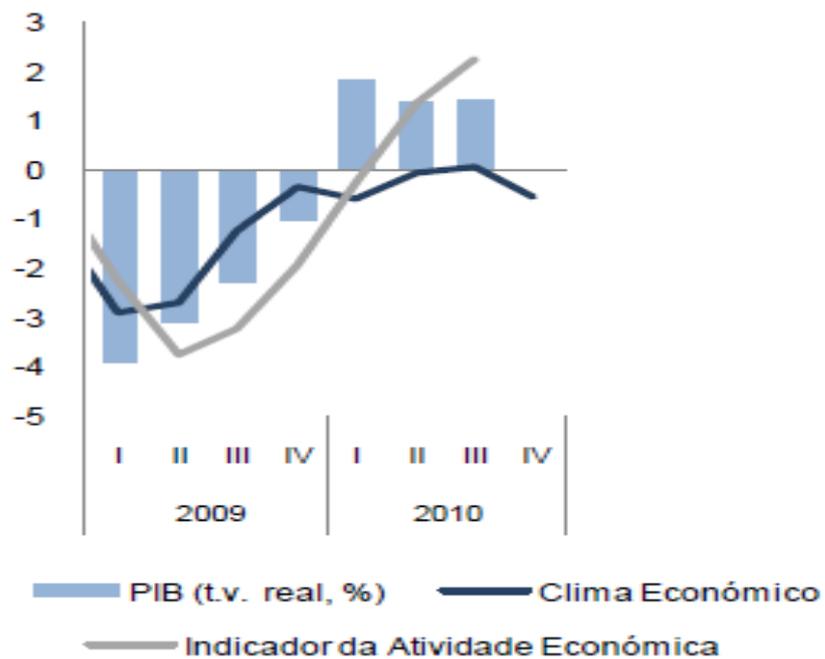
Fonte: Boletim Mensal Economia Portuguesa do GPEARI, Janeiro 2011

2.3 Conjuntura Nacional

PIB e Indicadores de Síntese Económica em Portugal

A conjuntura nacional que tanto depende da internacional pelo estado de confiança dos investidores a nível mundial reflecte-se, logo à partida, pelo decréscimo de indicador de clima económico entre 2009 e 2010. Também em 2010, o indicador de actividade económica abrandou tal como o indicador do consumo privado, comparativamente a 2009.

Figura 2.6 PIB e Indicadores de Síntese Económica

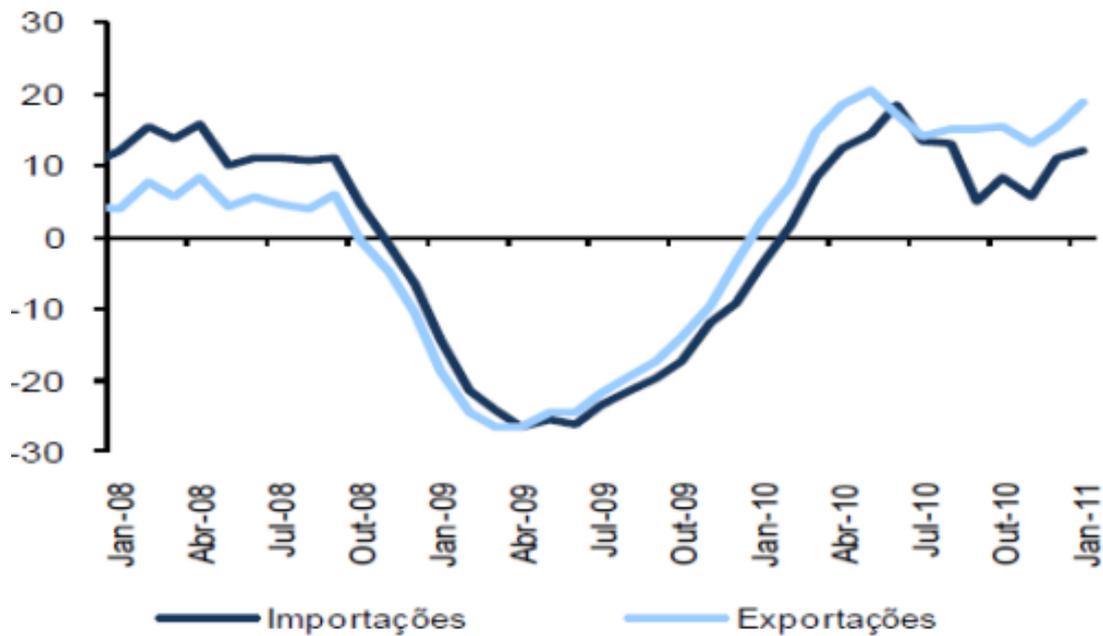


Fonte: Boletim Mensal Economia Portuguesa do GPEARI, Janeiro 2011

Fluxo de Comércio Internacional em Portugal

Relativamente à saída e entrada de mercadorias, segundo o Boletim da Economia Portuguesa, tem-se que foram registadas variações homólogas positivas em 2010. Não sendo, apesar de tudo, suficiente para evitar o agravamento de 1% do défice da balança comercial que, mesmo assim, recuperou cerca de 5% face a 2009. Igualmente, devido ao facto de, no ano de 2010, a taxa de cobertura das entradas pelas saídas de mercadorias ter sido superior em período homólogo. Onde as exportações foram para 2010 cerca de 65.1% das importações. Os produtos industriais transformados foram dos principais que levaram ao aumento das saídas em 2010 face ao período homólogo.

Figura 2.7 Fluxo de Comércio Internacional

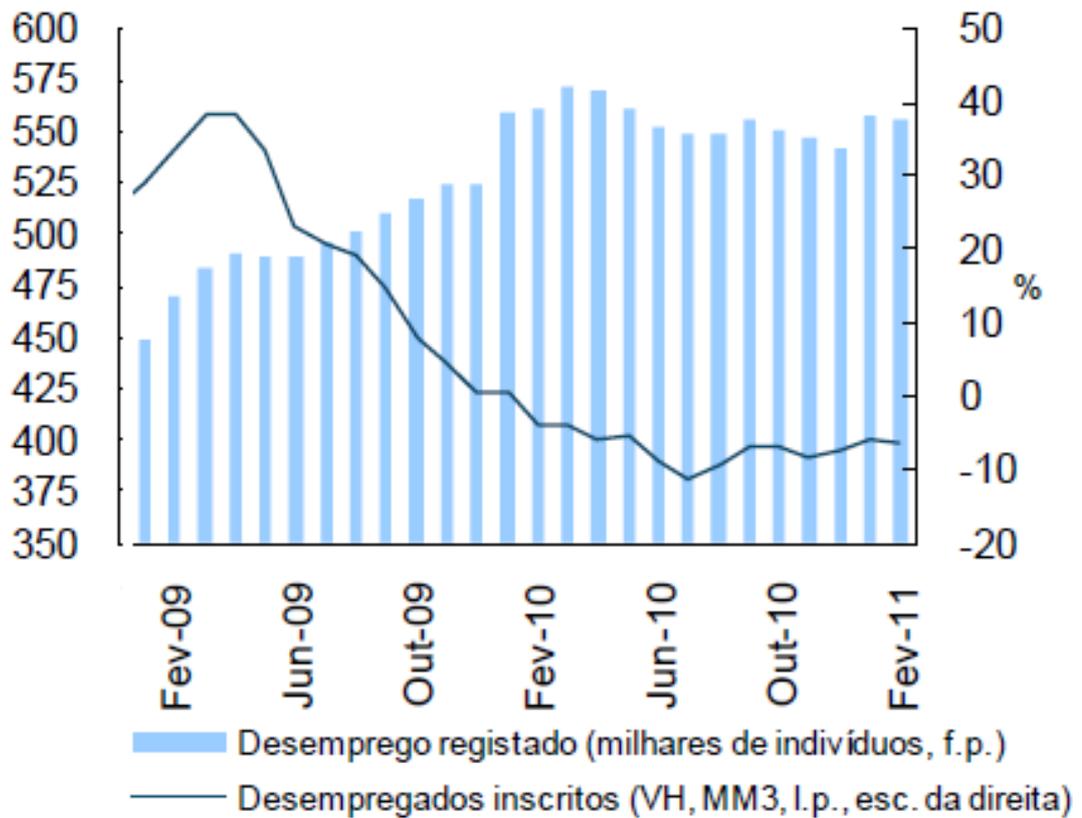


Fonte: Boletim Mensal Economia Portuguesa do GPEARI, Janeiro 2011

Desemprego

De acordo com os dados fornecidos pelo Boletim Mensal da Economia Portuguesa, em valor absoluto verifica-se uma diminuição de cerca de 6 mil desempregados registados face ao valor de Fevereiro de 2010. Também se observa uma redução de cerca de 2 mil desempregados em relação a Janeiro de 2011. O número de desempregados inscritos mensalmente apresentou em 2010 uma diminuição em comparação com 2009.

Figura 2.8 Desemprego

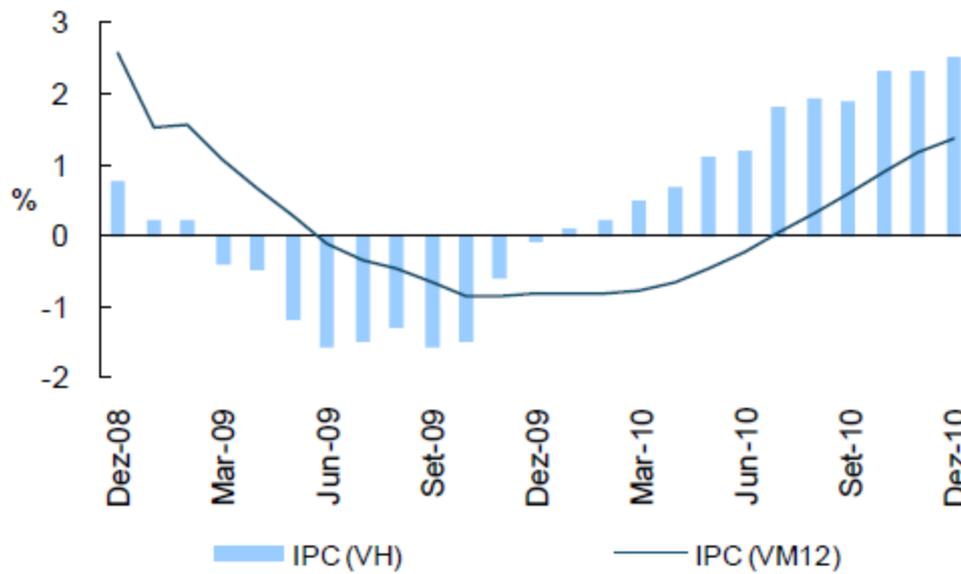


Fonte: Boletim Mensal Economia Portuguesa do GPEARI, Janeiro 2011

Preços no Consumidor

Tem-se também como factor determinante para afectação das actividades económicas os preços. Na figura seguinte pode-se verificar que no final do ano de 2010 se assistiu a uma aceleração do crescimento do Índice de Preços no Consumidor (IPC). Segundo o Boletim Mensal da Economia Portuguesa de Janeiro de 2011 este factor levou a que a taxa de inflação em 2010 se tenha situado cerca de 1.4% acima da inflação de 2009.

Figura 2.9 Taxa de Variação do IPC



Fonte: Boletim Mensal Economia Portuguesa, Janeiro 2011

Indicadores para os anos de 2009 e 2010, com maior relevância em termos económicos são mostrados de forma mais sucinta na seguinte tabela.

Tabela 2.1 Principais Indicadores da Conjuntura Nacional

	2009	2010
Receitas do Estado Cobradas (TOTAL)	34 715 900 000	36 307 000 000
Receitas do Estado – Execução (TOTAL)	48 773 000 000	50 555 600 000
Contas da Administração Central (Finanças Públicas)	Saldo = -13 252 700 000	Saldo = -11 985 100 000
Financiamento da Administração Central (Finanças Públicas)	Saldo = 98 405 100 000	Saldo = 99 890 000 000
PIB	168 586 800 000 VH(%) = -2.0	172 720 800 000 VH(%) = 2.5
Exportações	47 142 900 000 VH(%) = -15.5	53 462 300 000 VH(%) = 13.4
Importações	59 789 400 000 VH(%) = -18.2	65 828 000 000 VH(%) = 10.1
Formação Bruta de Capital	33 575 600 000 VH(%) = -15.7	32 812 000 000 VH(%) = -2.3
Agricultura, Silvicultura	3 501 000 000 VH(%) = -0.5	3 580 000 000 VH(%) = 2.2
Industria, Energia, Água e Saneamento	24 935 000 000 VH(%) = -3.7	25 732 000 000 VH(%) = 3.2
Construção	9 635 000 000 VH(%) = -11.5	9 813 000 000 VH(%) = 1.9
Serviços	110 220 300 000 VH(%) = 1.1	112 143 900 000 VH(%) = 1.7
Indicador de Confiança dos Consumidores (Máx = -4 e Mín = -54)	-39	-41
Evolução do Desemprego (Máx = 86 e Mín = 8)	64	56
Investimento Empresarial – Taxa de Crescimento Nominal Previsto (%)	-18.5	5.6
Construção-Obras Concluídas	34 254 VH(%) = -11.8	31 887 VH(%) = -6.9
Mercado de Trabalho – População (Milhares)	10 622 7 VH(%) = 0.1	10 635 8 VH(%) = 0
Taxa de Emprego (Percentagem)	47.5	46.8
Taxa Total de Desemprego (%)	9.5	10.8

Fonte: Indicadores Mensais de Conjuntura (GPEARl 2011)

Relativamente aos índices que vão ser estudados neste trabalho, é possível ter através da seguinte tabela uma perspectiva em termos homólogos entre 2009 e 2010. Para cada um dos índices são apresentados valores dos principais ramos industriais, como a indústria, o comércio a retalho, os serviços e a construção e obras públicas. Esta representação de resultados vai ajudar a justificar os agrupamentos de actividades obtidos entre o ano de 2009 e o ano de 2010.

Tabela 2.2 Variação dos Índices de Volume de Negócios, Emprego, Remunerações, e Horas trabalhadas nos serviços entre 2009 e 2010

VH (%)		Industria	Comércio a Retalho	Serviços	Construção e Obra Públicas
Índice de Emprego	2009	-5.6	-2.8	-2.9	-7.7
	2010	-2.8	-0.2	-0.8	-8.0
Índice de Horas trabalhadas nos serviços	2009	-6.7	-3.4	-3.3	-8.9
	2010	-2.9	-0.9	-0.8	-6.2
Índice de Remunerações	2009	-4.8	-1.1	-1.4	-8.3
	2010	-0.8	0.3	0	-6.0
Índice de Volume de Negócios no Comércio a Retalho	2009	-2.0			
	2010	-0.2			

Fonte: Indicadores Mensais de Conjuntura (GPEARI 2011)

3 Apresentação e Condições de Recolha dos Dados

3.1 Introdução

Os dados utilizados neste trabalho foram retirados de um estudo apresentado pelo Instituto Nacional de Estatística. Titulado como: “Índices de Volume de Negócios, Emprego, Remunerações e Horas Trabalhadas nos Serviços em destaque informação à comunicação social (1º Semestre de 2011)”. A actualização mensal torna os resultados mais enriquecidos dado que, com este nível de periodicidade, permite conhecer a evolução num curto prazo.

Estes dados são fornecidos mensalmente para todas as actividades económicas trabalhadas. A operação estatística que dá origem aos índices, resulta da fusão de cinco operações estatísticas relativas aos Indicadores de Curto Prazo na área dos Serviços (IVNES), Indústria (IVNEI e INEI), Comércio a Retalho (IVNECR) e Construção e Obras Públicas (IPCOP).

Segundo o INE o Índice de Volume de Negócios e Emprego: “Para o efeito foi realizado um inquérito amostral, de periodicidade mensal e com cobertura nacional, onde é obtida informação através da recolha directa, para cada uma das variáveis. Os resultados são divulgados, tendencialmente, 30 dias após o período de referência. No caso do Comércio a Retalho, 36 dias no caso da Indústria, 40 dias no caso dos Serviços e de Construção e Obras Públicas e de 38 dias no caso das Novas Encomendas na Indústria.”(INE, Índices de Volume de Negócios e Emprego, Versão 1.0, Setembro 2010)

3.2 Sazonalidade dos dados

Segundo o INE: “Os índices obtidos são ajustados de efeitos de calendário e sofrem o tratamento da sazonalidade com recurso a software específico demetra⁷ que implementa as metodologias X12 ARIMA⁸ e Tramo/Seats⁹.”(INE, Índices de Volume de Negócios e Emprego, 2010).

⁷ Seasonal Adjustment Software

⁸ Seasonal Adjustment Software desenvolvido por United States Census

⁹ Seasonal Adjustment Software desenvolvido pelo Banco de Espanha

3.3 Índices de Volume de Negócios, Emprego, Remunerações e Horas Trabalhadas nos Serviços

3.3.1 Introdução

Os índices utilizados são: «Índices de Volume de Negócios, Emprego, Remunerações e Horas Trabalhadas nos Serviços»¹⁰.

Estes índices permitem conhecer a evolução, no curto prazo, nestes sectores e respondem, igualmente, ao Regulamento CE nº 1165/98 de 10 de Maio de 1998, alterado pelo regulamento de 1158/05 de 6 de Julho de 2005.¹¹

Os índices de «Volume de Negócios, Emprego, Horas e Remunerações» resultam do apuramento de informação recolhida junto de empresas sediadas em território nacional cuja actividade principal se enquadre dentro das actividades trabalhadas. Por secções tem-se: A (Industria), B (Construção e Obras Públicas), C (Comércio e Retalho) e D (Serviços) do Regulamento (CE) nº 1165/98 do Conselho alterado pelo Regulamento 1158/05 de 6 de Julho de 2005 e pelo Regulamento CE nº 1893/2006 do Parlamento Europeu e do Conselho de 20 de Dezembro de 2006.

3.3.2 Objectivos

Estes índices têm como objectivo mostrar a evolução:

- Dos negócios nos mercados de bens e serviços .
- Do emprego.
- Dos salários e vencimentos.
- Do volume de trabalho efectuado.
- Do volume da produção em intervalos curtos e regulares.
- Da procura de produtos industriais, como indicação da produção futura, por mercados de origem.

¹⁰http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=106164961&DESTAQUESmodo=2

¹¹ INE. Índices de Volume de Negócios e Emprego. código:136, versão 1.0

3.3.3 Descrição

Volume de Negócios: Quantia líquida das vendas e prestações de serviços (abrangendo as indemnizações compensatórias) respeitantes às actividades normais das entidades, consequentemente após as reduções em vendas e não incluindo nem o imposto sobre o valor acrescentado nem outros impostos directamente relacionados com as vendas e prestações de serviços. (INE, 2011)¹²

Emprego: O emprego compreende todas as pessoas (tanto trabalhadores por conta de outrem como trabalhadores por conta própria) que exercem uma actividade produtiva abrangida pela definição de produção dada pelo sistema. (INE, 2011)¹³

Remunerações: As remunerações dos empregados definem-se como o total das remunerações, em dinheiro ou em espécie, a pagar pelos empregadores aos empregados como retribuição pelo trabalho prestado por estes últimos no período de referência. (INE, 2011)¹⁴

Horas trabalhadas nos serviços: Representa o número total de horas de trabalho efectivamente cumpridas por um trabalhador, por conta de outrem ou por conta própria, durante o período contabilístico, quando aquilo que produzem é abrangido pelo domínio da produção. (INE, 2011)¹⁵

¹² Directriz Contabilística n.º 22, DR 112, SÉRIE II, de 15-05-98; Transacções Sujeitas a Impostos Especiais sobre o Consumo

¹³ Regulamento (CEE) n.º 2223/96 do Conselho, de 25-06-96 - JO L 310 de 30-11-1996;

¹⁴ Regulamento (CEE) n.º 2223/96 do Conselho, de 25-06-96 - JO L 310 de 30-11-1996;

¹⁵ Regulamento (CEE) n.º 2223/96 do Conselho, de 25-06-96 - JO L 310 de 30-11-1996;

3.4 Secções e Actividades Económicas

3.4.1 Introdução

Neste subcapítulo serão apresentadas as secções económicas, tal como as actividades nestas contidas. Através da CAE.Rev.3 será apresentada a regulamentação na constituição das actividades económicas. Vista a importância da Classificação Portuguesa das Actividades Económicas, será apresentada uma pequena perspectiva histórica, tais como os seus principais objectivos e o melhor método de utilização desta classificação.

3.4.2 Descrição das Secções Económicas

TOTAL**¹⁶ – Secções G**, H, I, J, L, M, N da CAE-Ver.3

G**¹⁷ - Comércio por grosso; reparação de veículos automóveis e motociclos

H – Transportes e armazenagem

I – Alojamento, restauração e similares

J – Actividades de informação e de comunicação

L – Actividades imobiliárias

M – Actividades de consultoria, científicas, técnicas e similares

N - Actividades administrativas e dos serviços de apoio

¹⁶ , ¹⁷** Exclui Comércio a Retalho

3.4.3 Descrição das Actividades Económicas

A representação da descrição das actividades económicas de modo a compreender os dados, é feita da seguinte forma:

Tabela 3.1 Representação das Actividades Económicas

Variável	Identificação da actividade segundo CAE-REV.3	Secção	Nome da actividade
(01)	58	(J)	Actividades

Fonte: Criado pelo autor

- (01) 58 – (J) Actividades de Edição
- (02) 45 – (G) Comércio, manutenção e reparação de veículos automóveis e motociclos
- (03) 46 – (G) Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos
- (04) 49 – (H) Transportes terrestres, transportes por oleodutos e gasodutos
- (05) 50 - (H) Transportes por água
- (06) 51 - (H) Transportes aéreos
- (07) 52 - (H) Armazenagem e actividades auxiliares dos transportes
- (08) 53 - (H) Actividades postais e de «courier»
- (09) 55 – (I) Alojamento
- (10) 56 - (I) Restauração e Similares
- (11) 59 - (J) Actividades cinematográficas, de vídeo, de produção de programas de televisão
- (12) 60 - (J) Actividades de rádio e televisão
- (13) 61 – (J) Telecomunicações
- (14) 62 - (J) Consultoria e programação informática e actividades relacionadas
- (15) 63 - (J) Actividades dos serviços de informação
- (16) 68 – (L) Actividades imobiliárias
- (17) 69 – (M) Actividades jurídicas e de contabilidade e actividades de consultadoria para os negócios e gestão

- (18) 71 - (M) Actividades de arquitectura, de engenharia e técnicas afins, actividades de ensaios e de análises técnicas
- (19) 72 - (M) Actividades de investigação científica e de desenvolvimento
- (20) 73 - (M) Publicidade, estudos de mercado e sondagens de opinião
- (21) 78 - (N) Actividades de emprego
- (22) 79 - (N) Agências de viagem, operadores turísticos, outros serviços de reservas e actividades relacionadas
- (23) 80 - (N) Actividades de investigação e segurança
- (24) 77 - Actividades de Aluguer
- (25) 74 - (M) Outras actividades de consultadoria, científicas, técnicas e similares
- (26) 82 - (N) Actividades de serviços administrativos e de apoio prestados às empresas
- (27) 81 - (N) Actividades relacionadas com edifícios, plantação e manutenção de jardins

3.4.4 Classificação Portuguesa das Actividades Económicas CAE.Rev.3

“A classificação Portuguesa de Actividades Económicas, Revisão 3, do ano de 2007 abreviadamente designada por CAE-Rev.3, elaborada pelo Instituto Nacional de Estatística (INE) com a colaboração de cerca de duas centenas de entidades, envolvendo a Administração Pública, os Parceiros Sociais e, pontualmente, as Empresas, destina-se a substituir a CAE-Ver.2.1 de 2006.

A CAE-Rev.3, cuja estrutura foi publicada no Diário da República pelo Decreto-Lei de 2007, estabelece o novo quadro das actividades económicas portuguesas, harmonizado com a Nomenclatura Estatística das Actividades Económicas na Comunidade Europeia (NACE-Rev.2), no âmbito do Regulamento da CE nº 1893/2006 e do Parlamento Europeu e do Conselho, de 20 de Dezembro de 2006.

As alterações estruturais em relação à CAE-Rev.2.1 são significativas e decorrem da adaptação da NACE-Rev.2 ao Sistema Estatístico Nacional (SEN) e da necessidade de uma estruturação mais ajustada à actual organização económico-social nacional.

As tabelas de equivalência entre as CAE-Rev 2.1 e a CAE-Rev.3, incluídas em capítulo próprio desta publicação, procuram dar uma ideia das principais alterações entre as duas classificações, a nível da Subclasse (nível mais elementar). Para informações mais

desenvolvidas para este nível e os restantes níveis superiores aconselha-se consulta do site do INE.

As notas explicativas apresentam também desenvolvimentos de relevante importância para efeitos duma maior autonomia e coordenação na aplicação da CAE-Rev.3.e no apoio à análise estatística.”(CAE-Rev.3, 2007, p.3).

3.4.5 Objectivos CAE-Rev.3

“A presente nomenclatura de actividades económicas pretende dar resposta aos seguintes objectivos principais:

- Classificação e agrupamento das unidades estatísticas produtoras de bens e serviços (com ou sem fins lucrativos), segundo a actividade económica;
- Organização, de forma coordenada e coerente, da informação estatística económico-social, por ramo de actividade económica, em diversos domínios (produção, emprego, energia, investimento, etc.);
- Comparabilidade estatística a nível nacional, comunitário e mundial.

A estrutura, conceitos e notas explicativas da CAE-Rev.3 são, no essencial, o resultado, por um lado, da harmonização imposta pelo Regulamento NACE-Rev.2 e, por outro, da conciliação de interesses e de necessidades nacionais a satisfazer face às condições actuais de organização económica e à previsão da sua evolução no médio prazo.

Os objectivos da CAE-Rev.3 são essencialmente estatísticos, embora possam ser utilizados para fins não-estatísticos. Neste sentido, os princípios básicos da sua construção, o tipo de unidades estatísticas a que se aplica, as regras de classificação e a determinação da actividade principal, entre outros aspectos, estão subordinados aos objectivos estatísticos.” (CAE-Rev.3, 2007, p 10).

3.4.6 Evolução História CAE-Rev.3

“A primeira versão da classificação Portuguesa de Actividades Económicas (CAE) publicada remonta ao ano de 1953 e resultou de uma tradução, da responsabilidade do INE, da Classificação Internacional Tipo de Todos os Ramos de Actividade Económica, abreviadamente designada por CITA, editada em 1949 pelos Serviços de Estatística das Nações Unidas, (sob a cota ST/STAT/SER.M/4).

A CITA foi objecto de uma primeira revisão para uma melhor adaptação à realidade económica mundial, tendo a ONU publicado em 1958, sob a cota ST/STAT/SER.M/4/Rev.1, uma versão rectificada e actualizada (CITA-Rev.1).

A edição da CITA-Rev.1 levou o INE a empreender, de novo, a sua tradução para português, tendo-a submetido posteriormente à apreciação de várias entidades públicas e privadas. Das contribuições recebidas resultaram modificações ao projecto inicial do INE, que o publicou em 1961, após aprovação prévia dos Serviços de Estatística das Nações Unidas.

A experiência com a execução dos trabalhos estatísticos revelou que a tradução da CITA-Rev.1 era insuficiente para responder às necessidades nacionais, tendo publicado em 1964 a primeira CAE adaptada à realidade económica portuguesa, elaborada a partir da CITA-Rev.1, após aprovação de várias entidades públicas e privadas.

Em 1969, os Serviços de Estatística das Nações Unidas publicaram, sob a cota ST/STAT/SER.M./4/Rev.2, a segunda revisão da CITA (CITA-Rev.2), tendo o INE, sempre atento às necessidades portuguesas nesta matéria, decidido igualmente proceder à sua tradução e publicação em 1970, após aprovação dos Serviços de Estatística da ONU.

Como a tradução para português da CITA-Rev.2 não respondia às necessidades nacionais, o Conselho Nacional de Estatística (CNE) encarregou uma Comissão de conceber uma nova CAE a partir da CITA-Rev.2, tendo o INE, após aprovação do CNE, publicado em 1973 a CAE-Rev.1.

Em Outubro de 1978, face à necessidade de ajustar o sistema de informação estatística nacional às exigências em matéria estatística derivadas da futura adesão de Portugal à CEE, o CNE, através da 11ª Resolução, criou um Grupo de Trabalho com o mandato de rever a CAE-Rev.1/73 à luz da Nomenclatura Geral de Actividades Económicas da CEE (NACE) de 1970 e criar uma Classificação Nacional de Bens e Serviços (CNBS). Este Grupo de Trabalho,

transformado posteriormente pelo CNE em Subcomissão Especializada da CAE/CNBS, concluiu, em 1985, os trabalhos de uma nova CAE harmonizada com a NACE e da CNBS. Estes projectos não foram aprovados nem adoptados por as actividades do CNE terem sido suspensas no início de 1986.

Na sequência do Regulamento (CEE) n.º 3037/90 do Conselho, de 9 de Outubro, relativo à Nomenclatura Estatística das Actividades Económicas Europeias (NACE-Rev.1), o INE, em colaboração com cerca de centena e meia de entidades, elaborou a CAE-Rev.2, integrada na NACE-Rev.1 e harmonizada, tanto quanto possível, com a CAE-Rev.1/73 e com o projecto da CAE/85 que não chegou a ser aprovado pelo CNE.

A CAE-Rev.2, foi aprovada em Dezembro de 1991, pela 32ª Deliberação do Conselho Superior de Estatística (CSE), e pela Comissão da CEE (EUROSTAT), nos termos do n.º 3 do art.º 3 do Regulamento (CEE) n.º 3037/ 90, tendo sido publicada no Diário da República, a coberto do Decreto-Lei n.º 182/93, de 14 de Maio.

A CAE-Rev.2.1, aprovada em Novembro de 2002, pela 241ª Deliberação do CSE e pela Comissão nos termos do Regulamento n.º 29/2002, foi publicada posteriormente no Diário da República a coberto do Decreto-Lei n.º 197/2003, de 27 de Agosto.” (CAE-Rev.3, 2007, p.9).

3.4.7 Como utilizar melhor esta Classificação

“A CAE-Rev.3 apresenta desenvolvimentos importantes, em quase todas as partes, em relação à versão precedente e foi cuidadosamente preparada para ser utilizada por leitores mais ou menos familiarizados com as especificidades técnicas destas matérias.

A experiência tem demonstrado que as matérias aqui incluídas exigem uma leitura atenta para melhor satisfazer os interesses de cada utilizador, revestindo de particular importância os destaques incluídos neste capítulo da Apresentação Geral.

As matérias apresentam-se arrumadas em compartimentos o que leva muitas vezes os utilizadores a não terem uma preocupação de apreender a classificação como um todo, situação que não permite obter os melhores resultados.

Apesar de não se apresentar tarefa fácil propor um método de consulta para tirar o maior proveito desta classificação, a experiência revela que só a leitura repetida e aprofundada, em particular, dos princípios, conceitos e métodos incluídos na Apresentação Geral permite compreender e potenciar os resultados a alcançar com esta classificação.

A par da leitura da Apresentação Geral, recomenda-se a utilização do Índice Alfabético (concebido para assegurar uma maior coordenação na atribuição da CAE-Rev.3 às unidades estatísticas) e as Tabelas de Equivalência pela sua importância na ligação das séries estatísticas.” (CAE-Rev.3, 2007, p.10).

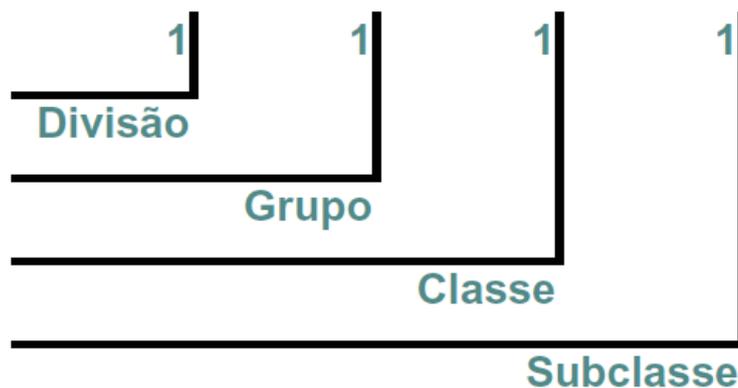
3.4.8 Sistemas de Codificação

“O sistema de codificação adoptado na CAE-Rev.3 pode dividir-se em duas partes: uma alfabética com um nível (Secção) e outra numérica com quatro níveis (Divisão, Grupo, Classe e Subclasse).

Na parte alfabética, as 21 Secções são codificadas com uma letra de A a U. A codificação numérica inicia-se no nível da Divisão com dois dígitos, desce ao Grupo (nível com 3 dígitos), segue-se o nível Classe (4 dígitos) e termina na Subclasse (nível com 5 dígitos).

A relação nível/número de dígitos depreende-se facilmente do esquema que a seguir se apresenta para a Subclasse 01111 (Cerealicultura)”. (CAE-Rev.3, 2007, p.14)

Figura 3.1 Sistema de Codificação



Fonte: CAE-Rev.3

4 Métodos de Classificação

4.1 Introdução

“Cluster Analysis is the art of finding groups in data.” (Kaufman e Rousseeuw, 2005, p.1)

A Classificação de variáveis e objectos semelhantes em grupos é uma importante actividade humana. Em todos os processos da vida a classificação faz parte do processo de aprendizagem principalmente quando esta tem um peso enorme tal como quando uma criança aprende a distinguir o preto do branco, o pai da mãe, tornando-se num processo do quotidiano onde no subconsciente estão a ser aplicadas classificações consoante as situações com que a criança se depara constantemente. Também nas áreas de economia, gestão e marketing a classificação segundo Gordon (1999) surge para a investigação de relações entre conjuntos de objectos ou variáveis o que poderá resultar ou não num resumo de dados sob a forma de grupo, de objectos ou variáveis semelhantes.

Em gestão e marketing tem-se como exemplo a segmentação de mercado. Tratando-se de uma das mais fundamentais estratégias no conceito de mercado.

“The Market is the set of all actual and potential buyers of a product or service.” (Armstrong and Kotler, 1999, p.179)

“Market Segmentation is dividing a market into distinct groups of buyers on the basis of needs, characteristics, or behavior who might require separate products or marketing mixes”(Armstrong and Kotler, 1999, p.180)

Ainda segundo Armstrong and Kotler, “uma empresa que pratica segmentação de mercado, reconhece que os consumidores têm necessidades diferentes, percepções e comportamentos próprios de compra. A empresa tenta isolar grandes segmentos, que

constituam um mercado e que este, se adapte mais perto das necessidades de um ou mais segmentos.” (Amstrong and Kotler, 1999, p.181)

Desta forma para o segmento de mercado pode ser utilizada a análise classificatória, para formar agrupamentos de pessoas de acordo com mesmo nível de semelhança dos produtos ou serviços em questão.

Já desde o século XVIII segundo Linnaeus e Sauvages “A classificação foi aplicada em animais, doenças, plantas e minerais, fazendo esta parte do desenvolvimento da história da ciência (Kaufman e Rousseeuw, 2005, p.1).

Segundo Saporta “A Classificação de variáveis é um assunto importante, mas menos abordado nas obras clássicas”(Saporta. 2005, p.2).

O seguinte exemplo de uma aplicabilidade mais conhecida permite uma melhor percepção da explicação do termo.

“Archacologists discover artifacts, such as ornaments and stone implements, at various sites. One of their aims is to detect if there are any artifact “types”, i.e. to establish whether some of the artifacts closely resemble one another. This could be a preliminary to an investigation of the spatial distribution of artifact “types” ”. (Hodson, Sneath and Doran, 1966, p.1)

Neste caso o objecto corresponde ao artefacto arqueológico. Os objectos, por sua vez, são geralmente descritos por um conjunto de variáveis, no exemplo apresentado trata-se, portanto, das propriedades físicas do artefacto. Os artefactos podem ser considerados como as variáveis a agrupar sendo os artefactos descritos por características (objectos).

Outro exemplo, sendo este dos mais famosos que surgiu, recorrendo aos mais modernos métodos de classificação, é segundo Gordon “A Tabela Periódica de elementos químicos”. (Gordon, 1999, p.2).

Em suma, através da análise classificatória, pretendem-se obter grupos de variáveis que têm em comum, entre si, o mesmo nível de semelhança.

“Une bonne classification produira des classes avec une grande similarité á l'intérieur de chaque classe et une petite similarité entre les différentes classes. Sa qualité dépend á la fois de la mesure de similarité utilisée par la méthode et de son implémentation.” (Nakache e Confais, 2005, p.15).

Este trabalho compara classificação de variáveis - as vinte e sete actividades económicas para cada um dos índices referidos, ao longo dos doze meses de 2009 e dos doze meses de 2010. Apesar de os dados serem temporais, são tratados como sendo dados seccionais, não utilizando a abordagem dada, por exemplo em Caiado (2010), pois não se possui observações temporais suficientes para cada caso de modo a que justifique uma abordagem classificatória de dados temporais, não considerando «abusivo» tratá-los como seccionais até porque como já se referiu em 3.2, os dados recolhidos apresentam-se limpos de sazonalidade.

4.2 Classificação Hierárquica

Existem duas formas de classificação hierárquica: a Ascendente e a Descendente. Neste capítulo vai ser apresentada uma explicação do método de classificação hierárquica descendente e uma explicação um pouco mais extensa do método de classificação hierárquica Ascendente, não só por ser o utilizado neste trabalho, mas também porque é o mais usado de uma forma geral em classificação.

“A hierarchical classification is a sequence of partitions of a set of n objects, starting with the partition into n object group, and successively merging two or more groups until the final partition consists of one n -object group” (Murtagh, 1984, p.35).

“The end result of a classification study is often a partition of the set of disjoint classes such that objects in same classes is similar to one another.” (Gordon, 1999, p.23).

Nos métodos de classificação hierárquica ascendente tem-se habitualmente necessidade de comparar as estruturas hierárquicas obtidas por diferentes metodologias com base nos mesmos indivíduos ou variáveis, mas também as partições extraídas dessas estruturas. Para o efeito recorre-se a coeficientes de comparação específicos.

4.2.1 Classificação Não Hierárquica

Apesar de não ser utilizada neste trabalho, vai ser apresentada uma breve descrição sobre métodos de classificação não hierárquica. Destes métodos os mais utilizados são do tipo k -means¹⁸ – estes métodos necessitam do conhecimento de k e consistem em duas etapas fundamentais:

1. Parte-se de k classes iniciais (de elementos), k aleatório ou pré-fixado.
2. Continua-se com um processo sequencial de reafectação dos elementos aos grupos, com o objectivo de se obterem grupos em que a semelhança entre os elementos de cada grupo seja máxima.

4.2.2 Classificação Hierárquica Ascendente

Neste método em cada etapa dois ou mais grupos são reunidos num só grupo: sequencialmente o número de grupos diminui, mas o número de elementos por grupo aumenta, até se obter um único grupo formado por todos os elementos como descrito anteriormente na citação de Murtagh, 1984.

“A Classificação Hierárquica Ascendente tem como resultado uma sucessão de partições encaixadas. Esta metodologia parte da partição mais fina, formada pelos elementos a classificar isolados, e em cada passo do algoritmo reúne as classes mais semelhantes, de acordo com o critério predefinido.”(Sousa e Nicolau, 2002)

¹⁸ Em Português designam-se por “ k -médias”

Neste trabalho, quando se falar apenas em métodos de classificação hierárquica, está-se a referir a métodos ACHA – métodos de análise classificatória hierárquica ascendente

Os métodos ACHA, sobre um conjunto de elementos (variáveis ou objectos), baseiam-se:

- i. na escolha de um índice ou coeficiente de comparação entre os pares de elementos – de acordo com a natureza dos dados em análise,
- ii. seguindo-se a escolha de um coeficiente de comparação entre os pares de grupos de elementos.

O coeficiente de comparação entre os grupos de elementos conduz a que esses grupos se reúnam em novos grupos em cada etapa da hierarquia – nos grupos encontram-se elementos semelhantes entre si e dissemelhantes em relação aos outros grupos.

Em ACHA obtém-se uma sucessão de partições encaixadas, representado a hierarquia que, graficamente, se pode representar sob a forma de árvore de classificação ou dendrograma.

Neste trabalho vão ser utilizadas apenas árvores binárias – árvores que apresentam $n-1$ nós.

4.3 Medidas Proximidade e de Distância

Conforme se pode verificar em Kaufman e Rousseeuw (2005), entre outros, existem vários métodos de obtenção de proximidade e distância. Neste trabalho vai ser utilizado apenas um, isto é, coeficiente de semelhança¹⁹ - «coeficiente de correlação de Pearson» ou «coeficiente de correlação de Bravais-Pearson» modificado, visto tratar-se de um coeficiente adequado a variáveis contínuas – tipo de variáveis tratadas neste estudo.

4.3.1 Coeficiente de Semelhança «Correlation-coefficient»

Em estatística o coeficiente de correlação é utilizado para dar a medida de intensidade de relação linear entre um par de variáveis.

$$r_p = s_{ij} = \frac{\sum_{k=1}^p (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\left[\sum_{k=1}^p (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^p (x_{jk} - \bar{x}_j)^2 \right]^{\frac{1}{2}}} \quad (4.1)$$

$$\text{Onde: } \bar{x}_i = \frac{1}{p} \sum_{k=1}^p x_{ik} \quad (i = 1, \dots, n)$$

Este coeficiente varia entre -1 e 1.

O coeficiente de dissemelhança que vai ser utilizado neste trabalho designar-se-á por r_p^* e é dado pela expressão

$$r_p^* = \frac{1 - r_p}{2} \quad (4.2)$$

tal que $0 \leq r_p^* \leq 1$, segundo sugestão de diversos autores, nomeadamente Gower (1985), Silva (2005), entre outros.

¹⁹ Neste caso transformado em dissemelhança

4.4 Coeficientes de comparação entre elementos de em conjunto de dados

Dado um conjunto de dados E , se for pretendido classificar os seus elementos, define-se o coeficiente de comparação (função de comparação) como sendo uma aplicação:

$$\begin{aligned} c : E \times E &\rightarrow \mathbb{R}_0^+ \\ (k, k') &\rightarrow c(k, k') \end{aligned} \quad (4.3)$$

estes coeficientes ou índices de comparação podem ser coeficientes de:

- semelhança - $c \equiv s$ (4.4)

ou

- dissemelhança - $c \equiv d$ (4.5)

(em particular d pode ser uma distância)

Diz-se que dois elementos são tão mais semelhantes entre si quanto maior for o valor do coeficiente de semelhança e consequentemente, menor for o valor do coeficiente de dissemelhança. E são tão mais dissemelhantes entre si quanto menor for o valor do coeficiente de semelhança e consequentemente, maior for o valor do coeficiente de dissemelhança.

Estes coeficientes têm sido estudados e descritos por diversos autores referidos ao longo do texto e exaustivamente em Gower (1985). As notações aqui utilizadas são as que se encontram em Bacelar-Nicolau (1980) e Silva (2005) entre outros.

4.4.1 Semelhança

Considere-se uma aplicação s entre dois elementos como definida em (4.3) e representada em (4.4).

A expressão $s_{kk'}$ representa o coeficiente de comparação, satisfazendo às seguintes condições (não necessariamente em simultâneo):

- i. $s_{kk} = s_{max}, \forall k \in E$ — onde s_{max} é o maior valor de s
- ii. $s_{kk'} = s_{k'k}, \forall k, k' \in E$
- iii. $s_{kk'} = s_{max} \Rightarrow k = k', \forall k, k' \in E$
- iv. $s_{max} + s_{kk'} \geq s_{kk''} + s_{k''k'}, \forall k, k', k'' \in E$
- v. $s_{kk'} \geq \min(s_{kk''}, s_{k''k'}), \forall k, k', k'', \in E$

- Se s satisfaz as condições **(i)** e **(ii)**, s diz-se um índice de semelhança;
- Se s satisfaz as condições **(i)**, **(ii)** e **(iii)**, s diz-se um índice de proximidade;
- Se o índice de proximidade satisfaz a propriedade **(iv)**, s diz-se uma proximidade;
- Se o índice de proximidade satisfaz as propriedades **(iv)** e **(v)**, s diz-se uma proximidade ultramétrica.

4.4.2 Dissemelhança

Considere-se, agora, uma aplicação d entre dois elementos, como definida em (4.3) e representada em (4.5), $d_{ij} = 1 - s_{ij}$ (Gower, 1985).

A expressão $d_{kk'}$ representa o coeficiente de comparação, satisfazendo às seguintes condições (não necessariamente em simultâneo):

- i. $d_{kk} = 0 \forall k \in E$
- ii. $d_{kk'} = d_{k'k}, \forall k, k' \in E$
- iii. $d_{kk'} = 0 \Rightarrow k = k', \forall k, k' \in E$
- iv. $d_{kk'} \leq d_{kk''} + d_{k''k'}, \forall k, k', k'' \in E$
- v. $d_{kk'} \leq \max(d_{kk''}, d_{k''k'}), \forall k, k', k'' \in E$

- Se d satisfaz as condições (i) e (ii), d diz-se um índice de dissemelhança;
- Se d satisfaz as condições (i), (ii) e (iii), d diz-se um índice de distância;
- Se o índice de distância satisfaz a propriedade (iv), d diz-se uma distância;
- Se o índice de distância satisfaz as propriedades (iv) e (v), d diz-se uma distância ultramétrica.

Silva(2005).

Segundo Lerman (1981): “duas medidas de semelhança (ou dissemelhança) são equivalentes sobre um conjunto E de dados, se e só se, as preordenações associadas sobre E , são idênticas”.

4.5 Alguns métodos de classificação hierárquica ascendente

Existem diversos métodos de classificação hierárquica ascendente que se baseiam em diferentes coeficientes de comparação entre grupos. Os utilizados neste trabalho são apresentados de seguida.

4.5.1 Vizinho mais próximo

Designado em inglês por «single linkage» ou «nearest neighbour» este método traduz-se pela expressão abaixo indicada onde segundo Seber “Se C_1 e C_2 formam dois grupos, então a distância entre eles é definida pela menor dissemelhança entre C_1 e C_2 . (Sneath (1957), Sokal e Sneath (1963))”

$$d_{(C_1)(C_2)} = \min\{d_{rs} : r \in C_1, s \in C_2\} \quad (4.6)$$

Neste método:

“Passa-se portanto de um dado nível da hierarquia ao nível seguinte, por reunião das duas classes que possuem o par de elementos mais próximos. Daí o chamado “efeito de cadeia” deste método, que consiste na formação de classes com um núcleo muito forte, ao qual se vão agregando os restantes elementos um a um” (Bacelar-Nicolau, 1980, p.2).

4.5.2 Média da distância entre os grupos

Designado em inglês por «average linkage between groups», tem um desempenho intermédio entre o Vizinho mais próximo e o Vizinho mais afastado, traduz-se pela Média da distância entre os grupos onde segundo Seber “A distância entre C_1 e C_2 é definida de forma a ser a média de $n_1 n_2$ dissemelhanças entre todos os pares. (Sokal e Michener (1958), McQuitty (1964))”.

$$d_{(C_1)(C_2)} = \frac{1}{n_1 n_2} \sum_{r \in C_1} \sum_{s \in C_2} d_{rs} \quad (4.7)$$

«Average linkage clustering was developed as an antidote to the extremes of both single linkage and complete linkage». (Sokal e Michener, 1958, p.57)

4.5.3 Vizinho mais afastado

Este método também designado em inglês por «complete linkage» ou «furthest neighbour», é o oposto do Vizinho mais próximo que, segundo Seber: “A distância entre dois grupos é definida pela maior dissemelhança entre C_1 e C_2 (Sokal e Sneath (1963), McQuitty (1964))”.

$$d_{(C_1)(C_2)} = \max\{d_{rs} : r \in C_1, \quad s \in C_2\} \quad (4.8)$$

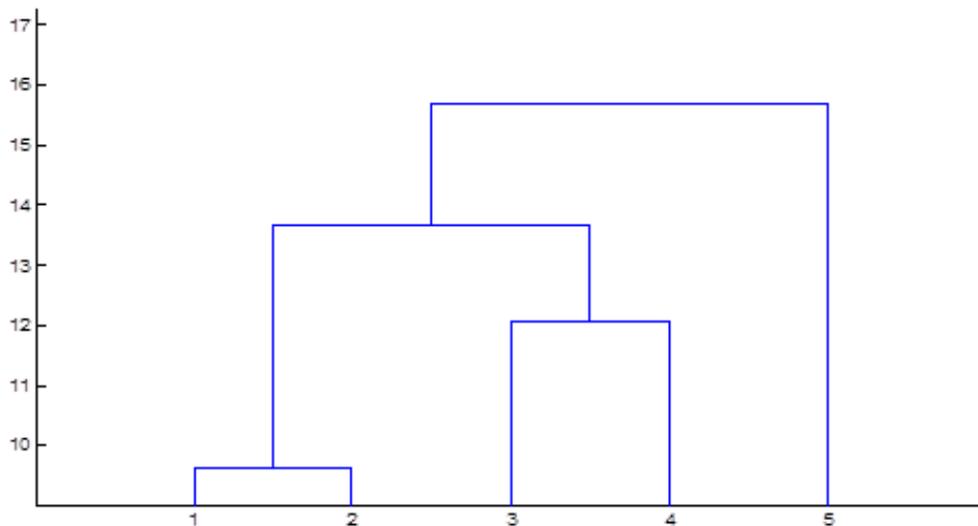
“The method is called complete linkage because all entities in a cluster are linked to each other at some maximum distance or minimum similarity” (Anderberg, 1973, p.18).

4.5.4 Matriz Cofenética e Matriz Ultramétrica

A Matriz ultramétrica é uma matriz do tipo quadrada, também designada por diversos autores como matriz cofenética. Nesta, o conhecimento dos elementos abaixo da diagonal principal é suficiente para o conhecimento da hierarquia, uma vez que é uma matriz simétrica e os elementos da diagonal principal são iguais a zero. A apresentação da matriz poderá ser representada com níveis de agregação ou ordem de agregação dos grupos (valores ordinais).

Conforme foi introduzido no subcapítulo 1.2 da Metodologia, através do software Matlab foi necessário construir totalmente a função para determinar as matrizes ultramétricas (Apêndice 9).

Figura 4.1 Exemplo de um Dendrograma com Cinco Variáveis.



Matriz Cofenética

Para apresentação dos níveis de agregação, tem-se a matriz cofenética, que é a matriz que agrupa as variáveis consoante os níveis de agregação.

Tabela 4.1 Matriz Cofenética

	1	2	3	4	5
1					
2	9.7266				
3	13.8484	13.8484			
4	13.8484	13.8484	12.1621		
5	15.6928	15.6928	15.6928	15.6928	

Matriz Ultramétrica

A correspondente matriz ultramétrica é formada por valores ordinais que demonstram a ordem de agregação dos grupos, isto é o primeiro grupo a ser obtido é o {4,5} a que vai corresponder valor ordinal 1 indicação do primeiro grupo, o segundo grupo a ser formado será {1,3} a que corresponde o valor ordinal 2.

Tabela 4.2 Matriz Ultramétrica

	1	2	3	4	5
1					
2	1				
3	3	3			
4	3	3	2		
5	4	4	4	4	

4.6 Comparação de Estruturas Hierárquicas e de Partições

Uma das questões importantes em classificação são os métodos como se comparam as estruturas hierárquicas e as partições resultantes dos cortes feitas nas mesmas. Esses métodos consistem na análise dos coeficientes de comparação que se descrevem nos pontos seguintes.

4.6.1 Coeficiente de correlação de Spearman e comparação de Estruturas Hierárquicas

“De todas as estatísticas baseadas em postos, o coeficiente de correlação por postos de Spearman foi a que surgiu primeiro e é talvez a mais conhecida hoje. É representada por r_s e é uma medida de associação que exige que ambas as variáveis se apresentem em escala de mensuração pelos menos ordinal, de modo que os objectos ou indivíduos em estudo possam dispor-se por postos em duas séries ordenadas”(Siegel, 2006, p.21).

Exemplo:

Suponham-se n indivíduos ordenados em postos segundo duas variáveis. Por exemplo um grupo de estudantes ordenado de acordo com suas médias de entrada para uma universidade e também de acordo com a média da classificação escolar ao fim do primeiro ano da universidade. Sendo que a grelha de classificações com a média de entrada para a universidade é representada por X_1, X_2, \dots, X_n , e a grelha de classificações ao fim do primeiro ano representada por Y_1, Y_2, \dots, Y_n , pode então determinar-se a relação entre X'_s e os Y'_s .”

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad (4.9)$$

Onde d_i é diferença entre os dois postos ($d_i = X_i - Y_i$) e X, Y são as respectivas variáveis, n é número de indivíduos.

Se forem consideradas as médias das classificações com uma ou duas casas decimais, poder-nos-emos deparar com indivíduos cujas médias são todas diferentes, sendo esta formula

adequada – continuando a sê-lo no caso do número de empates (não sendo certamente o número de classificações idênticas elevado). No entanto se forem consideradas as médias arredondadas, sem casas decimais, o número de classificações iguais em X e Y , será mais frequente, deve-se então considerar a forma de r_s corrigida tendo em consideração os empates, como se pode consultar em Siegel (2006), entre outros:

$$r_s = \frac{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 + \sum_{i=1}^n d_i^2}{2\sqrt{(\sum_{i=1}^n x_i^2) \times (\sum_{i=1}^n y_i^2)}} \quad (4.10)$$

onde:

$$\sum_{i=1}^n x_i^2 = \frac{n^3 - n}{12} - \sum_{i=1}^n T_x \quad (4.11)$$

$$\sum_{i=1}^n y_i^2 = \frac{n^3 - n}{12} - \sum_{i=1}^n T_y \quad (4.12)$$

tais que os factores de correcção T_x e T_y , são representados de um modo genérico por:

$$T = \frac{t^3 - t}{12} \quad (4.13)$$

para X e Y respectivamente; t é o número de observações empatadas.

A título de exemplo, considerando as médias das classificações arredondadas, tem-se:

Tabela 4.3 Média de Classificações Arredondadas

X	Y
12	12
12	15
14	13
14	15
15	13
15	15
16	17
16	18
17	17

Como habitualmente fixa-se um nível de significância α e recorre-se á tabela de coeficientes de correlação de Spearman, apresentada no anexo 1.

Tendo-se que para α fixo a região de rejeição será dada por $|r_s| > k$, (k – tabelado)

Aplicando o coeficiente de correlação Spearman tem-se $r_{s_{obs}} = 0.7435$

Utilizando-se um nível de significância de 5% ($\alpha = 0.05$) e tendo neste caso $n = 9$; pela tabela do coeficiente de correlação de Spearman tem-se $k = 0.648$ e portanto $r_{s_{obs}} > k$, logo observa-se que para esta amostra as médias de entrada na universidade estão correlacionadas com as médias obtidas no final do primeiro ano da universidade.

É este é o coeficiente que se vai utilizar para comparar as estruturas hierárquicas (os dendrogramas) com base nas matrizes das ultramétricas, apresentadas sob a forma de colunas à semelhança do exemplo anterior.

Neste trabalho de acordo com os dados irá utilizar-se a nomenclatura:

- Se $|r_s| > k$ as estruturas hierárquicas são semelhantes.

4.6.2 Índice de Rand e outros Índices de Semelhança entre Partições

Um dos problemas fundamentais da análise de classificatória, é comparar duas partições com origem nos mesmos dados.

“One way to compare two clusterings of the same data set is to calculate smilarity between such clusters using similarity indices”. (Albatineh, 2010).

Existem diversos índices de semelhança na literatura também designados por índices de concordância por alguns autores, como por exemplo em Cardoso (2010), tais como o índice de Rand em Rand (1971), bastante usado na literatura, bem como o índice de Rand ajustado, segundo Hubert e Arabie (1985) entre outros discutidos por exemplo em Hubalek

(1982), Albatineh (2006). Existem índices de semelhança direccionados nas mais diversas áreas tais como: «clustering validation» por Milligan (1983), «image segmentation and analysis», Unnikrishnan (2005), «bioinformaticas», Albatineh and Razeghifard (2008).

Coloca-se frequentemente a questão de saber interpretar os resultados fornecidos pelos diversos índices. Vários artigos baseando-se em simulações de Monte Carlo sugerem interpretações para os mesmos.

Neste trabalho utilizar-se-ão os índices de Rand, segundo Rand (1971), e Albatineh (2006), que generaliza a dedução de desvio padrão e variância para o índice de Rand originalmente deduzida por Fowlkes and Mallows (1983).

Em Albatineh (2010) é introduzida uma família L de índices de semelhança (nove índices), tais que os membros dessa família são funções lineares valores das células presentes na matriz de contingência $M = [m_{ij}]$, obtidas a partir de dois algoritmos de classificação, com I e J partições respectivamente P_i e P_j ($i = 1, 2, \dots, I$) e ($j = 1, 2, \dots, J$)

Tabela 4.4 Tabela de Contingência

	P_1	P_2	...	P_j	...	P_J	
P_1	m_{11}	m_{12}	...	m_{1j}	...	m_{1J}	m_{1+}
P_2	m_{21}	m_{22}	...	m_{2j}	...	m_{2J}	m_{2+}
...
P_i	m_{i1}			m_{ij}		m_{iJ}	...
...
P_I	m_{I1}	m_{I2}	...	m_{IJ}	...	m_{IJ}	m_{I+}
	m_{+1}	m_{+2}	...	m_{+j}	...	m_{+J}	m^{20}

²⁰ No trabalho de Albatineh (2010) m coincide com o n apresentado nas fórmulas este m nunca consta do trabalho, a matriz de contigência aqui apresentada representa a matriz $M = [m_{ij}]$, apresentada no texto.

Para os nove índices é deduzida uma expressão geral para a média e variância (de cada um deles) “under fixed marginal totals of the matching counts matrix and independence of clustering algorithms”(Albatineh,2010).

Para este trabalho seleccionaram-se três índices da família **L**, entre eles o índice de Rand como já foi referido e os índices PE e κ , utilizou-se também o índice de Rand Ajustado – para este último recorreu-se às propriedades estabelecidas em Steinley(2004).

“The article concludes by summarizing several properties of the adjusted Rand Index across many conditions and provides a method for the significance of observed adjusted Rand índices”(Steinley, 2004).

Albatineh(2010) estabelece a expressão geral dos índices de semelhança da família **L** com base em dois parâmetros α e β , SI^{21} fornece a expressão geral de qualquer índice de semelhança entre dois métodos de agrupamento de I e J partições desta família **L** e é escrito na seguinte forma:

$$SI = \alpha + \beta \left(\sum_{i=1}^I \sum_{j=1}^J m_{ij}^2 \right) \quad (4.14)$$

Número total de pares:

$$M = \frac{n(n-1)}{n} \quad (4.15)$$

Onde n é número de observações a agrupar (variáveis ou objectos).

²¹ SI representa uma família de índices de semelhança

Valor Esperado do índice de semelhança:

$$E(SI) = \alpha + \beta \left(\frac{PQ}{n(n-1)} + n \right) \quad (4.16)$$

Onde P e Q :

$$P = \sum_{i=1}^I m_{i+}^2 - n \quad (4.17)$$

$$Q = \sum_{j=1}^J m_{+j}^2 - n \quad (4.18)$$

Variância do índice de semelhança:

$$Var(SI) = \beta^2 \left(\frac{2PQ}{n(n-1)} + \frac{4P'Q'}{n(n-1)(n-2)} + \frac{(P^2-4P'-2P)(Q^2-4Q'-2Q)}{n(n-1)(n-2)(n-3)} - \frac{P^2Q^2}{n^2(n-1)^2} \right) \quad (4.19)$$

Onde P' e Q' :

$$P' = \sum_{i=1}^I m_{i+}(m_{i+} - 1)(m_{i+} - 2) \quad (4.20)$$

$$Q' = \sum_{j=1}^J m_{+j}(m_{+j} - 1)(m_{+j} - 2) \quad (4.21)$$

Intervalo de rejeição de semelhança entre partições

$$]E(SI) - 2\sqrt{Var(SI)}; E(SI) + 2\sqrt{Var(SI)}[\quad (4.22)$$

isto é,

“for pairs of samples that are similar, the SI falls outside the limits $E(SI) \pm 2\sqrt{Var(SI)}$ ”

Albatineh(2010).

Apresentam-se nos pontos seguintes os índices seleccionados para o estudo prático deste trabalho e já indicados em 4.6.2.

Índice de Rand

Com

$$\alpha = 1 - \frac{1}{2M}(P + Q + 2n) \quad (4.23)$$

e

$$\beta = \frac{1}{M} \quad (4.24)$$

Resultando em:

$$R = 1 - \frac{1}{2M}(P + Q + 2n) + \frac{1}{M} \left(\sum_{i=1}^I \sum_{j=1}^J m_{ij}^2 \right) \quad (4.25)$$

- $R \in [0,1]$ - Este índice aproxima-se da unidade quando o número de grupos aumenta.

Índice de Peirce

Peirce, (1884) - PE

Com

$$\alpha = \frac{2nM + PQ}{P(2M - P)} \quad 4.26$$

e

$$\beta = \frac{2M}{P(2M - P)} \quad 4.27$$

Resultando em:

$$PE = \frac{2nM + PQ}{P(2M - P)} + \frac{2M}{P(2M - P)} \left(\sum_{i=1}^I \sum_{j=1}^J m_{ij}^2 \right) \quad 4.28$$

- $PE \in [-1,1]$

Índice de CohenCohen, (1960) – κ

Com

$$\alpha = \frac{-2nM - PQ}{M(P + Q) - PQ} \quad 4.29$$

e

$$\beta = \frac{2M}{M(P + Q) - PQ} \quad 4.30$$

Resultando em:

$$\kappa = \frac{-2nM - PQ}{M(P + Q) - PQ} + \frac{2M}{M(P + Q) - PQ} \left(\sum_{i=1}^I \sum_{j=1}^J m_{ij}^2 \right) \quad 4.31$$

- $\kappa \in [-1,1]$

Segundo Albatineh, (2010). Estes dois índices apresentam padrões semelhantes de acordo com as simulações efectuadas.

Todos os índices estudados neste trabalho excepto PE e κ exibem uma distribuição assimétrica.

Índice de Rand Ajustado

O índice de Rand Ajustado segundo Hubert e Arabie $ARI_{HA}^{22, 23}$

$$ARI_{HA} = \frac{\sum_{i,j} \binom{n_{ij}}{2} - \frac{[\sum_i \binom{n_{i.}}{2}] \sum_j \binom{n_{.j}}{2}]}{\binom{n}{2}}}{\frac{1}{2} [\sum_i \binom{n_{i.}}{2} + \sum_j \binom{n_{.j}}{2}] - \frac{[\sum_i \binom{n_{i.}}{2} + \sum_j \binom{n_{.j}}{2}]}{\binom{n}{2}}} \quad (4.32)$$

- $ARI_{HA} \in [0,1]$

²² A forma geral de um índice com um valor esperado constante é dado pela expressão $\frac{\text{índice} - \text{valor esperado do índice}}{\text{valor máximo do índice} - \text{valor esperado do índice}}$, como descrito em Yeung e Ruzzo (2001)

²³ $n \equiv m$ da tabela de contigência $n_{i.} \equiv n_{i+}$ e $n_{.j} \equiv n_{+j}$

Alguns autores como Youness e Saporta (2010) preferem utilizar o índice de Rand pelo facto de não assumir valores negativos, de acordo com as condições do estudo. Outros autores como refere Cardoso (2010), concluem que ARI_{HA} “pode ser usado como vantagem sobre os restantes, no âmbito de avaliação de agrupamentos”.

Steinley (2004) apresenta os seguintes valores de referência após simulação para ARI_{HA} :

Tabela 4.5 «Quality of Cluster Recovery»

«Percentiles»	«Value»
95	0.86
90	0.77
85	0.60
80	0.60

Fonte: Steinley - *Properties of the Hubert-Arabic Adjusted Rand Index* (2004.,pp.392)

Segundo Steinley (2004):“Thus when one is validating clustering techniques, a set of heuristics for determining the quality of cluster recovery as related to ARI_{HA} could be:” (Steinley, 2004).

- i. «Values greater than 0.90 can be viewed as excellent recovery».
- ii. «Values greater than 0.80 can be considered good recovery».
- iii. «Values greater than 0.65 can be considered moderate recovery».
- iv. «Values less than 0.65 reflect poor recovery».

4.7 Metodologias de determinação do número de grupos

Neste ponto apresentam-se algumas, metodologias de particionamento resultantes de cortes nos dendrogramas, por uma linha paralela à base dos dendrogramas (visível ou não) que permitem determinar o número de grupos a utilizar.

Segundo Aldenderfer e Blashfield (1984) “Em alguns campos, mais precisamente das ciências sociais o problema de determinar o número de grupos não é o importante, visto que a análise incide na exploração das relações dos diferentes dados resultantes da hierarquia determinada”. No entanto para as ciências sociais existem duas formas de determinar o número de grupos:

- **Procedimentos Heurísticos:** São os métodos mais comuns; segundo os quais, a árvore hierárquica deve cortar-se segundo inspecção subjectiva dos diferentes níveis da árvore. Pode, no entanto, não ser totalmente satisfatório pois é baseado nas necessidades e opiniões do investigador e da respectiva estrutura hierárquica. Uma abordagem mais formal, mas também heurística, consiste por um lado na análise gráfica de onde resulta o número de grupos que essa análise sugere e por outro pela análise dos coeficientes de fusão, que em conjunto indicarão o possível nível do coeficiente de corte.

Gordon (1999) descreve uma metodologia muito semelhante à apresentada. Tem com objectivo alcançar o número de grupos obtidos segundo uma partição óptima em c classes para um intervalo de valores de c . De forma sucinta os grupos são formados fazendo a fusão de pares de grupos mais próximos, onde diferentes definições do que significa mais próximo, correspondem a diferentes critérios. O intervalo de partições em c grupos ($1 \leq c \leq n$) é examinado de forma a determinar o valor mais apropriado de c . Este autor também apresenta outros métodos mais formais para decidir o valor, designando-os de modo geral como regras de paragem («stopping rules»)

Em simultâneo recorreu-se a outro método baseado num trabalho desenvolvido pela orientadora desta dissertação, mas ainda não publicado:

- **Gráfico de frequências acumuladas:** baseia-se numa metodologia de representação gráfica. Sendo esta composta por comparações dos índices de agregação de grupo para grupo, entre outras representações gráficas, como por exemplo as apresentadas em Nakache e Confais (2005). Nesta metodologia selecionam-se os «maiores saltos» obtidos no gráfico de frequências acumuladas (relativas ou absolutas), baseadas nas frequências dos valores registados nas matrizes das ultramétricas ou das cofenéticas, á medida que se vão formando os grupos (considerando os singletões – grupos formados por um só elemento aos quais se atribui «valor nulo»). Não deixando de considerar o procedimento heurístico, como se mostra em seguida nas figuras 4.2. e 4.3.

Figura 4.2 Dendrograma de Horas com o Método Vizinho Mais Afastado no Ano 2009

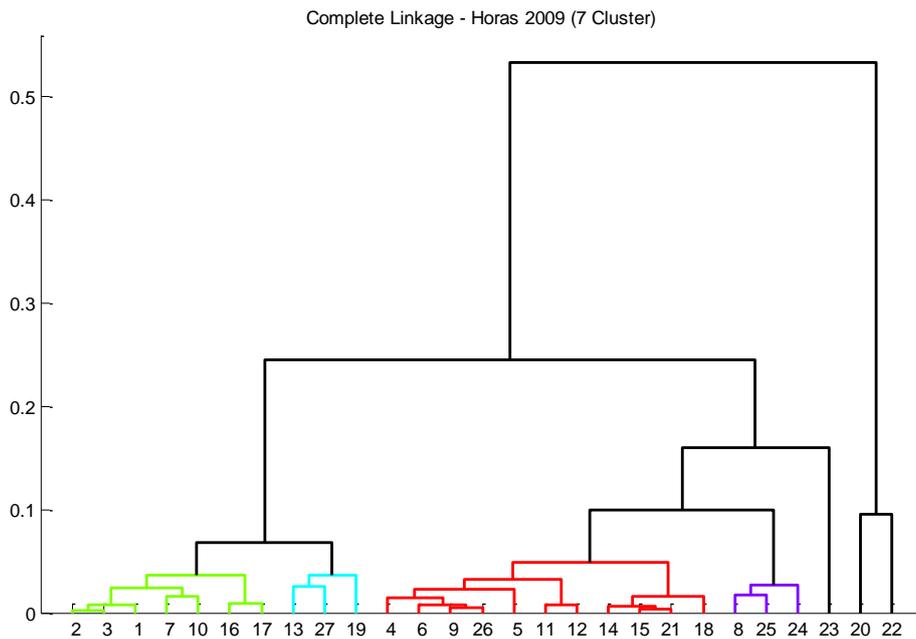
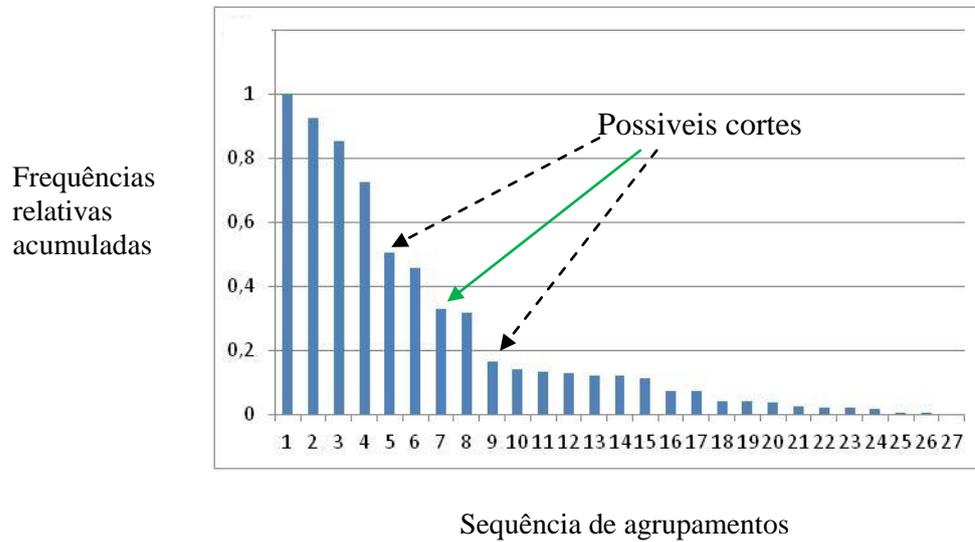


Figura 4.2 Gráfico de Frequências Relativas Acumuladas – Índice Horas com o Método Vizinho Mais Afastado no Ano 2009



Pela figura anterior pode-se verificar quais as possibilidades para selecção do nível de corte. Pelo gráfico de barras verifica-se um decréscimo mais acentuado para cinco, sete ou nove grupos.

5 Estudo Empírico e Resultados

5.1 Introdução

Neste trabalho procede-se à classificação de variáveis contínuas, ao longo dos doze meses de 2009 e dos doze meses de 2010²⁴ de acordo com o descrito no início do trabalho. São analisados quatro índices: Volume de Negócios, Emprego, Remunerações e Horas trabalhadas nos serviços. Para cada um deles é apresentado o comportamento mensal de vinte e sete actividades económicas que estão contidas em sete secções ou ramos distintos. Desta forma pretende-se verificar como se comportam áreas que contêm produtos ou serviços totalmente diferentes. Separado por índices o resultado surgirá mediante a formação de actividades económicas em grupos que se assemelham entre si. Apesar dos dados serem temporais, são tratados como sendo dados seccionais, não utilizando a abordagem dada por exemplo em Caiado (2010), pois não se possui observações temporais suficientes para cada caso, de modo que justifique uma abordagem classificatória de dados temporais, não considerando “abusivo” tratá-los como seccionais, até porque, como já se referiu em 3.2, os dados recolhidos apresentam-se limpos de sazonalidade.

Este capítulo começa por apresentar as diversas estruturas hierárquicas obtidas para os índices referidos, comparando-as dentro do mesmo ano e entre os dois anos com recurso ao coeficiente de correlação de Spearman. Em seguida procede-se a um corte nos dendrogramas seleccionados, obtendo-se partições que se comparam por meio de diversos índices de semelhança seleccionados da família L apresentada e do índice de Rand ajustado por Hubert e Arabie(1985).

Os primeiros resultados são apresentados sob a forma de dendrograma. Estas representações gráficas permitem, do ponto de vista do observador, a verificação de uma forma clara, como os grupos estão divididos. As diferentes cores representam um grupo com o mesmo nível de semelhança. De forma a mostrar outra perspectiva dos resultados, os mesmos serão mostrados sob forma de tabela. Por cada índice serão apresentadas conclusões dos resultados obtidos.

²⁴ Verificou-se a data de entrega desta dissertação, que apesar de recentemente, já estarem disponíveis os dados referentes ao ano de 2011.

5.2 Escolha do Método de Classificação Hierárquica a Utilizar

Inicialmente vão ser apresentados utilizando apenas um índice as diversas estruturas hierárquicas, baseadas nos métodos: Vizinho mais próximo, Média da distância entre os grupos e Vizinho mais afastado para o mesmo índice (Horas trabalhadas nos Serviços) no mesmo ano (2009).

Assim a preto tem-se um grupo de apenas um elemento (singletão), sendo que depois cada cor caracteriza grupos distintos com mais do que um elemento. No eixo dos xx tem-se as variáveis e no eixo dos yy os níveis de agregação.

Figura 5.1 Dendrograma de Horas com o Método do Vizinho Mais Próximo no Ano 2009

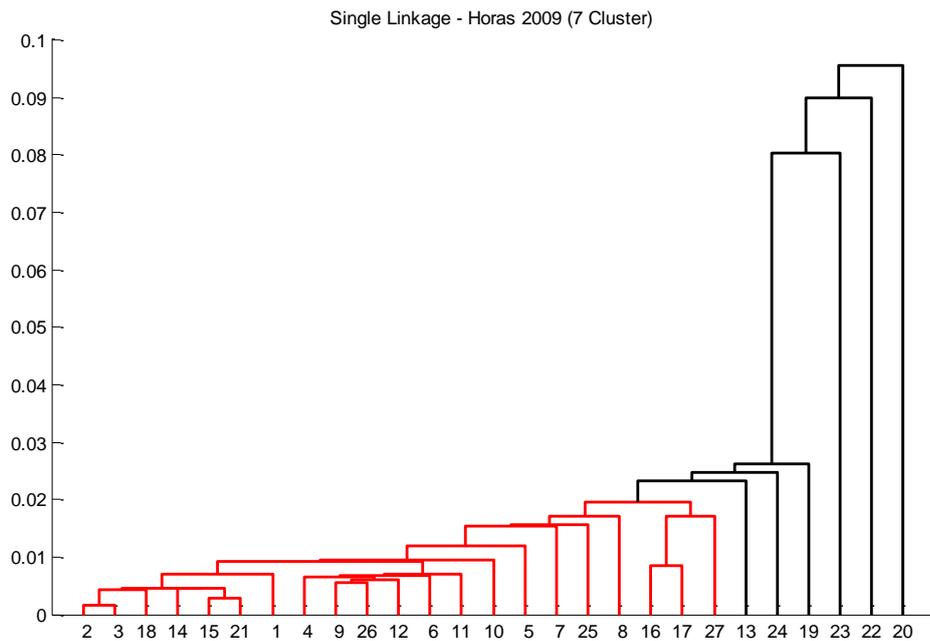


Figura 5.2 Dendrograma de Horas com o Método da Distância Entre os Grupos no Ano 2009

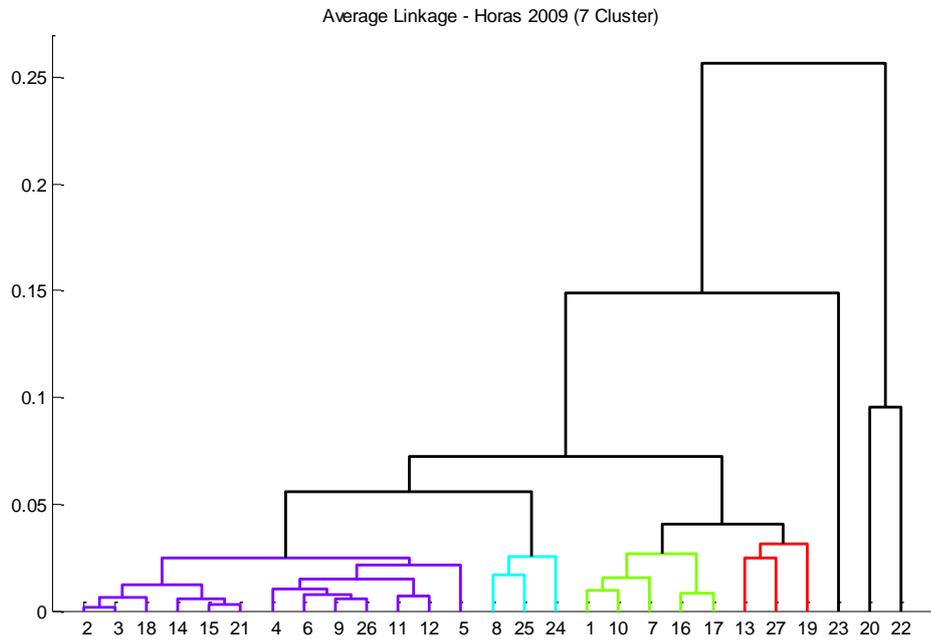
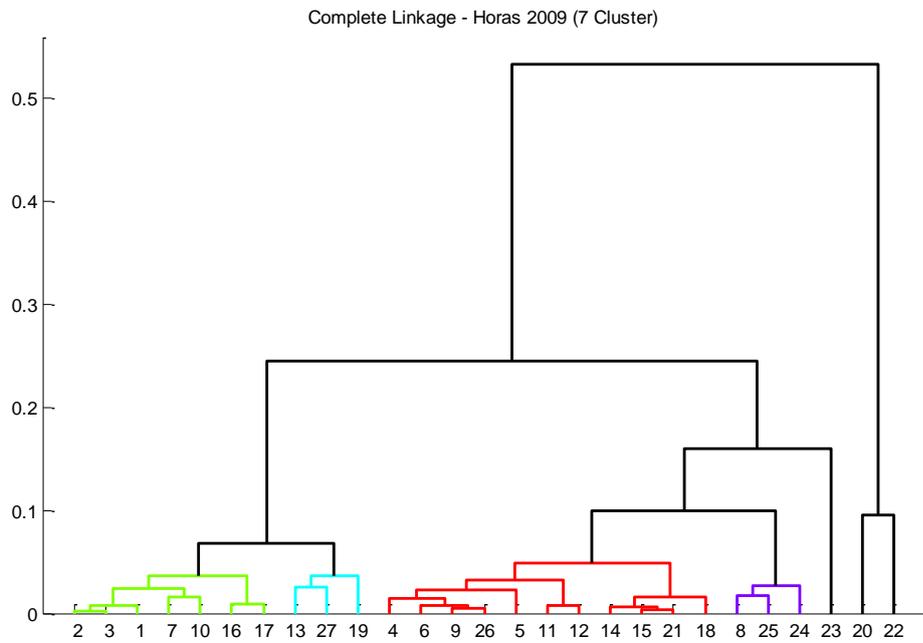


Figura 5.3 Dendrograma de Horas com o Método do Vizinho Mais Próximo no Ano 2009



Como se pode visualizar utilizando o método do vizinho mais próximo obtém-se o pior resultado para separação dos grupos, formando-se um efeito de cadeia característico deste método. O método da média da distância entre os grupos já permite visualizar melhor a separação dos grupos. No entanto o método do vizinho mais afastado permite obter a melhor estrutura. Desta forma o observador consegue, através deste método, uma clara visualização dos grupos. Por este motivo vai ser escolhido este método para estudo posterior, para além das justificações apresentadas no capítulo referente à descrição dos métodos.

5.3 Método de Particionamento

Através dos métodos apresentados em 4.7, foi seguida a metodologia baseada em gráficos de frequências acumuladas suportada também pelo método de observação (heurístico). Para este último o observador teve de aumentar a sensibilidade, através de um estudo mais aprofundado do tema desenvolvido e da análise das variáveis. Para isso foram determinantes os conhecimentos adquiridos ao longo de todo o percurso e conteúdo do 1º ano do Mestrado de Gestão de Empresas. A sensibilidade no tema é muito importante por se tratarem de áreas económicas diferentes.

Conforme foi mostrado em 4.7, as partições vão ser obtidas do ponto de vista do observador mediante a sensibilidade do tema que se está a trabalhar. Nesse sentido, após conclusão do método a utilizar, pode-se, pela figura 4.2, identificar sete grupos. A partição composta por sete grupos é aplicável a todas as estruturas hierárquicas apresentadas.

5.4 Coeficiente de Correlação de Spearman aplicado à comparação das Estruturas Hierárquicas

Tabela 5.1 Coeficiente de Correlação de Spearman – Resultados

	Volume Negócios		Emprego		Remunerações		Horas	
	2009	2010	2009	2010	2009	2010	2009	2010
Vizinho mais próximo vs. Média da distância entre grupos	0.9351	0.8948	0.862	0.953	0.405	0.956	0.788	0.950
Vizinho mais próximo vs. Vizinho mais afastado	0.0684	0.765	0.751	0.808	0.250	0.808	0.557	0.811
Média da distância entre grupos vs. Vizinho mais afastado	0.739	0.889	0.950	0.858	0.482	0.640	0.760	0.860
	2009/2010		2009/2010		2009/2010		2009/2010	
Vizinho mais próximo	0.887		0.892		0.463		0.730	
Média da distância entre grupos	0.919		0.805		0.711		0.565	
Vizinho mais afastado	0.456		0.637		0.214		0.477	

Através do subcapítulo 4.6.1 sabe-se que as estruturas hierárquicas são consideradas semelhantes (correlacionadas) quando se verifica a condição $r_s > k$.

O valor de k é determinado através da consulta da tabela de coeficiente de correlação de Spearman ²⁵ que neste trabalho para $n=27$ e $\alpha = 0.05$ assume o valor de $k=0.382$.

Os resultados obtidos na tabela 5.1. dariam indicação para selecionar como métodos a utilizar o Vizinho mais próximo e Média da distância entre os grupos. No entanto, para comparar os resultados de 2009/2010, conjugando os resultados obtidos com a representação gráfica e a experiência pessoal, levou à escolha do método do Vizinho mais afastado. Este encontra-se apresentado sob forma de tabela para cada um dos índices. Indicando se as estruturas hierárquicas resultantes são ou não semelhantes, isto é, se a “estrutura geral das duas classificações hierárquicas não é a mesma, mas as duas ultramétricas são “significativamente correlacionadas””(Silva, 2005).

Tabela 5.2 Comparação de Estruturas Hierárquicas 2009/2011 Utilizando o Método do Vizinho Mais Afastado para um Nível de Significância de 5%

Volume de Negócios	0.456 > 0.382	As estruturas hierárquicas são semelhantes
Emprego	0.637 > 0.382	As estruturas hierárquicas são semelhantes
Remunerações	0.214 < 0.382	As estruturas hierárquicas não são semelhantes – as duas estruturas de classificação não são correlacionadas
Horas	0.477 > 0.382	As estruturas hierárquicas são semelhantes

²⁵ Anexo 1-Tabela de coeficiente de correlação de Spearman

Apesar de, à partida, o resultado obtido relativamente a Remunerações em que as estruturas hierárquicas não são globalmente semelhantes, vai-se proceder à obtenção de uma partição em sete grupos de acordo com o critério descrito à semelhança do que fará para Volume de Negócios, Emprego e Horas trabalhadas nos serviços. Essas partições serão comparadas no ponto seguinte, tendo especial atenção às partições «associadas a Remunerações».

5.5 Índice de Rand e outros Índices de Comparação Aplicados à Comparação de Partições

São agora apresentados os cálculos analíticos para índice de Rand através das equação (4.14) e intervalo de rejeição através da equação (4.19) para cada uma das variáveis. No final será apresentada uma tabela sucinta.

Tabela 5.3 Comparação de Partições Através do Índice de Rand

	Índice Rand	Intervalo Rejeição	Conclusão
Volume	0.7350]0.6262,0.6961[Partições Correlacionadas
Emprego	0.6724]0.4241,0.6338[Partições Correlacionadas
Remunerações	0.7236]0.6017,0.6908[Partições Correlacionadas
Horas	0.6667]0.5076,0.6396[Partições Correlacionadas

Tabela 5.4 Comparação de Partições Através do Índice PE

	PE	Intervalo Rejeição	Conclusão
Volume	0.1923] - 0.3516, -0.1698[Partições Correlacionadas
Emprego	0.3254] - 0.6817, -0.2057[Partições Correlacionadas
Remunerações	0.2124] - 0.4402, -0.1952[Partições Correlacionadas
Horas	0.2598] - 0.6253, -0.2566[Partições Correlacionadas

Tabela 5.5: Comparação de Partições Através do Índice k

	K	Intervalo Rejeição	Conclusão
Volume	0.2180] - 0.1030, 0.1030[Partições Correlacionadas
Emprego	0.3044] - 0.2225, 0.2225[Partições Correlacionadas
Remunerações	0.2186] - 0.1260, 0.1260[Partições Correlacionadas
Horas	0.2182] - 0.1548, 0.1548[Partições Correlacionadas

Note-se que os resultados obtidos em relação a Remunerações e os índices utilizados podem questionar-se em relação às estruturas obtidas pelo método de ACHA adoptado no trabalho (como já se referiu). Deve no entanto ter-se em consideração que o método de comparação das estruturas hierárquicas é um «método global».

Tabela 5.6 Comparação de Partições Através do Índice de Rand Ajustado por Hubert e Arabie

	Índice Rand Ajustado	Valores de Referência Stanley(2004)	Conclusões
Volume	0.218	<0.65	recuperação de estrutura pobre
Emprego	0.6724	>0.65	recuperação de estrutura moderada
Remunerações	0.2182	<0.65	recuperação de estrutura pobre
Horas	0.2187	<0.65	recuperação de estrutura pobre

Conclusões da tabela 5.6 foram obtidas de acordo com Stenley(2004).

Pode-se concluir que, de acordo com o conjunto de índices utilizados, de uma forma geral, os resultados mais coerentes são obtidos na tabela 5.6 através do índice de Rand Ajustado ARI_{HA} . O que corrobora a teoria dos autores que defendem ser este o mais adequado para comparação de partições nestas condições. Esta indica-nos que «globalmente» as partições não estão correlacionadas. As actividades entre 2009 e 2010 mudam substancialmente de grupo, ou por outra forma, as actividades apresentadas em 2009 tendem a não manter-se nos mesmo grupos quanto comparando com ano de 2010, segundo o mesmo nível de semelhança. Tal facto vem reforçar a opinião da maioria dos autores que no conjunto

de índices de semelhança defendem a utilização do índice de Rand ajustado. No entanto do ponto de vista da classificação das actividades económicas faria mais sentido o resultado obtido pelos três índices da família L utilizados.

5.6 Demonstração dos Resultados

Neste subcapítulo após escolha do método Vizinho mais afastado, onde se viu através do ponto 5.2 ter sido o método de ACHA mais adequado, e também pelo ponto 5.3 onde foi determinado que seriam sete o número de grupos a utilizar, vai ser mostrada a formação dos grupos sob forma de dendrograma e sob forma de tabela. Estes terão respectiva cor para cada um dos grupos, tratando-se da mesma, tanto no dendrograma como na tabela. A demonstração será por índice, com a seguinte ordem: Volume Negócios, Emprego, Remunerações, Horas trabalhadas nos serviços. Em cada um dos índices a apresentação para o ano de 2009 e 2010 permitirá verificar e visualizar de que forma as estruturas dos grupos são alteradas.

De forma a completar a informação dada pelos dendrogramas, sendo que esta está representada sob diagrama de árvore contendo os grupos separados com cores consoante o seu nível de semelhança, também irão ser apresentadas tabelas com mesma informação que estes, mas com a particularidade de conter a descrição de cada uma das actividades que estarão representadas com a respectiva correspondência às secções.

Desta forma pretende-se visualizar do ponto de vista das actividades, a constituição dos grupos após a aplicação do método de particionamento. Estes são formados por actividades de secções distintas. Tal factor vai permitir retirar algumas conclusões pois consegue-se ter uma melhor observação de que forma as actividades com a correspondente secção podem ou não manter-se no mesmo grupo entre 2009 e 2010. As cores entre os grupos nos dendrogramas, sendo as mesmas nos grupos das tabelas, facilitam a observação e comparação da informação demonstrada.

5.6.1 Volume Negócios

Figura 5.4 Dendrograma de Volume Negócios com o Método do Vizinho Mais Afastado no Ano 2009

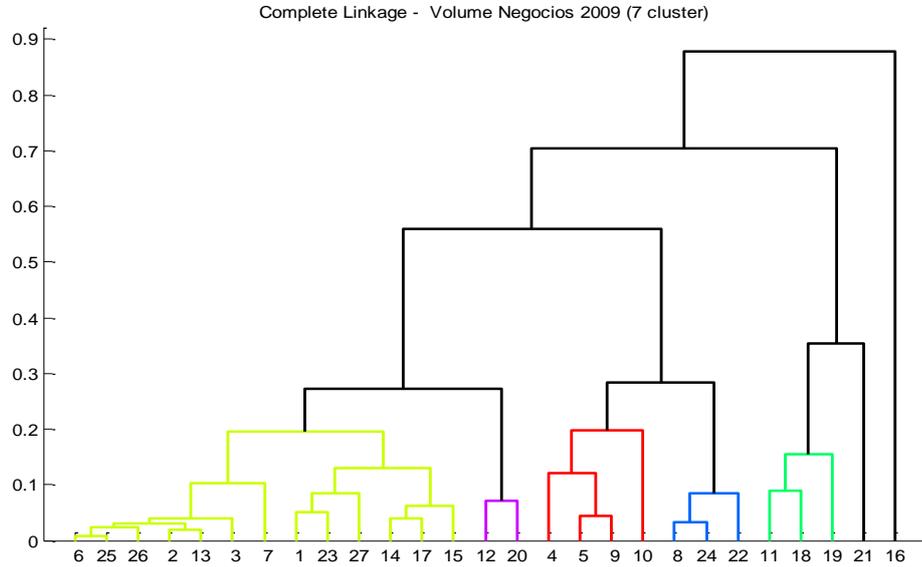


Figura 5.5 Dendrograma de Volume Negócios com o Método do Vizinho Mais Afastado no Ano 2010

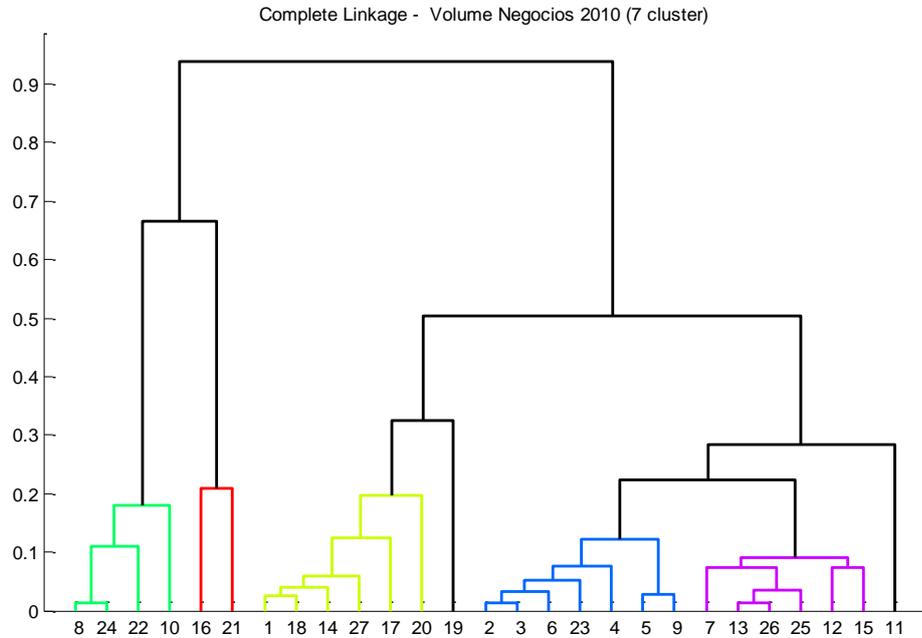


Tabela 5.7 Grupo por Volume de Negócios

	2009	2010
G-Comércio por grosso; reparação de veículos automóveis e motociclos	45-Comércio, manutenção e reparação de veículos automóveis e motociclos 46-Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos	45-Comércio, manutenção e reparação de veículos automóveis e motociclos 46-Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos
H-Transportes armazenagem	49-Transportes terrestres, transportes por oleodutos e gasodutos (pipe-lines) 50-Transportes por água 51-Transportes aéreos 52-Armacenagem e actividades auxiliares dos transportes (inclui manuseamento) 53-Actividades postais e de courier	49-Transportes terrestres, transportes por oleodutos e gasodutos (pipe-lines) 50-Transportes por água 51-Transportes aéreos 52-Armacenagem e actividades auxiliares dos transportes (inclui manuseamento) 53-Actividades postais e de courier
I- Alojamento, restauração e similares	55-Alojamento 56-Restauração e Similares	55-Alojamento 56-Restauração e Similares
J-Informação comunicação	58-Actividades de Edição 59-Actividades cinematográficas, de vídeo, de produção de programas de televisão 60-Actividades de rádio e televisão 61-Telecomunicações 62-Consultoria e programação informática e actividades relacionadas 63-Actividades dos serviços de informação	58-Actividades de Edição 59-Actividades cinematográficas, de vídeo, de produção de programas de televisão 60-Actividades de rádio e televisão 61-Telecomunicações 62-Consultoria e programação informática e actividades relacionadas 63-Actividades dos serviços de informação
L-Imobiliárias	68-Actividades imobiliárias	68-Actividades imobiliárias
M-Consultoria, científicas, técnicas e similares	69-Actividades jurídicas e de contabilidade e actividades de consultadoria para os negócios e gestão 71-Actividades de arquitectura, de engenharia e técnicas afins , actividades de ensaios e de análises técnicas 72-Actividades de investigação científica e de desenvolvimento 73-Publicidade, estudos de mercado e sondagens de opinião 74-Outras actividades de consultadoria, científicas, técnicas e similares	69-Actividades jurídicas e de contabilidade e actividades de consultadoria para os negócios e gestão 71-Actividades de arquitectura, de engenharia e técnicas afins , actividades de ensaios e de análises técnicas 72-Actividades de investigação científica e de desenvolvimento 73-Publicidade, estudos de mercado e sondagens de opinião 74-Outras actividades de consultadoria, científicas, técnicas e similares
N-Administrativas e dos serviços de apoio	78-Actividades de emprego 79-Agencias de viagem, operadores turísticos, outros serviços de reservas e actividades relacionadas 80-Actividades de investigação e segurança 77-Actividades de Aluguer 82-Actividades de serviços administrativos e de apoio prestados às empresas 81-Actividades relacionadas com edifícios, plantação e manutenção de jardins	78-Actividades de emprego 79-Agencias de viagem, operadores turísticos, outros serviços de reservas e actividades relacionadas 80-Actividades de investigação e segurança 77-Actividades de Aluguer 82-Actividades de serviços administrativos e de apoio prestados às empresas 81-Actividades relacionadas com edifícios, plantação e manutenção de jardins

Quanto ao Volume de Negócios, segundo o resultado obtido com o coeficiente de correlação de Spearman, na tabela 5.2 tem-se para estruturas apresentadas como sendo estruturas hierárquicas semelhantes. Porém verifica-se que se obteve uma mudança expressiva das variáveis entre grupos, do ponto de vista da visualização dos dois dendrogramas.

Pode-se concluir que existiu uma alteração significativa das actividades por grupo entre 2009 e 2010. Esta conclusão é justificada pela «recuperação pobre da estrutura» dos grupos conforme o índice de Rand Ajustado mostra através da tabela 5.5. Apesar de, mesmo após o corte dos dendrogramas, os índices da família L utilizados não corroborarem esta conclusão. Para além desta justificação, esta conclusão pode também ser obtida através da visualização da tabela atrás, ou seja, 5.5. Através desta pode-se verificar que para 2009 existe a formação de um grande grupo (verde lima) composta por actividades de todas as secções.

No entanto para 2010 pode-se verificar que esse grupo foi alterado e subdividido por 3 grupos médios, isto é, contendo cerca de seis actividades cada um.

Conclui-se portanto que em 2009 cerca de metade das actividades se agruparam num só grupo (verde lima) e em 2010 foram formados três com dimensão média conforme referido no parágrafo anterior

Por um lado o grande grupo formado no ano de 2009 contém actividades de todas as secções, mas para 2010 tem-se para os três grupos formados:

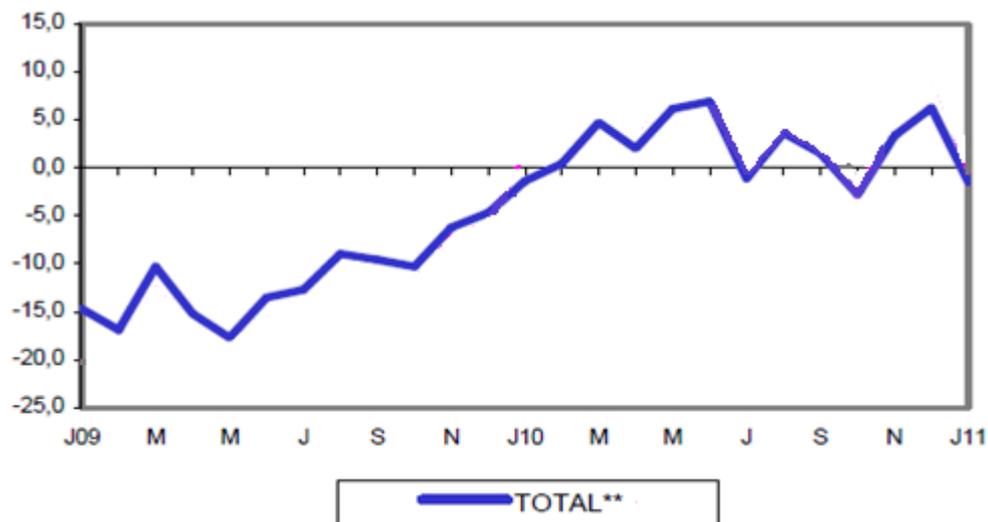
- Azul, tem actividades essencialmente das secções de comércio por grosso; reparação de veículos automóveis e motociclos; transportes e armazenagem; alojamento; restauração e similares.

- Violeta, contém essencialmente actividades das secções de informação comunicação; imobiliárias; consultoria; científicas; técnicas e similares; administrativas e dos serviços de apoio.

- Verde, contém actividades das secções de informação comunicação; consultoria científica técnica e similares; administrativas e serviços de apoio.

Comparando com estudo efectuado pelo INE (destaque de informação à comunicação social²⁶, de estatística descritiva) pode-se retirar que uma das razões principais para a alteração das actividades dos respectivos grupos se verifica através gráfico 5.7, isto é, pela variação do índice TOTAL do volume de negócios entre 2009 e 2010. Este índice em 2010 teve valores superiores aos de 2009. Conforme se pode verificar pelo (destaque de informação à comunicação social do INE²⁷) “O índice de volume de negócios apresentou uma variação homóloga de -1.5% em Janeiro, o que compara com o aumento de 6.2% observado em Dezembro.”(destaque de informação à comunicação social do INE, 2011)

Figura 5.6 Volume de Negócios – Índice Total Variações Homólogas, %



TOTAL**²⁸

Fonte: INE, Janeiro de 2011. Destaque de informação à comunicação social – Índices de volume de negócios, emprego, remunerações e horas trabalhadas nos serviços

No capítulo 2, sobre conjuntura nacional e internacional, foram apresentadas as justificações económicas que induziram a estas variações entre 2009 e 2010, pelo comportamento dos indicadores económicos para o período estudado.

²⁶ INE, Janeiro de 2011. Destaque de informação à comunicação social – Índices de volume de negócios, emprego, remunerações e horas trabalhadas nos serviços

²⁷ INE, Janeiro de 2011. Destaque de informação à comunicação social – Índices de volume de negócios, emprego, remunerações e horas trabalhadas nos serviços

²⁸ Total de todas as secções

5.6.2 Emprego

Figura 5.7 Dendrograma de Emprego com Método do Vizinho Mais Afastado no Ano 2009

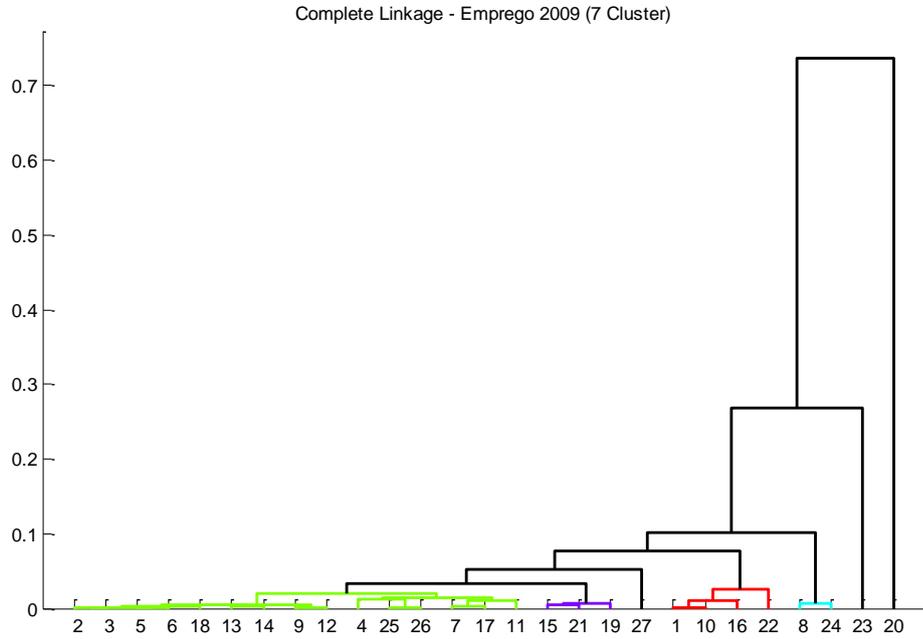


Figura 5.8 Dendrograma de Emprego com Método do Vizinho Mais Afastado no Ano 2010

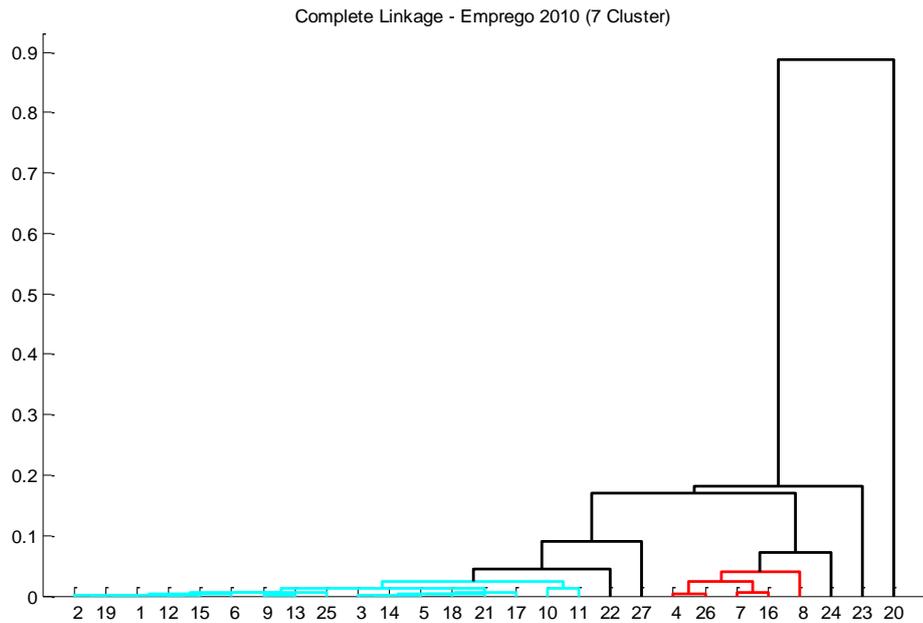


Tabela 5.8 Grupo por Emprego

	2009	2010
G-Comércio por grosso; reparação de veículos automóveis e motociclos	45-Comércio, manutenção e reparação de veículos automóveis e motociclos 46-Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos	45-Comércio, manutenção e reparação de veículos automóveis e motociclos 46-Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos
H-Transportes armazenagem	49-Transportes terrestres, transportes por oleodutos e gasodutos (pipe-lines) 50-Transportes por água 51-Transportes aéreos 52-Armazenagem e actividades auxiliares dos transportes (inclui manuseamento) 53-Actividades postais e de courier	49-Transportes terrestres, transportes por oleodutos e gasodutos (pipe-lines) 50-Transportes por água 51-Transportes aéreos 52-Armazenagem e actividades auxiliares dos transportes (inclui manuseamento) 53-Actividades postais e de courier
I- Alojamento, restauração e similares	55-Alojamento 56-Restauração e Similares	55-Alojamento 56-Restauração e Similares
J-Informação comunicação	58-Actividades de Edição 59-Actividades cinematográficas, de vídeo, de produção de programas de televisão 60-Actividades de rádio e televisão 61-Telecomunicações 62-Consultoria e programação informática e actividades relacionadas 63-Actividades dos serviços de informação	58-Actividades de Edição 59-Actividades cinematográficas, de vídeo, de produção de programas de televisão 60-Actividades de rádio e televisão 61-Telecomunicações 62-Consultoria e programação informática e actividades relacionadas 63-Actividades dos serviços de informação
L-Imobiliárias	68-Actividades imobiliárias	68-Actividades imobiliárias
M-Consultoria, científicas, técnicas e similares	69-Actividades jurídicas e de contabilidade e actividades de consultadoria para os negócios e gestão 71-Actividades de arquitectura, de engenharia e técnicas afins, actividades de ensaios e de análises técnicas 72-Actividades de investigação científica e de desenvolvimento 73-Publicidade, estudos de mercado e sondagens de opinião 74-Outras actividades de consultadoria, científicas, técnicas e similares	69-Actividades jurídicas e de contabilidade e actividades de consultadoria para os negócios e gestão 71-Actividades de arquitectura, de engenharia e técnicas afins, actividades de ensaios e de análises técnicas 72-Actividades de investigação científica e de desenvolvimento 73-Publicidade, estudos de mercado e sondagens de opinião 74-Outras actividades de consultadoria, científicas, técnicas e similares
N-Administrativas e dos serviços de apoio	78-Actividades de emprego 79-Agencias de viagem, operadores turísticos, outros serviços de reservas e actividades relacionadas 80-Actividades de investigação e segurança 77-Actividades de Aluguer 82-Actividades de serviços administrativos e de apoio prestados às empresas 81-Actividades relacionadas com edifícios, plantação e manutenção de jardins	78-Actividades de emprego 79-Agencias de viagem, operadores turísticos, outros serviços de reservas e actividades relacionadas 80-Actividades de investigação e segurança 77-Actividades de Aluguer 82-Actividades de serviços administrativos e de apoio prestados às empresas 81-Actividades relacionadas com edifícios, plantação e manutenção de jardins

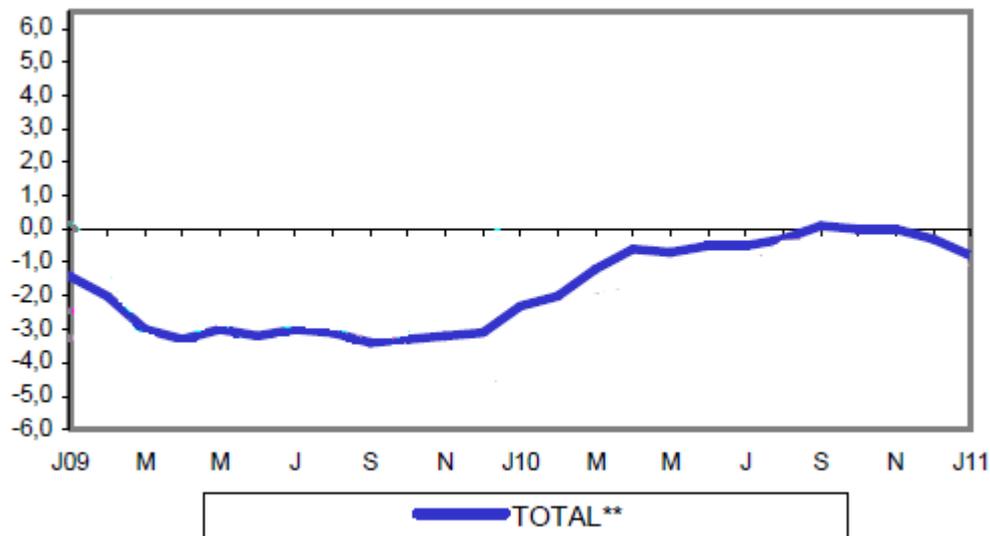
Quanto ao Emprego, segundo o resultado obtido com o coeficiente de correlação de Spearman pela tabela 5.2, tem-se para as estruturas apresentadas como sendo estruturas hierárquicas semelhantes. Conclui-se que existiu uma alteração das actividades por grupo entre 2009 e 2010. Esta conclusão é justificada pelos resultados obtidos através dos índices de Rand, PE e κ e também pelo índice de Rand Ajustado conforme tabela 5.6. «recuperação de estrutura moderada»

Destaca-se a formação, em 2010, de várias variáveis separadas unitariamente representadas na tabela atrás a branco e no dendrograma a preto, enquanto que em 2009 apenas três actividades se encontram distintamente separadas sendo essas secções: administrativas e dos serviços de apoio e consultoria científicas, técnicas e similares.

Em 2009 existe um grande grupo com cerca de quinze actividades (Verde), sendo estas de todas as secções. Comparativamente com 2010 esse grande grupo não é muito alterado, conforme se visualiza melhor na tabela 5.6. No entanto a esse grupo em 2010 são acrescentadas algumas actividades, nomeadamente, das secções de informação comunicação e consultoria científicas, técnicas e similares. Desta forma em 2010 para além das actividades isoladas já demonstrado (branco na tabela e preto no dendrograma) e do grande grupo (azul) surge um pequeno grupo (vermelho) contendo praticamente uma actividade de cada secção.

Comparando com estudo efectuado pelo INE (destaque de informação à comunicação social²⁹, de estatística descritiva) pode-se retirar que uma das razões principais para a «movimentação» das actividades dos respectivos grupos, se verifica através gráfico 5.11, isto é, pela variação do índice TOTAL do volume de negócios entre 2009 e 2010. Este índice em 2010 teve um lento aumento mas gradual até que no final de 2010. Nota-se a maior diferença em termos homólogos do ano inteiro. Conforme se pode verificar pelo destaque de informação à comunicação social do INE³⁰ “O índice de Emprego diminuiu 0.8% em Janeiro de 2011 comparando com Janeiro de 2010, existe também uma variação de -0.3% em Dezembro” (destaque de informação à comunicação social do INE, 2011).

Figura 5.9 Emprego – Índice Total Variações Homólogas, %



TOTAL**³¹

Fonte: INE, Janeiro de 2011. Destaque de informação à comunicação social – Índices de volume de negócios, emprego, remunerações e horas trabalhadas nos serviços

²⁹ INE, Janeiro de 2011. Destaque de informação à comunicação social – Índices de volume de negócios, emprego, remunerações e horas trabalhadas nos serviços

³¹ Total de todas as secções

No capítulo 2, sobre conjuntura nacional e internacional, foram apresentadas as justificações económicas que induziram a estas variações entre 2009 e 2010, pelo comportamento dos indicadores económicos para o período estudado.

5.6.3 Remunerações

Figura 5.10 Dendrograma de Remunerações com o Método do Vizinho Mais Afastado no Ano 2009

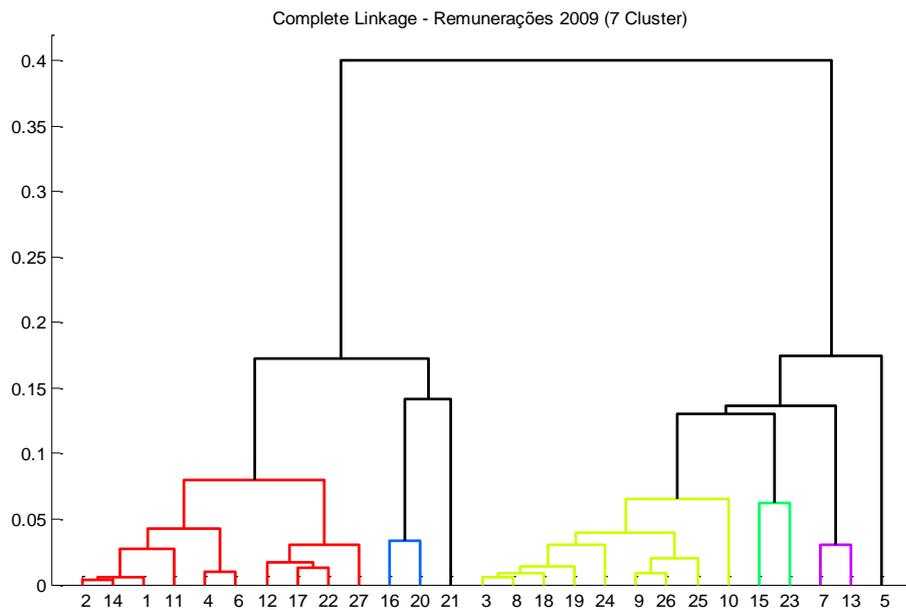


Figura 5.11 Dendrograma de Remunerações com o Método do Vizinho Mais Afastado no Ano 2010

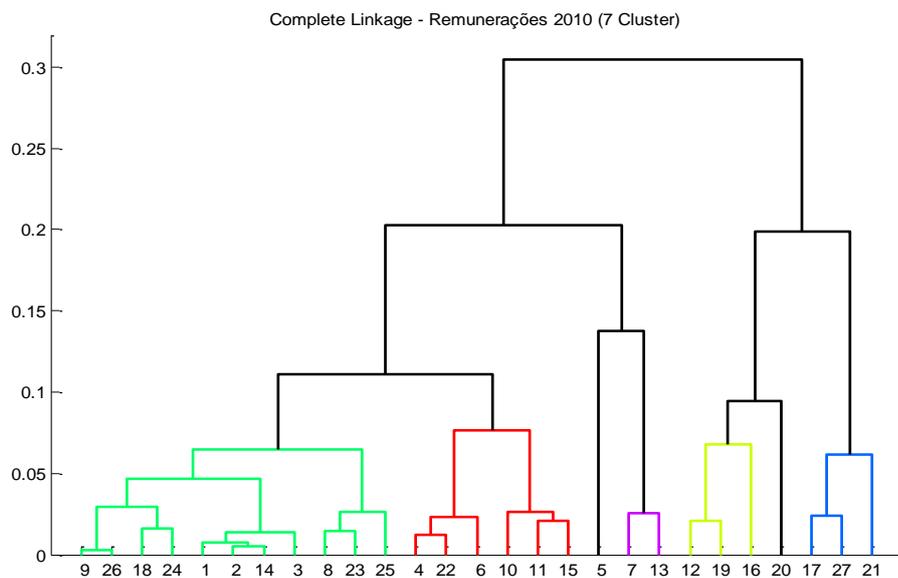


Tabela 5.9 Grupo por Remunerações

	2009	2010
G-Comércio por grosso; reparação de veículos automóveis e motociclos	45-Comércio, manutenção e reparação de veículos automóveis e motociclos	45-Comércio, manutenção e reparação de veículos automóveis e motociclos
	46-Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos	46-Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos
H-Transportes armazenagem	49-Transportes terrestres, transportes por oleodutos e gasodutos (pipe-lines)	49-Transportes terrestres, transportes por oleodutos e gasodutos (pipe-lines)
	50-Transportes por água	50-Transportes por água
	51-Transportes aéreos	51-Transportes aéreos
	52-Armacenagem e actividades auxiliares dos transportes (inclui manuseamento)	52-Armacenagem e actividades auxiliares dos transportes (inclui manuseamento)
	53-Actividades postais e de courier	53-Actividades postais e de courier
I- Alojamento, restauração e similares	55-Alojamento	55-Alojamento
	56-Restauração e Similares	56-Restauração e Similares
J-Informação comunicação	58-Actividades de Edição	58-Actividades de Edição
	59-Actividades cinematográficas, de vídeo, de produção de programas de televisão	59-Actividades cinematográficas, de vídeo, de produção de programas de televisão
	60-Actividades de rádio e televisão	60-Actividades de rádio e televisão
	61-Telecomunicações	61-Telecomunicações
	62-Consultoria e programação informática e actividades relacionadas	62-Consultoria e programação informática e actividades relacionadas
	63-Actividades dos serviços de informação	63-Actividades dos serviços de informação
L-Imobiliárias	68-Actividades imobiliárias	68-Actividades imobiliárias
M-Consultoria, científicas, técnicas e similares	69-Actividades jurídicas e de contabilidade e actividades de consultoria para os negócios e gestão	69-Actividades jurídicas e de contabilidade e actividades de consultoria para os negócios e gestão
	71-Actividades de arquitectura, de engenharia e técnicas afins, actividades de ensaios e de análises técnicas	71-Actividades de arquitectura, de engenharia e técnicas afins, actividades de ensaios e de análises técnicas
	72-Actividades de investigação científica e de desenvolvimento	72-Actividades de investigação científica e de desenvolvimento
	73-Publicidade, estudos de mercado e sondagens de opinião	73-Publicidade, estudos de mercado e sondagens de opinião
	74-Outras actividades de consultoria, científicas, técnicas e similares	74-Outras actividades de consultoria, científicas, técnicas e similares
N-Administrativas e dos serviços de apoio	78-Actividades de emprego	78-Actividades de emprego
	79-Agencias de viagem, operadores turísticos, outros serviços de reservas e actividades relacionadas	79-Agencias de viagem, operadores turísticos, outros serviços de reservas e actividades relacionadas
	80-Actividades de investigação e segurança	80-Actividades de investigação e segurança
	77-Actividades de Aluguer	77-Actividades de Aluguer
	82-Actividades de serviços administrativos e de apoio prestados às empresas	82-Actividades de serviços administrativos e de apoio prestados às empresas
	81-Actividades relacionadas com edifícios, plantação e manutenção de jardins	81-Actividades relacionadas com edifícios, plantação e manutenção de jardins

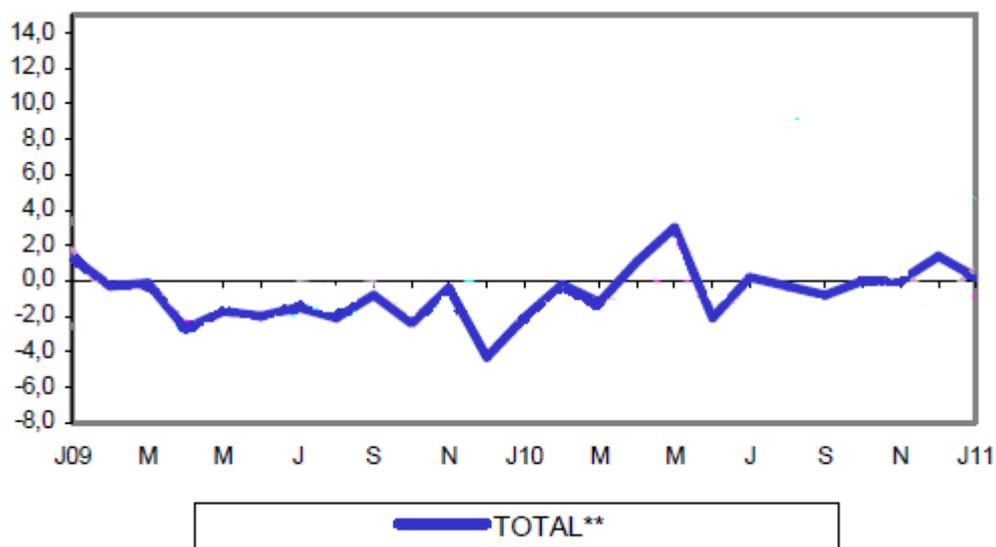
Quanto ao índice Remunerações segundo o resultado obtido com o coeficiente de correlação de Spearman pela tabela 5.2, tem-se para as estruturas apresentadas como sendo não semelhantes – as duas estruturas de classificação hierárquicas não são correlacionadas. Como já foi referido, utilizou-se o método de particionamento tendo em conta os índices da família L utilizados, que fornecem os mesmos resultados que nos casos anteriores. No entanto o índice de Rand Ajustado indica que a recuperação da estrutura é pobre. As estruturas apresentam uma boa diferenciação em relação à formação dos grupos das actividades. Pode-se concluir que existiu uma alteração significativa das actividades por grupo entre 2009 e 2010.

Estas estruturas diferem das apresentadas dos outros índices porque não têm um grande grupo de actividades formado. Tem de outra forma as actividades distribuídas pelos grupos de uma forma mais uniforme, no entanto estas actividades mudam bastante de grupo comparando os anos 2009 e 2010, apesar dos grupos terem aproximadamente o mesmo número de actividades entre os anos referidos. Através da tabela 5.7, atrás mostrada pode-se de uma forma mais clara verificar o quanto as actividades mudam de grupo.

Conclui-se portanto que o índice Remunerações foi o que mais sofreu alterações quanto ao nível de semelhança entre as actividades entre 2009 e 2010, não deixando de ter sempre em atenção que foi a estrutura hierárquica que demonstrou à partida não ter estrutura semelhante entre os dois anos.

Comparando com estudo efectuado pelo INE (destaque de informação à comunicação social³²), de estatística descritiva pode-se retirar que uma das razões principais para a alteração das actividades dos respectivos grupos se verifica através gráfico 5.15, isto é, pela variação do índice TOTAL do volume de negócios entre 2009 e 2010. Este índice no primeiro semestre de cada ano tem uma diferença maior pelas subidas abruptas do início de 2010. Apenas no segundo semestre de cada ano existe uma variação menor. Conforme se pode verificar pelo (destaque de informação à comunicação social do INE³³) “Em Janeiro de 2011, as remunerações aumentaram 0.2% em termos homólogos, tendo uma variação de 1.4% em Dezembro.” (destaque de informação à comunicação social do INE, 2011).

Figura 5.12 Remunerações – Índice Total e Secções Variações Homólogas, %



TOTAL**³⁴

Fonte: INE, Janeiro de 2011. Destaque de informação à comunicação social – Índices de volume de negócios, emprego, remunerações e horas trabalhadas nos serviços

No capítulo 2, sobre conjuntura nacional e internacional, foram apresentadas as justificações económicas que induziram a estas variações entre 2009 e 2010, pelo comportamento dos indicadores económicos para o período estudado.

³² INE, Janeiro de 2011. Destaque de informação à comunicação social – Índices de volume de negócios, emprego, remunerações e horas trabalhadas nos serviços

³⁴ Total de todas as secções

5.6.4 Horas trabalhadas nos serviços

Figura 5.13 Dendrograma de Horas com o Método do Vizinho Mais Afastado no Ano 2009

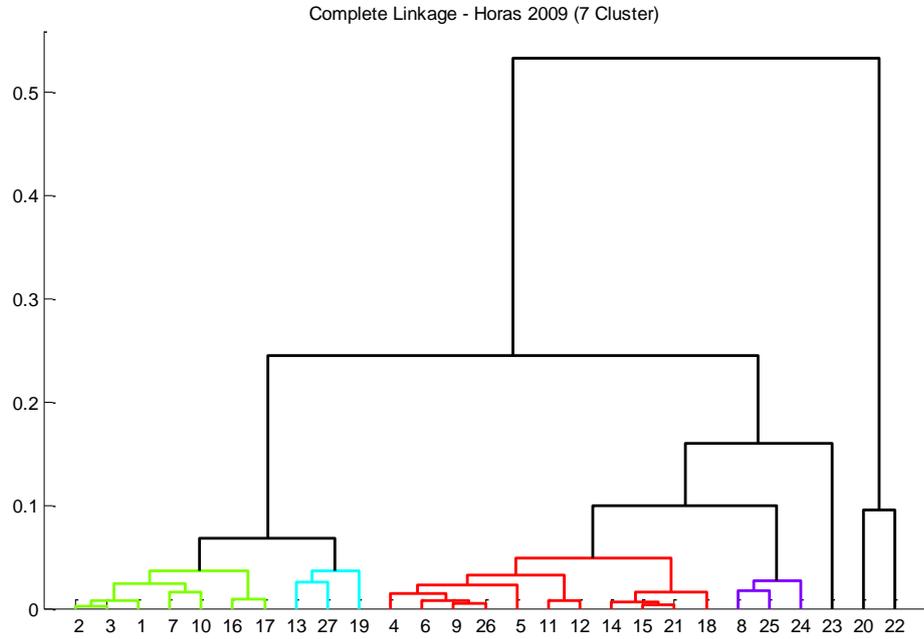


Figura 5.14 Dendrograma de Horas com o Método do Vizinho Mais Afastado no Ano 2010

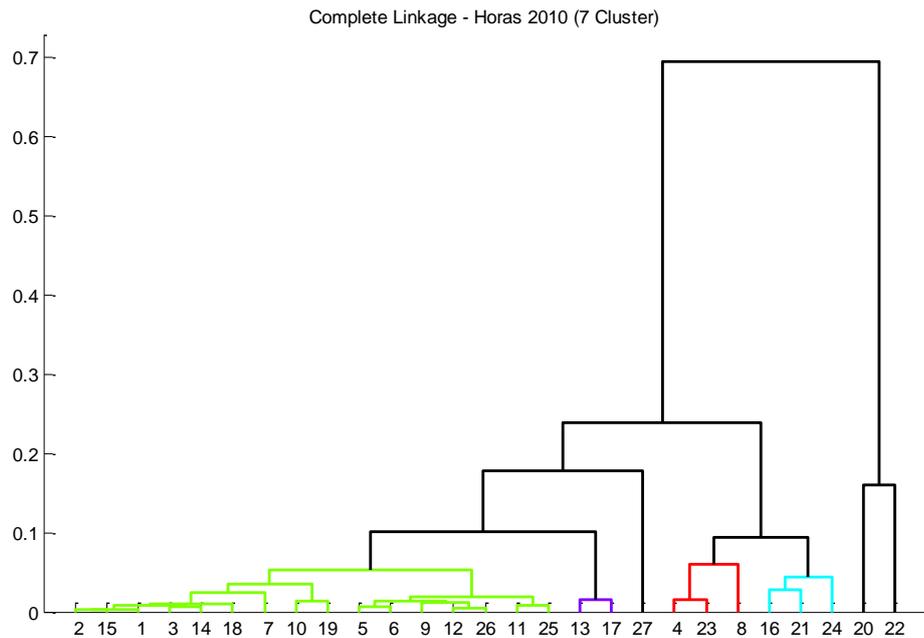


Tabela 5.10 Grupo por Horas trabalhadas nos serviços

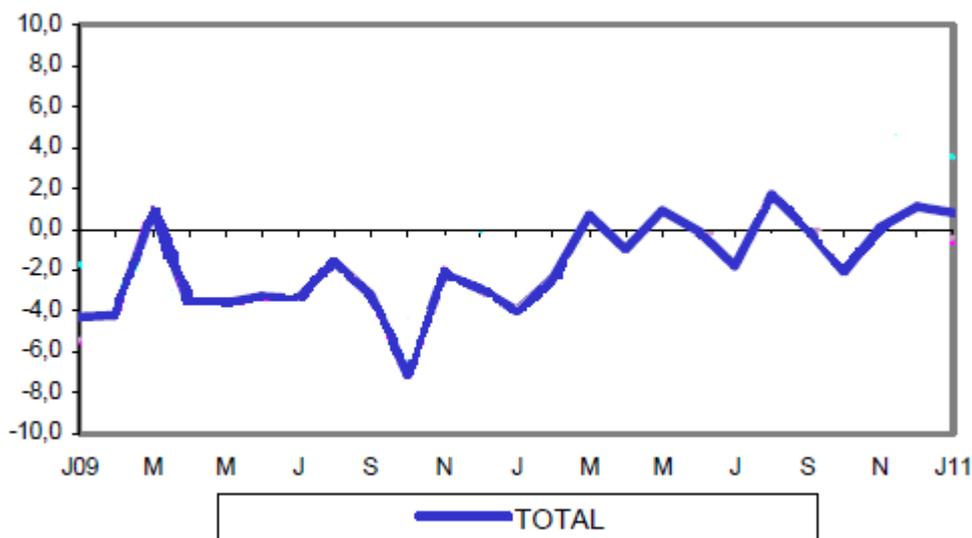
	2009	2010
G-Comércio por grosso; reparação de veículos automóveis e motociclos	45-Comércio, manutenção e reparação de veículos automóveis e motociclos 46-Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos	45-Comércio, manutenção e reparação de veículos automóveis e motociclos 46-Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos
H-Transportes armazenagem	49-Transportes terrestres, transportes por oleodutos e gasodutos (pipe-lines) 50-Transportes por água 51-Transportes aéreos 52-Armazenagem e actividades auxiliares dos transportes (inclui manuseamento) 53-Actividades postais e de courier	49-Transportes terrestres, transportes por oleodutos e gasodutos (pipe-lines) 50-Transportes por água 51-Transportes aéreos 52-Armazenagem e actividades auxiliares dos transportes (inclui manuseamento) 53-Actividades postais e de courier
I- Alojamento, restauração e similares	55-Alojamento 56-Restauração e Similares	55-Alojamento 56-Restauração e Similares
J-Informação comunicação	58-Actividades de Edição 59-Actividades cinematográficas, de vídeo, de produção de programas de televisão 60-Actividades de rádio e televisão 61-Telecomunicações 62-Consultoria e programação informática e actividades relacionadas 63-Actividades dos serviços de informação	58-Actividades de Edição 59-Actividades cinematográficas, de vídeo, de produção de programas de televisão 60-Actividades de rádio e televisão 61-Telecomunicações 62-Consultoria e programação informática e actividades relacionadas 63-Actividades dos serviços de informação
L-Imobiliárias	68-Actividades imobiliárias	68-Actividades imobiliárias
M-Consultoria, científicas, técnicas e similares	69-Actividades jurídicas e de contabilidade e actividades de consultadoria para os negócios e gestão 71-Actividades de arquitectura, de engenharia e técnicas afins, actividades de ensaios e de análises técnicas 72-Actividades de investigação científica e de desenvolvimento 73-Publicidade, estudos de mercado e sondagens de opinião 74-Outras actividades de consultadoria, científicas, técnicas e similares	69-Actividades jurídicas e de contabilidade e actividades de consultadoria para os negócios e gestão 71-Actividades de arquitectura, de engenharia e técnicas afins, actividades de ensaios e de análises técnicas 72-Actividades de investigação científica e de desenvolvimento 73-Publicidade, estudos de mercado e sondagens de opinião 74-Outras actividades de consultadoria, científicas, técnicas e similares
N-Administrativas e dos serviços de apoio	78-Actividades de emprego 79-Agencias de viagem, operadores turísticos, outros serviços de reservas e actividades relacionadas 80-Actividades de investigação e segurança 77-Actividades de Aluguer 82-Actividades de serviços administrativos e de apoio prestados às empresas 81-Actividades relacionadas com edifícios, plantação e manutenção de jardins	78-Actividades de emprego 79-Agencias de viagem, operadores turísticos, outros serviços de reservas e actividades relacionadas 80-Actividades de investigação e segurança 77-Actividades de Aluguer 82-Actividades de serviços administrativos e de apoio prestados às empresas 81-Actividades relacionadas com edifícios, plantação e manutenção de jardins

Quanto ao índice de Horas trabalhadas nos serviços, segundo o resultado obtido com o coeficiente de correlação de Spearman pela tabela 5.2 tem-se para as estruturas apresentadas como sendo estruturas correlacionadas. Pode-se concluir que existiu uma alteração significativa das actividades por grupo entre 2009 e 2010. Esta conclusão é justificada tal como nos três casos anteriores não só pelos índices da família L, mas também pelo resultado obtido com o índice de Rand Ajustado de acordo com a tabela 5.6.

As estruturas estão representadas de forma clara e distinta como se pode ver nos últimos dendrogramas, 5.14 e 5.15. Como facto mais significativo, tem-se para o índice de Horas trabalhadas nos serviços em 2010, a formação de um grande grupo de actividades que abrangem todas as secções (verde lima). Tal resultado neste caso observa-se melhor através dos dendrogramas. Para o ano 2009 as actividades estão mais divididas pelos grupos, porém os dois grupos maiores em 2009 contêm actividades de todos os ramos/secções (verde lima e vermelho). Este último resultado verifica-se de uma forma mais concreta pela tabela 5.8, pois permite verificar que efectivamente os grupos médios contêm actividades de praticamente todas as secções.

Comparando com estudo efectuado pelo INE (destaque de informação à comunicação social³⁵), de estatística descritiva pode-se retirar que uma das razões principais para a alteração das actividades dos respectivos grupos se verifica através do gráfico 5.16, isto é, pela variação do índice TOTAL do volume de negócios entre 2009 e 2010. Este índice apenas tem um pico semelhante na transição do 1º para o 2º trimestre de cada ano. Tem-se, para o final de 2009, devido à descida apresentada, a maior diferença relativamente ao mesmo período de 2010. Nos restantes períodos no ano de 2010 existiu sempre um valor relativamente maior comparando com período homólogo. Conforme se pode verificar pelo (destaque de informação à comunicação social do INE³⁶) “Em Janeiro de 2011, face a igual período de 2010, o volume de trabalho medido em horas, registou um aumento de 0.8% , 1,1 relativamente ao final de 2010”.(destaque de informação à comunicação social do INE, 2011)

Figura 5.15 Horas – Índice Total e Secções Variações Homólogas, %



TOTAL**³⁷

Fonte: INE, Janeiro de 2011. Destaque de informação à comunicação social – Índices de volume de negócios, emprego, remunerações e horas trabalhadas nos serviços

No capítulo 2, sobre conjuntura nacional e internacional, foram apresentadas as justificações económicas que induziram a estas variações entre 2009 e 2010, pelo comportamento dos indicadores económicos para o período estudado.

³⁵ INE, Janeiro de 2011. Destaque de informação à comunicação social – Índices de volume de negócios, emprego, remunerações e horas trabalhadas nos serviços

³⁷ Total de todas as secções

Conclusão e Projecto Futuro

Conclusão

Os indicadores evoluem de acordo com as actividades económicas que se desenvolvem no país. Assim progredem ou regredem em função do desempenho dos sectores de actividades económicas. Mediante essas evoluções, foi cumprido neste estudo, o objectivo de verificar de que forma as actividades económicas portuguesas são agrupadas entre si tendo em conta o estado da evolução económica nacional entre os anos de 2009 e 2010. No capítulo 2 sobre conjuntura nacional e internacional foram apresentadas as justificações económicas que induziram estas variações entre 2009 e 2010, pelo comportamento dos indicadores económicos para o período estudado. Para além das diferentes áreas onde estão inseridos os vários sectores económicos, estudou-se o agrupamento segundo quatro índices: Volume de Negócios, Emprego, Remunerações, Horas trabalhadas nos serviços.

Utilizando como base metodologias de Análise Multivariada de dados ACHA e «partições» foi possível, com os dados fornecidos pelo INE, estudar de que forma se agrupam as actividades económicas de secções diferentes nos anos 2009 e 2010. Na Classificação Hierárquica utilizou-se como coeficiente de dissemelhança r^* como indicado no subcapítulo 4.4.2. Após testes verificou-se que o Vizinho mais afastado foi o método que permitiu obter as melhores conclusões. Mediante a sensibilidade no tema trabalhado, para determinação do número de grupos foi escolhido o método de tabelas de frequências acumuladas mas também tendo em conta os procedimentos heurísticos. Desta forma foram determinados sete grupos a comparar.

A representação dos resultados foi diversificada entre dendrogramas, isto é, estruturas em árvores e também através de tabelas. A apresentação dos grupos é distinguida por cores, sendo que as actividades tanto no dendrograma como na tabela têm mesma cor correspondente. Desta forma é mais fácil a interpretação dos resultados. A análise dos resultados é efectuada através do coeficiente de correlação de Spearman, do índice de Rand, índice de Perce, índice de Cohen e índice de Rand Ajustado segundo Hubert e Arabie.

Concluindo os índices, segundo os resultados obtidos através do coeficiente de correlação de Spearman estão representados sob a forma de estruturas hierárquicas semelhantes, excepto para o índice Remunerações.

Quanto ao Volume de Negócios pode-se concluir que existiu uma alteração significativa das actividades por grupo entre 2009 e 2010. Esta situação é justificada pela «recuperação de estrutura» conforme o índice de Rand Ajustado. Para o ano de 2009 cerca de metade das actividades agruparam-se num só grupo. Já em 2010 foram formados 3 grupos consideráveis. Um tem actividades essencialmente das secções de: comércio por grosso; reparação de veículos automóveis e motociclos; transportes; armazenagem; alojamento; restauração e similares. Outro com actividades essencialmente das secções de: informação comunicação; imobiliárias; consultoria; científicas; técnicas e similares; administrativas e dos serviços de apoio. Por fim um terceiro grupo que contem poucas, no entanto actividades de praticamente todas as secções.

Quanto ao Emprego segundo os resultados obtidos através do coeficiente de correlação de Spearman também se tratam de estruturas hierárquicas semelhantes. Em 2009 existe um grande grupo com cerca de quinze actividades, sendo este de todas as secções. Comparativamente com 2010, esse grande grupo não é muito alterado

O índice Remunerações segundo os resultados obtidos através do coeficiente de correlação de Spearman, mostra as estruturas apresentadas como não sendo «globalmente» semelhantes – as duas estruturas hierárquicas de classificação não são correlacionadas. Por esse motivo foi dada especial atenção a este índice. No entanto, de todos os dendrogramas, este é o que apresenta a melhor estrutura do ponto de vista da separação de grupos. As estruturas apresentam uma boa diferenciação em relação à formação dos grupos de actividades. Estas diferem das apresentadas pelos outros índices porque não têm um grande grupo de actividades formado. Têm actividades distribuídas de uma forma mais uniforme pelos grupos. No entanto estas actividades mudam bastante de grupo comparando os anos 2009 e 2010, apesar dos grupos terem aproximadamente o mesmo número de actividades por cada um destes.

O índice Horas trabalhadas nos serviços segundo os resultados obtidos através do coeficiente de correlação de Spearman também se tratam de estruturas hierárquicas semelhantes. Apresenta, em 2010, a formação de um grande grupo de actividades de todas as secções. Tal factor neste caso verifica-se melhor através dos dendrogramas. Para o ano 2009

as actividades estão mais divididas pelos grupos, porém os dois grupos maiores em 2009 contêm actividades de todos os ramos/secções.

O índice de Rand Ajustado segundo Hubert e Arabie foi o que se revelou mais adequado a este estudo. Não se apresentou as Hipóteses de simulação através das quais foram obtidos os resultados que nos permitiram concluir sobre a comparação de partições. Tal facto deve-se a estes estarem exaustivamente descritos nos artigos referenciados.

Projecto Futuro

Pensa-se utilizar outras metodologias para este trabalho. Por exemplo análise factorial e em componentes principais. Estas não implicam à partida a mudança de variáveis entre secções, combinadas com outras metodologias, inclusive com métodos de classificação, utilizando mais observações que vão sendo fornecidas pelo INE.

Inspirado nos resultados obtidos neste trabalho e de forma a ser utilizada a mesma metodologia deste trabalho, ou seja, Classificação Hierárquica Ascendente e de Particionamento tem-se, como projecto futuro, a aplicação desta noutros temas. A escolha de outro tema está directamente relacionado com a profissão do autor, isto é, telecomunicações móveis.

Tem-se como hipótese de projecto a área de «Gestão da Qualidade do serviço de Voz nos principais Aglomerados Urbanos e Eixos Rodoviários de Portugal Continental das três Operadoras Móveis». Desta forma pretende-se unir duas áreas que me são familiares, isto é, gestão e telecomunicações móveis. Este estudo teria o objectivo de agrupar os distritos de Portugal, consoante o nível de semelhança da qualidade do serviço de voz, para cada uma das três grandes operadoras nacionais.

Desta forma seria permitido mostrar um estudo alternativo sobre qualidade das três operadoras por distrito. Por outro lado seriam apresentados resultados de um ponto de vista diferente do «tradicional»: estatística descritiva. A diversidade na utilização da Classificação Hierárquica em várias áreas iria resultar no reforço da utilidade das técnicas utilizadas.

Bibliografia

Albatineh, A. N. (2010). *Means and variances for a family of similarity indices used in cluster analysis* em: Journal of Statistical Planning and Inference. (Vol. 140., pp.2828-2838). Elsevier.

Albatineh, A. N. Razeghifard, R. (2008). *Clustering amino acids using maximum clustering similarity*. Em: Proceedings of the International Conference on Bioinformatics, Computational Biology, Genomics and Chemoinformatics. (pp., 89-92).

Aldenderfer, Mark S and Blashefield, Roger K. (1984). *Analysis Cluster – Quantitative Applications in the Social Sciences*. Sage University Paper.

Armstrong, Gary and Kotler, Philip, (1999). *Marketing : an Introduction*, (5ªEd., pp.78), Prentice Hall.

Bacelar-Nicolau, H. (1980). **Contribuições ao Estudo dos Coeficientes de Comparação em Análise Classificatória**, DEIOC, FC, UL, Lisboa. Dissertação de Doutoramento,.

Cardoso, M (2010). Estabilidade de uma Estrutura de Agrupamento – Segmentos de Clientes de uma Instituição Cultural, em: Revista de Ciências da Computação, (Vol 3., nº5., pp 1-9).

Fowlkes. E. B e Mallows. C. L (1983), *A Method for Comparing Two Hierarchical Clusterings*. Journal of the American Statistical Association. (Vol. 78., nº 383., pp. 553-569). American Statistical Association.

GPEARI, Gabinete de Planeamento, Estratégia, Avaliação e Relações Internacionais. (2009). **Economia Portuguesa**. Ministério das Finanças e da Administração Pública.. ISSN – 0874-8195. [Consult. 1º Semestre de 2011]. Disponível em: WWW:[URL:http://gpeari.min-financas.pt](http://gpeari.min-financas.pt)

GPEARI, Gabinete de Planeamento, Estratégia, Avaliação e Relações Internacionais. (2011). **Indicadores Mensais de Conjuntura**. Ministério das Finanças e da Administração Pública. [Consult. 1º Semestre de 2011]. Disponível em: WWW:[URL:http://gpeari.min-financas.pt](http://gpeari.min-financas.pt)

Gordon, A. D. (1999). *Classification*. (2ª ed, pp.1, 20-40). Chapman & Hall/CRC.

Gower, J. C. (1985). *Measures of similarity, dissimilarity, and distance*.em: Encyclopedia of Statistical Sciences. (Vol. 5., pp.397-405). John Wiley and Sons.

Hodson, F.R. Sneath, P. H. A. e Doran, J.E. (1966). *Some experiments in the numerical analysis of archaeological data*. Biometrics.

Hubalek, Z. (1982). *Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation*. (pp., 732-764). Biological Review

Hubert. L. Arabie. (1985). Comparing partitions em Journal of Classification (nº2., pp.193-218).

INE, Instituto Nacional de Estatística. (2007). **Classificação Portuguesa das Actividades Económicas** (Ver.3.) Lisboa. ISBN: 978-972-673-919-7. [Consult. 1º Semestre de 2011]. Disponível em: [WWW:URL:http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=10376048&PUBLICACOESmodo=2](http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=10376048&PUBLICACOESmodo=2).

INE, Instituto Nacional de Estatística. (2011). **Índices de Volume de Negócios, Emprego, Remunerações e Horas Trabalhadas nos Serviços em destaque informação à comunicação social**. [Consult. 1º Semestre de 2011]. Disponível em: [WWW:URL:http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=106165006&DESTAQUESmodo=2](http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=106165006&DESTAQUESmodo=2).

Kaufman, Leonard e Rousseeuw, Peter. (2005). *Finding Groups in Data* (1ª ed., pp.24-31). John Wiley and Sons.

Krzanowski, W. J. (2008). *Principles of Multivariate Analysis* (2ª ed., pp.31). Oxford University press.

McQuitty, L. L. (1966). *Single and Multiple Hierarchical Classification by Reciprocal Pairs and Rank Order Types Educational and Psychological Measurement* (Vol 26.,nº2.,pp.253-265). Michigan State University. DOI: 10.1177/001316446602600201

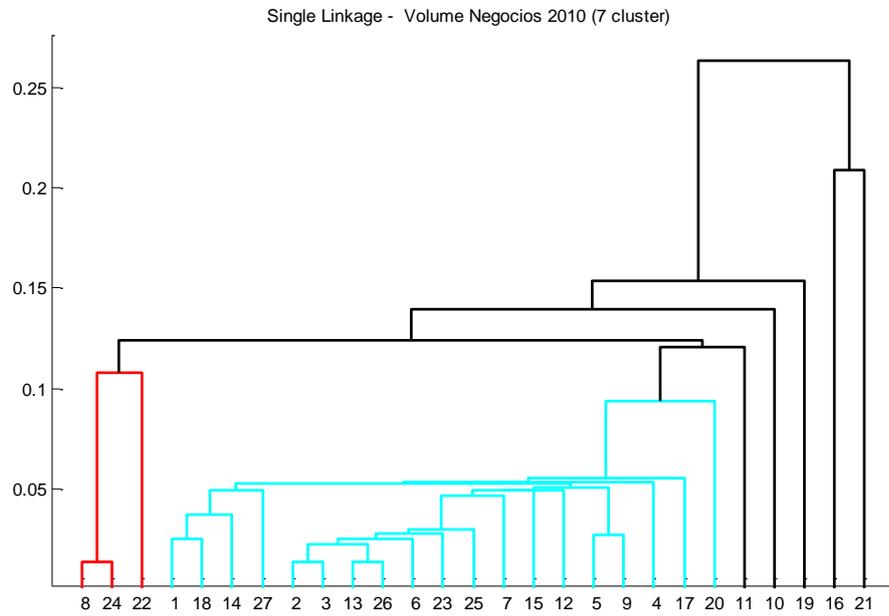
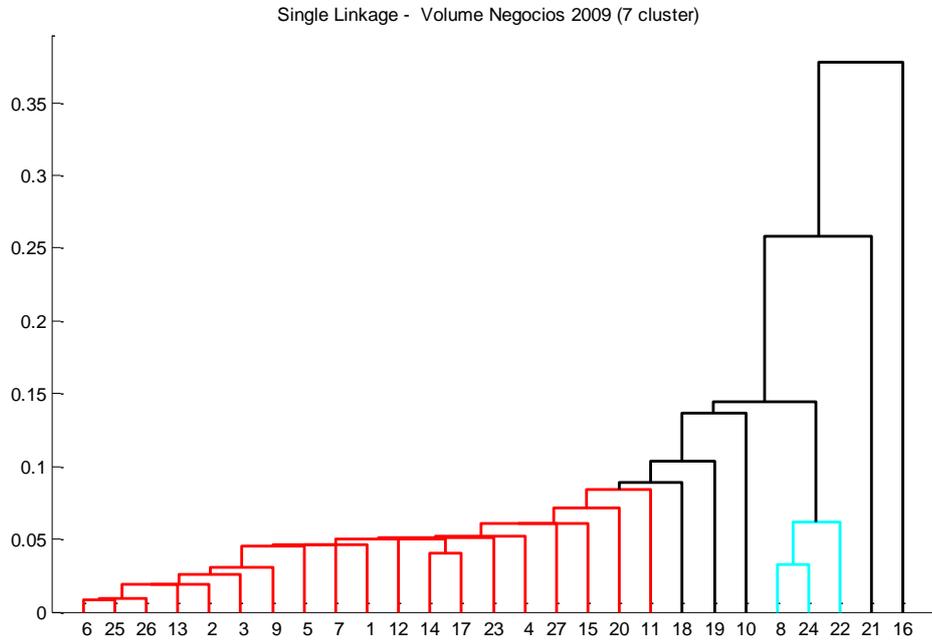
Milligan, G., Soon, S., Sokol, L. (1983). *The effect of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure, IEEE Transactions on Pattern Analysis and Machine Intelligence* (pp. 40-47).

Murtagh, F. (1984). Counting dendrograms: a survey. *Discrete Applied Mathematics*. Elsevier

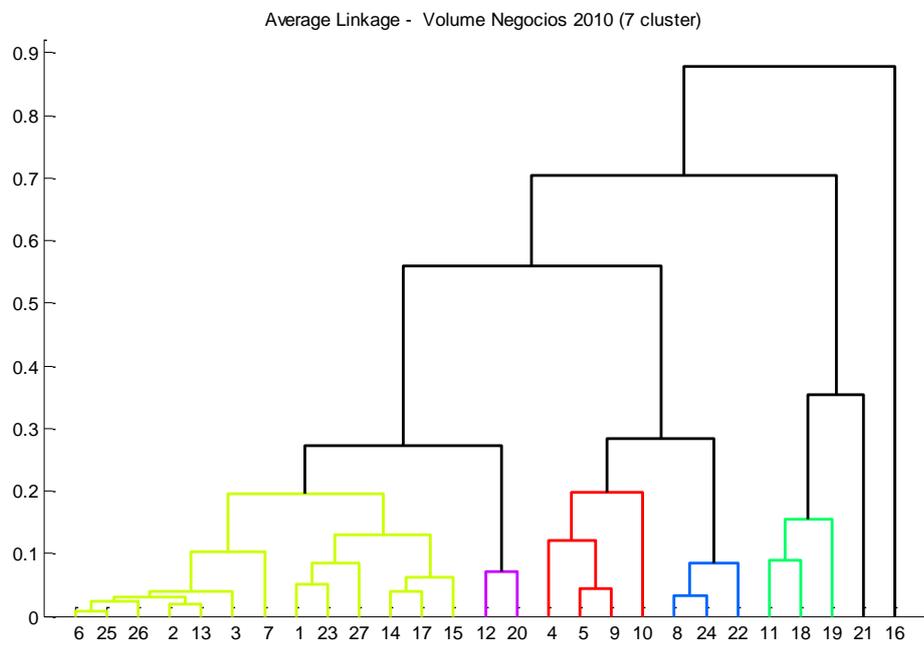
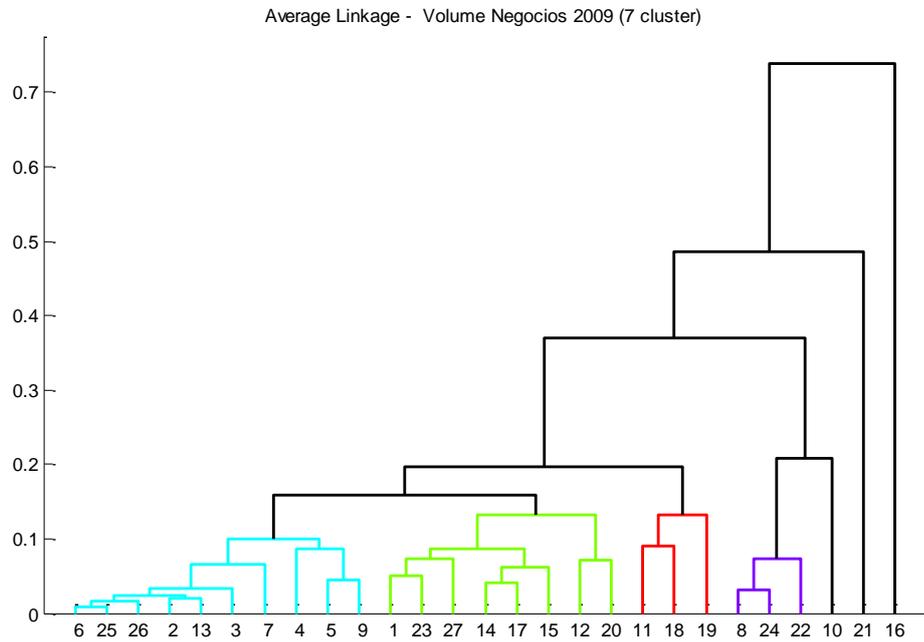
- Nacache, J-P e Confais, J. (2005). *Approche Pragmatique de la Classification*. Editions Technip.
- Saporta, Gilbert. (2006). *Probabilités Analyse Des Données Et Statistique* (2^a ed., pp.142). Technip Editions.
- Saporta, Gilbert. (2009). *Comparing partitions of two sets of units based on the same variables*. Springer-Verlag.
- Seber, G.A.F. (1984). *Multivariate Observations*. (2^a ed., pp.360-363). John Wiley & Sons.
- Siegel, Sidney. (2006). *Estatística Não-Paramétrica Para as ciências do Comportamento*. (2^a ed., pp.228). McGraw-Hill.
- Silva, A. L. (2005). *Tratamento de Dados Omissos e Métodos de Imputação em Classificação*, ISEG, UTL e CNAM Paris. Dissertação de Doutoramento.
- Sokal R.R. and Michener C.D. (1958). *A Statistical Method for Evaluating Systematic Relationships*. The University of Kansas Scientific Bulletin (Vol38., pp.1409-1438).
- Sokal R R & Sneath PH A (1963). *Principles of numerical taxonomy*(pp.359).
- Sousa, F., Nicolau, F. C. (2002). *Validação em Classificação Hierárquica Ascendente – Alguns Resultados*, em *Novos Rumos em Estatística* (L. Carvalho, F. Brilhante e F. Rosado, eds) (pp. 403-413). Sociedade Portuguesa de Estatística.
- Steinley, D. (2004). *Properties of the Hubert-Arabic Adjusted Rand Index* in *American Psychological Association* (pp.386-396).
- Unnikrishnan, R., Pantofaro, C., Hebert, M. (2005). *A measure for objective evaluation of image segmentation algorithms*.em : *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1063-6919).
- Yeung, K. Y. e Ruzzo, W. L. (2001). *An empirical Study on Principal Component Analysis for clustering gene expression data* in *Details of the Adjusted Rand Index and Clustering Algorithms Supplement Bioinformatics* (pp.763-774).
- Youness, G. e Saporta, G. (2010) *Comparing partitions of two sets of units based on the same variables* in *Advances in Data Analysis and Classification* (pp.53-64)

Apêndices

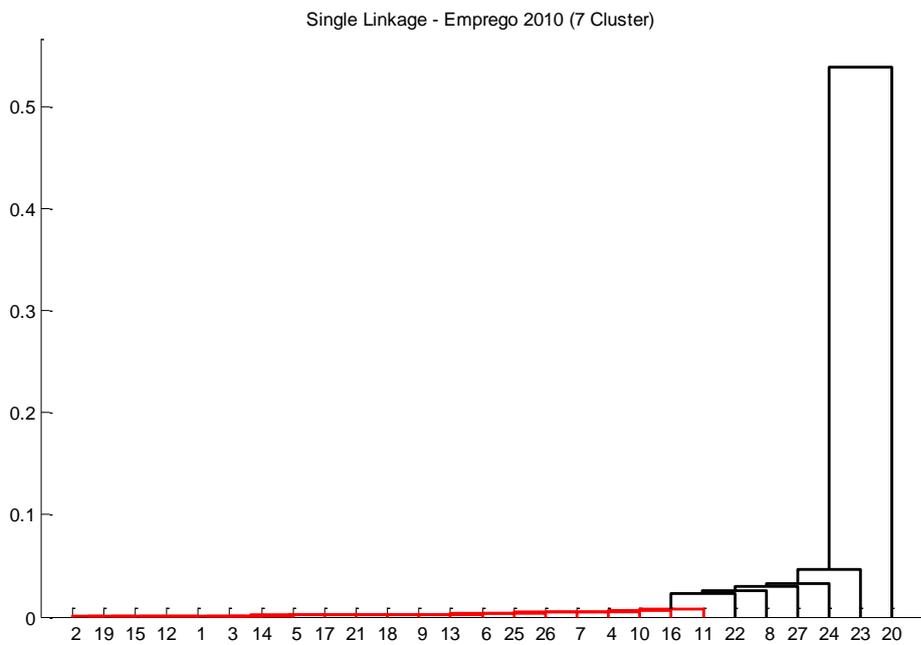
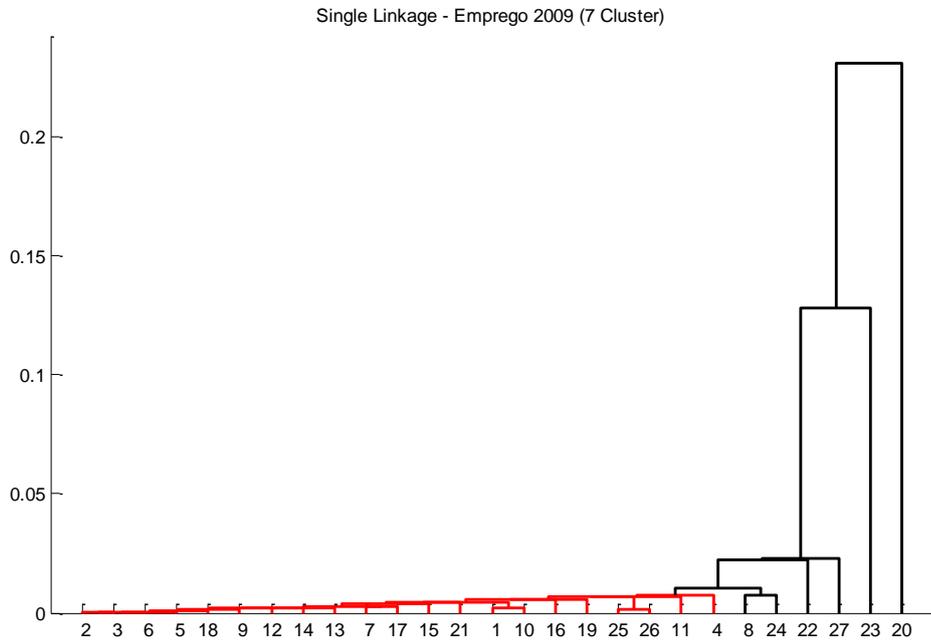
Apêndice 1 Dendrograma do Índice Volume Negócios Utilizando o Método do Vizinho mais próximo para 2009 e 2010



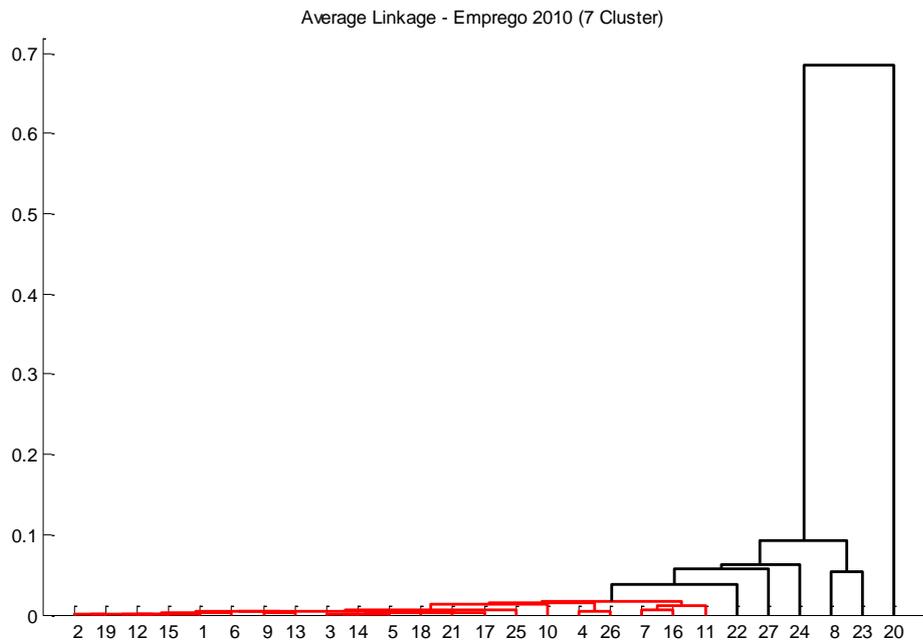
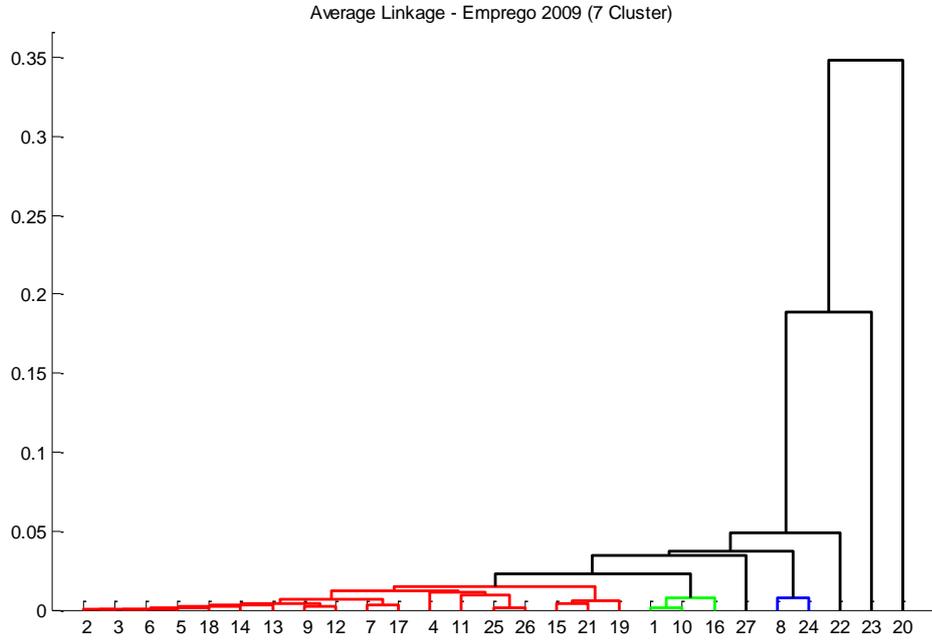
Apêndice 2 Dendrograma do Índice Volume Negócios Utilizando o Método da Média da Distância Entre os Grupos Para 2009 e 2010



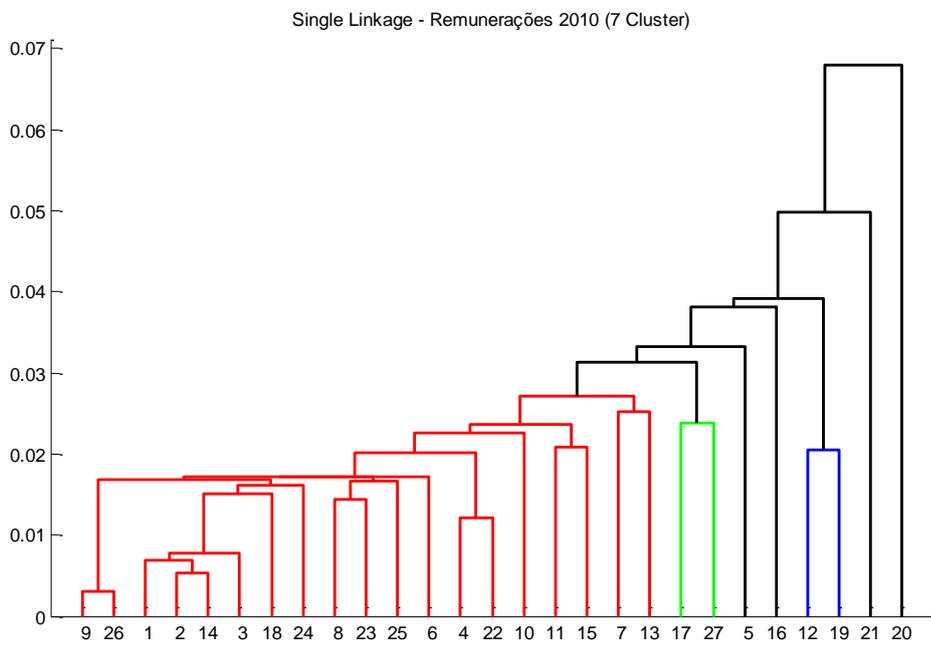
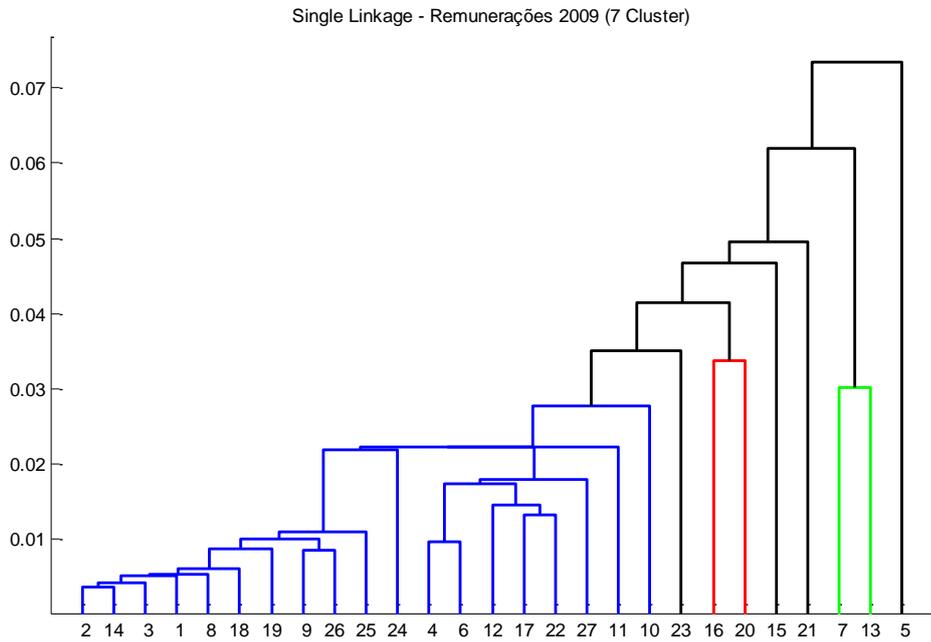
Apêndice 3 Dendrograma do Índice Emprego Utilizando o Método do Vizinho Mais Próximo para 2009 e 2010



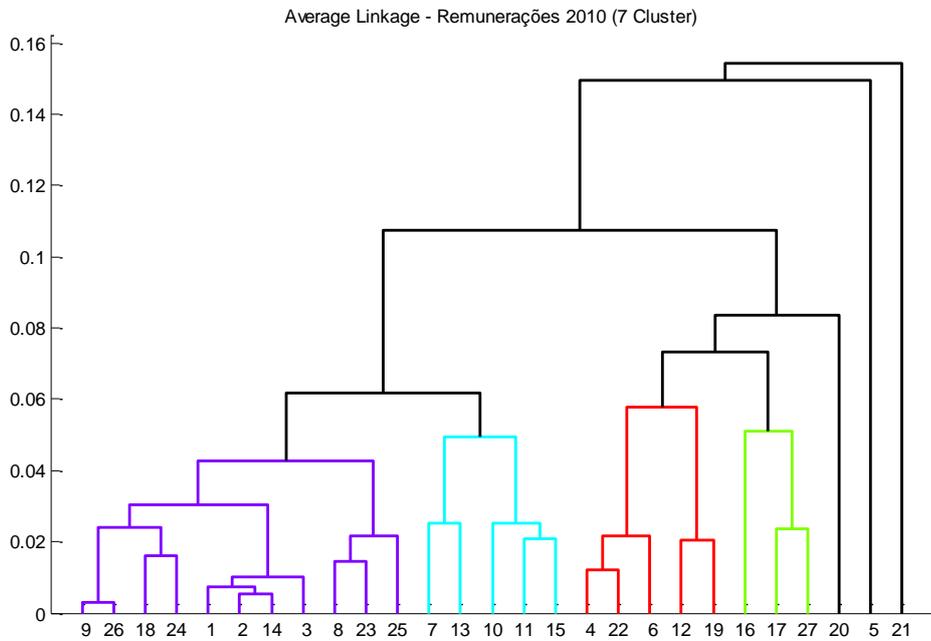
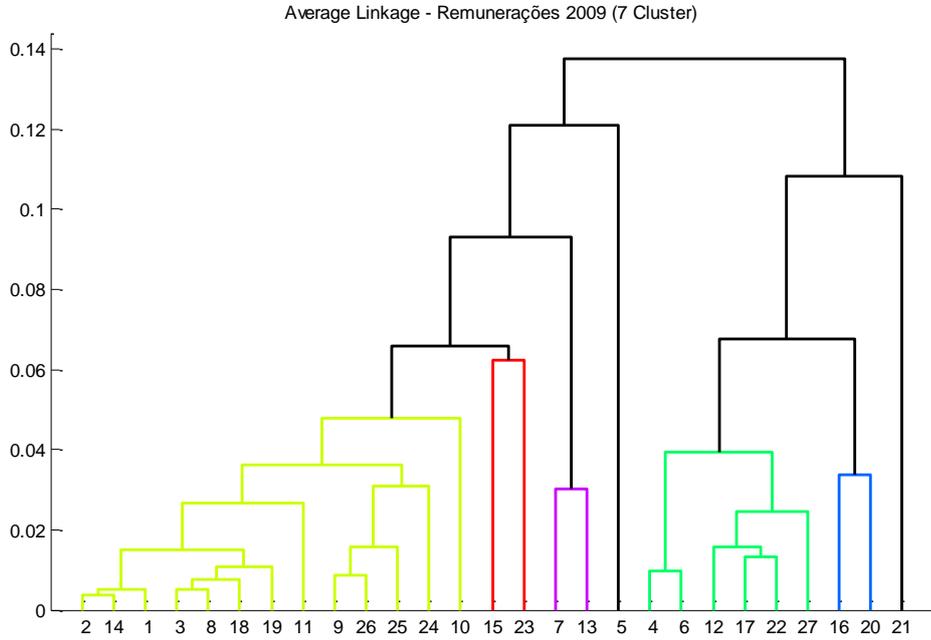
Apêndice 4 Dendrograma do Índice Emprego Utilizando o Método da Média da Distância Entre os Grupos para 2009 e 2010



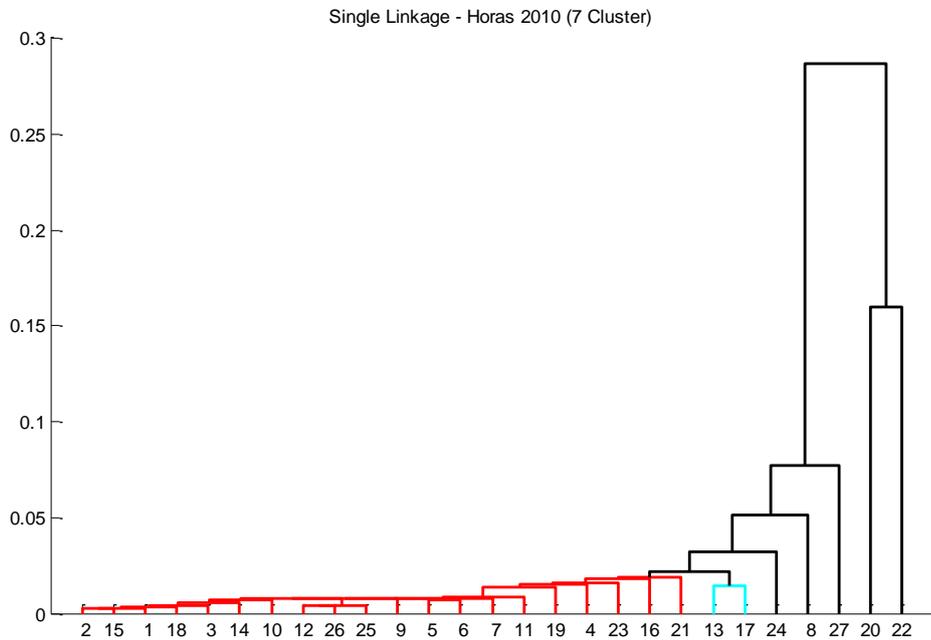
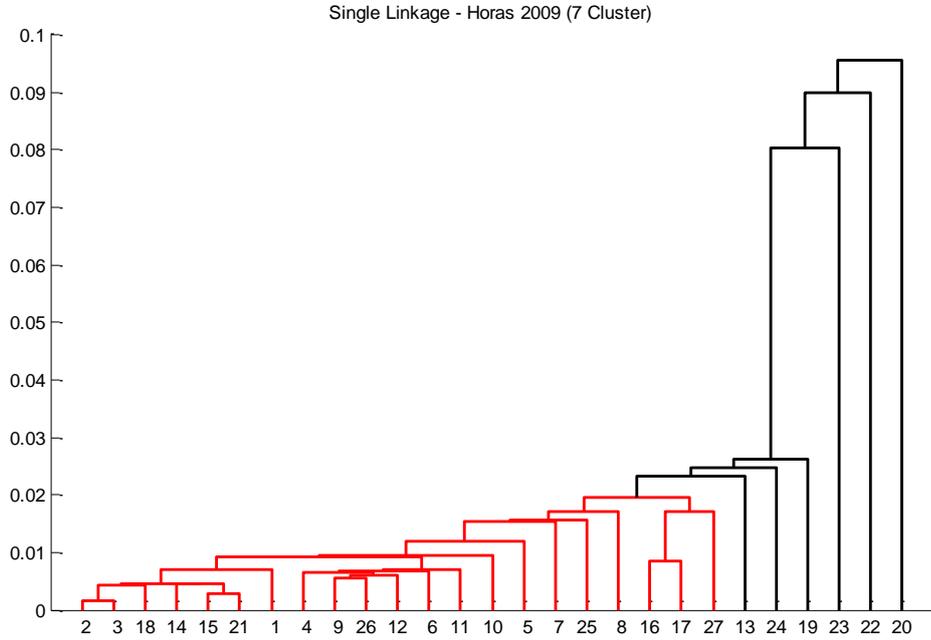
Apêndice 5 Dendrograma do Índice Remunerações Utilizando o Método do Vizinho Mais Próximo para 2009 e 2010



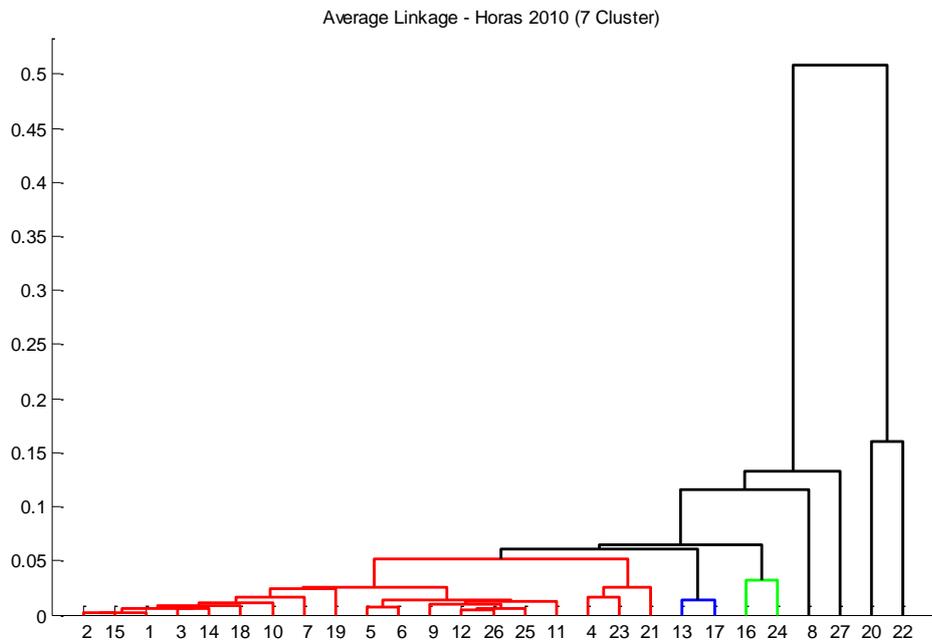
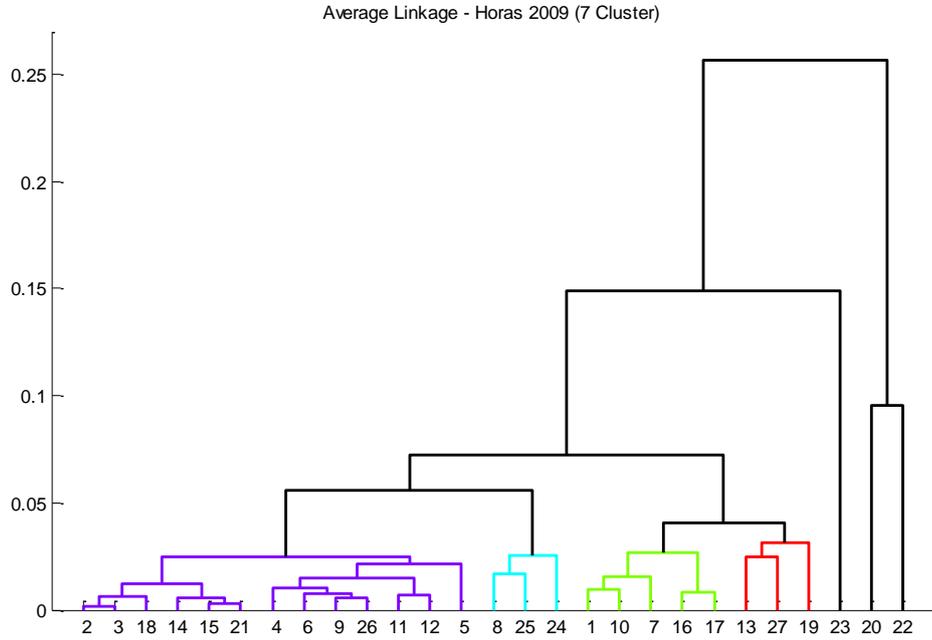
Apêndice 6 Dendrograma do Índice Remunerações Utilizando o Método da Média da Distância Entre os Grupos, para 2009 e 2010



Apêndice 7 Dendrograma do Índice Horas Utilizando o Método do Vizinho mais próximo para 2009 e 2010



Apêndice 8 Dendrograma do Índice Horas Utilizando o Método da Média da Distância Entre os Grupos para 2009 e 2010



Apêndice 9 Algoritmo das Ultramétricas

INPUT: elementos que se encontram abaixo da diagonal principal das matrizes das cofenéticas. Este resultado surge em linha sendo transposto e inserido sob a forma de coluna.

OUTPUT: matriz com duas colunas. A primeira é o valor ordinal e segunda é o índice de agregação correspondente ao grupo.

```

%%%%%%%%%%ULTRAMETRICAS%%%%%%%%%%
b=INPUT;
min=20;
lastmin=0;
index=1;
%%Achar o min
for i=1:length(b);
    if b(i) > lastmin && b(i) < min
        min=b(i);
    end
end
lastmin=min;
min=20;

for j=1:length(b);
    for i=1:length(b);
        %% --atribui o index ao ultimo min
        if b(i) == lastmin
            m=[[b(i)] [index]];
            matrixvalor(i)=[b(i)];
            matrixindex(i)=[index];
        end
        %% --encontra novo min
        if b(i) > lastmin && b(i) < min
            min=b(i);
        end

    end
    index=index+1;
    lastmin=min;
    min=20;

end
%OUTPUT=[[matrixvalor'] [matrixindex']]
% xlswrite('teste.xls', matrixfinal);

```

Apêndice 10 Código Fonte Matlab Ano 2009 (Exemplo: Volume Negócios)

INPUT: ficheiro com os dados a trabalhar, com extensão excel.

OUTPUT1: dendrogramas.

OUTPUT2: índice das ultramétricas.

OUTPUT3. Coeficiente de correlação de Spearman.

```

clc
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[tabela] = xlsread('INPUT');
Y = pdist(tabela,'corr')/2;
Zsl = linkage(Y,'single');
[Hsl] = dendrogram(Zsl,'colorthreshold',0.086);
Tmaxclustersl2009 = cluster(Zsl,'maxclust',7);
%title('Single Linkage - Volume Negocios 2009 (7 cluster) ')
%OUTPUT1set(Hsl,'LineWidth',2)
[csl,OUTPUT2] = cophenet(Zsl,Y);
pause
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Zal = linkage(Y,'average');
[Hal] = dendrogram(Zal,'colorthreshold',0.15);
Tmaxclusteral2009 = cluster(Zal,'maxclust',7);
%set(Hal,'LineWidth',2)
%title('Average Linkage - Volume Negocios 2009 (7 cluster)')
[cal,Dal2009] = cophenet(Zal,Y);
pause
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Zcl = linkage(Y,'complete');
[Hcl] = dendrogram(Zcl,'colorthreshold',0.25);
Tmaxclustercl2009 = cluster(Zcl,'maxclust',7);
set(Hcl,'LineWidth',2)
title('Complete Linkage - Volume Negocios 2009 (7 cluster)')
[ccl,Dcl2009] = cophenet(Zcl,Y);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
matrixcluster2009=[Tmaxclustersl2009 Tmaxclusteral2009
Tmaxclustercl2009];
xlswrite('mcluster2009.xls', matrixcluster2009);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
OUTPUT3= corr(Dsl2009',Dal2009','type','spearman');
rslcl= corr(Dsl2009',Dcl2009','type','spearman');
ralcl= corr(Dal2009',Dcl2009','type','spearman');
coefcorrsp2009=[rslal rslcl ralcl];
xlswrite('coeficientecorrsperman2009.xls', coefcorrsp2009);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
b=Dsl2009';
min=20;
lastmin=0;
index=1;
%%Achar o min
for i=1:length(b);
    if b(i) > lastmin && b(i) < min
        min=b(i);
    end
end

```

```
end
lastmin=min;
min=20;

for j=1:length(b);
    for i=1:length(b);
        %%% --atribui o index ao ultimo min
        if b(i) == lastmin
            m=[[b(i)] [index]];
            matrixvalor(i)=[b(i)];
            matrixindex(i)=[index];
        end
        %%% --encontra novo min
        if b(i) > lastmin && b(i) < min
            min=b(i);
        end

    end
    index=index+1;
    lastmin=min;
    min=20;

end
%matrixfinal=[[matrixvalor'] [matrixindex']]
% xlswrite('teste.xls', matrixfinal);
```

Apêndice 11 Cálculo Analítico do Índice de Rand e Respective Intervalos de Rejeição**Volume Negócios:**

$$R = 1 - \frac{1}{27(26 - 1)}(143 + 209) + \frac{2}{27(26 - 1)}(83) = 0.7350$$

Intervalo Rejeição:]0.6262 , 0.6961[

Emprego:

$$R = 1 - \frac{1}{27(26 - 1)}(319 + 257) + \frac{2}{27(26 - 1)}(173) = 0.6724$$

Intervalo Rejeição:]0.4241 , 0.6338[

Remunerações:

$$R = 1 - \frac{1}{27(26 - 1)}(181 + 195) + \frac{2}{27(26 - 1)}(91) = 0.7236$$

Intervalo Rejeição:]0.6017 , 0.6909[

Horas:

$$R = 1 - \frac{1}{27(26 - 1)}(281 + 191) + \frac{2}{27(26 - 1)}(119) = 0.6667$$

Intervalo Rejeição:]0.5076 , 0.6396[

Apêndice 12 Cálculos Analíticos Relativos aos Índices k, PE e R**Volume Negócios:**

	k	PE	R
A	-0,47991	-0,42335	0,498575
B	0,008409	0,007418	0,002849
SI	0,218001	0,192308	0,735043
E(SI)	0	-0,26075	0,66118
VAR(SI)	0,002656	0,002067	0,000305
Intervalo Inferior	-0,10307	-0,35167	0,626257
Intervalo Superior	0,103071	-0,16982	0,696102

Emprego:

	k	PE	R
A	-0,74197	-0,79324	0,179487
B	0,006048	0,006466	0,002849
SI	0,304423	0,325461	0,672365
E(SI)	0	-0,44375	0,528973
VAR(SI)	0,012385	0,014156	0,002748
Intervalo Inferior	-0,22258	-0,68171	0,424132
Intervalo Superior	0,22258	-0,20579	0,633814

Remunerações:

	k	PE	R
A	-0,51435	-0,49967	0,464387
B	0,008055	0,007825	0,002849
SI	0,218657	0,212413	0,723647
E(SI)	0	-0,31774	0,64631
VAR(SI)	0,003973	0,003749	0,000497
Intervalo Inferior	-0,12607	-0,44021	0,601722
Intervalo Superior	0,126065	-0,19528	0,690898

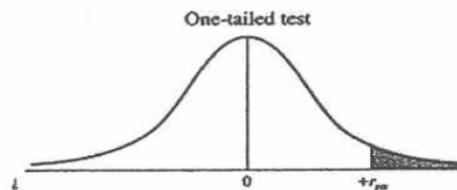
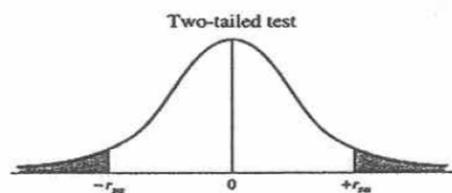
Horas:

	k	PE	R
A	-0,5769	-0,68694	0,327635
B	0,006682	0,007956	0,002849
SI	0,218233	0,25986	0,666667
E(SI)	0	-0,44096	0,573615
VAR(SI)	0,005991	0,008495	0,001089
Intervalo Inferior	-0,15481	-0,6253	0,507607
Intervalo Superior	0,154809	-0,25662	0,639624

Anexos

Anexo 1 Tabela Coeficiente de Spearman

Critical Values of the Spearman Rank Order Correlation Coefficients: The r_s Tables



Level of Significance

N	Level of Significance	
	$\alpha = .05$	$\alpha = .01$
5	1.000	–
6	.886	1.000
7	.786	.929
8	.738	.881
9	.700	.833
10	.648	.794
11	.618	.755
12	.587	.727
13	.560	.703
14	.538	.675
15	.521	.654
16	.503	.635
17	.485	.615
18	.472	.600
19	.460	.584
20	.447	.570
21	.435	.556
22	.425	.544
23	.415	.532
24	.406	.521
25	.398	.511
26	.390	.501
27	.382	.491
28	.375	.483
29	.368	.475
30	.362	.467
31	.356	.459
32	.350	.452
33	.345	.446
34	.340	.439
35	.335	.433
36	.330	.427
37	.325	.421
38	.321	.415
39	.317	.410
40	.313	.405
41	.309	.400
42	.305	.395
43	.301	.391
44	.298	.386
45	.294	.382
46	.291	.378
47	.288	.374
48	.285	.370
49	.282	.366
50	.279	.363

Level of Significance

N	Level of Significance	
	$\alpha = .05$	$\alpha = .01$
5	.900	1.000
6	.829	.943
7	.714	.893
8	.643	.833
9	.600	.783
10	.564	.745
11	.536	.709
12	.503	.671
13	.484	.648
14	.464	.622
15	.443	.604
16	.429	.582
17	.414	.566
18	.401	.550
19	.391	.535
20	.380	.520
21	.370	.508
22	.361	.496
23	.353	.486
24	.344	.476
25	.337	.466
26	.331	.457
27	.324	.448
28	.317	.440
29	.312	.433
30	.306	.425
31	.301	.418
32	.296	.412
33	.291	.405
34	.287	.399
35	.283	.394
36	.279	.388
37	.275	.383
38	.271	.378
39	.267	.373
40	.264	.368
41	.261	.364
42	.257	.359
43	.254	.355
44	.251	.351
45	.248	.347
46	.246	.343
47	.243	.340
48	.240	.336
49	.238	.333
50	.235	.329

Adapted from Zar, J. H. (1972). Significance testing of the Spearman rank correlation. *Journal of the American Statistical Association*, 67, 578 – 580.

Anexo 2 Índice de Rand Adaptado por Hubert e Arabie (1985)

Adaptado por Hubert e Arabie (1985) tem-se

O índice de Rand apresentado originalmente sob a forma de concordâncias fornece a proporção de concordâncias para as duas partições I e J de n variáveis ou indivíduos, encontrar-se-ão quatro tipos de pares:

a - número de pares que pertencem simultaneamente aos mesmos grupos de I e J

b - número de pares que pertencem a diferentes grupos de I e aos mesmos de J

c - número de pares que pertencem a diferentes grupos de I e aos mesmos de J

d - número de pares que pertencem a diferentes grupos de I e J .

Tem-se assim:

- $A=a+d$ - representa o número total de concordâncias
- $D=b+c$ - representa o número total de discordâncias

e
$$A + D = \frac{n(n - 1)}{2}.$$

O índice de Rand clássico é dado pela expressão:

$$I_R = \frac{2A}{n(n - 1)}.$$