



Universidade de Aveiro Departamento de Matemática
2013

Dânia Fidalgo Carlos

**“O CONSUMO DAS FAMÍLIAS PORTUGUESAS”
ATRAVÉS DA ANÁLISE MULTIVARIADA**



Dânia Fidalgo Carlos

**“O CONSUMO DAS FAMÍLIAS PORTUGUESAS”
ATRAVÉS DA ANÁLISE MULTIVARIADA**

Relatório de estágio apresentado à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Matemática e Aplicações, realizado sob a orientação científica da Doutora Adelaide Freitas, Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro

Dedico este trabalho ao meu marido e aos meus filhos, pois são a luz da minha vida. E principalmente, aos meus pais, os meus primeiros e grandes professores.

O júri

Presidente

Prof. Doutora Isabel Maria Simões Pereira
Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro

Prof. Doutora Helena Maria Pereira Pinto Dourado e Alvelos
Professora Auxiliar do Departamento de Gestão da Universidade de Aveiro

Prof. Doutora Adelaide de Fátima Baptista Valente Freitas
Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro

Agradecimentos

Neste relatório pretendo agradecer a todos aqueles que me ajudaram, apoiaram e participaram na minha vida, aqueles que sempre acreditaram em mim, na minha força para concretizar este objetivo.

Para já ficam os agradecimentos:

À minha orientadora do GEMEO, Prof. Dra. Valentina Ribau, pela forma como me acolheu, por tentado sempre acompanhar-me da melhor forma e pelas experiências que me proporcionou. Agradeço também à minha orientadora científica, a Prof. Auxiliar Adelaide Freitas, pela sua disponibilidade, pela sua orientação, pela sua paciência em ajudar-me a tirar todas as dúvidas.

Ao meu marido e aos meus filhos pelo grande apoio, compreensão, ajuda, carinho, amor e incentivo dados durante a realização deste trabalho.

Aos meus pais que me trouxeram com todo o amor e carinho a este mundo, pelos seus esforços, por incentivarem o meu desenvolvimento pessoal e profissional, e sobretudo pelo seu apoio incondicional.

A todos os meus amigos e família que sempre me apoiaram para conquistar mais uma vitória.

Palavras-chave

Pesquisa de Mercado; Análise Fatorial; Análise de Conglomerados

Resumo

Durante os seis meses em que decorreu o Estágio no GEMEO (Gabinete de Estudos de Mercado e Opinião do IPAM) foram desenvolvidas diversas atividades na área do Marketing, entre elas o estudo sobre o consumo das famílias portuguesas. Este relatório baseia-se nas Pesquisas de Orçamento Familiar (HBS – *Household Budget Survey*) da União Europeia que fornecem informações sobre gastos de consumo das famílias com bens e serviços, entre outros. Deste modo, o estudo desenvolvido no âmbito do estágio no GEMEO teve como objetivo colocar em prática metodologias estatísticas multivariadas adquiridas a nível académico, nomeadamente análise factorial e análise de *clusters*. O consumo das famílias portuguesas foi desenvolvido como tema do estágio, pois nos tempos que decorrem e com a crise europeia, achou-se relevante ver que alterações ocorreram desde o último HBS até agora no consumo.

Keywords

Marketing Research; Factor Analysis; Cluster Analysis

Abstract

During the six months that elapsed the curricular training in GEMEO (Bureau of Market Research and Opinion of IPAM) several activities were developed in the field of marketing, including the study about the consumption of Portuguese families. This report is based on the Household Budget Surveys (HBS) of the European Union to provide information on household consumption expenditures on consumer assets and services, among others. Thus, the study developed under the stage in GEMEO had as objective to put in practice multivariate statistical methodologies acquired in academic masters, including factorial analysis and clusters analysis. Household consumption was developed as a Portuguese theme stage, because in the times that arise and the European crisis, it was felt important to see what changes have occurred since the last HBS so far in consumption.

Índice

Agradecimentos	v
Resumo.....	vi
Abstract	vii
LISTA DE FIGURAS.....	xi
LISTA DE TABELAS.....	xiii
LISTA DE VARIÁVEIS.....	xiv
ABREVIATURAS.....	xv
CAPÍTULO 1: INTRODUÇÃO	1
1.1 Enquadramento.....	1
1.1.1 O consumo das famílias.....	1
1.1.2 A empresa	3
1.2 Objetivos.....	5
1.3 Estrutura do Relatório	6
CAPÍTULO 2: ENQUADRAMENTO TEÓRICO.....	7
2.1. Tipos de entrevista	7
2.1.1 Entrevista pessoal	7
2.1.2 Entrevista telefónica	8
2.1.3 Entrevista por correio.....	11
2.1.4 Entrevista por <i>e-mail</i>	12
2.2 Técnicas de Amostragem.....	13
2.2.1. Amostragem não-probabilística.....	13
2.2.1.1. Amostragem intencional.....	13
2.2.1.2. Amostragem <i>Snowball</i> ou “Bola de Neve”	14
2.2.1.4. Amostragem por quotas.....	14
2.2.1.5. Amostragem <i>Random Route</i>	15
2.3. Trabalho de campo.....	15
2.3.1. Seleção dos entrevistadores	16
2.3.2. Formação dos entrevistadores.....	16
2.3.3. Supervisão dos entrevistadores	16
2.3.3.1. Controlo da qualidade e edição	16

2.3.3.2.	Controlo da amostragem	17
2.3.3.3.	Controlo de fraudes.....	17
2.3.3.4.	Controlo do escritório central	17
2.3.4.	Validação do trabalho de campo	17
2.3.5.	Avaliação dos entrevistadores	18
2.4.	Cálculo da dimensão da amostra em amostragem não-probabilística.....	18
2.5.	Elaboração do questionário.....	20
2.6.	Análise Fatorial	22
2.6.1.	Modelo Fatorial Ortogonal.....	25
2.6.2.	Métodos de estimação dos fatores.....	26
2.6.2.1.	Método das componentes principais.....	27
2.6.2.2.	Método da máxima verosimilhança	28
2.6.2.3.	Método dos mínimos quadrados	28
2.6.2.4.	Método da factorização do eixo principal.....	29
2.6.2.5.	Método de factorização alfa	29
2.6.3.	Rotação de fatores	29
2.6.3.1.	Rotação ortogonal.....	30
2.6.3.2.	Rotação oblíqua.....	30
2.6.4.	Métodos de estimação dos valores dos fatores (<i>factor scores</i>)	31
2.6.4.1.	Método de Bartlett (ou método dos mínimos quadrados ponderados)	31
2.6.4.2.	Método de Thompson (ou método de regressão)	33
2.6.4.3.	Método de Anderson-Rubin.....	34
2.6.5.	Avaliação da qualidade do modelo fatorial.....	34
2.7.	Análise de <i>Clusters</i>	36
2.7.1.	Medidas de dissemelhança	37
2.7.2.	Critérios de agregação hierárquicos.....	39
2.7.3.	Quantos <i>clusters</i> se devem reter?.....	41
CAPÍTULO 3: CASO DE ESTUDO		43
3.1.	Metodologia (Ficha Técnica)	43
3.2.	Análise de Resultados.....	45
3.2.1.	Caracterização sociodemográfica da amostra	45
3.2.2.	Análise comparativa Portugal vs Europa.....	47

3.2.3. Análise comparativa entre o consumo atual das famílias com os consumos de anos anteriores	60
3.2.4. Análise da existência de factores latentes associados ao consumo	65
CAPÍTULO 4: CONCLUSÃO	71
BIBLIOGRAFIA.....	73
ANEXOS.	75
ANEXO A. Exemplo de um relatório de ligações.....	76
ANEXO B. Tabelas de agrupamento da Análise de <i>Clusters</i>	77
ANEXO C. Tabelas e <i>Scree Plot</i> da Análise Fatorial	83
ANEXO D. Alpha de Cronbach	89
ANEXO E. Questionário	94

LISTA DE FIGURAS

Figura 1. Processo de elaboração de um questionário (adaptado de Malhotra, 2011)	21
Figura 2. Dendrograma da família de um adulto sem crianças	48
Figura 3. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de um adulto sem crianças	49
Figura 4. Distribuição geográfica das famílias de um adulto sem crianças segundo os <i>clusters</i> do dendrograma obtido na Figura 2.	49
Figura 5. Dendrograma da família de um adulto com crianças.....	50
Figura 6. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de um adulto com crianças	51
Figura 7. Distribuição geográfica das famílias de um adulto com crianças segundo os <i>clusters</i> do dendrograma obtido na Figura 5.	51
Figura 8. Dendrograma da família de dois adultos sem crianças	52
Figura 9. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de dois adultos sem crianças	53
Figura 10. Distribuição geográfica das famílias de dois adultos sem crianças segundo os <i>clusters</i> do dendrograma obtido na Figura 8.	53
Figura 11. Dendrograma da família de dois adultos com crianças	54
Figura 12. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de dois adultos com crianças	55
Figura 13. Distribuição geográfica das famílias de dois adultos com crianças segundo os <i>clusters</i> do dendrograma obtido na Figura 11.	55
Figura 14. Dendrograma da família de três ou mais adultos sem crianças	56
Figura 15. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de três ou mais adultos sem crianças	57

Figura 16. Distribuição geográfica das famílias de três ou mais adultos sem crianças segundo os <i>clusters</i> do dendrograma obtido na Figura 14.....	57
Figura 17. Dendrograma da família de três ou mais adultos com crianças	58
Figura 18. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de três ou mais adultos com crianças.....	59
Figura 19. Distribuição geográfica das famílias de três ou mais adultos com crianças segundo os <i>clusters</i> do dendrograma obtido na Figura 17.....	59
Figura 20. Representação gráfica da despesa (em euros) com os tipos de consumo da família de um adulto de 1988 a 2012	60
Figura 21. Representação gráfica da despesa com os tipos de consumo da família de um adulto com crianças dependentes de 1988 a 2012	61
Figura 22. Representação gráfica da despesa com os tipos de consumo da família de dois adultos de 1988 a 2012.....	62
Figura 23. Representação gráfica da despesa com os tipos de consumo da família de dois adultos com crianças dependentes de 1988 a 2012.....	63
Figura 24. Representação gráfica da despesa com os tipos de consumo da família de três ou mais adultos de 1988 a 2012	64
Figura 25. Representação gráfica da despesa com os tipos de consumo da família de três ou mais adultos com crianças dependentes de 1988 a 2012	65

LISTA DE TABELAS

Tabela 1. Resultados de uma ligação telefónica e recomendações para lidar com eles (adaptado de Aaker, 2004).	9
Tabela 2. População dividida por NUTS II e pelo número de pessoas que habitam uma residência	44
Tabela 3. Amostra por quotas segundo os NUTS II e o número de pessoas que habitam uma residência.	44
Tabela 4. Características sociodemográficas da amostra.	46
Tabela 5. Agrupamento das variáveis em quatro fatores com a rotação <i>varimax</i>	67

LISTA DE VARIÁVEIS

C1 - Alimentos e Bebidas não Alcoólicas

C2 - Bebidas Alcoólicas, Tabaco e Narcóticos

C3 - Vestuário e Calçado

C4 - Habitação, água, eletricidade, gás e outros combustíveis

C5 - Acessórios, equipamento doméstico e manutenção corrente da habitação

C6 - Saúde

C7 - Transporte

C8 - Comunicações

C9 - Lazer e Cultura

C10 - Educação

C11 - Restaurantes e Hotéis

C12 - Bens e serviços diversos

T1 – Um Adulto

T2 - Um Adulto com Crianças dependentes

T3 - Dois adultos

T4 - Dois adultos com Crianças dependentes

T5 - Três ou mais adultos

T6 - Três ou mais adultos com crianças dependentes

ABREVIATURAS

AC - *Análise de Clusters*

ACP - *Análise de Componentes Principais*

AF - *Análise Fatorial*

AGFI - *Adjusted Goodness of Fit Index*

CATI - *Computer Assisted Telephone Interviewing*

ERC - *Entidade Reguladora das Comunicações*

ESOMAR - *European Society for Opinion and Marketing Research*

GEMEO - *Gabinete de Estudos de Mercado e Opinião*

GFI - *Goodness of Fit Index*

HBS - *Household Budget Survey*

INE - *Instituto Nacional de Estatística*

IPAM - *Instituto Português de Administração de Marketing*

KMO - *Kaiser-Meyer-Olkin*

NUTS - *Unidades Territoriais para fins Estatísticos*

RMSR - *Root Mean Square Residual*

UE - *União Europeia*

CAPÍTULO 1: INTRODUÇÃO

No primeiro capítulo é apresentado um enquadramento do tema do relatório de estágio, descrevendo o consumo das famílias. É também exposta uma breve descrição da empresa onde decorreu o estágio. Por fim, apresentam-se os objetivos do estudo desenvolvido e a estrutura global do relatório.

1.1 Enquadramento

1.1.1 O consumo das famílias

Apesar das diferenças sociais que há de país para país, houve mudanças radicais na segunda metade do século XX, até mesmo na noção de 'família'. A conceção de família é baseada na ideologia, mitologia popular e convenções que têm por base a história, a política, a economia e as tradições culturais de cada país.

Os governos europeus têm informação sobre as famílias que é obtida regularmente:

Informação sócio-demográfica:

- Taxa de nascimento;
- Taxa de fertilidade;
- Taxa de divórcio.

Formas de Família:

- Tamanho;
- Estrutura;
- Organização.

A família tradicional está em declínio. As pessoas colocam mais ênfase nos amigos chegados para fornecer companhia e suporte social (Solomon, 2010).

Para fins estatísticos, o Eurostat implementou uma definição de agregado familiar baseada na concepção de família conjugal. Assim, 'família' é definida como se segue: 'Pessoas que vivem numa casa ou instituição e/ou que estão relacionadas como casal (casados ou solteiros) e/ou com filhos de sangue ou de adoção' (Solomon, 2010).

A definição tenta abranger inúmeras possibilidades, casais casados e casais que vivem em união de facto.

Recentemente, para a Comunidade Europeia, um agregado familiar é amplamente definido como uma residência compartilhada com serviço de limpeza e manutenção comum.

As mudanças nas estruturas familiares, tais como coabitação, casamento, parto retardado, o retorno das mães ao mercado de trabalho e a turbulência causada pelo divórcio, muitas vezes representam oportunidades para os profissionais de marketing, pois essas mudanças levam as pessoas a fazerem escolhas sobre novos produtos e marcas.

As crianças também são de enorme importância para o consumo, pois têm um grande poder de persuasão sobre os seus pais. Um terço dos pais mudou os seus hábitos de compra depois de terem os filhos, uma vez que muitas decisões são tomadas em família (Solomon, 2010).

As Pesquisas de Orçamento Familiar (HBS – *Household Budget Survey*) da União Europeia são inquéritos por amostragem de famílias privadas realizados regularmente sob a responsabilidade do Instituto Nacional de Estatística (INE), em cada um dos Estados-Membros (Sistema Europeu de Estatística). Essencialmente, a HBS fornece informações sobre gastos de consumo das famílias com bens e serviços, com detalhes consideráveis nas categorias utilizadas, informações sobre o rendimento, de bens de consumo duráveis e carros, informações básicas sobre a habitação e muitas características demográficas e socioeconómicas (Eurostat, 2003).

Segundo o Eurostat, existem seis tipos de família: de um adulto; de um adulto com crianças dependentes; de dois adultos; de dois adultos com crianças dependentes; de três ou mais adultos; de três ou mais adultos com crianças dependentes. Também existem doze tipos de consumo: Alimentos e bebidas não alcoólicas (C1); Bebidas alcoólicas, tabaco e narcóticos (C2); Vestuário e calçado (C3); Habitação, água, eletricidade, gás e outros combustíveis (C4); Acessórios, equipamento doméstico e manutenção corrente da habitação (C5); Saúde (C6); Transporte (C7); Comunicações (C8); Lazer e Cultura (C9); Educação (C10); Restaurantes e Hotéis (C11); Bens e serviços diversos (C12).

Ao contrário de outros domínios estatísticos europeus, a HBS é voluntária e não existe regulamentação da UE. Portanto, existe uma considerável liberdade para cada Estado-Membro decidir os objetivos, metodologia, programação e atribuição de recursos para a sua respetiva HBS.

As HBS são estudos polivalentes que servem para um grande número de utilizações e utilizadores, uma vez que fornecem informação sobre as condições de vida económica e social das famílias e dos indivíduos nos Estados-Membros da UE (Eurostat, 2003).

1.1.2 A empresa

O Gabinete de Estudos de Mercado e Opinião do IPAM (GEMEO) foi criado em Maio de 1996, por iniciativa conjunta da Direção e de um grupo de docentes da Escola Superior de Aveiro do IPAM.

A génese do projeto e o seu desenvolvimento posterior deveram-se em larga medida, e numa primeira instância, à necessidade de operacionalizar no interior da organização, de um modo consistente e organizado, os conceitos chave que no Curso Superior de Gestão de Marketing se encontram consagrados à Investigação em Marketing.

Numa segunda instância, sendo embora um projeto de matriz académica, o GEMEO pretendeu constituir-se numa oficina com ferramentas científicas de gestão e análise.

se social e de mercado ao serviço da comunidade local, nomeadamente dos seus atores mais intervenientes, dos Media ao Meio Político e Empresarial.

A partir de 2001, o GEMEO constituiu-se como empresa juridicamente independente, embora integrando o universo ENSIGEST a que sempre pertenceu.

O GEMEO pretende usar a Pesquisa de Mercado para antecipar, explorar, auditar e analisar as novas tendências e mudanças no consumo e opinião.

Na conjugação e articulação destas aspirações, o GEMEO define a sua missão como:

“Capitalizar e instrumentalizar os recursos humanos, técnicos e científicos do IPAM na compreensão da Comunidade e do Mercado.”

O GEMEO realiza todos os seus estudos de mercado ou de opinião de acordo com as regras estabelecidas pela ESOMAR – *European Society for Opinion and Marketing Research*. Este é, aliás, um requisito obrigatório para se ser acreditado pela Entidade Reguladora das Comunicações (ERC) para a realização de estudos de opinião e sondagens que se destinem a ser divulgados na Comunicação Social.

Na atividade que desenvolve, entrevistar pessoas é a tarefa principal do GEMEO.

Como entende as pessoas na dupla condição de cidadãos e consumidores, mede intenções de compra ou intenções de voto, comportamentos, atitudes, hábitos, valores, interesses ou necessidades.

Mercê das dezenas de projetos realizados ou em fase de execução, o GEMEO adquiriu já um elevado nível de competências no que concerne à condução de:

- Estudos eleitorais;
- Estudos de opinião;
- Estudos de audiência de Imprensa e Rádio;
- Estudos de publicidade;
- Estudos de satisfação do cliente.

No âmbito dos estudos qualitativos, a estrutura interna do GEMEO permite desenvolver:

- Focus group;
- Hall test.

Paralelamente, o GEMEO disponibiliza ainda a realização de pesquisa documental, bem como um amplo leque de serviços na área do processamento de dados e análise estatística univariada e multivariada.

(A informação contida nesta secção foi escrita com base na informação contida na página de Internet do GEMEO, <http://gemo.ipam.pt/>, acedida em 2012).

1.2 Objetivos

O objetivo geral deste estágio foi colocar em prática o conhecimento de metodologias estatísticas adquiridas durante a licenciatura e parte curricular do mestrado. O tema escolhido para estudo foi o consumo das famílias portuguesas, pois nos tempos que decorrem e com a crise europeia, achou-se relevante ver que alterações ocorreram desde o último HBS até agora no consumo das famílias portuguesas.

Assim, tem-se por objetivos específicos:

- Planeamento (escolher a metodologia e criar um questionário);
- Recolha dos dados (Levantamento dos dados e supervisão do estudo);
- Comparar Portugal com os outros países da UE através da análise de *clusters*;
- Comparar o consumo atual das famílias com os consumos anteriores;
- Responder à pergunta “Será que existe um fator social ou económico que influencie o consumo?” através da análise factorial (AF).

1.3 Estrutura do Relatório

Para além desta introdução, este relatório é constituído por mais três capítulos.

No Capítulo 2 faz-se um enquadramento teórico das técnicas das ciências sociais e estatísticas selecionadas para o desenvolvimento e concretização dos objetivos definidos.

Assim, começa-se por definir os principais conceitos relacionados com as Ciências Sociais que foram necessários para a realização do estágio: os tipos de entrevista que se realizam, pormenorizando-se a entrevista por telefone; como se constrói um questionário; as técnicas de amostragem, nomeadamente as técnicas de amostragem não-probabilísticas, uma vez que foram as aplicadas neste estudo; como se realiza o trabalho de campo, sendo detalhada a descrição da função de supervisão, visto que durante o estágio foi exercida esta função.

De seguida, define-se o conceito de Análise Fatorial (AF) como técnica destinada à redução e sumarização dos dados. Como se formula um problema, a determinação do número de fatores, a rotação e a interpretação dos fatores, também são pontos a serem vistos.

Neste segundo capítulo é feita ainda uma abordagem teórica à Análise de *Clusters* (AC) como técnica de agrupamento de dados. Apresenta-se uma breve descrição desta técnica, definem-se as principais medidas de semelhança ou dissemelhança e os principais métodos de agrupamento, realizando-se uma abordagem mais pormenorizada dos métodos hierárquicos aglomerativos por serem os aplicados no âmbito deste trabalho. Apresentam-se formas de validação dos resultados obtidos e ainda formas de interpretação dos *clusters*.

No Capítulo 3 começa-se por apresentar uma breve descrição dos dados em análise, seguindo-se a aplicação de todas as técnicas estatísticas descritas no Capítulo 2 e faz-se a análise e discussão dos resultados obtidos.

No quarto capítulo, são apresentadas as principais conclusões obtidas no estudo do capítulo anterior.

CAPÍTULO 2: ENQUADRAMENTO TEÓRICO

Neste capítulo é apresentado o enquadramento teórico dos conceitos utilizados nas Ciências Sociais e das técnicas estatísticas implementadas neste trabalho. Faz-se uma análise detalhada das seguintes técnicas de Estatística Multivariada: análise fatorial e análise de *clusters*.

2.1. Tipos de entrevista

Existem tantos métodos de entrevista quanto diferentes formas de tecnologias de comunicação. À medida que as tecnologias de comunicação evoluem, também proliferam os métodos de entrevista. Os avanços nas tecnologias de *web* e correio eletrónico trouxeram novas possibilidades para as entrevistas isentas de erros.

2.1.1 Entrevista pessoal

O processo de entrevista pessoal é caracterizado pela interação de quatro elementos: o investigador, o entrevistador, o entrevistado e o ambiente (Aaker, 2004).

Os métodos de entrevista pessoal podem ser classificados em entrevistas feitas em residências, em centros comerciais, na rua ou assistidas por computador (Malhotra, 2011).

Vantagens:

- Versatilidade;
- Interação entre o entrevistado e o entrevistador;
- Registo de informações adicionais que poderão ser importantes para a pesquisa (Gomes, 2005).

Desvantagens:

- Custo elevado;
- Tempo de pesquisa extenso;
- Necessidade de pessoas qualificadas e treinadas (Gomes, 2005).

2.1.2 Entrevista telefónica

A entrevista telefónica é o método dominante para a obtenção de informações de grandes amostras.

Existem três abordagens básicas para a seleção dos telefones. Pode-se utilizar uma lista pré-especificada, uma lista telefónica ou um procedimento de discagem aleatória. As listas pré-especificadas – membros de associações, clientes de uma empresa, ou listas vendidas por empresas especializadas – são utilizadas, às vezes, para grupos específicos de pessoas.

A abordagem mais tradicional é a utilização de uma lista telefónica. No entanto, esta lista pode não ser adequada para se conseguir uma amostra representativa dos consumidores ou domicílios, pois há pessoas que não têm os seus telefones na lista. As pessoas que, voluntariamente, decidem não constar na lista telefónica, costumam ter características diferenciadas em relação àqueles que têm os seus números nela.

Para evitar a falta de representatividade, muitos investigadores usam, atualmente, a discagem telefónica aleatória (ver tópico 2.2). De maneira geral, este procedimento significa a seleção aleatória de todos os números de um telefone. Ainda que esta técnica ofereça a hipótese de todos os números serem selecionados, ela tem várias limitações: o seu custo de implementação é alto, tanto em custo como em tempo, uma vez que nem todos os números possíveis são de telefones em uso ou de interesse para o estudo (Aaker, 2004).

Uma vez seleccionados os números de telefone, as ligações são feitas. Quando um telefonema é feito, podem-se obter oito resultados. A Tabela 1 mostra esses diversos resultados e apresenta recomendações de como lidar com cada um deles (Aaker, 2004).

Resultado da Ligação	Recomendações
O telefone está fora de serviço	Eliminar o número de outras considerações.
A ligação dá sinal de ocupado	Chamar novamente mais tarde.
Ninguém atende	Chamar novamente mais tarde.
O número discado é de fax	Enviar um fax pedindo o seu número de telefone para agendar uma entrevista.
A ligação cai num atendedor de chamadas	Deixar uma mensagem dizendo quem ligou e qual o motivo. Ligar mais tarde novamente.
A ligação é atendida por alguém que não o respondente	Perguntar quando o respondente estará disponível e ligar novamente.
A pessoa contactada está fora da amostra	Eliminar o número de outras considerações.
A ligação é atendida pelo respondente esperado	Realizar a entrevista.

Tabela 1. Resultados de uma ligação telefónica e recomendações para lidar com eles (adaptado de Aaker, 2004).

Um dos aspetos mais importantes de uma entrevista telefónica, e a chave para o seu sucesso, é a introdução. Para que a entrevista seja completada com sucesso, o entrevistador precisa de obter empatia imediata com o potencial respondente. Para que isso aconteça, é necessária uma voz agradável ao telefone e uma boa introdução. Esta deve ser breve e explicar os tópicos da pesquisa. Uma introdução muito longa tende a reduzir a cooperação e levar à recusa em participar.

Para facilitar a obtenção de uma amostra representativa dos participantes, as entrevistas telefónicas devem ser feitas nos horários em que é mais provável que os respondentes estejam disponíveis. Para as entrevistas com consumidores, as ligações devem ser feitas entre as 18h e as 21h nos dias úteis, e entre as 10h e as 20h nos fins-de-semana. As ligações antes das 18h nos dias úteis reduzem a probabilidade de encontrar pessoas que trabalham, e depois das 21h podem causar aborrecimentos àqueles que costumam dormir cedo. Por outro lado, o melhor horário para contactar as donas de casa ou as pessoas nas empresas é entre as 9h e as 16h30m.

O relatório das ligações (ANEXO A) é um formulário em que constam os números de telefone a serem chamados e colunas em que os entrevistadores relatam as suas tentativas de comunicação – o dia e o horário das ligações, o resultado, a duração de cada uma delas, e assim por diante. Estes relatórios oferecem o registo das experiências com as ligações e são úteis para gerir a recolha de dados (Aaker, 2004).

As entrevistas telefónicas podem ser tradicionais ou assistidas por computador. A entrevista telefónica tradicional consiste em telefonar para uma amostra de entrevistados e o entrevistador usa um questionário em papel e regista as respostas a lápis.

Os avanços nas telecomunicações e na tecnologia tornaram muito práticas as entrevistas por telefone em todo o país a partir de um *Call Center*.

A entrevista telefónica assistida por computador (CATI – *Computer Assisted Telephone Interviewing*) realizada a partir de um *Call Center* é agora mais utilizada do que a tradicional. Este tipo de entrevista usa um questionário computadorizado que é aplicado aos participantes por telefone. O entrevistador senta-se em frente a um computador e usa um minifone de ouvido. O computador substitui o questionário de papel e o lápis. Quando é feito o contacto, o entrevistador lê as perguntas formuladas no monitor do computador e regista as respostas do entrevistado diretamente na base de dados do computador. O computador guia sistematicamente o entrevistador. A recolha de dados flui de forma natural e suave. O tempo de entrevista é reduzido, a qualidade dos dados é melhorada e eliminam-se as etapas laboriosas no processo de recolha de dados, codificação de questionários e introdução dos dados no computador. Como as respostas são in-

trozidas diretamente no computador, é possível obter quase instantaneamente relatórios atualizados sobre a recolha de dados ou resultados (Malhotra, 2011).

Vantagens:

- Rapidez na obtenção de informações;
- Maior flexibilidade para o entrevistador;
- Custos baixos quando aplicado numa região restrita (Gomes, 2005).

Desvantagens:

- Questionário deve ser objetivo;
- Restrito a pessoas que possuem telefone;
- Dificuldade para encontrar pessoas: deve-se planejar o melhor horário para a realização dos telefonemas (Gomes, 2005).

2.1.3 Entrevista por correio

Neste tipo de entrevista, o questionário é enviado por correio a entrevistados pré-selecionados. Um pacote típico deste tipo de entrevista consiste num envelope de remessa, carta explicativa, questionário, envelope para retorno e, possivelmente, algum incentivo. Os entrevistados preenchem e devolvem os questionários. Não há interação verbal entre o investigador e o entrevistado (Aaker, 2004).

Vantagens:

- Ampla cobertura geográfica;
- Baixo custo por questionário;
- Oferece maior tempo de resposta para o entrevistado (Gomes, 2005).

Desvantagens:

- Retorno baixo: geralmente 1 a 2%, sendo necessário oferecer incentivos, como sorteios e prêmios com o intuito de aumentar o retorno;
- Lentidão de retorno;
- Total dependência do entrevistado (Gomes, 2005).

2.1.4 Entrevista por e-mail

Para se realizar este tipo de entrevista, é preciso obter uma lista de endereços eletrônicos. No corpo da mensagem do *e-mail* redige-se o *link* que direciona para o questionário e é enviado aos entrevistados pela Internet. Os entrevistados digitam, nos lugares indicados, as respostas a questões fechadas ou abertas (Malhotra, 2011).

Vantagens:

- Ampla cobertura;
- Baixo custo;
- Oferece maior tempo de resposta para o entrevistado (Gomes, 2005).

Desvantagens:

- Retorno baixo: necessário incentivo, como brindes, sorteios;
- Pode ser interpretado como *spam* (*e-mail* indesejado);
- Restrito às pessoas que possuem acesso à Internet (Gomes, 2005).

2.2 Técnicas de Amostragem

As técnicas de amostragem podem ser genericamente classificadas como não-probabilísticas (não aleatórias) e probabilísticas (aleatórias). A amostragem não-probabilística confia no julgamento pessoal do investigador. O investigador pode, arbitrariamente ou conscientemente, decidir os elementos a serem incluídos na amostra. As amostras não-probabilísticas podem oferecer boas estimativas das características da população, mas não permitem uma avaliação objetiva da precisão dos resultados amostrais.

Na amostragem probabilística, as unidades amostrais são escolhidas ao acaso, em que cada elemento da população tem probabilidade conhecida de ser incluído na amostra. Veja-se vários procedimentos de amostragem probabilística: Amostragem Aleatória simples, Amostragem aleatória Sistemática, Amostragem Estratificada, Amostragem por Cluster (ou Cachos ou Grupos), Amostragem Multi-etapas, Amostragem Multi-fásica (Malhotra, 2011).

2.2.1. Amostragem não-probabilística

Existem cinco tipos de procedimentos de amostragem não-probabilística: intencional, *Snowball* ou “Bola de Neve”, por conveniência, por quotas e *Random Route*.

2.2.1.1. Amostragem intencional

Neste tipo de amostragem, o investigador usa o seu julgamento ou a sua experiência para identificar as amostras representativas. O investigador escolhe os elementos a serem incluídos na amostra, pois considera-os representativos da população em estudo. Por exemplo, os clientes de um centro comercial podem servir de representantes dos habitantes de uma cidade, ou várias cidades podem ser selecionadas para representar um país (Aaker, 2004).

2.2.1.2. Amostragem *Snowball* ou “Bola de Neve”

A amostragem deste tipo é uma forma de amostragem intencional muito indicada para quando se necessita atingir populações pequenas e especializadas. Por meio do método “Bola de Neve”, cada entrevistado, depois de entrevistado, é solicitado a indicar uma ou mais pessoas do ramo. O resultado pode ser uma amostra bastante útil. Um problema é que aqueles com maior visibilidade social têm maior probabilidade de serem selecionados (Aaker, 2004).

2.2.1.3. Amostragem por conveniência

Para a obtenção rápida e barata de informações, pode-se empregar a amostragem por conveniência. A seleção das unidades amostrais é deixada a cargo do entrevistador. Com frequência, os entrevistados são escolhidos porque se encontram no lugar exato no momento certo.

As amostras por conveniência são frequentemente utilizadas para o pré-teste de questionário (Aaker, 2004).

2.2.1.4. Amostragem por quotas

A amostragem por quotas é uma técnica de amostragem não-probabilística utilizada, fundamentalmente, pela impossibilidade de obtenção de listagens da População. Assim, é a técnica de amostragem mais utilizada em estudos de mercado e opinião.

Na amostra, a proporção de elementos que possuem uma determinada característica é, aproximadamente, igual à proporção de indivíduos na População que possuem essa mesma característica. Deste modo, as amostras são obtidas dividindo a População em categorias ou estratos e, em seguida, seleciona-se um certo número (quota) de indivíduos de cada categoria, de modo não aleatório.

O processo da obtenção de uma amostra por quotas está dividido em dois estágios. O primeiro estágio consiste em definir as quotas, isto é, dividir a População em cate-

gorias. A escolha das variáveis é feita, na maioria dos casos, com base no recenseamento da População, quando se trata de variáveis sociodemográficas.

No segundo estágio, selecionam-se os elementos da amostra com base na conveniência ou julgamento. Uma vez atribuídas as quotas, há considerável liberdade na escolha dos elementos a serem incluídos na amostra. A única exigência é que os elementos selecionados respeitem as quotas estabelecidas no plano de amostragem. Assim sendo, é uma amostragem intencional, com a imposição de que a amostra inclua um número mínimo de cada um dos subgrupos específicos da população. A amostragem por quotas é sempre baseada em dados demográficos, como localização geográfica, idade, sexo, nível educacional e renda.

Este método é análogo ao método de amostragem estratificada (método probabilístico), mas em vez de se escolher uma amostra aleatória dentro de cada um dos estratos, escolhe-se uma amostra não-aleatória de tamanho determinado pela fração de amostragem (Hill, 2008).

2.2.1.5. Amostragem *Random Route*

O método de amostragem que utiliza a amostra *Random Route* é também designado, em português, por Método dos Itinerários Aleatórios. É muito utilizado, quando o estudo é feito dentro de cidades ou localidades, para orientar o entrevistador na seleção dos entrevistados, sempre que as entrevistas decorram, porta-a-porta. Em cada cidade ou localidade é realizado um percurso com viragens pré-definidas à direita e à esquerda, a partir de um ponto inicial previamente estabelecido (Câmara Municipal, Igreja, Escola ou outro).

2.3. Trabalho de campo

Todo o trabalho de campo envolve a seleção, a formação e a supervisão das pessoas que recolhem os dados. A validação do trabalho de campo e a avaliação dos entrevistadores também fazem parte do processo (Malhotra, 2011).

2.3.1. Seleção dos entrevistadores

O primeiro passo no processo do trabalho de campo é a seleção dos entrevistadores. O investigador deve: elaborar os requisitos dos postos de trabalho para o projeto, levando em conta a forma de recolha dos dados; decidir que características devem ter os entrevistadores; recrutar indivíduos qualificados. As características dos entrevistadores, a sua experiência, as suas opiniões, perceções, expectativas e atitudes podem influenciar nas respostas dos entrevistados.

2.3.2. Formação dos entrevistadores

A formação dos entrevistadores é decisiva para a qualidade dos dados recolhidos, pois assegura que todos os entrevistadores apresentem o questionário da mesma forma, garantindo assim uma recolha padronizada dos dados. A formação deve abranger o contato inicial, a formulação de perguntas, a sondagem, o registo de respostas e o encerramento da entrevista.

2.3.3. Supervisão dos entrevistadores

A supervisão dos entrevistadores significa ter a certeza de que eles estão a seguir os processos e as técnicas para que foram treinados. A supervisão abrange o controlo de qualidade e de edição, o controlo de amostras, o controlo de fraudes e o controlo do escritório central (Malhotra, 2011).

2.3.3.1. Controlo da qualidade e edição

O controlo da qualidade dos entrevistadores implica verificar se os processos de trabalho estão a ser implementados adequadamente. Uma vez detetado qualquer problema, o supervisor deve discuti-lo com os entrevistadores e fornecer formação adicional, se necessário. Para entender os problemas dos entrevistadores, o supervisor também deve fazer pessoalmente algumas entrevistas. O supervisor precisa de recolher questionários e alguns formulários e editá-los diariamente. Precisa de examinar os questionários

a fim de se certificar de que todos os requisitos foram preenchidos, de que as perguntas insatisfatórias ou incompletas não foram aceites e de que os dados estão legíveis.

2.3.3.2. Controlo da amostragem

Um aspeto importante da supervisão é o controlo da amostragem, que procura garantir que os entrevistadores estão a seguir exatamente o plano de amostragem, em vez de escolherem unidades de amostragem com base na conveniência ou na facilidade de acesso.

2.3.3.3. Controlo de fraudes

A fraude envolve a falsificação de parte de uma questão ou de todo um questionário. Um entrevistador pode falsificar parte de uma resposta para torná-la aceitável, ou falsificar respostas inteiras. A fraude pode ser evitada mediante formação e supervisão adequada, e pela validação do trabalho de campo.

2.3.3.4. Controlo do escritório central

Os supervisores fornecem ao escritório central informações sobre controlo de qualidade e de custo, a fim de manter um relato do progresso total. O controlo do escritório central inclui o tabelamento de variáveis de quota, características demográficas importantes e respostas de variáveis-chave (Malhotra, 2011).

2.3.4. Validação do trabalho de campo

A validação do trabalho de campo exige verificar se os entrevistadores estão a apresentar questionários autênticos. Para confirmar o estudo, os supervisores telefonam para 10 a 25% dos entrevistados a perguntar se os entrevistadores fizeram efetivamente o seu trabalho.

2.3.5. Avaliação dos entrevistadores

É importante avaliar os entrevistadores, de modo a lhes fornecer elementos para o seu trabalho, bem como identificar os melhores entre eles, para criar uma força de trabalho de alta qualidade.

2.4. Cálculo da dimensão da amostra em amostragem não-probabilística

A amostra é uma parcela da população N com a qual se faz uma pesquisa a partir de um universo previamente definido. Na pesquisa quantitativa, a pesquisa por amostragem utiliza conceitos estatísticos que indicam o número representativo de pesquisas a serem realizadas segundo o universo de que se dispõe.

O método de determinação da dimensão da amostra, quando o procedimento da sua recolha é não aleatório é, segundo Weiers (Vicente et al, 2001), possível de fazer-se de três modos:

- Decidir a dimensão da amostra, tendo em conta o orçamento disponível para o estudo e os custos envolvidos;
- Adotar a dimensão já utilizada, com sucesso, em estudos anteriores das mesmas características;
- Utilizar as fórmulas apresentadas para as amostras aleatórias, sendo a dimensão, assim obtida, meramente indicativa.

Deste modo, estima-se uma proporção com base na amostragem aleatória simples. Ora, a estimação de uma proporção toma, por base, uma População de Bernoulli, na qual uma observação ou pertence ou não pertence à categoria de interesse. A proporção de elementos que caem na categoria que interessa estudar denota-se por p . O estimador deste parâmetro é denotado por \hat{p} .

Na amostra aleatória simples, as observações amostrais são definidas por $X_i = 0$, se o i -ésimo elemento da amostra não possui a característica em estudo e por $X_i = 1$ se a possui. Portanto, o número total de elementos na amostra de dimensão n , a possuir uma especificada característica, é $\sum_{i=1}^n X_i$.

A proporção \hat{p} é a fração de elementos da amostra que possuem a característica de interesse, ou seja, $\hat{p} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$, ou ainda, \hat{p} é a média de valores 0 e 1 da amostra e pode pensar-se na proporção p , como a média de valores 0 e 1 de toda a População, isto é, $p = \mu$.

Pretende determinar-se a dimensão da amostra n , de tal modo que \hat{p} seja um bom estimador de p com precisão não superior a E e grau de confiança $1 - \alpha$, isto é, tal que se verifique, $P(|\hat{p} - p| \leq E) = 1 - \alpha$, $0 < \alpha < 1$, ou seja, $P(\hat{p} - E \leq p \leq \hat{p} + E) = 1 - \alpha$, $0 < \alpha < 1$.

Determina-se o valor de n resolvendo a equação $Z_{\alpha/2} \sqrt{V(\hat{p})} = E$, o que exige o conhecimento de $V(\hat{p})$, ou seja, $V(\bar{X})$.

Como se trata de uma população de Bernoulli, X_i só assume os valores 0 e 1 com probabilidade de, respetivamente, $q = 1 - p$ e p . Assim, a variância de uma população de Bernoulli é $\sigma^2 = pq$, pelo que $V(\hat{p}) = \frac{pq}{n} \left(\frac{N-n}{N-1} \right)$. Então vem $Z_{\alpha/2} \sqrt{\frac{pq}{n} \left(\frac{N-n}{N-1} \right)} = E$. Isto implica $\left(Z_{\alpha/2} \right)^2 \frac{pq}{n} \left(\frac{N-n}{N-1} \right) = E^2$, que resolvendo em ordem a n e tomando $x = \left(Z_{\alpha/2} \right)^2 pq$, conduz a $n = \frac{Nx}{(N-1)E^2 + x}$. No entanto, quando N grande, o fator de correção $\frac{N-n}{N-1}$ de populações finitas torna-se negligenciável. Deste modo, resulta a expressão $n = \frac{x}{E^2}$.

Mas a variância da população N é desconhecida, pelo que p e pq também são desconhecidos. Um valor de n aproximado pode ser encontrado através da substituição de p por um valor estimado \hat{p} .

A distribuição das respostas demonstra o grau de variação das respostas na pesquisa, isto é, o grau de homogeneidade da população. Uma população mais homogénea corresponde a uma população que possua características semelhantes. Assim, uma distribuição das respostas de 50/50 indica muita variação entre as respostas dos entrevistados (população mais heterogénea). Já uma distribuição de 80/20 indica uma menor variação nas respostas (população mais homogénea). Quando não se tem noção do grau de homogeneidade da população, deve-se considerar uma distribuição de 50/50 para se alcançar um maior grau de confiança nas respostas. Deste modo, toma-se $\hat{p} = 0,5$, pois este valor

maximiza a variância da População, conduzindo a um n maior do que o necessário, o que garante o cumprimento da precisão fixada.

2.5. Elaboração do questionário

Um questionário, seja ele chamado de formulário de entrevista ou instrumento de medida, é um conjunto formal de perguntas cujo objetivo é obter informações dos entrevistados.

Qualquer questionário tem três objetivos específicos. Em primeiro lugar, deve transformar a informação desejada num conjunto de perguntas específicas que os entrevistados tenham condições de responder. Em segundo lugar, um questionário precisa motivar e incentivar o entrevistado a deixar envolver-se pela entrevista, a cooperar e a completar a entrevista. Ao planejar um questionário, o investigador precisa sempre de minimizar o cansaço e tédio do entrevistado, esforçando-se também para minimizar as respostas incompletas e a falta de respostas. Em terceiro lugar, um questionário deve sempre minimizar o erro de resposta (Malhotra, 2011).

Segundo Malhotra (2011), o principal ponto fraco da elaboração de um questionário é a falta de teoria. Como não existem princípios científicos que garantam um questionário ótimo ou ideal, a concepção de um questionário é uma competência que se adquire com a experiência.

Numa investigação onde se aplica um questionário, a maioria das variáveis (frequentemente todas as variáveis) são medidas a partir das perguntas do questionário, e portanto, os métodos de investigação incluem os tipos de perguntas usadas, os tipos de respostas associadas a essas perguntas e as escalas de medida dessas respostas.

Assim sendo, aquando da elaboração do questionário, é necessário decidir não só que perguntas utilizar para medir as variáveis a elas associadas, mas também que tipo de resposta é o mais adequado para cada pergunta, que tipo de escala de medida está associado às respostas e que métodos são os corretos para analisar os dados (Hill, 2008).

A concepção de um questionário é apresentada na seguinte série de etapas:

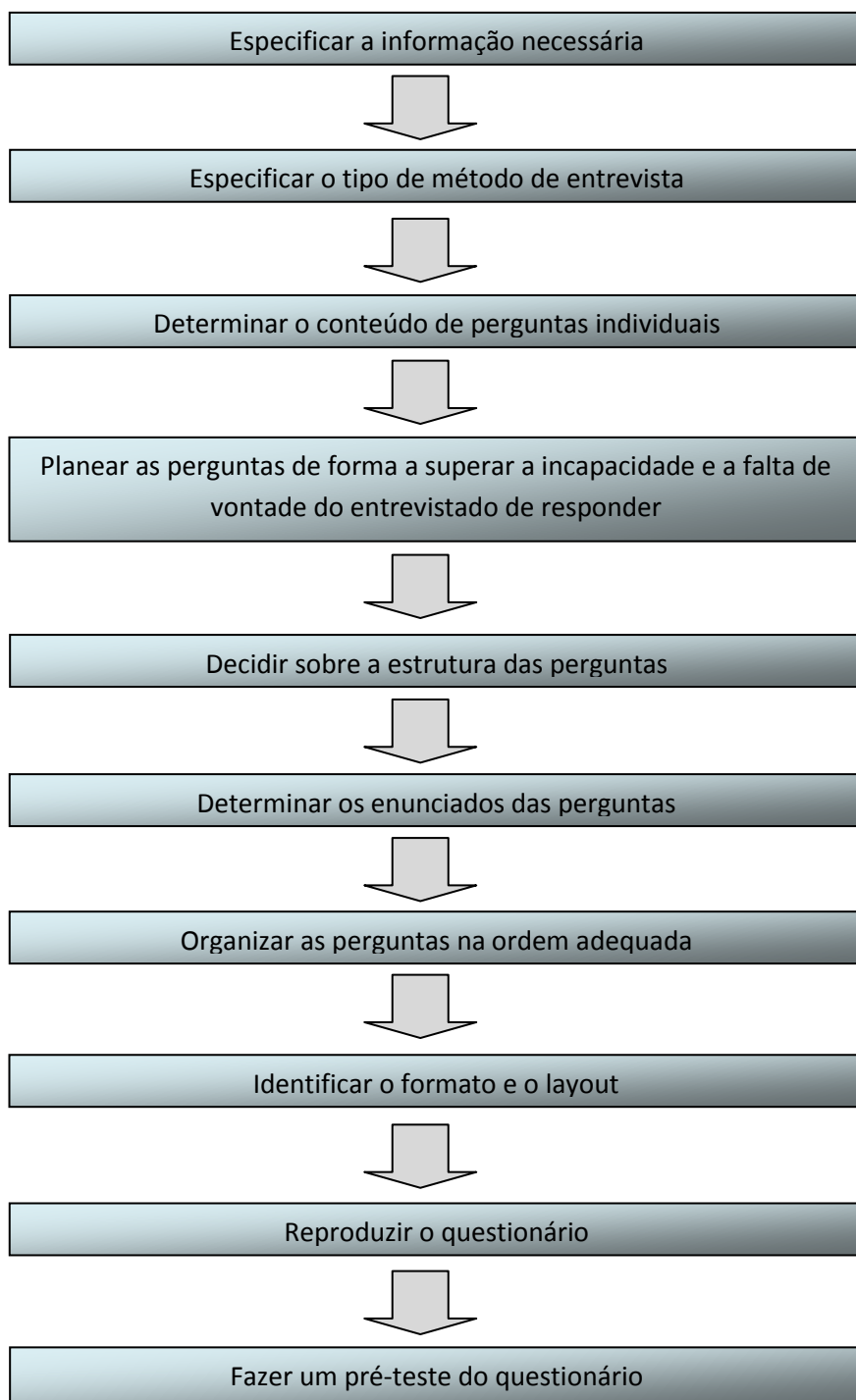


Figura 1. Processo de elaboração de um questionário (adaptado de Malhotra, 2011)

2.6. Análise Fatorial

O objetivo da análise fatorial é analisar a estrutura das correlações entre um grande número de variáveis definindo um conjunto de dimensões comuns subjacentes (fatores). Estes fatores são novas variáveis definidas por combinações lineares das variáveis em análise, as quais, em teoria, vão explicar como é que as variáveis iniciais estão correlacionadas. Grande parte do valor de cada uma das correlações entre as variáveis pode ser explicado em termos das influências dos fatores. O número de fatores é menor (normalmente, muito menor) do que o número de variáveis (Maroco, 2011).

O investigador pode identificar as dimensões da estrutura e então determinar o quanto cada variável é explicada por cada dimensão. Depois disto estar feito, pode-se alcançar as grandes utilidades da análise fatorial, que são: resumir e reduzir dados. Resumir porque as dimensões quando interpretadas e entendidas descrevem os dados num número muito mais pequeno de conceitos do que as variáveis originais. Reduzir porque se calculam os scores para cada dimensão subjacente e substituem-se as variáveis originais.

Existem diversos critérios para decidir o número de fatores a extrair, tais como: valores próprios superiores a 1 (critério de Kaiser); a *priori* por conhecimento do investigador; percentagem de variância (percentagem cumulativa do total de variância explicada pelos fatores sucessivos: ciências naturais (95%) e ciências sociais (60%)); *scree plot* (gráfico dos valores próprios vs. número de fatores por ordem de extração).

A interpretação dos fatores é feita através dos *pesos* dos fatores, que são um meio de interpretar o papel que cada variável tem na definição de cada fator, ou seja, são a correlação de cada variável e fator. Indicam o grau de correspondência entre a variável e o fator (*pesos* elevados tornam a variável representativa do fator).

A rotação dos eixos simplifica a estrutura dos fatores e torna mais fácil a sua interpretação.

Antes da realização da análise fatorial propriamente dita, deve-se investigar se os dados obtidos obedecem a uma distribuição normal e se existe correlação estatística entre as variáveis.

Para verificar se a distribuição é normal, utiliza-se um teste estatístico conhecido como Kolmogorov-Smirnov. É importante realçar que, mesmo que a distribuição não seja normal, a análise fatorial pode ser realizada. Assim, segundo este teste, tem-se duas hipóteses, H_0 : a distribuição é normal e H_1 : a distribuição não é normal. Quando a significância é superior aos α teóricos (0.01, 0.05 e 0.1), aceita-se H_0 . Caso contrário, rejeita-se H_0 .

De modo a descobrir a existência de fortes correlações entre as variáveis, utilizam-se dois testes estatísticos: o de KMO (Kaiser-Meyer-Olkin) e de Esfericidade de Bartlett.

Assim, para se verificar se é adequada a utilização de uma análise fatorial, é comum o cálculo do índice KMO. Este teste apresenta valores normalizados (entre 0 e 1) e mostra qual a proporção da variância que as variáveis apresentam em comum ou a proporção desta que é devida a fatores comuns. Segundo Pereira (2006), se este índice for inferior a 0.5, então é inaceitável a utilização de uma análise fatorial, se for superior a 0.8 a adequação é boa.

O segundo teste, o de Esfericidade de Bartlett, é baseado na distribuição estatística de “qui-quadrado” e testa a hipótese nula (H_0) de que a matriz de correlação é uma matriz identidade (cuja diagonal é 1 e todas as outras as outras iguais a zero), isto é, que não há correlação entre as variáveis (Pereira, 2006). Valores de significância maiores que 0.1 indicam que os dados não são adequados para o tratamento com o método em questão e que a hipótese nula não pode ser rejeitada. Já valores menores que 0.1 permitem rejeitar a hipótese nula (Hair et al, 1998).

Primeiro há que ver se a análise é confirmatória ou exploratória. Se for confirmatória utiliza-se um modelo de equações estruturais. Se for exploratória aplica-se a análise fatorial (por variáveis ou por indivíduos).

Na análise fatorial exploratória não existe qualquer ideia prévia sobre a estrutura dos dados, isto é, sobre o número de fatores comuns, sobre se os fatores são ortogonais

ou oblíquos, sobre o número de indicadores de cada fator e, sobre os indicadores que representam cada fator.

Na análise fatorial confirmatória, pelo contrário, existe já alguma ideia ou teoria sobre a estrutura dos dados: sobre o número de fatores, se são ortogonais ou não, sobre o número de indicadores de cada fator, sobre os indicadores que representam cada fator, etc.

Também é necessário verificar se os dados estão enviesados. Ou seja, pode acontecer que, perante uma afirmação, os sujeitos tenham diversas opiniões. Pode acontecer, ainda, que as respostas tenham alta variabilidade, não porque a assertiva esteja confusa ou mesmo que gere margem a diversas interpretações, mas simplesmente porque os respondentes têm diferentes opiniões a esse respeito, sendo impossível estabelecer algum tipo de padrão nas suas respostas. Este problema é denominado de consistência interna e um dos principais testes que medem este parâmetro é o Alfa de Cronbach (Pestana, 2008).

O Alfa de Cronbach reflete o grau de covariância dos itens entre si. Valores próximos de 1 indicam uma boa consistência interna.

A análise fatorial pode parecer muito semelhante à análise em componentes principais, contudo apresentam algumas diferenças fundamentais nos seus objetivos. Ambas as técnicas podem ser usadas como métodos de redução de dados, mas a análise fatorial não foi concebida para isso. O objetivo da ACP é a redução do número de variáveis para que cada componente explique o máximo da variância dos dados (não há modelo estatístico, apenas uma transformação geométrica). O objetivo da AF é a identificação dos fatores subjacentes aos dados que explicam a estrutura de correlações dos dados (há modelo estatístico). As componentes também são observáveis, enquanto os fatores não, uma vez que são variáveis latentes.

2.6.1. Modelo Fatorial Ortogonal

Seja X uma variável aleatória p -dimensional observável, $X = [X_1 \ X_2 \ \dots \ X_p]^T$, $X_i \in \mathbb{R}$, $i = 1, \dots, p$, com média μ e matriz de covariâncias Σ_X , F uma variável aleatória m -dimensional não observável (fatores comuns ou variáveis latentes), $F = [F_1 \ F_2 \ \dots \ F_m]^T$, $F_i \in \mathbb{R}$, $i = 1, \dots, m$ e ε uma variável aleatória p -dimensional de erros não observável (fatores específicos), $\varepsilon = [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_p]^T$, $\varepsilon_i \in \mathbb{R}$, $i = 1, \dots, p$.

O modelo fatorial assume que as variáveis observáveis X são linearmente dependentes dos fatores comuns F , com $m < p$:

$$\begin{aligned} X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \varepsilon_p. \end{aligned}$$

Ou seja, o modelo é da forma:

$$\begin{matrix} X & - & \mu & = & L & F & + & \varepsilon \\ (p \times 1) & & (p \times 1) & & (p \times m) & (m \times 1) & & (p \times 1) \end{matrix}$$

onde L é a matriz de coeficientes, chamados pesos (contribuições ou *loadings*) dos fatores e l_{ij} é o peso da variável X_i para o fator j , com $i = 1, \dots, p$ e $j = 1, \dots, m$.

Neste modelo assume-se que as variáveis aleatórias não observáveis F e ε são tais que F tem esperança $E[F] = 0$ e matriz de covariâncias identidade, isto é, $\Sigma_F = I_m$ e ε tem esperança $E[\varepsilon] = 0$ e matriz de covariâncias Σ_ε diagonal.

Como consequência, a matriz de covariâncias das variáveis observáveis pode ser expressa em função das contribuições dos fatores, pois

$$\begin{aligned}\Sigma_X &= E[(X - \mu)(X - \mu)^T] = E[(LF + \varepsilon)(LF + \varepsilon)^T] \\ &= LE[FF^T]L^T + E[\varepsilon F^T]L^T + LE[F\varepsilon^T] + E[\varepsilon\varepsilon^T] \\ &= LL^T + \Sigma_\varepsilon\end{aligned}$$

e a matriz de covariâncias entre as variáveis observáveis e os fatores comuns coincide com a matriz de pesos, pois

$$Cov[X, F] = E[(X - \mu)F^T] = LE[FF^T] + E[\varepsilon\varepsilon^T] = L.$$

Note-se que a variância de cada componente observada é composta por duas componentes:

$$V[X_i] = \underbrace{l_{i1}^2 + \dots + l_{im}^2}_{\text{Devida aos fatores comuns}} + \underbrace{\sigma_{\varepsilon i}^2}_{\text{Devida ao fator específico}}.$$

A variância devida aos fatores comuns é também designada por comunalidade, enquanto a variância devida ao fator específico é designada por variância específica (Johnson, 1992).

2.6.2. Métodos de estimação dos fatores

A estimação dos fatores depende do modelo de ajustamento a utilizar e da determinação dos fatores a serem considerados na representação da informação inicial.

A matriz de covariâncias das observações (da população), Σ_X , pode ser estimada pela matriz de covariâncias amostrais, S_X . Se os elementos fora da diagonal principal forem muito pequenos, não há interesse em usar o modelo AF. Se forem significativamente diferentes de 0, a AF faz sentido e interessa começar por estimar os pesos dos fatores comuns e as variâncias específicas.

Existem vários métodos de estimação (ou de extração de fatores), de entre os quais, o método das componentes principais, o método da máxima verosimilhança, o método dos mínimos quadrados, o método da factorização do eixo principal e o método alfa.

2.6.2.1. Método das componentes principais

Segundo Johnson (1992), o método das componentes principais tem como base a decomposição espectral da matriz de covariâncias das observações (da população), Σ_X .

Os valores próprios e os vetores próprios de Σ_X são estimados pelos valores e pelos vetores próprios da matriz de covariâncias amostrais, S_X .

Reescrevendo $\Sigma_X = LL^T + \Sigma_\varepsilon$ na forma

$$\Sigma_X = \lambda_1 e_1 e_1^T + \dots + \lambda_p e_p e_p^T + \Sigma_\varepsilon,$$

com $\lambda_1 \geq \dots \geq \lambda_p$, considera-se que

$$S_X = \hat{\Sigma}_X = \hat{\lambda}_1 \hat{e}_1 \hat{e}_1^T + \dots + \hat{\lambda}_m \hat{e}_m \hat{e}_m^T + \dots + \hat{\lambda}_p \hat{e}_p \hat{e}_p^T + \hat{\Sigma}_\varepsilon$$

é bem aproximada por

$$\hat{\lambda}_1 \hat{e}_1 \hat{e}_1^T + \dots + \hat{\lambda}_m \hat{e}_m \hat{e}_m^T + \hat{\Sigma}_\varepsilon.$$

Sejam X uma variável aleatória p -dimensional observável, $X = [X_1 \ X_2 \ \dots \ X_p]^T$, $X_i \in \mathbb{R}$, $i = 1, \dots, p$, com matriz de covariâncias amostrais S_X e

matriz de correlações R_X , S_X com valores próprios e vetores próprios $(\hat{\lambda}_1, \hat{e}_1), \dots, (\hat{\lambda}_p, \hat{e}_p)$, com $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p$, e $m < p$ o número de fatores comuns.

Então a matriz das estimativas dos pesos dos fatores é

$$\hat{L} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix},$$

as variâncias específicas são estimadas por $\hat{\Sigma}_\varepsilon = \text{diag}(\hat{\sigma}_{\varepsilon_1}^2, \dots, \hat{\sigma}_{\varepsilon_p}^2)$, com

$$\hat{\sigma}_{\varepsilon_i}^2 = S_{ii} - \sum_{1 \leq j \leq m} \hat{l}_{ij}^2,$$

e as variâncias devidas aos fatores comuns são estimadas por $\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots + \hat{l}_{im}^2$.

A AF da matriz de correlações amostrais faz-se substituindo S_X por R_X .

2.6.2.2. Método da máxima verosimilhança

O método da máxima verosimilhança recorre a um algoritmo iterativo para ponderar a matriz das correlações, a partir de uma amostra com distribuição normal multivariada. A função de verosimilhança depende de L e de Σ_ε através de $\Sigma_X = LL^T + \Sigma_\varepsilon$.

Para obter solução única para L , impõe-se a condição $L^T \Sigma_\varepsilon^{-1} L = \Delta$, com Δ diagonal.

Os resultados são obtidos por maximização numérica e os cálculos são realizados computacionalmente.

2.6.2.3. Método dos mínimos quadrados

Este método abrange dois tipos de métodos: mínimos quadrados não ponderados e mínimos quadrados generalizados.

O método dos mínimos quadrados não ponderados é um dos métodos de extração que minimiza a soma das diferenças quadráticas entre a matriz de dados e a matriz de correlação reproduzida, ignorando as diagonais.

O método dos mínimos quadrados generalizados é idêntico ao método anterior, mas neste caso a correlação é pesada pelo inverso das suas singularidades, assim como as variáveis com altas singularidades são tomadas com peso menor do que aquelas com menores singularidades.

2.6.2.4. Método da factorização do eixo principal

De acordo com Maroco (2011), o método da factorização do eixo principal parte do pressuposto de que cada variável tem duas partes constituintes. Uma parte que é comum à estrutura fatorial e uma outra que é específica da variável. Este método, através do quadrado dos coeficientes de correlação, estima novas comunalidades, que substituem a diagonal da matriz, usando iterações sucessivas até satisfazer o critério de convergência da extração. É o método recomendado quando os dados apresentam um desvio severo à normalidade.

2.6.2.5. Método de factorização alfa

É um método de extração que considera as variáveis na análise como uma amostra do universo potencial de variáveis. Ele maximiza a confiabilidade ou fidedignidade alfa de Cronbach dos fatores.

2.6.3. Rotação de fatores

O ideal seria encontrar um padrão nos pesos dos fatores comuns, de modo a que cada variável tivesse um grande peso num só fator comum e pesos muito mais reduzidos nos restantes. Isto nem sempre é possível.

Deste modo, frequentemente recorre-se à rotação dos eixos dos fatores, pois os valores dos coeficientes depois da rotação tenderão a aproximar-se de zero ou de um, facilitando a interpretação dos fatores.

Existem vários métodos de rotação que podemos agrupar em dois tipos: rotações ortogonais e oblíquas.

2.6.3.1. Rotação ortogonal

Devido a questões de identificabilidade, todos os fatores comuns obtidos por transformações ortogonais dos originais, têm a mesma capacidade para reproduzir a matriz de covariâncias (ou de correlações).

Uma transformação ortogonal corresponde a uma rotação dos eixos coordenados que mantêm o seu ângulo de 90°.

Se \hat{L} é a matriz $p \times m$ de pesos fatoriais estimados obtida por um método qualquer (componentes principais, máxima verosimilhança, etc.), então $\hat{L}^* = \hat{L}T$, onde $TT^T = T^T T = I$, é uma matriz $p \times m$ de pesos fatoriais “rodados”.

Após uma rotação dos fatores, a matriz dos resíduos coincide com a original:
$$S_n - \hat{L}^* \hat{L}^{*T} - \hat{\Sigma}_\varepsilon = S_n - \hat{L} \hat{L}^T - \hat{\Sigma}_\varepsilon.$$

Existem três métodos de rotação ortogonal mais referenciados: *varimax*, *quartimax* e *equamax*.

A *varimax* tem como objetivo maximizar a variância dos quadrados dos ponderadores dos fatores nas variâncias, mantendo, todavia, constante a variância comum de cada variável. Deste modo, minimiza o número de variáveis que têm um *peso* elevado para cada fator, simplificando a interpretação dos fatores. Este tipo de rotação é o mais vulgarmente utilizado.

A *quartimax* tem como objetivo maximizar a variância dos ponderadores dos fatores mantendo, todavia, constante a variância comum de cada variável. Minimiza o número de fatores necessários para explicar cada variável. Simplifica a interpretação das variáveis observadas. Tende a produzir soluções com mais variáveis bem correlacionadas com um fator do que a *varimax*.

A *equamax* é uma combinação do método *varimax* que simplifica os fatores e do método *quartimax* que simplifica as variáveis. Minimiza as variáveis que têm um *peso* elevado num fator e o número de fatores necessários para explicar a variável.

2.6.3.2. Rotação oblíqua

Os métodos de rotação oblíquos fazem com que se perca o pressuposto de independência entre os fatores, contudo permitem que estes rodem livremente de maneira a

simplificarem o agrupamento das variáveis e a interpretação dos fatores. Neste tipo de rotação os eixos não mantêm o ângulo de 90°.

Existem dois métodos de rotação oblíqua mais referenciados: *direct oblimin* e *promax*.

No caso da rotação *direct oblimin*, o grau no qual é permitida a correlação dos fatores é determinado pelo valor de uma constante chamada delta. Por norma, toma-se este valor por zero, o que assegura que uma alta correlação entre os fatores não seja permitida.

A *promax* permite que os fatores sejam correlacionados. É útil para conjuntos de dados grandes porque pode ser calculada mais rapidamente do que uma rotação *direct oblimin*.

2.6.4. Métodos de estimação dos valores dos fatores (*factor scores*)

Na AF, a ênfase é geralmente colocada na estimação dos parâmetros do modelo fatorial, isto é, na estimação dos pesos fatoriais e das comunalidades. No entanto, os valores estimados dos fatores comuns, denominados por *factor scores*, também podem ser necessários. Estes valores são muitas vezes utilizados para efeitos de diagnóstico, bem como *inputs* para uma análise posterior.

Como os fatores não são observáveis e correspondem a uma parcela das observações, a sua estimação é complexa. Existem três soluções heurísticas para este problema que são frequentemente utilizadas: o método de Bartlett (ou método dos mínimos quadrados ponderados), o método de Thompson (ou método de regressão) e o método de Anderson-Rubin. Os três métodos utilizam os valores estimados dos pesos fatoriais e das singularidades como se fossem os valores populacionais.

2.6.4.1. Método de Bartlett (ou método dos mínimos quadrados ponderados)

Segundo Johnson (1992), o método de Bartlett utiliza-se com as estimativas dos parâmetros obtidas por máxima verosimilhança. Depois de resolvido o modelo de AF

$$\begin{matrix} X & - & \mu & = & L & F & + & \varepsilon \\ (p \times 1) & & (p \times 1) & & (p \times m) & (m \times 1) & & (p \times 1) \end{matrix}$$

as estimativas dos fatores são obtidas por

$$\hat{f}_j = (\hat{L}^T \hat{\Sigma}_\varepsilon^{-1} \hat{L})^{-1} \hat{L}^T \hat{\Sigma}_\varepsilon^{-1} (x_j - \hat{\mu}) = \hat{\Delta}^{-1} \hat{L}^T \hat{\Sigma}_\varepsilon^{-1} (x_j - \bar{x}), 1 \leq j \leq n,$$

ou, para observações padronizadas,

$$\hat{f}_j = (\hat{L}_Z^T \hat{\Sigma}_{\varepsilon_Z}^{-1} \hat{L}_Z)^{-1} \hat{L}_Z^T \hat{\Sigma}_{\varepsilon_Z}^{-1} z_j = \hat{\Delta}_Z^T \hat{L}_Z^T \hat{\Sigma}_{\varepsilon_Z}^{-1} z_j, 1 \leq j \leq n,$$

onde

$$z_j^T = \left[\frac{x_{1,j} - \bar{x}_1}{\sqrt{s_{11}}} \quad \frac{x_{2,j} - \bar{x}_2}{\sqrt{s_{22}}} \quad \dots \quad \frac{x_{p,j} - \bar{x}_p}{\sqrt{s_{pp}}} \right] (1 \leq j \leq n)$$

$$\text{e } \hat{\rho}_Z = \hat{L}_Z \hat{L}_Z^T + \hat{\Sigma}_{\varepsilon_Z}.$$

As pontuações dos fatores obtidas com observações padronizadas (decomposição da matriz de correlações amostrais) têm média e correlações amostrais nulas.

Se houver rotação T dos eixos, com $\hat{L}^* = \hat{L}T$, então os *factor scores* são dados por $\hat{f}_j^* = T^T \hat{f}_j$.

Se os pesos fatoriais forem estimados pelo método das componentes principais, então as pontuações dos fatores são calculadas por $\hat{f}_j = (\hat{L}^T \hat{L})^{-1} \hat{L}^T (x_j - \hat{\mu})$, $1 \leq j \leq n$, usando a matriz de covariâncias (dados originais), ou $\hat{f}_j = (\hat{L}_Z^T \hat{L}_Z)^{-1} \hat{L}_Z^T z_j$, $1 \leq j \leq n$, usando a matriz de correlações (dados padronizados).

Como a matriz dos pesos fatoriais estimados é $\hat{L} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix}$, os *factor scores* coincidem com as m componentes principais (reduzidas):

$$\hat{f}_j = \begin{bmatrix} \hat{e}_1^T (x_j - \bar{x}) / \sqrt{\hat{\lambda}_1} \\ \hat{e}_2^T (x_j - \bar{x}) / \sqrt{\hat{\lambda}_2} \\ \vdots \\ \hat{e}_m^T (x_j - \bar{x}) / \sqrt{\hat{\lambda}_m} \end{bmatrix}$$

Estes *factor scores* têm média e correlações amostrais nulas.

2.6.4.2. Método de Thompson (ou método de regressão)

Este método também se utiliza com estimativas dos parâmetros obtidas por máxima verosimilhança (supondo distribuições normais).

As estimativas dos fatores obtidas pelo método de regressão para observações originais são da forma $\hat{f}_j = \hat{L}^T S^{-1} (x_j - \bar{x})$, $1 \leq j \leq n$, e para observações padronizadas são da forma $\hat{f}_j = \hat{L}_Z^T R_X^{-1} z_j$, $1 \leq j \leq n$.

2.6.4.3. Método de Anderson-Rubin

Neste método os *factor scores* são dados por $\hat{f}_j = Wz_j$, onde $U^2 = R - \hat{R}$ e $W = (L^T U^{-2} R U^{-2} L)^{-1/2} L^T U^{-2}$, sendo \hat{R} a matriz de correlações estimadas pelo modelo fatorial (Maroco, 2011).

2.6.5. Avaliação da qualidade do modelo fatorial

O modelo fatorial obtido após uma análise fatorial exploratória explícita, teoricamente, a estrutura de fatores latentes responsáveis pelas correlações observadas entre as variáveis originais. Naturalmente, o modelo assume que existe um número de fatores inferiores ao número de variáveis originais que são capazes de explicar uma percentagem elevada da variância total das variáveis originais. As regras do valor próprio superior a 1 e do *scree plot* (escolher apenas os fatores acima do ponto de inflexão) são geralmente utilizadas para decidir qual o número mínimo de fatores necessários para explicar uma proporção considerável da variância total das variáveis originais. Contudo, estas regras apenas ajudam a selecionar os fatores necessários para explicar a covariância observada entre as variáveis originais e nada dizem sobre a qualidade do modelo fatorial deduzido, isto é, se o modelo ajustado explica ou não as correlações observadas entre as variáveis originais. Deste modo, é necessário averiguar quão bem a estrutura fatorial deduzida se ajusta aos dados.

A avaliação da qualidade do modelo pode ser feita de forma heurística, comparando as correlações observadas entre as variáveis originais com as correlações estimadas pelo modelo fatorial retido. As correlações estimadas pelo modelo, entre duas variáveis X_i e X_k , são dadas pela soma dos produtos dos pesos fatoriais dessas variáveis com todos os m fatores retidos:

$$\hat{\rho}_{ik} = \sum_{j=1}^m l_{ij} l_{kj}$$

Assim, pode-se comparar as correlações observadas (r_{ik}) com as correlações re-produzidas pelo modelo ($\hat{\rho}_{ik}$). Se as diferenças entre ambas as correlações forem muito elevadas, o modelo fatorial retido reproduz mal a estrutura correlacional observada. O SPSS calcula a matriz dos resíduos (e) e, de uma forma empírica, considera-se que uma percentagem elevada de resíduos inferiores a 0.05 (mais de 50%) é indicador de um modelo fatorial com bom ajustamento (Maroco, 2011).

Um outro índice que é vulgarmente utilizado é o *Goodness of Fit Index (GFI)*. A fórmula generalizada do *GFI* é:

$$GFI = 1 - 0.5tr^*(e)^2,$$

onde tr^* representa a função traço da matriz de correlações sem a diagonal principal de 1's. O *GFI* varia normalmente entre 0 e 1 e pode interpretar-se como sendo a fração das correlações entre as variáveis originais que são explicadas pelo modelo. De uma forma geral, considera-se que o modelo se ajusta razoavelmente bem aos dados quando o *GFI* é superior a 0.9. Valores de *GFI* acima de 0.95 são indicadores de um ajustamento muito bom. No entanto, o *GFI* tende a sobrestimar o verdadeiro valor da qualidade do ajustamento pelo que deve ser ajustado pelos graus de liberdade (gl) relativamente ao número de covariâncias não redundantes (k) do modelo fatorial ajustado. O *GFI* ajustado é dado por:

$$AGFI = 1 - \left(\frac{k}{gl}\right)(1 - GFI)$$

onde, no método da máxima verosimilhança, $k = p(p + 1)/2$ e $gl = [(p - m)^2 - p - m]/2$.

Um outro índice que é também comum usar na avaliação da qualidade do ajustamento é o *Root Mean Square Residual (RMSR)* calculado a partir do quadrado médio entre as diferenças das covariâncias estimadas pelo modelo e as covariâncias observadas:

$$RMSR = \sqrt{\frac{\sum_{i=1}^p \sum_{j=1}^p (\sigma_{ij} - s_{ij})^2}{\frac{p(p+1)}{2}}}$$

Estandarizando as covariâncias obtém-se as correlações estimadas e observadas. Contudo, as variâncias estandarizadas resultam na correlação de cada variável com ela própria ($r_{ii} = 1$), pelo que eliminando esses elementos da análise, e usando apenas as médias das correlações não redundantes, a expressão anterior escreve-se da forma:

$$RMSR^* = \sqrt{\frac{\sum_{i=1}^{p-1} \sum_{j=i+1}^p (\rho_{ij} - r_{ij})^2}{p(p-1)/2}}$$

Desta forma, considera-se o ajustamento como inaceitável para $RMSR^*$ superior a 0.1, bom para $RMSR^*$ inferior a 0.1 e muito bom para $RMSR^*$ inferior a 0.05 (Maroco, 2011).

2.7. Análise de Clusters

A análise de *clusters* ou de conglomerados tem como objetivo principal o agrupamento de indivíduos ou variáveis com base num conjunto de variáveis, de modo a que os indivíduos ou variáveis dentro de cada grupo (*cluster* ou conglomerado) sejam similares entre si e que indivíduos ou variáveis de diferentes grupos sejam pouco semelhantes. Interessa encontrar grupos naturais que façam sentido para o utilizador.

O número de *clusters* (k) não pode ser maior do que o número de variáveis (p), e, normalmente, k é muito menor do que p .

A dificuldade inicial reside no facto de não existir uma única via de definição de grupos, isto é, um único critério de partição ou agrupamento dos indivíduos ou variáveis com base numa única medida de semelhança ou dissemelhança.

2.7.1. Medidas de dissemelhança

A identificação de agrupamentos naturais de indivíduos ou casos exige que a semelhança entre estes possa ser medida de uma forma mais ou menos explícita, e o menos subjetiva possível. Existem várias medidas de semelhança cuja utilização implica algum conhecimento da matriz de dados. De um modo geral, os indivíduos são agrupados segundo algum tipo de distância métrica (medida de dissemelhança), enquanto o agrupamento de variáveis é baseado em medidas de correlação ou associação apropriadas (Maroco, 2011).

Existem numerosas medidas de semelhança e de distância (dissemelhança) que dependem da natureza das características que são observadas nos objetos, sendo que as mais utilizadas são:

1. Distância euclidiana

Esta medida de dissemelhança métrica mede o comprimento da reta que une duas observações num espaço p -dimensional. Para p -variáveis, a distância euclidiana entre os indivíduos i e j é dada por:

$$D_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$
$$= \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}$$

onde x_{ik} é o valor da variável k no sujeito i e x_{jk} é o valor da variável k para o sujeito j .

De igual modo, de forma matricial, $D_{ij} = \sqrt{(x_i - x_j)^T (x_i - x_j)}$, onde x_i e x_j representam os vetores p -dimensionais das observações correspondentes aos indivíduos i e j . A distância euclidiana pode ser calculada a partir das variáveis originais ou das variáveis padronizadas.

2. Distância de Minkowski

Esta medida de dissimilaridade métrica pode considerar-se uma generalização da distância euclidiana, dada por (Johnson, 1992):

$$D_{ij} = \sqrt[m]{\sum_{k=1}^p |x_{ik} - x_{jk}|^m}$$

onde para $m = 1$, D_{ij} é o módulo da distância absoluta entre os sujeitos i e j relativamente às p -variáveis medidas, denominada por distância *city-block*; e para $m = 2$ obtém-se a distância euclidiana.

3. Distância de Mahalanobis

Esta medida de dissimilaridade métrica mede a distância estatística entre dois indivíduos i e j (Maroco, 2011):

$$D_{ij} = \sqrt{(x_i - x_j)^T S^{-1} (x_i - x_j)}$$

onde S é uma estimativa amostral da matriz de covariâncias Σ dentro dos grupos ou *clusters*. Esta medida tem em conta a correlação entre as variáveis e é particularmente útil quando a matriz Σ é diferente da matriz identidade, caso contrário, a distância de Mahalanobis reduz-se à distância euclidiana.

Muitas vezes interessa agrupar variáveis em vez de indivíduos. Para isso basta efetuar a AC sobre as linhas da transposta da matriz de dados. As variáveis tomam o lugar dos indivíduos e as medidas de proximidade entre indivíduos já podem ser usadas. No entanto, para análise de variáveis, as medidas mais adequadas são as de correlação e de associação. Nestes casos, por vezes, à correlação negativa é atribuído o valor 0 (Maroco, 2011).

2.7.2. Critérios de agregação hierárquicos

Há dois tipos principais de métodos de análise de conglomerados: métodos hierárquicos e métodos não hierárquicos. Os métodos hierárquicos dividem-se em dois métodos: aglomerativos e divisivos. Os métodos aglomerativos partem de um número de grupos igual ao número de observações e vão, sucessivamente, aglomerando os dados. Em cada etapa do método, cada indivíduo ou variável ou cada grupo é atribuído a um grupo. Nos métodos divisivos, o processo é oposto ao anterior, isto é, um grupo inicial divide-se em dois subgrupos, de modos a que os indivíduos ou variáveis de um estejam afastados dos indivíduos ou variáveis do outro, e assim sucessivamente, até haver tantos grupos como indivíduos ou variáveis.

A estrutura hierárquica proveniente destes procedimentos pode representar-se por um gráfico bidimensional ao qual se dá o nome de dendrograma.

Os métodos não hierárquicos são aplicáveis em mais situações do que os anteriores. Nestes métodos os indivíduos são divididos em k grupos. O número de k grupos pode ser definido à partida ou ao longo do próprio processo.

Neste trabalho apenas se descrevem os métodos hierárquicos aglomerativos.

Os métodos com utilização mais frequente são (Maroco, 2011):

1. Menor distância (*Single linkage* ou *Nearest neighbor*)

Neste método, após a formação do primeiro *cluster*, a distância deste aos restantes indivíduos ou variáveis é a menor das distâncias de cada um dos elementos constituintes deste *cluster* a cada um dos restantes indivíduos ou variáveis, isto é,

$$d_{(i-j)-k} = \min \{d_{i-k}; d_{j-k}\}.$$

2. Maior distância (*Complete linkage* ou *Farthest neighbor*)

Neste método, após a formação do primeiro *cluster*, a distância deste aos restantes indivíduos ou variáveis é a maior das distâncias de cada um dos elementos constituintes deste *cluster* a cada um dos restantes indivíduos ou variáveis, isto é,

$$d_{(i-j)-k} = \max \{d_{i-k}; d_{j-k}\}.$$

3. Distância média entre *clusters* (*Average linkage between groups*)

Neste método, após a formação do primeiro *cluster*, a distância deste aos restantes indivíduos ou variáveis é a média das distâncias de cada um dos elementos constituintes deste *cluster* a cada um dos restantes indivíduos ou variáveis, isto é,

$$d_{(i-j)-k} = \text{média} \{d_{i-k}; d_{j-k}\}.$$

4. Distância média entre *clusters* (*Average linkage within groups*)

Este método é semelhante ao método anterior, mas aqui os *clusters* são unidos de modo a que a soma de quadrados dos erros (isto é, a variabilidade dentro dos *clusters*) seja a menor possível.

5. Distância mediana (*Median linkage*)

Neste método, após a formação do primeiro *cluster*, a distância deste aos restantes indivíduos ou variáveis é a mediana das distâncias de cada um dos elementos constituintes deste *cluster* a cada um dos restantes indivíduos ou variáveis, isto é,

$$d_{(i-j)-k} = \text{mediana} \{d_{i-k}; d_{j-k}\}.$$

6. Método do centróide

Neste método, o novo *cluster* formado é representado por um ponto cujas coordenadas são a média dos indivíduos ou variáveis que fazem parte do cluster para cada uma das variáveis (isto é, pelo centróide). As novas distâncias são então calculadas usando estas médias e os valores dos restantes indivíduos ou variáveis.

7. Método de Ward

Neste método não são calculadas distâncias e os *clusters* são formados de modo a minimizar a soma dos quadrados dos erros. Em cada um dos passos do algoritmo, os *clusters* retidos são aqueles, de todos os possíveis, com a menor soma de quadrados dos erros, isto é, a menor perda de informação resultante da aglutinação.

Não existe um melhor critério de agregação em AC. É prática comum utilizar vários critérios e para cada um deles experimentar várias (dis)semelhanças e comparar os resultados obtidos. Se todos eles produzirem soluções interpretáveis semelhantes, é possível concluir que a matriz de dados apresenta agrupamentos naturais (os resultados são fidedignos).

2.7.3. Quantos *clusters* se devem reter?

A aplicação dos métodos hierárquicos permite a apresentação dos resultados sob a forma de um dendrograma. A questão que se coloca é por onde cortar o dendrograma por forma a obter um número de grupos ótimo.

Existem alguns métodos mais ou menos heurísticos que permitem avaliar a solução de *clusters* escolhida e o número de *clusters* a reter. Um dos que podem ser calculados com alguma facilidade a partir dos outputs produzidos pelo SPSS é a distância entre *clusters*.

A distância entre *clusters* pode ser obtida da coluna de coeficientes na tabela “*Agglomeration Schedule*” que representa a medida de proximidade usada para juntar os *clusters*. Se a distância entre dois *clusters* é pequena, estes devem ser agregados, caso contrário, devem manter-se separados. Um gráfico do número de *clusters* em função dos seus coeficientes de aglomeração permite sugerir o número de *clusters* naturais a reter quando o declive da reta que une a distância entre dois *clusters* for relativamente pequeno, não esquecendo que o número de *clusters* é oposto ao número de passos (Maroco, 2011).

CAPÍTULO 3: CASO DE ESTUDO

No início deste capítulo será apresentada uma descrição da base de dados em estudo. De seguida é apresentado todo o tratamento estatístico efetuado, com aplicação das técnicas estatísticas referidas anteriormente. O *software* utilizado no tratamento de dados foi o *SPSS*.

3.1. Metodologia (Ficha Técnica)

1. UNIVERSO

Constituído por indivíduos com mais de 16 anos, residentes em lares com telefone das regiões NUTS II: Norte, Centro, Lisboa, Alentejo e Algarve.

2. AMOSTRA/AMOSTRAGEM

A amostra é constituída por 601 entrevistas.

Amostragem multietápica: seleção aleatória sistemática para a determinação da 1ª unidade de amostragem (lar), a partir da Lista Telefónica; para a seleção do inquirido, 2ª unidade de amostragem, foi utilizado o método de “Quotas” ou proporcional, tendo como variáveis de controlo a “Região Nuts II” e o “Número de Pessoas que residem na habitação”.

Cálculo da dimensão da amostra:

Pretende-se uma amostra que conduza à obtenção de um intervalo de confiança de 95% para a proporção com uma margem de erro de 4%.

Sendo o tamanho da população $N = 3\ 869\ 531$ famílias portuguesas e tomando

$$x = Z_{\alpha/2}^2 pq = 1.96^2 \times 0.5(1 - 0.5) = 0.9604,$$

tem-se que o tamanho da amostra é

$$n = \frac{Nx}{(N - 1)E^2 + x} = \frac{3869531 \times 0.9604}{(3869531 - 1) \times 0.04^2 + 0.9604} = 600.1571.$$

Logo, tendo um grau de confiança 95% e um erro de 4%, a dimensão da amostra é 601.

Os dados sobre a população em estudo foram retirados da página de Internet do INE.

Nº de Pessoas NUTS II	Uma Pessoa	Duas Pessoas	Três Pessoas	Quatro Pessoas	Cinco ou Mais Pessoas
Norte	229 097	390 623	349 815	256 371	105 160
Centro	195 411	300 976	209 069	149 560	49 730
Lisboa	293 394	376 934	255 811	160 520	61 252
Alentejo	71 680	103 674	68 481	44 372	14 737
Algarve	45 581	60 709	40 602	26 387	9 591

Tabela 2. População dividida por NUTS II e pelo número de pessoas que habitam uma residência

Nº de Pessoas NUTS II	Uma Pessoa	Duas Pessoas	Três Pessoas	Quatro Pessoas	Cinco ou Mais Pessoas
Norte	36	61	54	40	16
Centro	30	47	33	23	8
Lisboa	45	58	40	25	10
Alentejo	11	16	11	7	2
Algarve	7	9	6	4	2

Tabela 3. Amostra por quotas segundo os NUTS II e o número de pessoas que habitam uma residência.

3. MÉTODO DE RECOLHA DE DADOS

A recolha foi feita por entrevista telefónica assistida por computador (CATI), utilizando um questionário semi-estruturado (ie, questionário com questões abertas e questões fechadas).

4. TRABALHOS DE CAMPO

A recolha de informação foi realizada por uma equipa de entrevistadores do GE-MEO, com formação específica sobre a forma de abordagem do entrevistado e a condução da própria entrevista. Todos os entrevistadores receberam “*briefing*” adequado às especificidades do estudo. Os trabalhos de recolha de informação decorreram entre os dias 18 e 24 de Maio de 2012.

5. SUPERVISÃO E CONTROLE DE QUALIDADE

A supervisão em campo foi realizada “*in loco*” pela autora. Todas as entrevistas foram objeto de revisão e posterior codificação. A informação foi gravada em suporte informático em tempo real, sendo objeto de testes automáticos de consistência no programa *Sphinx*, especialmente concebido para Estudos de Mercado.

3.2. Análise de Resultados

A presente análise de resultados foi realizada com base em inquéritos aos habitantes das regiões NUTS II com mais de 16 anos (ANEXO E).

3.2.1. Caracterização sociodemográfica da amostra

A amostra que se recolheu é constituída por 601 famílias portuguesas, tendo as características que se encontram na seguinte tabela:

Características	N(601)	%
Idade do entrevistado		
< 20	6	1,0
20 – 40	91	15,1
40 – 60	203	33,8
≥ 60	301	50,1
Sexo do entrevistado		
Masculino	177	29,5
Feminino	424	70,5
Estado civil do entrevistado		
Solteiro	82	13,6
Casado	365	60,7
União de facto	9	1,5
Divorciado	32	5,3
Viúvo	113	18,8
Pessoas na habitação		
Uma	129	21,5
Duas	191	31,8
Três	144	24,0
Quatro	99	16,5
Cinco ou mais	38	6,3
Adultos		
Um	141	23,5
Dois	270	44,9
Três ou mais	190	31,6
Crianças		
Nenhuma	472	78,5
Uma	76	12,6
Duas	43	7,2
Três ou mais	10	1,7
Rendimento mensal do agregado familiar (€)		
< 475	139	23,1
475 – 1.000	228	37,9
1.001 – 1.500	95	15,8
1.501 – 2.000	64	10,6
2.001 – 2.500	28	4,7
2.501 – 3.000	26	4,3
> 3.000	21	3,5

Tabela 4. Características sociodemográficas da amostra.

3.2.2. Análise comparativa Portugal vs Europa

Procedeu-se de seguida a uma Análise de Conglomerados (AC) com o objetivo de identificar países com consumos semelhantes ao nível do tipo de família, através do SPSS.

Os dados sobre os países que constituem a União Europeia foram obtidos através da página de Internet do Eurostat.

Dado que se pretende um agrupamento de indivíduos utilizou-se uma medida de dissemelhança. Para os vários métodos aplicados efetuou-se a análise com diferentes distâncias e a alteração do tipo de distância não provocou grandes alterações nos resultados. Optou-se pelo quadrado da distância euclidiana visto ser a mais usual.

Foram aplicados vários métodos hierárquicos aglomerativos e os resultados obtidos na composição dos *clusters* foram semelhantes para os métodos da ligação completa, da ligação média e de Ward, pelo que é possível concluir que os resultados obtidos são fidedignos. Utilizou-se o método de Ward para identificar os grupos por ser aquele que tem em consideração a menor perda de informação resultante da aglutinação.

Como existem seis tipos de famílias, agruparam-se os países segundo as características (variáveis de consumo) de cada uma delas. Deste modo, obtiveram-se seis dendrogramas.

No caso das famílias de um adulto sem crianças, a análise do dendrograma (método empírico) na Figura 2 sugere que existem pelo menos cinco *clusters* distintos de indivíduos, neste caso, países, como se pode verificar na Figura 4.

Outro método de avaliar o número de *clusters* ideal baseia-se no valor da medida de proximidade usada para juntar os *clusters*, que se encontra na coluna “coeficientes” da tabela de agrupamento do Anexo B. O gráfico da Figura 3, onde estão representados para cada etapa os respetivos coeficientes de aglomeração, permite visualizar as diferenças das distâncias entre duas etapas consecutivas e verifica-se que o primeiro grande aumento destas distâncias ocorre da etapa 21 para a 22. Seguindo a junção de indivíduos assinalada na coluna “*Cluster Combined*” da tabela de agrupamento do ANEXO B até à etapa 21 obtém-se 5 *clusters*. Deste modo, a nível de consumo, Portugal assemelha-se à Grécia, à Áustria, à República Checa, à Hungria e ao Chipre. Os países da Europa do Norte têm um comportamento semelhante entre si.

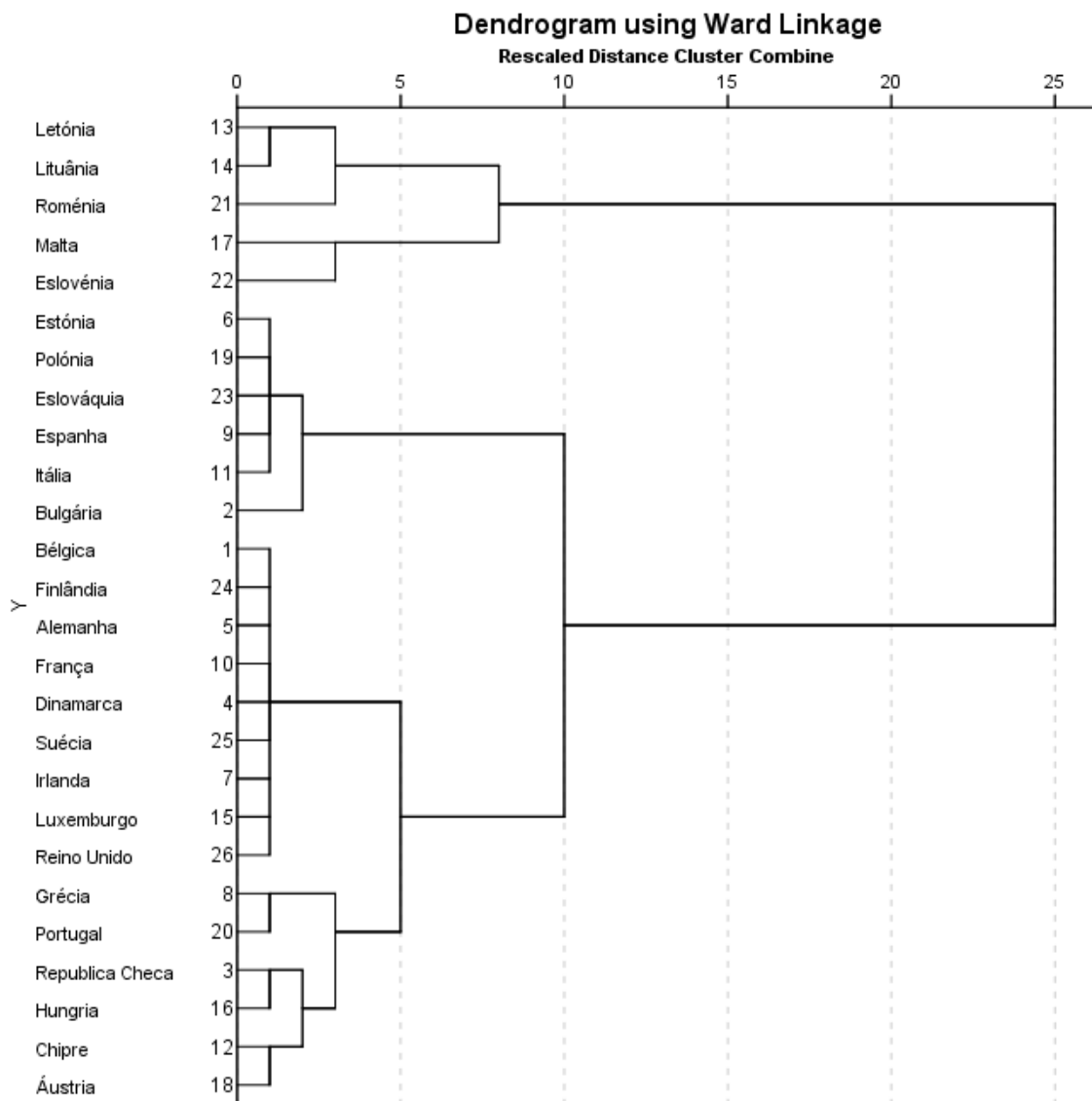


Figura 2. Dendrograma da família de um adulto sem crianças

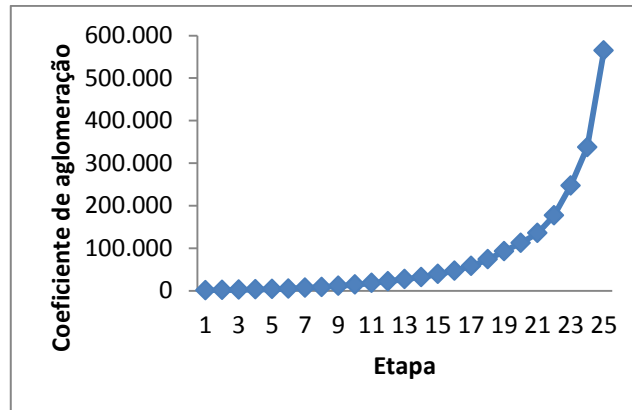


Figura 3. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de um adulto sem crianças

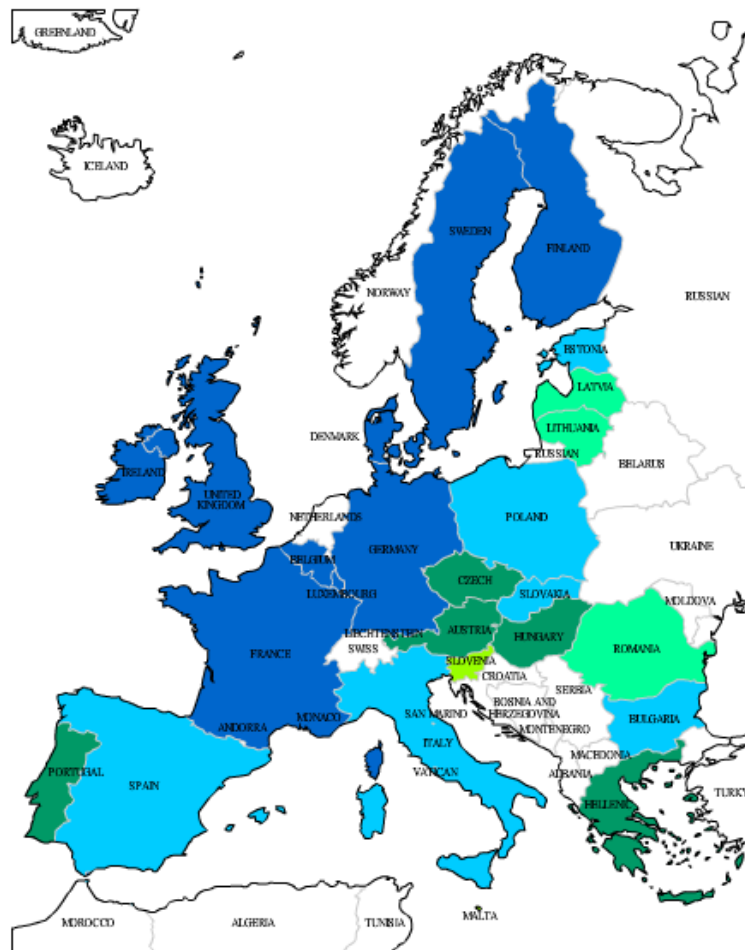


Figura 4. Distribuição geográfica das famílias de um adulto sem crianças segundo os *clusters* do dendrograma obtido na Figura 2.

No caso das famílias de um adulto com crianças, a análise do dendrograma na Figura 5 sugere que também existem pelo menos quatro *clusters* distintos de indivíduos, neste caso, países, como se pode verificar na Figura 7. Efetuando também o método da distância entre *clusters*, no gráfico da Figura 6, verifica-se que o primeiro grande aumento destas distâncias ocorre da etapa 22 para a 23. Seguindo a junção de indivíduos assinalada na coluna “*Cluster Combined*” da tabela de agrupamento do ANEXO B até à etapa 22 obtêm-se 4 *clusters*. Sendo assim, verifica-se que toda a União Europeia, a nível de consumo, tem comportamento semelhante, exceto a Europa do Leste.

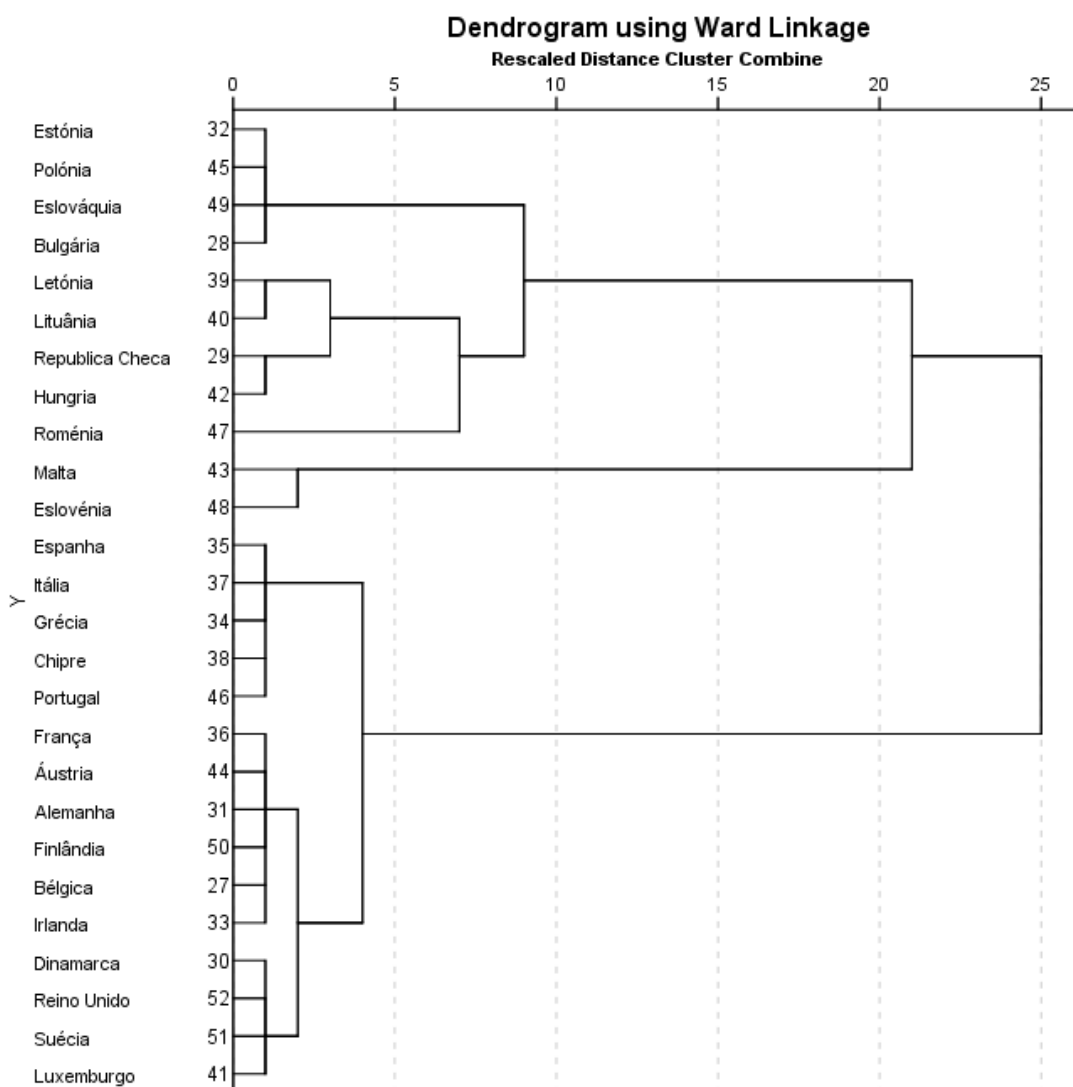


Figura 5. Dendrograma da família de um adulto com crianças

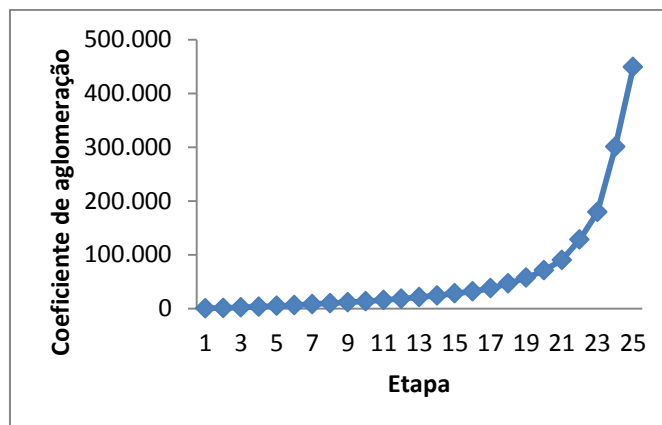


Figura 6. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de um adulto com crianças

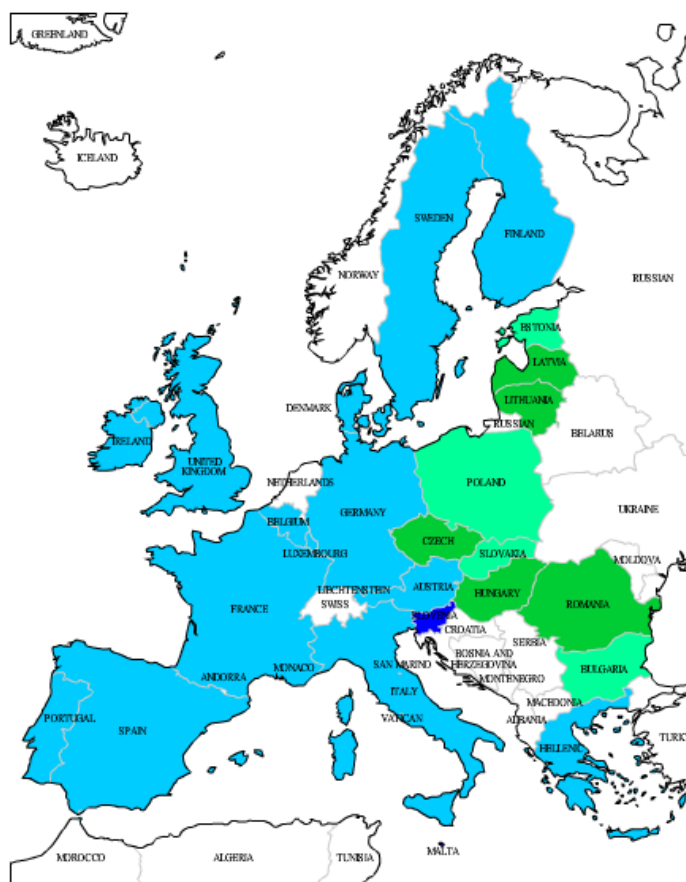


Figura 7. Distribuição geográfica das famílias de um adulto com crianças segundo os *clusters* do dendrograma obtido na Figura 5.

A análise do dendrograma na Figura 8, relativo a famílias de dois adultos sem crianças, sugere que existem pelo menos cinco *clusters* distintos de indivíduos, neste caso, países, como se pode verificar na Figura 10. Efetuando de novo o método da distância entre *clusters*, no gráfico da Figura 9, verifica-se que o primeiro grande aumento destas distâncias ocorre da etapa 21 para a 22. Seguindo a junção de indivíduos assinalada na coluna “*Cluster Combined*” da tabela de agrupamento do ANEXO B até à etapa 21 obtêm-se 5 *clusters*. Deste modo, a nível de consumo, Portugal assemelha-se à Grécia, à Malta, à Republica Checa, à Hungria e ao Chipre. Os países da Europa do Norte têm um comportamento semelhante entre si.

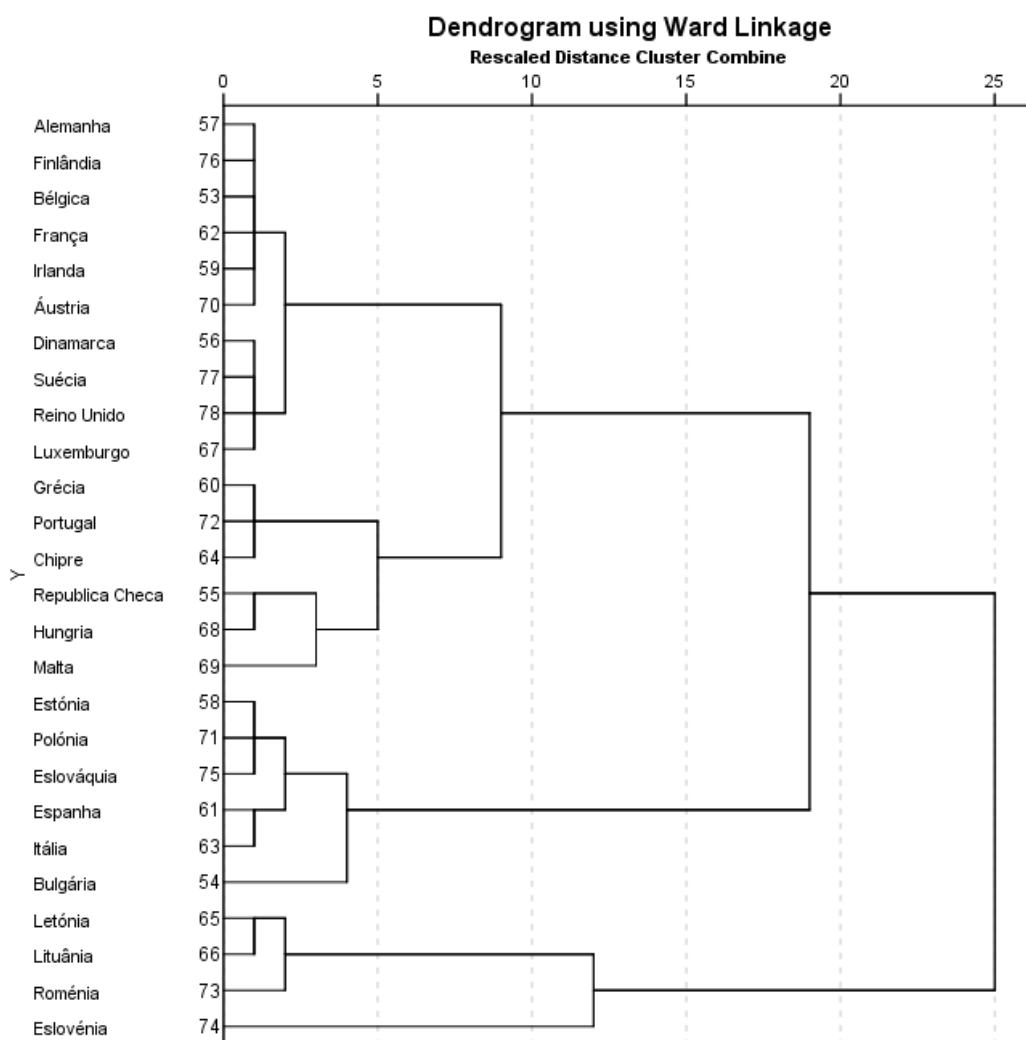


Figura 8. Dendrograma da família de dois adultos sem crianças

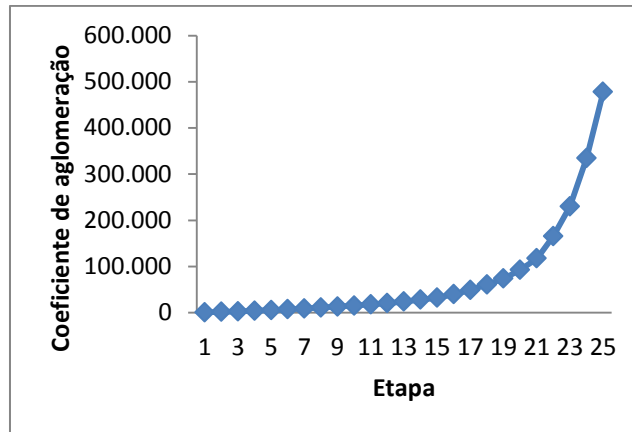


Figura 9. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de dois adultos sem crianças

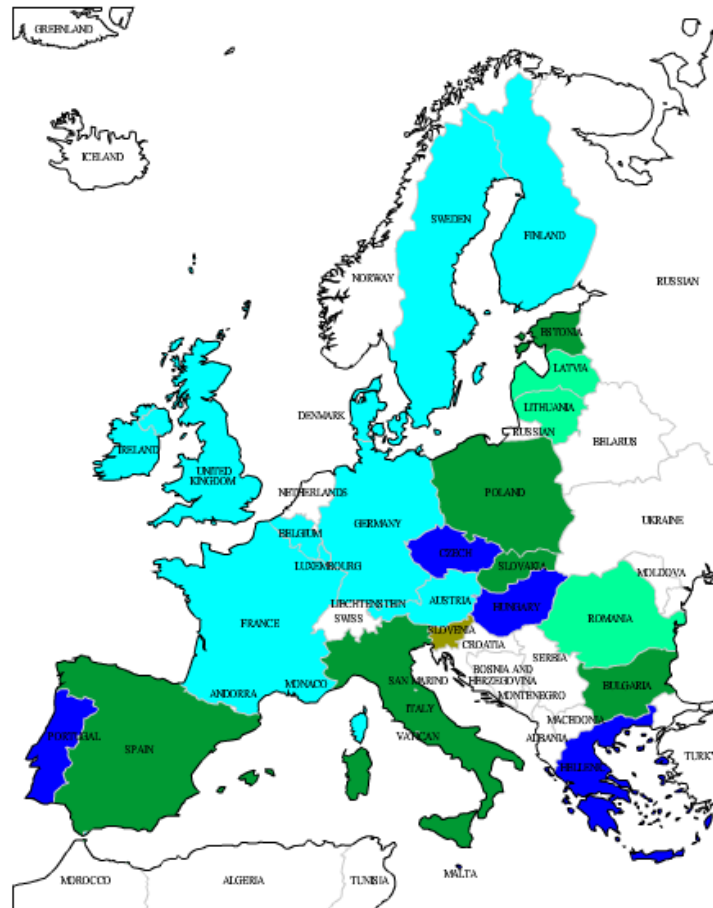


Figura 10. Distribuição geográfica das famílias de dois adultos sem crianças segundo os *clusters* do dendrograma obtido na Figura 8.

No caso das famílias de dois adultos com crianças, a análise do dendrograma na Figura 11 sugere que existem pelo menos quatro *clusters* distintos de indivíduos, neste caso, países, como se pode verificar na Figura 13. O método da distância entre *clusters* sugere o mesmo, visto que no gráfico da Figura 12 verifica-se que o primeiro grande aumento destas distâncias ocorre da etapa 22 para a 23. Seguindo a junção de indivíduos assinalada na coluna “*Cluster Combined*” da tabela de agrupamento do ANEXO B até à etapa 22 obtêm-se 4 *clusters*. Sendo assim, verifica-se que toda a União Europeia, a nível de consumo, tem comportamento semelhante, exceto a Europa do Leste.

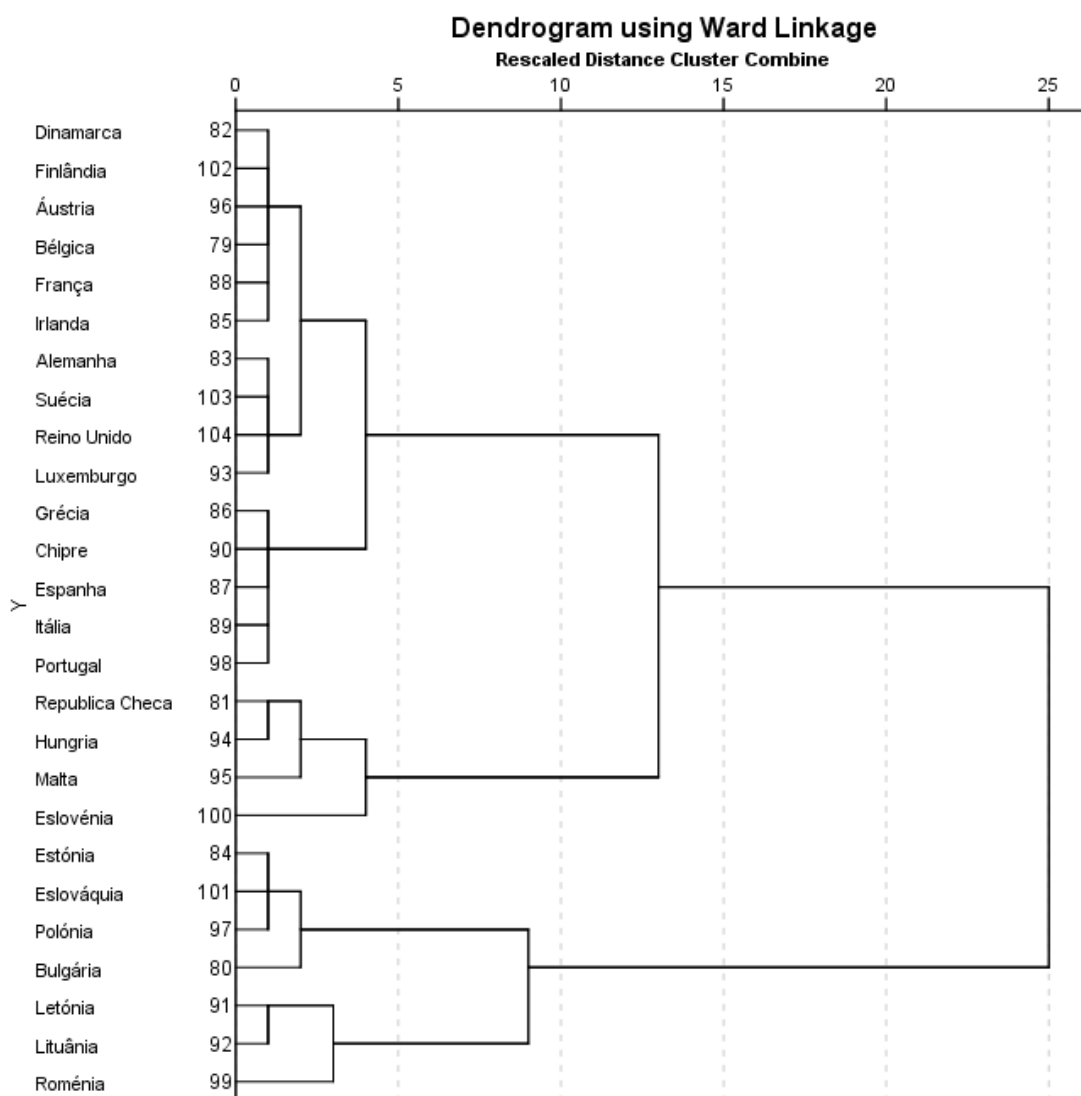


Figura 11. Dendrograma da família de dois adultos com crianças

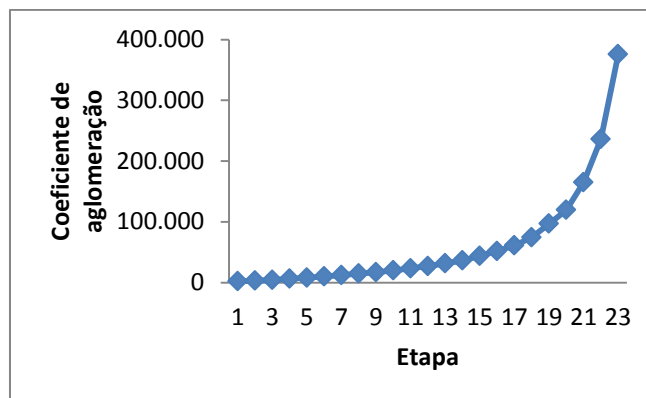


Figura 12. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de dois adultos com crianças

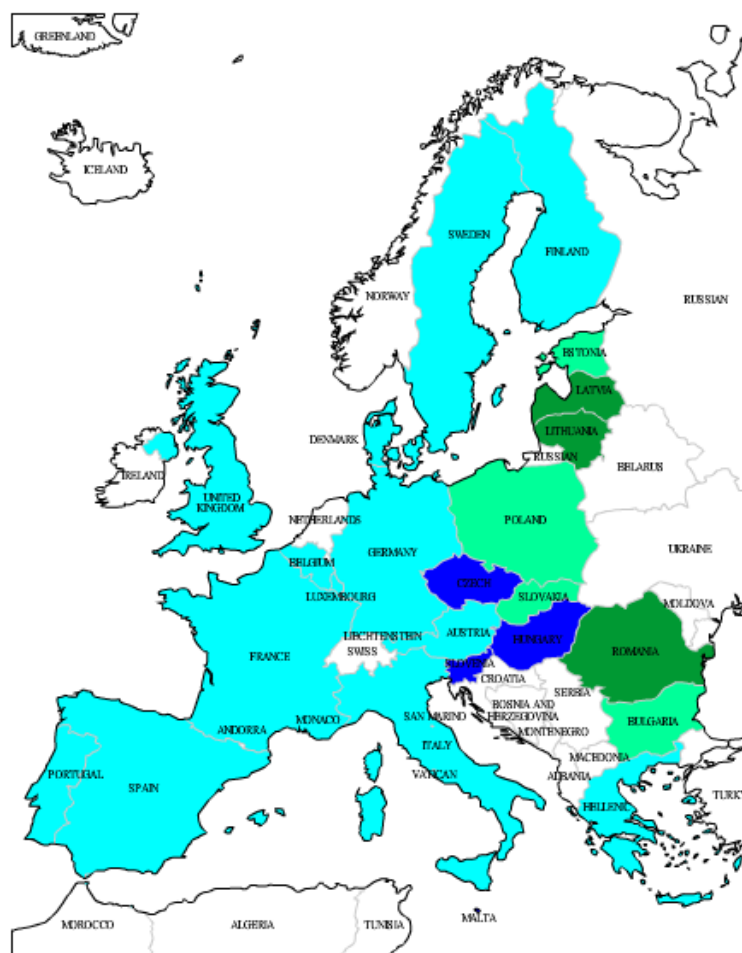


Figura 13. Distribuição geográfica das famílias de dois adultos com crianças segundo os *clusters* do dendrograma obtido na Figura 11.

No caso das famílias de três ou mais adultos sem crianças, a análise do dendrograma na Figura 14 sugere que existem pelo menos cinco *clusters* distintos de indivíduos, neste caso, países, como se pode verificar na Figura 16. Através do método da distância entre *clusters*, também se obtêm 5 *clusters*, uma vez que no gráfico da Figura 15 verifica-se que o primeiro grande aumento destas distâncias ocorre da etapa 21 para a 22 (ANEXO B). Deste modo, a nível de consumo, Portugal assemelha-se à Grécia, à Espanha, à Estónia, à Eslováquia, à Itália, à Bélgica e ao Chipre. Os países da Europa do Norte têm um comportamento semelhante entre si.

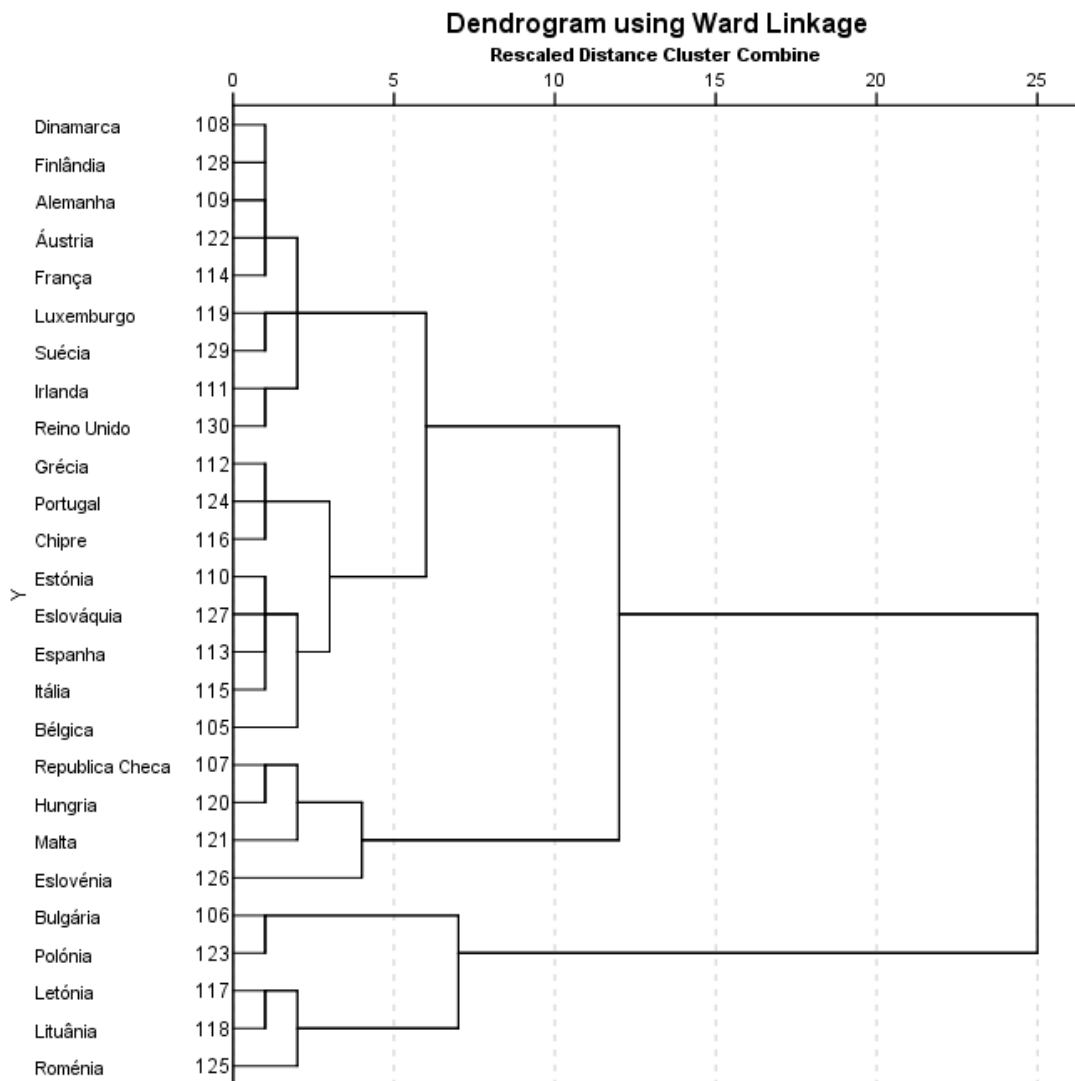


Figura 14. Dendrograma da família de três ou mais adultos sem crianças

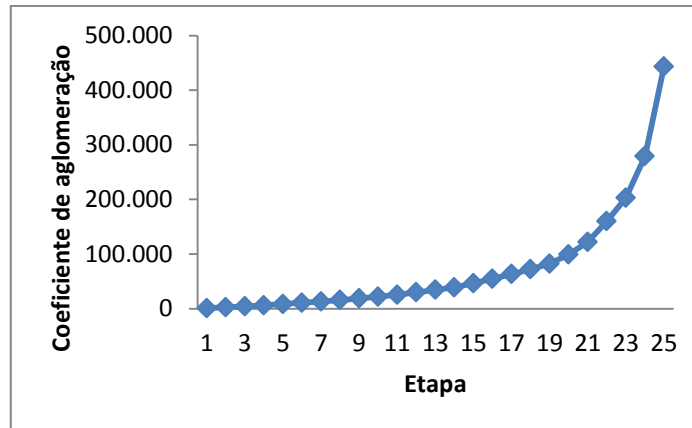


Figura 15. Representação dos coeficientes de aglomeração para cada etapa, no caso das famílias de três ou mais adultos sem crianças



Figura 16. Distribuição geográfica das famílias de três ou mais adultos sem crianças segundo os *clusters* do dendrograma obtido na Figura 14.

Quanto às famílias de três ou mais adultos com crianças, a análise do dendrograma na Figura 17 sugere que existem pelo menos quatro *clusters* distintos de indivíduos, neste caso, países, como se pode verificar na Figura 19. Efetuando, mais uma vez, o método da distância entre *clusters*, também se obtêm 4 *clusters*, uma vez que no gráfico da Figura 18 verifica-se que o primeiro grande aumento destas distâncias ocorre da etapa 22 para a 23 (ANEXO B). Sendo assim, verifica-se que toda a União Europeia, a nível de consumo, tem comportamento semelhante, exceto a Europa do Leste.

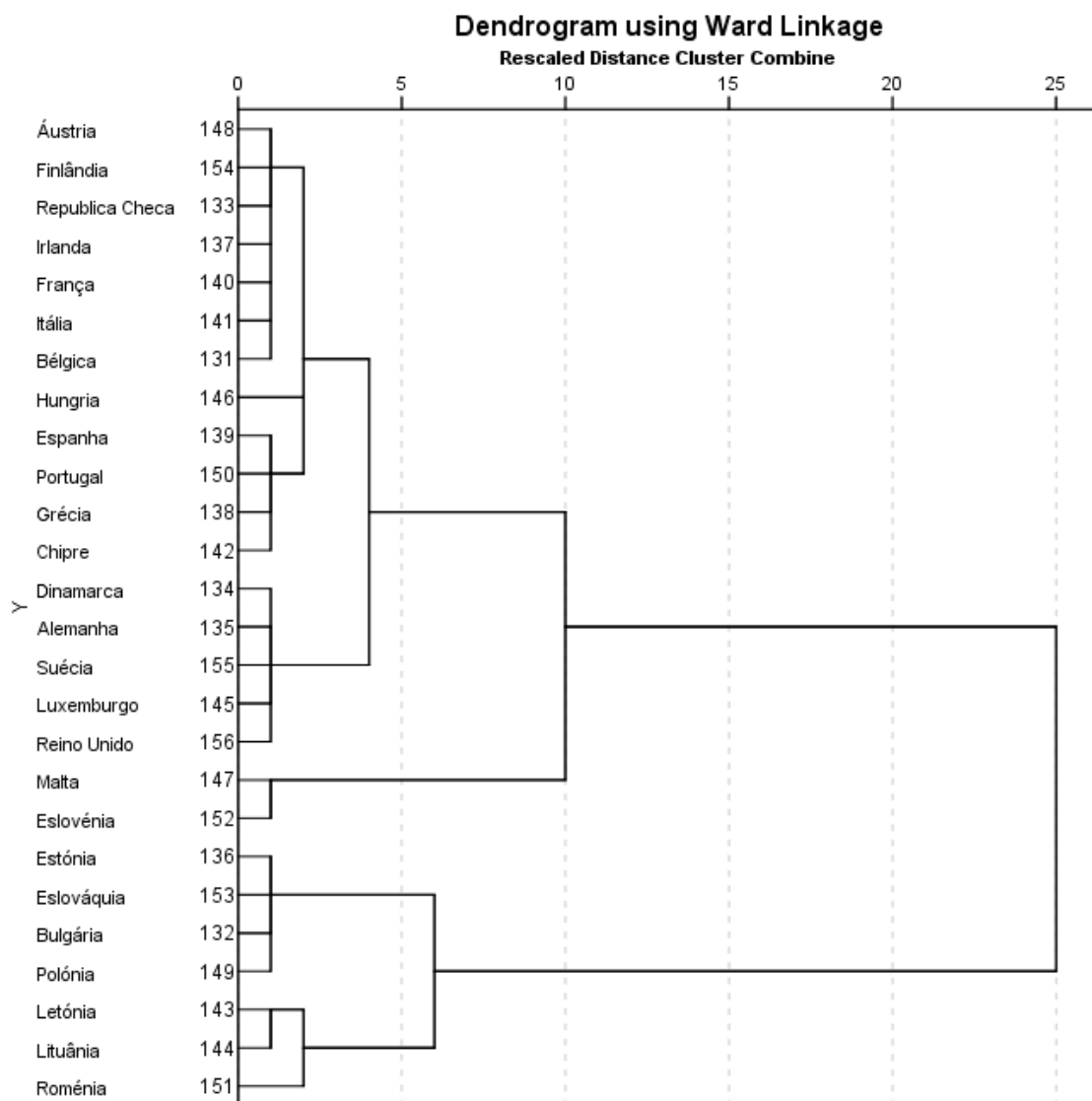


Figura 17. Dendrograma da família de três ou mais adultos com crianças

3.2.3. Análise comparativa entre o consumo atual das famílias com os consumos de anos anteriores

Através da análise do gráfico da Figura 20, verifica-se que nas famílias inquiridas de um só adulto sem crianças, os únicos consumos que aumentaram em 2012 face aos anos anteriores, foram dos alimentos e bebidas não alcoólicas (C1) e da educação (C10). As despesas com os outros tipos de consumo desceram, mas é de realçar a descida abrupta do consumo com habitação, água, eletricidade, gás e outros combustíveis (C4) e do consumo com restaurantes e hotéis (C11).

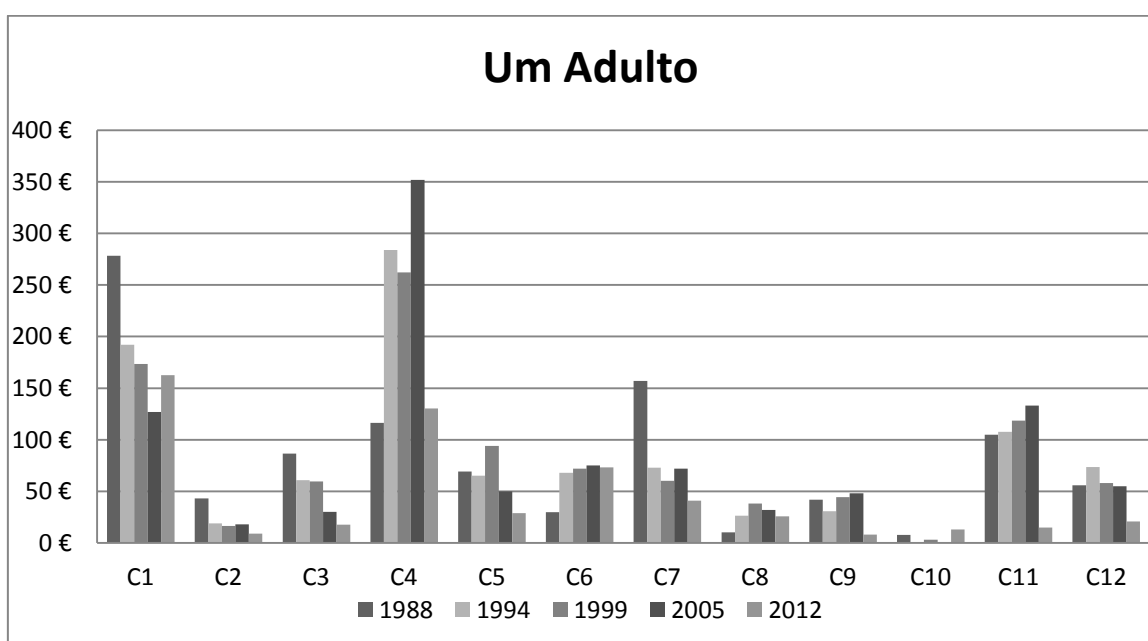


Figura 20. Representação gráfica da despesa (em euros) com os tipos de consumo da família de um adulto de 1988 a 2012

No que concerne ao consumo das famílias de um adulto com crianças dependentes, através dos inquiridos, verifica-se na Figura 21 que as despesas em 2012 com o consumo de alimentos e bebidas não alcoólicas (C1), bebidas alcoólicas, tabaco e narcóticos (C2), vestuário e calçado (C3), habitação, água, eletricidade, gás e outros combustíveis

(C4), saúde (C6) e educação (C10) aumentaram face aos anos anteriores. As despesas com os restantes tipos de consumo diminuíram.

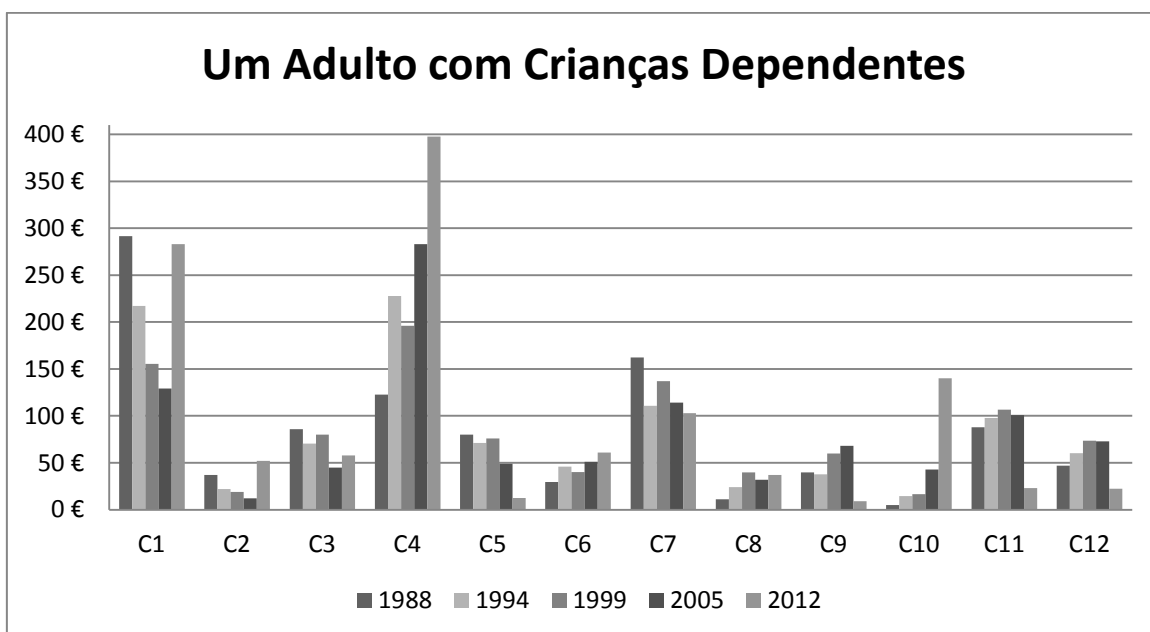


Figura 21. Representação gráfica da despesa com os tipos de consumo da família de um adulto com crianças dependentes de 1988 a 2012

No que respeita às despesas das famílias compostas por dois adultos (Figura 22), segundo os inquiridos, estas aumentaram em 2012 com o consumo de alimentos e bebidas não alcoólicas (C1), saúde (C6), comunicações (C8) e educação (C10). A despesa com os restantes consumos diminuiu, exceto com o vestuário e calçado (C3), que se manteve igual.

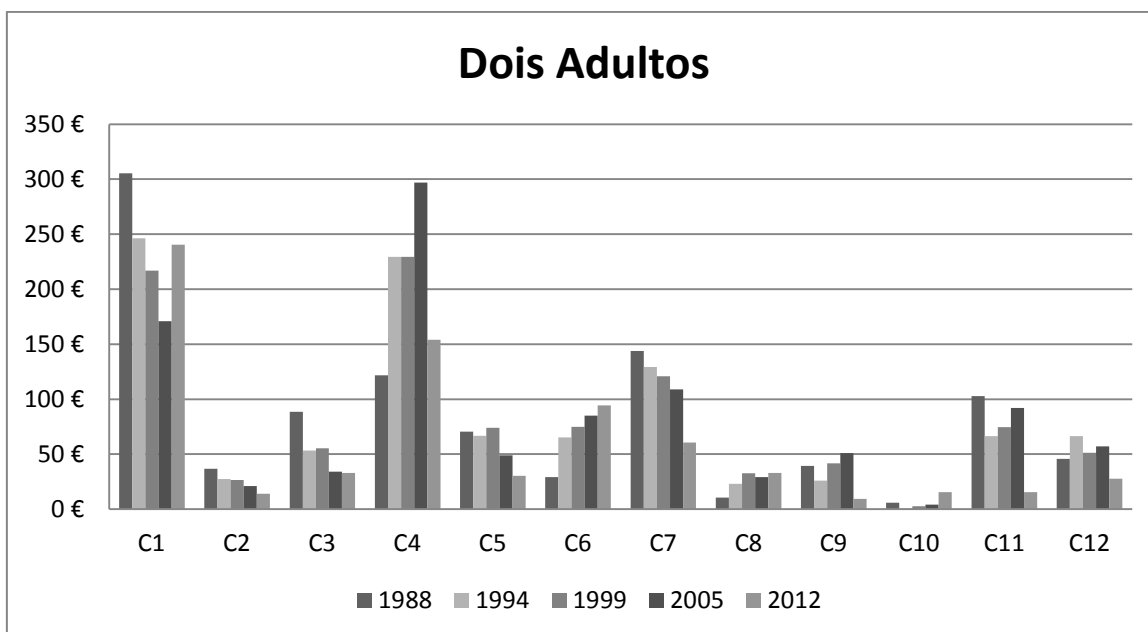


Figura 22. Representação gráfica da despesa com os tipos de consumo da família de dois adultos de 1988 a 2012

Quanto ao consumo das famílias inquiridas de dois adultos com crianças dependentes (Figura 23), todos os consumos em 2012 aumentaram face ao ano 2005, exceto cinco que diminuíram: acessórios, equipamento doméstico e manutenção corrente da habitação (C5), transporte (C7), lazer e cultura (C9), restaurantes e hotéis (C11) e bens e serviços diversos (C12). É ainda de salientar o grande aumento da despesa com o consumo de alimentos e bebidas não alcoólicas (C1) e da educação (C10).

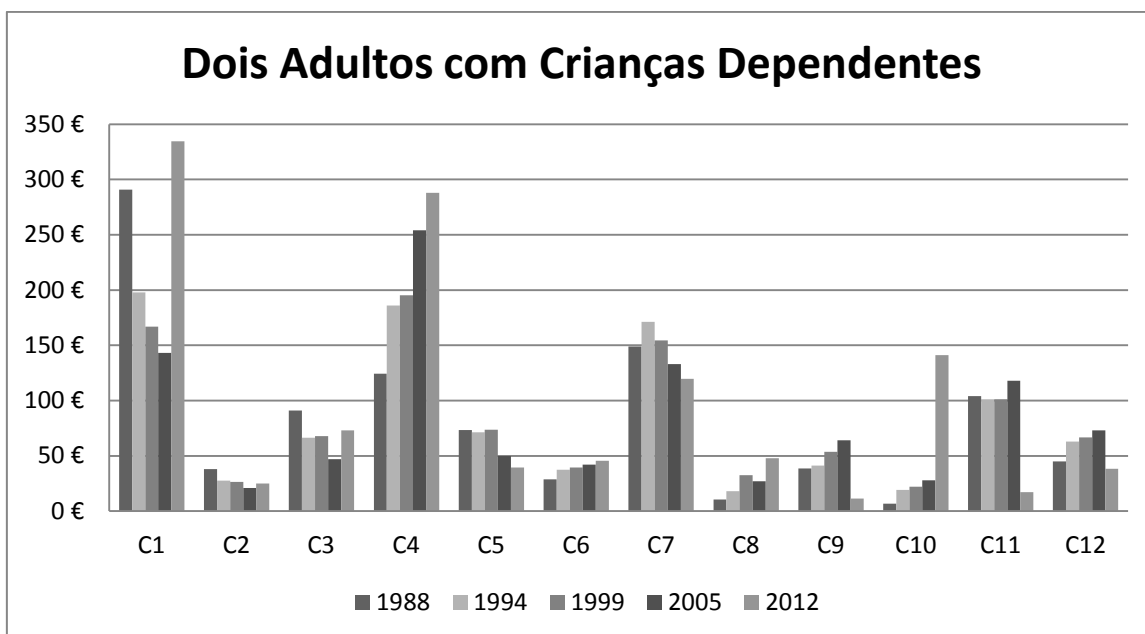


Figura 23. Representação gráfica da despesa com os tipos de consumo da família de dois adultos com crianças dependentes de 1988 a 2012

No que concerne ao consumo das famílias inquiridas de três ou mais adultos (Figura 24), apenas cinco tipos de consumo aumentaram em 2012: alimentos e bebidas não alcoólicas (C1), vestuário e calçado (C3), saúde (C6), comunicações (C8) e educação (C10). Todos os outros consumos diminuíram.

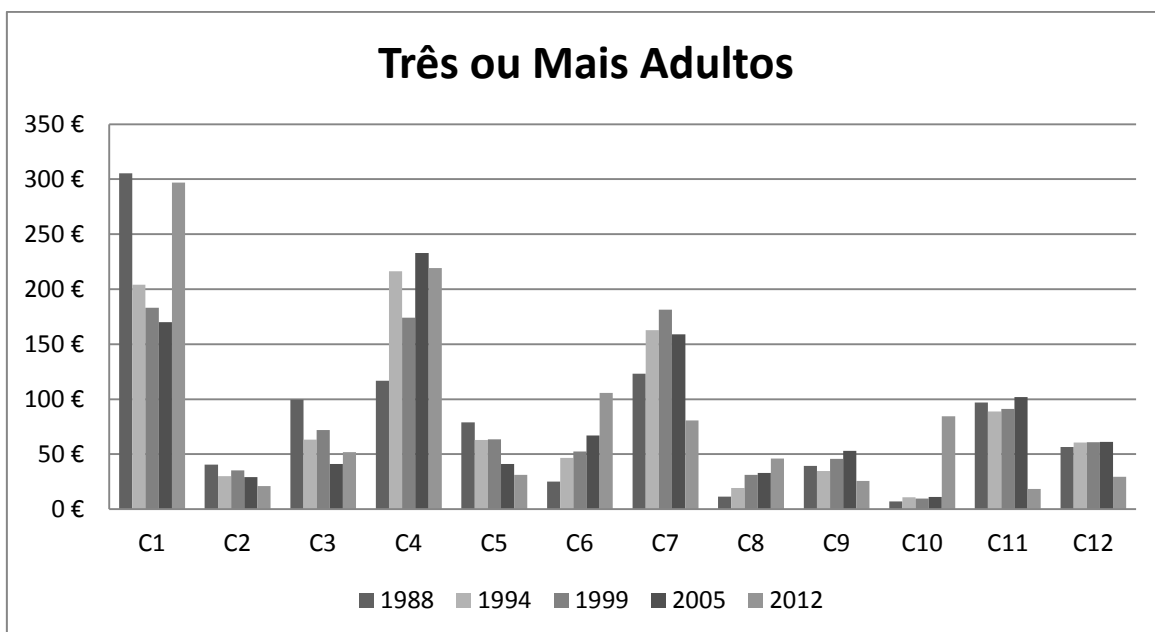


Figura 24. Representação gráfica da despesa com os tipos de consumo da família de três ou mais adultos de 1988 a 2012

No que respeita ao consumo das famílias inquiridas de três ou mais adultos com crianças dependentes (Figura 25), metade dos tipos de consumo aumentaram em 2012, sendo que outra metade diminuiu. Os consumos que aumentaram foram: alimentos e bebidas não alcoólicas (C1), vestuário e calçado (C3), acessórios, equipamento doméstico e manutenção corrente da habitação (C5), saúde (C6), comunicações (C8) e educação (C10).

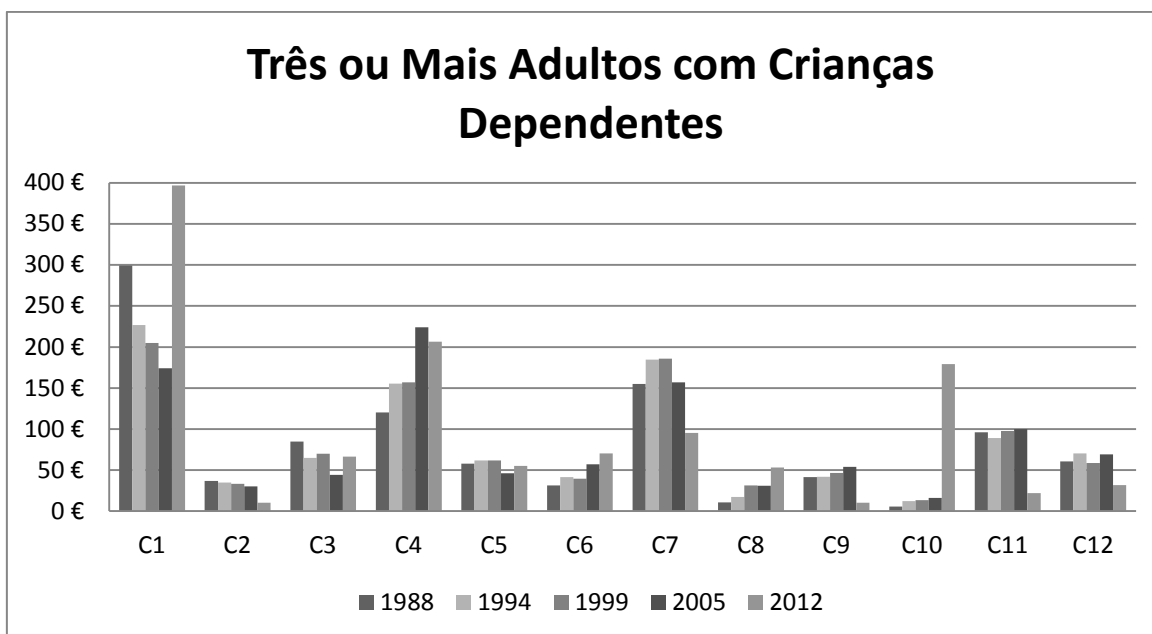


Figura 25. Representação gráfica da despesa com os tipos de consumo da família de três ou mais adultos com crianças dependentes de 1988 a 2012

3.2.4. Análise da existência de factores latentes associados ao consumo

Com o intuito de responder à pergunta “Será que existe um fator social ou económico que influencie o consumo?”, realizou-se a análise fatorial com o método das Componentes Principais. Para facilitar a interpretação, procedeu-se à rotação subsequente de fatores, que visa aumentar o poder explicativo das principais componentes, maximizando as correlações entre as variáveis e as componentes.

Na rotação ortogonal a solução inicial sofre uma rotação de 90°, enquanto na oblíqua não se impõe restrições sobre o ângulo de rotação. O tipo de rotação mais comum é o ortogonal, uma vez que permite manter os componentes em separado.

Dentro dos métodos ortogonais (*varimax*, *quartimax* e *equamax*) o mais amplamente utilizado é o *varimax*. Assim, as componentes foram extraídas e rodadas utilizando o método de rotação *varimax*. Os *scores* foram guardados usando o método de regressão.

O primeiro passo foi analisar a matriz de correlações (ANEXO C). Este tipo de procedimento é necessário, e o seu objetivo é examinar a correlação e associação linear en-

tre as variáveis. Se as correlações obtidas fossem elevadas, indicaria que existe informação redundante e que as dimensões do conjunto de variáveis podem ser reduzidas. Existem vários indicadores ou métodos que ajudam a avaliar esta matriz de correlação como o teste de Bartlett da esfericidade. Além disso, o teste de Kaiser-Meyer-Olkin (KMO) tenta medir a adequação da aplicação da análise de componentes principais.

O teste de esfericidade de Bartlett não aceita a hipótese nula de que a matriz de correlação é uma matriz de identidade e o teste de Kaiser-Meyer-Olkin (KMO), com um valor de 0,804, aponta para a presença de fatores comuns (ANEXO C). Assim, estes resultados indicam que a amostra é adequada para a análise fatorial.

Numa análise inicial, verificou-se que quatro fatores obedeceram ao critério de Kaiser do valor próprio (“*eigenvalue*”) maior que 1 e explicaram 57,29% da variância. O *scree plot* mostrou que apenas dois fatores estão posicionados antes da inflexão (ANEXO C). Considerando o tamanho da amostra e o critério de Kaiser, decidiu-se manter quatro fatores na análise final.

Como o peso fatorial representa a correlação entre a variável original e o fator, é importante determinar o nível de significância para a interpretação dos pesos fatoriais. Segundo Hair (1998), como regra prática, tem-se que o peso fatorial mínimo significativo é 0,30. Pesos em 0,40 são considerados mais importantes e acima de 0,50 são considerados significantes. Deste modo, optou-se por utilizar o corte em 0,490 (ANEXO C). De acordo com a matriz de componentes rodada, isto permitiu dar uma interpretação dos quatro fatores já selecionados.

Rotated Component Matrix^a

	Component			
	1	2	3	4
Alimentos e Bebidas não Alcoólicas em €	,648			
Bebidas Alcoólicas, Tabaco e Narcóticos em €			,744	
Vestuário e Calçado em €			,496	
Habituação, água, electricidade, gás e outros combustíveis em €	,744			
Acessórios, equipamento doméstico e manutenção corrente da habitação em €		,584		
Saúde em €				,886
Transporte em €	,534			
Comunicações em €	,590			
Lazer e Cultura em €			,732	
Educação em €	,538			
Restaurantes e Hóteis em €		,687		
Bens e serviços diversos em €		,528		

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Tabela 5. Agrupamento das variáveis em quatro fatores com a rotação *varimax*

Analisando o quadro da matriz das correlações entre as variáveis originais estimadas pelo modelo fatorial com os 4 fatores retidos, bem como a matriz dos resíduos, verifica-se a existência de 44 resíduos (66%) com valor absoluto superior a 0,05 (ANEXO C). Assim, a qualidade do modelo pode ser classificada como boa. Contudo, podemos recorrer a outros critérios de qualidade, nomeadamente ao *GFI* e ao *RMSR**. Usando os resíduos não redundantes calculados pelo SPSS:

$$GFI = 1 - 0,5(tr^*(e))^2$$

$$\begin{aligned}
&= 1 - 0,5 \times (0,41 - 0,129 - 0,018 + 0,161 - 0,091 + 0,064 - 0,096 - 0,026 \\
&\quad + 0,135 - 0,013 - 0,144)^2 \\
&= 0,9680
\end{aligned}$$

Estimando os graus de liberdade pelo método da máxima verosimilhança, o *GFI* ajustado é:

$$\begin{aligned}
AGFI &= 1 - \left(\frac{k}{gl}\right) (1 - GFI) \\
&= 1 - \left(\frac{p(p+1)/2}{[(p-m)^2 - p - m]/2}\right) (1 - GFI) \\
&= 1 - \left(\frac{12(12+1)/2}{[(12-4)^2 - 12 - 4]/2}\right) (1 - 0,9680) \\
&= 0,8960
\end{aligned}$$

De modo semelhante,

$$\begin{aligned}
RMSR^* &= \sqrt{\frac{\sum_{i=1}^{p-1} \sum_{j=i+1}^p (\rho_{ij} - r_{ij})^2}{p(p-1)/2}} \\
&= \sqrt{\frac{\sum_{i=1}^{p-1} \sum_{j=i+1}^p (e_{ij})^2}{p(p-1)/2}} \\
&= \sqrt{\frac{0,041^2 + (-0,071)^2 + \dots + (-0,059)^2 + (-0,144)^2}{12 \times (12-1)/2}} \\
&= 0,0940
\end{aligned}$$

De acordo com os valores do *GFI*, *AGFI* e *RMSR**, e ponderando a elevada percentagem de resíduos com valor absoluto superior a 0,05, pode ainda afirmar-se que o modelo apresenta um bom ajustamento aos dados.

Um outro aspeto que deve de ser analisado no modelo refere-se ao grau de consistência interna. Esta medida dá a proporção da variabilidade nas respostas que resulta das diferenças entre as respostas dos sujeitos, devido a diferentes opiniões. A consistência interna pode ser verificada pelo Alpha de Cronbach que reflete o grau de covariância dos itens entre si. Valores próximos de 1 indicam uma boa consistência interna. Para uma análise exploratória, aceitam-se valores acima de 0,6.

O valor de *alpha* encontrado nesta pesquisa, tendo em consideração todas as variáveis, foi o de 0,756, mostrando-se adequado, uma vez que está acima de 0,6. Por outro lado, também se calculou este parâmetro para as variáveis que compõem cada um dos fatores (ANEXO D). Quanto ao Fator 4, como se trata apenas de uma variável (saúde), tem-se em conta o seu peso no fator.

Seguindo este critério, elimina-se o fator 2, pois ao calcular-se o *alpha* para este fator, obteve-se um valor inferior a 0,6.

Desta forma, após a análise quantitativa dos dados, procura-se caracterizar cada fator por uma propriedade que possa representar a síntese de cada agrupamento estatístico.

Pode-se então interpretar, de forma já objetiva e através das definições dos diferentes tipos de Bens de Consumo, os dados apresentados anteriormente, afirmando que o consumo é influenciado pelos Bens de Consumo Saciado (quando se aumenta o salário, a procura mantém-se inalterada, pois já fazia parte das antigas aquisições do indivíduo – Fator 1), pelos Bens Normais (ao se aumentar o salário, a procura também aumenta – Fator 3) e, por fim, pelos Bens e Serviços de Saúde (Fator 4).

CAPÍTULO 4: CONCLUSÃO

Este trabalho teve por objetivos específicos:

- Planeamento (escolher a metodologia e criar um questionário);
- Recolha dos dados (Levantamento dos dados e supervisão do estudo);
- Comparar Portugal com os outros países da EU através da análise de *clusters*;
- Comparar o consumo atual das famílias com os consumos anteriores;
- Responder à pergunta “Será que existe um fator social ou económico que influencie o consumo?” através da análise fatorial.

A recolha dos dados reporta-se apenas ao ano 2012.

Os dados utilizados de anos anteriores foram recolhidos do Eurostat.

Para a concretização destes objetivos foram usadas diferentes técnicas de Estatística.

A análise comparativa entre o consumo atual das famílias com os consumos de anos anteriores permitiu verificar que em todos os tipos de família o gasto com o consumo de alimentos e bebidas não alcoólicas aumentou face a 2005, ao contrário de outros.

A análise de *clusters* permitiu um agrupamento dos países consoante o tipo de família, de acordo com os tipos de consumo. Relativamente às famílias sem crianças, esse agrupamento foi realizado em 5 *clusters*, enquanto o agrupamento dos países relativo às famílias com crianças foi realizado em 4 *clusters*. Também se verificou em todos os tipos de família que os Países Nórdicos da União Europeia têm um comportamento idêntico, a nível de consumo.

A análise fatorial permitiu uma redução da dimensionalidade do problema tendo em consideração as variáveis do consumo analisadas. A aplicação desta técnica resultou em quatro fatores, os quais explicam, aproximadamente, 57,29% da variabilidade total dos dados. No entanto, devido ao valor do Alpha de Cronbach de um dos fatores, consideraram-se apenas três fatores.

Deste modo, concluiu-se que o consumo é influenciado pelos Bens de Consumo Saciado, pelos Bens Normais e pelos Bens e Serviços de Saúde. Pode-se assim dizer que estes quatro factores contêm uma característica económica.

BIBLIOGRAFIA

Aaker, David A.; Day, George S.; Kumar, V.: *Pesquisa de Marketing*, 2ª edição, Editora Atlas, Brasil (2004);

Eurostat: *Household Budget Surveys in the EU: Methodology and Recommendations for Harmonization – 2003*, European Communities, Luxemburgo (2003);

Frank, Robert H.: *Microeconomia e Comportamento*, 6ª edição, McGraw-Hill, Lisboa (2006);

Gomes, Isabela Motta: *Como Elaborar uma Pesquisa de Mercado*, SEBRAE, Brasil (2005);

Hair, J.F. Jr.; Anderson, R.E.; Tatham, R.L.; Black, W.C.: *Multivariate Data Analysis*, 5ª edição, Prentice Hall, New Jersey (1998);

Härdle, W.; Simar, L.: *Applied Multivariate Statistical Analysis*, Springer, Alemanha (2003);

Hill, Manuela Magalhães; Hill, Andrew: *Investigação por Questionário*, 2ª edição, Edições Sílabo, Portugal (2008);

Johnson, Richard A.; Wichern, Dean W.: *Applied Multivariate Statistical Analysis*, 3ª edição, Prentice-Hall International Editions (1992);

Malhotra, Naresh K.: *Pesquisa de Marketing: Uma Orientação Aplicada*, 6ª edição, Bookman, Brasil (2011);

Maroco, João: *Análise Estatística com o SPSS Statistics*, 5ª edição, Report Number (2011);

Pereira, Alexandre: *SPSS: Guia Prático de Utilização: Análise de Dados para Ciências Sociais e Psicologia*, 7ª edição, Edições Sílabo, Lisboa (2006);

Pestana, Maria Helena; Gageiro, João Nunes: *Análise de Dados para Ciências Sociais, A complementaridade do SPSS*, 5ª edição, Edições Sílabo, Lisboa (2008);

Sharma, Subhash: *Applied Multivariate Techniques*, John Wiley & Sons, USA (1996);

Solomon, Michael; Bamossy, Gary; Askegaard, Saren; Hogg, Margaret K.: *Consumer Behavior: A European Perspective*, 4ª edição, Financial Times/Prentice-Hall (2010);

Vicente, Reis e Ferrão: *Sondagens. A amostragem como factor decisivo de qualidade*, 2ª edição, Edições Sílabo, Lisboa (2001).

ANEXOS.

ANEXO A. Exemplo de um relatório das ligações

Estudo _____

Data:..... Entrevistador:.....

Folha nº.....

	Nº Telefone	Não Atende	Interrompido	Atendedor de chamadas	marcar para mais tarde	Fax	Não Atribuído	Avariado	Recusa	Fora de quota	Realizada	Notas
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												
31												
32												
33												
34												
35												
Total												

ANEXO B. Tabelas de agrupamento da Análise de *Clusters*

Famílias de um adulto sem crianças

Agglomeration Schedule^a

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	13	14	590,500	0	0	19
2	1	24	1222,000	0	0	3
3	1	5	2017,833	2	0	8
4	4	25	2824,833	0	0	10
5	6	19	3643,333	0	0	6
6	6	23	4808,167	5	0	15
7	3	16	6501,167	0	0	17
8	1	10	8334,083	3	0	16
9	8	20	11046,083	0	0	20
10	4	7	14247,083	4	0	13
11	9	11	17888,583	0	0	15
12	15	26	22173,583	0	0	13
13	4	15	26937,383	10	12	16
14	12	18	31855,883	0	0	17
15	6	9	38869,050	6	11	18
16	1	4	46984,667	8	13	22
17	3	12	57711,417	7	14	20
18	2	6	73469,083	0	15	24
19	13	21	92829,250	1	0	23
20	3	8	112230,833	17	9	22
21	17	22	135408,833	0	0	23
22	1	3	176874,867	16	20	24
23	13	17	246598,600	19	21	25
24	1	2	337347,924	22	18	25
25	1	13	564324,423	24	23	0

a. Tipo_Família = 1

Famílias de um adulto com crianças

Agglomeration Schedule^a

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	32	45	631,500	0	0	8
2	31	50	1310,500	0	0	4
3	34	38	2267,500	0	0	14
4	27	31	3493,167	0	2	11
5	39	40	4937,167	0	0	20
6	35	37	6514,167	0	0	17
7	30	52	8229,667	0	0	12
8	32	49	9987,500	1	0	15
9	36	44	11867,500	0	0	13
10	29	42	13793,000	0	0	20
11	27	33	16039,583	4	0	13
12	30	51	18571,417	7	0	16
13	27	36	21516,667	11	9	18
14	34	46	24468,333	3	0	17
15	28	32	28220,500	0	8	23
16	30	41	32044,917	12	0	18
17	34	35	37845,650	14	6	21
18	27	30	46894,400	13	16	21
19	43	48	57308,400	0	0	24
20	29	39	70937,150	10	5	22
21	27	34	90471,617	18	17	25
22	29	47	128424,967	20	0	23
23	28	29	179252,533	15	22	24
24	28	43	300964,776	23	19	25
25	27	28	449002,731	21	24	0

a. Tipo_Família = 2

Famílias de dois adultos sem crianças

Agglomeration Schedule^a

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	57	76	750,500	0	0	10
2	56	77	1695,000	0	0	6
3	58	71	2788,000	0	0	5
4	53	62	4011,500	0	0	10
5	58	75	5649,833	3	0	16
6	56	78	7347,333	2	0	13
7	55	68	9188,333	0	0	19
8	61	63	11129,833	0	0	16
9	65	66	13101,833	0	0	18
10	53	57	15425,833	4	1	12
11	60	72	17867,333	0	0	15
12	53	59	20460,133	10	0	14
13	56	67	24112,383	6	0	17
14	53	70	28139,750	12	0	17
15	60	64	32447,583	11	0	21
16	58	61	39914,350	5	8	20
17	53	56	49201,133	14	13	22
18	65	73	60861,133	9	0	23
19	55	69	73846,800	7	0	21
20	54	58	92825,533	0	16	24
21	55	60	117401,033	19	15	22
22	53	55	165529,271	17	21	24
23	65	74	230003,521	18	0	25
24	53	54	334612,023	22	20	25
25	53	65	478077,154	24	23	0

a. Tipo_Família = 3

Famílias de dois adultos com crianças

Agglomeration Schedule^a

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	82	102	696,000	0	0	10
2	84	101	1434,000	0	0	7
3	86	90	2323,000	0	0	16
4	79	88	3249,500	0	0	8
5	83	103	4713,500	0	0	9
6	87	89	6396,000	0	0	12
7	84	97	8186,667	2	0	17
8	79	85	10247,500	4	0	15
9	83	104	12474,833	5	0	14
10	82	96	14795,500	1	0	15
11	81	94	17161,000	0	0	19
12	87	98	20128,500	6	0	16
13	91	92	23708,500	0	0	20
14	83	93	27334,667	9	0	18
15	79	82	31692,833	8	10	18
16	86	87	36496,633	3	12	22
17	80	84	44218,717	0	7	23
18	79	83	52082,550	15	14	22
19	81	95	61313,717	11	0	21
20	91	99	74635,050	13	0	23
21	81	100	96998,383	19	0	24
22	79	86	119613,417	18	16	24
23	80	91	165265,619	17	20	25
24	79	81	236440,180	22	21	25
25	79	80	375617,231	24	23	0

a. Tipo_Família = 4

Famílias de três ou mais adultos sem crianças

Agglomeration Schedule^a

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	108	128	962,000	0	0	2
2	108	109	2528,000	1	0	6
3	110	127	4187,500	0	0	12
4	112	124	6269,000	0	0	8
5	106	123	8573,500	0	0	23
6	108	122	10884,750	2	0	10
7	117	118	13281,250	0	0	18
8	112	116	15901,750	4	0	20
9	113	115	18680,750	0	0	12
10	108	114	22007,900	6	0	16
11	107	120	25571,400	0	0	15
12	110	113	29917,150	3	9	17
13	119	129	34452,150	0	0	16
14	111	130	39103,650	0	0	19
15	107	121	46610,150	11	0	21
16	108	119	54733,036	10	13	19
17	105	110	63180,786	0	12	20
18	117	125	72529,619	7	0	23
19	108	111	82374,500	16	14	22
20	105	112	99480,250	17	8	22
21	107	126	122153,250	15	0	24
22	105	108	160247,304	20	19	24
23	106	117	203081,471	5	18	25
24	105	107	279272,952	22	21	25
25	105	106	443392,154	24	23	0

a. Tipo_Família = 5

Famílias de três ou mais adultos com crianças

Agglomeration Schedule^a

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	148	154	1303,000	0	0	16
2	134	135	2965,000	0	0	11
3	136	153	4771,000	0	0	14
4	139	150	6731,500	0	0	13
5	143	144	8808,000	0	0	21
6	138	142	10997,000	0	0	13
7	132	149	13316,000	0	0	14
8	133	137	15806,000	0	0	15
9	140	141	18459,500	0	0	10
10	131	140	21564,667	0	9	15
11	134	155	24976,000	2	0	17
12	145	156	29496,500	0	0	17
13	138	139	34662,750	6	4	20
14	132	136	40864,750	7	3	23
15	131	133	48000,083	10	8	16
16	131	148	55896,226	15	1	19
17	134	145	63893,993	11	12	22
18	147	152	72834,993	0	0	24
19	131	146	82536,975	16	0	20
20	131	138	97579,433	19	13	22
21	143	151	112908,267	5	0	23
22	131	134	138909,216	20	17	24
23	132	143	184055,454	14	21	25
24	131	147	255877,203	22	18	25
25	131	132	450002,846	24	23	0

a. Tipo_Família = 6

ANEXO C. Tabelas e *Scree Plot* da Análise Fatorial

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Alimentos e Bebidas não Alcoólicas em €	261,19	203,395	598
Bebidas Alcoólicas, Tabaco e Narcóticos em €	16,65	48,159	598
Vestuário e Calçado em €	42,42	76,300	598
Habituação, água, electricidade, gás e outros combustíveis em €	190,75	200,267	598
Acessórios, equipamento doméstico e manutenção corrente da habitação em €	32,91	105,432	598
Saúde em €	83,90	152,751	598
Transporte em €	72,26	93,382	598
Comunicações em €	38,08	34,032	598
Lazer e Cultura em €	13,53	53,402	598
Educação em €	61,79	169,205	598
Restaurantes e Hóteis em €	16,89	45,693	598
Bens e serviços diversos em €	28,32	66,510	598

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,804
Bartlett's Test of Sphericity	Approx. Chi-Square	1433,135
	df	66
	Sig.	,000

Correlation Matrix

	Alimentos e Bebidas não Alcoólicas em €	Bebidas Alcoólicas, Tabaco e Narcóticos em €	Vestário e Calçado em €	Habituação, água, electricidade, gás e outros combustíveis em €	Accesórios, equipamento doméstico e manutenção corrente da habitação em €	Saúde em €	Transporte em €	Comunicações em €	Lazer e Cultura em €	Educação em €	Restaurantes e Hóteis em €	Bens e serviços diversos em €
Correlation	1,000	,199	,362	,288	,237	,069	,372	,411	,252	,220	,206	,218
Alimentos e Bebidas não Alcoólicas em €												
Bebidas Alcoólicas, Tabaco e Narcóticos em €	,199	1,000	,292	,195	,063	-,061	,217	,257	,389	,039	,242	,069
Vestário e Calçado em €	,362	,292	1,000	,263	,220	,061	,464	,445	,578	,196	,338	,379
Habituação, água, electricidade, gás e outros combustíveis em €	,288	,195	,263	1,000	,044	,024	,221	,349	,154	,112	,091	,060
Accesórios, equipamento doméstico e manutenção corrente da habitação em €	,237	,063	,220	,044	1,000	,022	,166	,234	,167	,069	,285	,080
Saúde em €	,069	-,061	,061	,024	,022	1,000	-,018	-,019	,033	-,040	,071	,023
Transporte em €	,372	,217	,464	,221	,166	-,018	1,000	,415	,312	,311	,369	,222
Comunicações em €	,411	,257	,445	,349	,234	-,019	,415	1,000	,407	,231	,346	,219
Lazer e Cultura em €	,252	,389	,578	,154	,167	,033	,312	,407	1,000	,052	,331	,159
Educação em €	,220	,039	,196	,112	,069	-,040	,311	,231	,052	1,000	,238	,070
Restaurantes e Hóteis em €	,206	,242	,338	,091	,285	,071	,369	,346	,331	,238	1,000	,244
Bens e serviços diversos em €	,218	,069	,379	,060	,080	,023	,222	,219	,159	,070	,244	1,000

Communalities

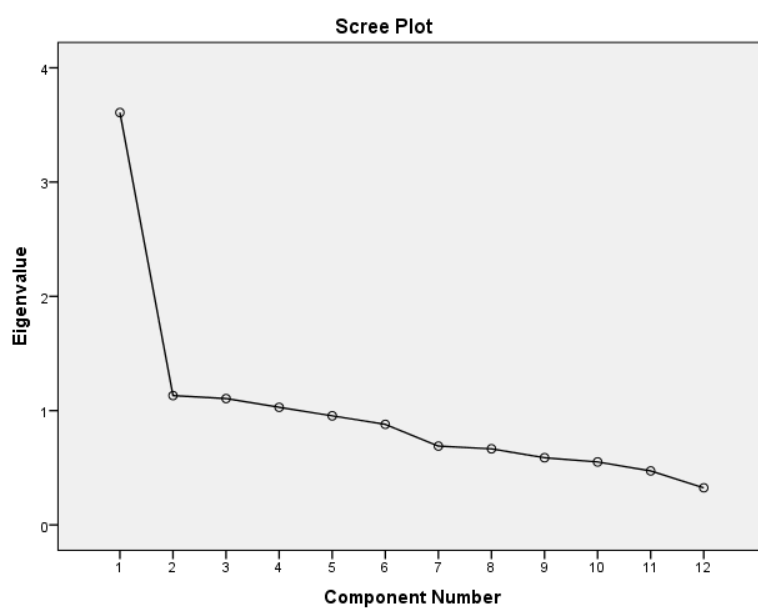
	Initial	Extraction
Alimentos e Bebidas não Alcoólicas em €	1,000	,512
Bebidas Alcoolicas, Tabaco e Narcóticos em €	1,000	,601
Vestuário e Calçado em €	1,000	,633
Habituação, água, electricidade, gás e outros combustíveis em €	1,000	,703
Acessórios, equipamento doméstico e manutenção corrente da habitação em €	1,000	,344
Saúde em €	1,000	,821
Transporte em €	1,000	,532
Comunicações em €	1,000	,556
Lazer e Cultura em €	1,000	,678
Educação em €	1,000	,644
Restaurantes e Hóteis em €	1,000	,538
Bens e serviços diversos em €	1,000	,313

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,608	30,068	30,068	3,608	30,068	30,068	2,127	17,722	17,722
2	1,132	9,432	39,500	1,132	9,432	39,500	1,977	16,471	34,193
3	1,106	9,215	48,715	1,106	9,215	48,715	1,714	14,283	48,476
4	1,029	8,575	57,290	1,029	8,575	57,290	1,058	8,814	57,290
5	,955	7,955	65,245						
6	,880	7,331	72,576						
7	,689	5,746	78,321						
8	,666	5,548	83,870						
9	,588	4,899	88,769						
10	,551	4,592	93,361						
11	,472	3,936	97,297						
12	,324	2,703	100,000						

Extraction Method: Principal Component Analysis.



Component Matrix^a

	Component			
	1	2	3	4
Alimentos e Bebidas não Alcoólicas em €	,614	,191	-,125	,286
Bebidas Alcoolicas, Tabaco e Narcóticos em €	,472	-,579	-,157	-,141
Vestuário e Calçado em €	,771	-,140	,137	,014
Habitação, água, electricidade, gás e outros combustíveis em €	,427	-,095	-,493	,517
Acessórios, equipamento doméstico e manutenção corrente da habitação em €	,379	,227	,353	-,161
Saúde em €	,043	,087	,546	,718
Transporte em €	,682	,210	-,126	-,083
Comunicações em €	,724	,015	-,163	,070
Lazer e Cultura em €	,654	-,464	,160	-,094
Educação em €	,369	,614	-,331	-,149
Restaurantes e Hóteis em €	,598	,144	,296	-,267
Bens e serviços diversos em €	,422	,123	,340	-,066

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Reproduced Correlations

Reproduced Correlation	Alimentos e Bebidas não Alcoólicas em €	Bebidas Alcoólicas, Tabaco e Narcóticos em €	Vestúlio e Calçado em €	Habituação, água, electricidade, gás e outros combustíveis em €	Acessórios, equipamento doméstico e manutenção corrente da habitação em €	Saúde em €	Transporte em €	Comunicação em €	Lazer e Cultura em €	Educação em €	Restaurantes e Hotéis em €	Bens e serviços diversos em €
Alimentos e Bebidas não Alcoólicas em €	.511 ^a	.158	.433	.454	.185	.180	.451	.488	.266	.343	.281	.221
Bebidas Alcoólicas, Tabaco e Narcóticos em €	.602 ^a	.421	.421	.261	.015	.232	.232	.348	.565	.185	.389	.084
Vestúlio e Calçado em €	.433	.421	.632 ^a	.282	.306	.105	.478	.535	.678 ^a	.251	.407	.167
Habituação, água, electricidade, gás e outros combustíveis em €	.454	.261	.282	.702 ^a	.117	.291	.424	.511	.434	.319	.368	.248
Acessórios, equipamento doméstico e manutenção corrente da habitação em €	.185	.015	.306	.345 ^a	.113	.275	.407	.209	.214	.423	.280	.318
Saúde em €	.180	.217	.105	.112	.822 ^a	.081	.008	.006	.008	.218	.008	.167
Transporte em €	.451	.232	.478	.291	.275	.532 ^a	.532 ^a	.511	.336	.435	.423	.276
Comunicações em €	.488	.348	.535	.424	.209	.511	.556 ^a	.434	.434	.319	.368	.248
Lazer e Cultura em €	.266	.565	.590	.196	.214	.008	.678 ^a	.434	.678 ^a	.083	.397	.280
Educação em €	.343	.108	.151	.185	.186	.218	.083	.319	.083	.645 ^a	.251	.129
Restaurantes e Hotéis em €	.281	.389	.478	.389	.407	.008	.397	.368	.397	.251	.538 ^a	.389
Bens e serviços diversos em €	.221	.084	.354	.033	.318	.167	.276	.248	.280	.129	.389	.313 ^a
Residual ^b	.041	.041	-.071	-.166	.052	-.111	-.079	-.076	-.013	-.123	-.075	-.003
Alimentos e Bebidas não Alcoólicas em €	.041			-.066	.048	.156	-.014	-.091	-.176	.148	.052	-.014
Bebidas Alcoólicas, Tabaco e Narcóticos em €		.041		-.066	.048	.156	-.014	-.091	-.176	.148	.052	-.014
Vestúlio e Calçado em €	-.071	-.129		-.018	-.086	-.045	-.013	-.089	-.012	.045	-.139	.026
Habituação, água, electricidade, gás e outros combustíveis em €	-.166	-.066			.161	-.088	-.069	-.076	-.042	-.072	.133	.093
Acessórios, equipamento doméstico e manutenção corrente da habitação em €	.052	.048		.161	-.091	-.091	-.108	.026	-.046	-.117	-.121	-.238
Saúde em €	-.111	.156		-.088	-.091	-.111	.064	-.013	.025	.178	.063	-.145
Transporte em €	-.079	-.014		-.069	-.108	.064	.064	-.096	-.024	-.124	-.054	-.055
Comunicações em €	-.076	-.091		-.076	.026	-.013	-.096	-.026	-.026	-.088	-.022	-.028
Lazer e Cultura em €	-.013	-.176		-.042	-.046	.025	-.024	-.026	-.026	.135	-.066	-.120
Educação em €	-.123	.148		-.072	-.117	.178	-.124	-.088	.135	-.059	-.013	-.059
Restaurantes e Hotéis em €	-.075	.052		.389	.407	-.008	.397	.368	.397	.251	.538 ^a	-.144
Bens e serviços diversos em €	-.003	-.014		.093	.313 ^a	-.145	-.055	-.028	-.120	-.013	-.144	-.144

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 44 (66,0%) nonredundant residuals with absolute values greater than 0.05.

Rotated Component Matrix^a

	Component			
	1	2	3	4
Alimentos e Bebidas não Alcoólicas em €	,648	,254	,103	,125
Bebidas Alcoólicas, Tabaco e Narcóticos em €	,151	,014	,744	-,158
Vestuário e Calçado em €	,385	,480	,496	,083
Habituação, água, electricidade, gás e outros combustíveis em €	,744	-,274	,234	,135
Acessórios, equipamento doméstico e manutenção corrente da habitação em €	,049	,584	,007	,038
Saúde em €	,075	,131	-,121	,886
Transporte em €	,534	,435	,156	-,179
Comunicações em €	,590	,308	,331	-,057
Lazer e Cultura em €	,151	,342	,732	,046
Educação em €	,538	,303	-,342	-,383
Restaurantes e Hóteis em €	,149	,687	,196	-,078
Bens e serviços diversos em €	,095	,528	,108	,117

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Component Transformation Matrix

Component	1	2	3	4
1	,644	,590	,486	-,026
2	,303	,380	-,868	-,102
3	-,545	,612	,010	,574
4	,444	-,365	-,100	,812

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

ANEXO D. Alpha de Cronbach

Valor de *alpha* tendo em consideração todas as variáveis

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,631	,756	12

Item Statistics

	Mean	Std. Deviation	N
Alimentos e Bebidas não Alcoólicas em €	260,96	202,917	601
Bebidas Alcoólicas, Tabaco e Narcóticos em €	16,57	48,053	601
Vestuário e Calçado em €	42,30	76,145	601
Habituação, água, electricidade, gás e outros combustíveis em €	190,50	199,959	601
Acessórios, equipamento doméstico e manutenção corrente da habitação em €	32,75	105,194	601
Saúde em €	83,94	152,481	601
Transporte em €	72,15	93,195	601
Comunicações em €	37,99	33,973	601
Lazer e Cultura em €	13,46	53,277	601
Educação em €	61,73	168,858	601
Restaurantes e Hóteis em €	16,81	45,594	601
Bens e serviços diversos em €	28,21	66,365	601

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Alimentos e Bebidas não Alcoólicas em €	596,41	261976,652	,495	,286	,554
Bebidas Alcoólicas, Tabaco e Narcóticos em €	840,80	387190,299	,275	,202	,620
Vestuário e Calçado em €	815,07	349009,823	,568	,516	,579
Habituação, água, electricidade, gás e outros combustíveis em €	666,87	300337,894	,299	,180	,619
Acessórios, equipamento doméstico e manutenção corrente da habitação em €	824,62	364867,609	,236	,138	,617
Saúde em €	773,43	376542,685	,033	,033	,668
Transporte em €	785,22	342218,230	,505	,349	,577
Comunicações em €	819,38	380936,892	,569	,388	,609
Lazer e Cultura em €	843,91	377017,786	,399	,431	,609
Educação em €	795,64	330036,486	,244	,155	,624
Restaurantes e Hóteis em €	840,56	380274,337	,420	,294	,611
Bens e serviços diversos em €	829,16	380325,631	,259	,188	,617

Valor de *alpha* tendo em consideração as variáveis do Fator 1

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,550	,675	5

Item Statistics

	Mean	Std. Deviation	N
Alimentos e Bebidas não Alcoólicas em €	260,96	202,917	601
Habituação, água, electricidade, gás e outros combustíveis em €	190,50	199,959	601
Transporte em €	72,15	93,195	601
Comunicações em €	37,99	33,973	601
Educação em €	61,73	168,858	601

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Alimentos e Bebidas não Alcoólicas em €	362,37	114052,879	,424	,242	,418
Habituação, água, electricidade, gás e outros combustíveis em €	432,83	129422,717	,306	,149	,514
Transporte em €	551,18	169947,449	,452	,260	,464
Comunicações em €	585,34	196542,070	,521	,296	,530
Educação em €	561,60	149703,763	,269	,116	,523

Valor de *alpha* tendo em consideração as variáveis do Fator 2

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,350	,434	3

Item Statistics

	Mean	Std. Deviation	N
Acessórios, equipamento doméstico e manutenção corrente da habitação em €	32,75	105,194	601
Restaurantes e Hóteis em €	16,81	45,594	601
Bens e serviços diversos em €	28,21	66,365	601

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Acessórios, equipamento doméstico e manutenção corrente da habitação em €	45,02	7962,391	,206	,082	,372
Restaurantes e Hóteis em €	60,96	16594,417	,359	,131	,136
Bens e serviços diversos em €	49,56	15886,603	,156	,060	,345

Valor de *alpha* tendo em consideração as variáveis do Fator 3

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,669	,685	3

Item Statistics

	Mean	Std. Deviation	N
Bebidas Alcoolicas, Tabaco e Narcóticos em €	16,57	48,053	601
Vestuário e Calçado em €	42,30	76,145	601
Lazer e Cultura em €	13,46	53,277	601

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Bebidas Alcoolicas, Tabaco e Narcóticos em €	55,76	13326,791	,372	,158	,704
Vestuário e Calçado em €	30,03	7140,096	,531	,340	,558
Lazer e Cultura em €	58,87	10247,021	,620	,387	,418

ANEXO E. Questionário

Estudo do Orçamento Familiar

Maio de 2012 - GEMEO

Bom dia/ tarde/ noite!

Estou a ligar do GEMEO (Gabinete de Estudos de Mercado e Opinião do IPAM) e estamos a realizar um estudo sobre o orçamento familiar. Agradecia a sua colaboração respondendo a algumas questões.

O número do seu telefone foi sorteado completamente ao acaso e os dados que fornecer apenas servirão para fins académicos ou de controle da realização da entrevista, e deles será guardada estrita confidencialidade.

Dados da Caracterização Familiar

Para cumprir o meu plano de estudo, e saber se posso seleccioná-lo(a), responda-me por favor às seguintes questões:

1. Quantas pessoas vivem na habitação?

1. Uma 2. Duas 3. Três
 4. Quatro 5. Cinco ou Mais

A resposta é obrigatória.

2. Edessas pessoas quantas é que são adultas?

1. Um 2. Dois 3. Três ou Mais

A resposta é obrigatória.

3. E quantas é que são crianças?

1. Nenhuma 2. Uma 3. Duas 4. Três ou Mais

A resposta é obrigatória.

4. Sexo do Entrevistado

1. Masculino 2. Feminino

A resposta é obrigatória.

5. Qual é a sua idade?

A resposta é obrigatória.

6. E o seu estado civil?

1. Solteiro 2. Casado 3. União de Facto
 4. Divorciado 5. Viúvo

A resposta é obrigatória.

7. Qual é o grau de parentesco das pessoas que habitam a residência face ao Entrevistado?

1. Pai/ Mãe 2. Esposo/a 3. Filho/os
 4. Irmão/ Irmã 5. Avó/ Avô 6. Primo/a
 7. Tio/a 8. Sobrinho/a 9. Cunhado/a
 10. Colega

Você pode marcar diversas casas.

A resposta é obrigatória.

8. Profissão dos Adultos

A resposta é obrigatória.

Dados Sócio Demográficos

9. Distrito

1. Viana do Castelo 2. Braga
 3. Vila Real 4. Bragança
 5. Porto 6. Aveiro
 7. Viseu 8. Guarda
 9. Coimbra 10. Castelo Branco
 11. Leiria 12. Santarém
 13. Portalegre 14. Lisboa
 15. Évora 16. Setúbal
 17. Beja 18. Faro

A resposta é obrigatória.

10. Qual é o rendimento mensal líquido do agregado familiar?

1. Menos de 475€ 2. De 475€ a 1.000€
 3. De 1.001 a 1.500€ 4. De 1.501€ a 2.000€
 5. De 2.001€ a 2.500€ 6. De 2.501€ a 3.000€
 7. De de 3.001€ a 3.500€ 8. De 3.501€ a 4.000€
 9. De 4.001€ a 4.500€ 10. De 4.501€ a 5.000€
 11. Mais de 5.000€

A resposta é obrigatória.

Dados sobre a distribuição do rendimento mensal pelas seguintes categorias

Avalie de 0% a 100% ou diga a quantia em euros que gasta com cada categoria, não esquecendo que a soma perceptual de todas as categorias deve de corresponder a 100%.

11. Alimentos e Bebidas não Alcoólicas em %

12. Alimentos e Bebidas não Alcoólicas em €

13. Há um ano atrás gastava:

1. Menos 2. Igual 3. Mais

A resposta é obrigatória.

14. Bebidas Alcoólicas, Tabaco e Narcóticos em %

15. Bebidas Alcoólicas, Tabaco e Narcóticos em €

16. Há um ano atrás gastava:

1. Menos 2. Igual 3. Mais

A resposta é obrigatória.

17. Vestuário e Calçado em %

18. Vestuário e Calçado em €

19. Há um ano atrás gastava:
 1. Menos 2. Igual 3. Mais
A resposta é obrigatória.

20. Habitação, água, electricidade, gás e outros combustíveis em %

21. Habitação, água, electricidade, gás e outros combustíveis em €

22. Há um ano atrás gastava:
 1. Menos 2. Igual 3. Mais
A resposta é obrigatória.

23. Acessórios, equipamento doméstico e manutenção corrente da habitação em %

24. Acessórios, equipamento doméstico e manutenção corrente da habitação em €

25. Há um ano atrás gastava:
 1. Menos 2. Igual 3. Mais
A resposta é obrigatória.

26. Saúde em %

27. Saúde em €

28. Há um ano atrás gastava:
 1. Menos 2. Igual 3. Mais
A resposta é obrigatória.

29. Transporte em %

30. Transporte em €

31. Há um ano atrás gastava:
 1. Menos 2. Igual 3. Mais
A resposta é obrigatória.

32. Comunicações em %

33. Comunicações em €

34. Há um ano atrás gastava:
 1. Menos 2. Igual 3. Mais
A resposta é obrigatória.

35. Lazer e Cultura em %

36. Lazer e Cultura em €

37. Há um ano atrás gastava:
 1. Menos 2. Igual 3. Mais
A resposta é obrigatória.

38. Educação em %

39. Educação em €

40. Há um ano atrás gastava:
 1. Menos 2. Igual 3. Mais
A resposta é obrigatória.

41. Restaurantes e Hóteis em %

42. Restaurantes e Hóteis em €

43. Há um ano atrás gastava:
 1. Menos 2. Igual 3. Mais
A resposta é obrigatória.

44. Bens e serviços diversos em %

45. Bens e serviços diversos em €

46. Há um ano atrás gastava:
 1. Menos 2. Igual 3. Mais
A resposta é obrigatória.

Dados de Supervisão e Controle

47. Telefone
O código tem 9 caracteres.

48. Data da entrevista
A resposta é obrigatória.

49. Nome do Entrevistador
A resposta é obrigatória.

50. Região de Portugal:
 1. Norte 2. Centro 3. Lisboa
 4. Alentejo 5. Algarve
A preencher pela supervisão.