

Instituto Superior de Ciências do Trabalho e da Empresa



## MODELOS DE MISTURA EM CRM

**Uma aplicação à segmentação no sector bancário**

Claudia da Encarnação Calado

Dissertação submetida como requisito parcial para obtenção do grau de  
Mestre em Prospecção e Análise de Dados

Orientador:

Prof. Doutor José Manuel Gonçalves Dias

Co-orientadora:

Prof<sup>ª</sup>. Doutora Manuela Magalhães Hill

Março de 2008

## RESUMO

Actualmente, os modelos de Mistura são considerados um dos métodos de segmentação mais eficientes na área de Marketing para o estudo das estruturas de preferências.

Com base numa amostra de clientes de uma Instituição Financeira, numa primeira fase, será realizada uma segmentação com base nos modelos de mistura finita, de forma a perceber as estruturas de necessidades de produtos financeiros. Com base nas perfilagens dos segmentos, será possível efectuar a avaliação da necessidade ou não de desenvolvimento de estratégias diferenciadas consoante os segmentos obtidos de forma a aumentar o valor da rendibilidade dos clientes já existentes, adequando assim a oferta de produtos. Nesta fase, serão obtidas as probabilidades de pertença *a posteriori* para classificar novos clientes nos segmentos mais adequados, permitindo contactar o cliente da melhor forma e com a melhor oferta.

Numa segunda fase, será utilizado o modelo de mistura de regressões para perceber o impacto das acções de Marketing nos produtos detidos pelos clientes. Admitindo a existência de heterogeneidade das necessidades financeiras dos clientes, e colocando a hipótese de que as mesmas são explicadas com base nas acções de marketing realizadas, pretende obter-se um conjunto de estimativas de regressão para cada segmento identificado. A obtenção dessas estimativas de regressão, consoante a significância estatística, irá fornecer um maior conhecimento sobre a adequação da actual estratégia de marketing definida, e perceber a necessidade de afinação ou não da mesma consoante o segmento.

**Palavras Chave:** *Customer Relationship Management*, segmentação, modelos de mistura finita, regressão multinomial, variáveis concomitantes.

**JEL:** C25, C52, M31

## ABSTRACT

Nowadays, finite mixture models are one of the most efficient segmentation technique in the marketing field, in order to analyse structures of preferences in a given population.

Based on a sample of clients of a given Financial Institution, the first step of this study applies a finite mixture model to understand the existing structures of financial needs of the clients. Based on the profiled segments, the need of developing different marketing strategies for each segment will be assessed, in order to increase the profit of the actual clients due to a correct contact strategy and offer of products. The probabilities of belonging to a certain segment will be obtained in order to allocate new clients in the most adequate segment, allowing to reach the clients with the best contact and offer of products strategy.

In a second step, a regression mixture model will be applied to understand the impact of the actual marketing strategy in the portfolio of products of the clients. Assuming the existence of the heterogeneity in the financial needs of the clients and the fact that these needs can be explained by the acquisition campaigns, a set of regression models are estimated for each segment. Depending on the significance of this regression estimates, one understands the adequacy of the actual defined marketing strategy and decides if there is the need of improvement depending on the segment.

**Keywords:** Customer Relationship Management, segmentation, mixture models, multinomial regression, concomitants variables.

**JEL:** C25, C52, M31

## SUMÁRIO EXECUTIVO

No ramo bancário, os produtos desenvolvidos são cada vez mais complexos e diversificados por forma a dar resposta a qualquer necessidade financeira do cliente, garantindo assim a sua fidelização à Instituição Financeira. Uma das formas de o conseguir é através de acções de *cross-selling*, onde são propostos para os clientes já existentes, novos produtos. Mas para que as ofertas sejam bem sucedidas, é preciso “acertar” no conjunto de produtos a oferecer. Surge, assim, uma nova problemática de marketing, que é a de conhecer os milhares de clientes existentes na Instituição Financeira. O CRM – *Customer Relationship Management* pretende, desta forma, responder a algumas destas questões:

- quais são as novas necessidades financeiras dos clientes de forma a proceder ao desenvolvimento de novos produtos?
- que acções de *cross-selling* devem ser desenvolvidas de forma a antecipar as necessidades futuras dos clientes?
- quais serão os clientes mais rentáveis para apostar neles e criar ofertas mais dirigidas?
- quais os canais mais adequados para interagir com os clientes?

Uma das técnicas utilizadas para responder a tais perguntas é recorrendo à segmentação. Esta última é considerada uma das peças fundamentais para a implementação numa empresa de uma estratégia de CRM.

Os modelos de mistura ganharam protagonismo na área de marketing pelas suas vantagens face aos métodos tradicionais de segmentação. Os modelos de mistura assumem que os dados provêm de diversos segmentos homogêneos de clientes que foram misturados em proporções desconhecidas formando uma população. Não é dado a conhecer o segmento a que pertence o

cliente, ou quantos segmentos existiam inicialmente, sendo estes não observáveis ou usualmente designados por latentes. O objectivo destes modelos é o de desfazer essas misturas de forma a ser revelado o número de segmentos e a definição das suas respectivas estruturas que compõem a população em análise.

Os benefícios da utilização dos modelos de mistura para segmentar os clientes de uma dada população, serão demonstrados pela aplicação dos mesmos, considerando uma amostra de clientes que iniciaram a sua relação financeira numa dada Instituição Financeira em Janeiro de 2007. Esta amostra contém a informação dos produtos detidos em Outubro de 2007, das características demográficas, tais como, as variáveis *Sexo*, *Idade* e a *Zona Markttest* do cliente, e das acções de marketing realizadas. Foram, assim, aplicados os modelos de mistura finita e modelos de mistura de regressões com e sem variáveis concomitantes.

A aplicação do modelo de mistura finita sem variáveis concomitantes identificou 4 segmentos. No entanto a aplicação do referido modelo com variáveis concomitantes permitiu identificar dois sub-segmentos com perfis diferentes, apesar das suas preferências, em termos de produtos detidos serem semelhantes. Tais sub-segmentos estavam integrados num só na segmentação derivada da não utilização das variáveis concomitantes. Tais resultados evidenciam a importância da utilização das variáveis concomitantes, pois as mesmas permitiram identificar dois tipos de clientes para os quais a forma de contactar deverá ser diferenciada.

A aplicação do modelo de mistura de regressões com variáveis concomitantes não permitiu evidenciar qualquer impacto das acções de vendas dirigidas efectuadas aos clientes nos produtos detidos pelos mesmos.

## AGRADECIMENTOS

Finalmente a tese de mestrado está finalizada e um tempo de “luto” se avizinha. O emprego da palavra “luto” pode parecer exagerada, mas quando se acaba um projecto que sempre foi tão desejado e para o qual foi dado tanta dedicação e energia, compreenderão que não se pode ficar indiferente ao espaço por ele deixado. É por isso, que quero que fique neste espaço, a memória de todos, que de alguma forma, consciente ou inconscientemente, participaram comigo nesta travessia, que teve início no final de 2006, e ajudaram-me a chegar a bom porto em 2008.

Em primeiro lugar, agradeço a Instituição Financeira por ter acedido ao meu pedido de utilização dos dados. Sem esta informação, certamente que a tese não teria tido o mesmo sabor.

Uma palavra de apreço vai para a Rachel Ferreira, que me ajudou na formalização do pedido de utilização dos dados da Instituição Financeira, e no apoio dado numa fase em que a disponibilização dos mesmos parecia inviável.

Agradeço o meu Orientador, o Prof. Doutor José Manuel Gonçalves Dias, pela sua disponibilidade constante e pelo apoio dado às minhas ideias, mesmo quando estas eram demasiadas ambiciosas face ao prazo estipulado. Por isso, não posso deixar de admirar a sua constante paciência nos momentos mais difíceis, nomeadamente na sua postura sempre adequada face aos meus picos de desespero. Sem ele, certamente que não estaria agora a escrever estes agradecimentos. Se existisse um prémio do orientador mais paciente, de certeza que o teria ganho por ter conseguido aguentar todas as minhas neuroses.

Agradeço à minha Co-orientadora, a Prof. Doutora Manuela Hill, que graças à sua capacidade generosa de comunicação, transmitiu-me a confiança e a tranquilidade necessária para encarar o desafio da elaboração de uma tese de mestrado.

Não posso esquecer os meus peixes. Olhar para eles, perder-me nos seus movimentos sem nexos, assim como assistir ao nascimento de alguns deles, transportava-me, por breves momentos, para estados de meditação, conseguindo trazer calma e tranquilidade, facilitando o passar do tempo à frente do computador a escrever a dissertação.

A minha gratidão vai com os meus Irmãos, Nelson e Jaime e os meus Pais por terem conseguido ouvir com paciência, vezes sem conta, as minhas lamúrias, angústias e desesperos.

E por fim, mas nunca esquecidos, agradeço também os meus amigos; a Teresa Magalhães, a Carla Cardoso, a Anabela Barros Spencer e o Diogo Loureiro, que sempre me apoiaram, compreenderam e não se esqueceram da minha existência, durante o meu período de cativo, a que elaboração da tese me obrigou.

## ÍNDICE

1. Introdução .....	1
2. Os conceitos de CRM e Segmentação de Mercado .....	5
2.1. CRM - <i>Customer Relationship Management</i> .....	5
2.2. Segmentação de Mercado .....	12
3. Modelo de mistura finita .....	19
3.1. Introdução .....	19
3.2. Modelos de mistura finita .....	21
3.3. Modelos de mistura de regressões .....	23
3.4. Modelo de mistura de regressões para dados binários .....	26
3.5. Modelo de mistura de regressões com variáveis concomitantes .....	29
3.6. Identificação do modelo .....	32
3.7. Algoritmo EM .....	34
3.8. Determinação do número de segmentos .....	37
4. Uma aplicação à Banca .....	40
4.1. Introdução .....	40
4.2. Os dados .....	41
4.2.1. Indicadores de posse .....	42
4.2.2. Variáveis demográficas .....	45
4.2.3. Variáveis de marketing .....	47
4.3. Segmentação baseada nos indicadores de posse de produtos .....	48
4.3.1. Modelo de mistura finita .....	48
4.3.2. Estimação dos parâmetros do modelo .....	51
4.3.3. Relações estruturais existentes nos segmentos .....	54
4.3.4. Perfis dos segmentos .....	57
4.4. Segmentação baseada nos indicadores de posse e variáveis demográficas .....	61
4.4.1. Modelo de mistura finita com variáveis concomitantes .....	61
4.4.2. Estimação dos parâmetros do modelo com variáveis concomitantes .....	63
4.4.3. Estimação dos parâmetros do modelo não considerando o indicador de posse Seguro de Capitalização .....	64
4.4.4. Relações estruturais existentes nos segmentos .....	67
4.4.5. Perfis dos segmentos .....	70
4.4.6. Classificação dos clientes .....	75
4.5. Segmentação e regressão em simultâneo com recurso a variáveis de marketing .	79
4.5.1. Modelo de mistura de regressões .....	79
4.5.2. Análises bivariadas.....	80
4.5.3. Preparação dos dados para aplicação do modelo.....	83
4.5.4. Modelo de mistura de regressões com concomitantes .....	84
4.6. Discussão .....	85
5. Conclusão .....	93
6. Referências .....	97
7. Apêndices .....	103
Apêndice A – Modelo de mistura finita com concomitantes .....	103
Apêndice B – Modelo de mistura finita com concomitantes não considerando o indicador de posse de <i>Seguros de Capitalização</i> .....	104
Apêndice C – Modelo de mistura de regressões – Análise bivariada .....	106
Apêndice D – Modelo de mistura de regressões com concomitantes .....	111

## ÍNDICE DE FIGURAS

Figura 1 – Fases de maturidade do CRM .....	6
Figura 2 – Zonas Markttest .....	46
Figura 3 – Número de contactos efectuados por produto e canal de contacto .....	47
Figura 4 – Critérios de informação .....	50
Figura 5 – Distribuição da variável <i>Sexo</i> nos segmentos .....	58
Figura 6 – Distribuição da variável <i>Zona Markttest</i> nos segmentos .....	58
Figura 7 – Distribuição da variável <i>Idade</i> nos segmentos .....	60
Figura 8 – Variável dependente para o modelo de mistura de regressões .....	83
Figura 9 – Pirâmide etária da população residente, Portugal, 2001 e 2006 .....	86

## ÍNDICE DE TABELAS

Tabela 1 – Critérios de informação .....	38
Tabela 2 – Taxas de posse de produto .....	44
Tabela 3 – Distribuição da variável <i>Sexo</i> .....	45
Tabela 4 – Distribuição dos clientes pelas <i>Zonas Marktest</i> .....	46
Tabela 5 – Estimativas das constantes do modelo sem concomitantes .....	52
Tabela 6 – Estimativas dos efeitos do modelo sem concomitantes .....	52
Tabela 7 – Constantes de regressão dos pesos relativos do modelo sem concomitantes ...	53
Tabela 8 – Resíduos bivariados do modelo sem concomitantes .....	54
Tabela 9 – Dimensão dos segmentos do modelo sem concomitantes .....	54
Tabela 10 – Probabilidades condicionadas de posse do modelo sem concomitantes .....	55
Tabela 11 – Resíduos bivariados do modelo com concomitantes .....	64
Tabela 12 – Constantes de regressão dos pesos relativos do modelo com concomitantes .	66
Tabela 13 – Efeitos de regressão dos pesos relativos do modelo com concomitantes .....	67
Tabela 14 – Dimensão dos segmentos do modelo com concomitantes .....	67
Tabela 15 – Probabilidades condicionadas de posse do modelo com concomitantes .....	68
Tabela 16 – Probabilidades condicionadas ao segmento .....	71
Tabela 17 – Resumo dos perfis dos segmentos obtidos .....	74
Tabela 18 – Probabilidades condicionadas de pertença .....	76
Tabela 19 – Valor da estatística de teste do qui-quadrado .....	81
Tabela 20 – Resultado das regressões logísticas realizadas para todos os indicadores de posse .....	82
Tabela 21 – Distribuição da variável dependente na nova amostra .....	85

## 1 Introdução

Os modelos de mistura têm ganho protagonismo na área de estudos de mercado pelas suas vantagens face aos métodos tradicionais de agrupamento. Os modelos de mistura assumem que os clientes provêm de diversos segmentos homogéneos do mercado misturados em proporções desconhecidas. Não é conhecido o segmento a que pertence o cliente, ou quantos segmentos existiam inicialmente, sendo estes não observáveis ou usualmente designados por latentes. Assim, o objectivo destes modelos é o de desfazer essas misturas de forma a ser revelado o número de segmentos e a definição das suas respectivas estruturas que compõem a população em análise. Com efeito, os modelos de mistura podem ser uma ferramenta de segmentação na implementação de uma estratégia CRM – *Customer Relationship Management*.

A segmentação consiste em descrever a heterogeneidade existente no mercado que são uma consequência dos diferentes comportamentos e preferências existentes, identificando um número de mercados com menor dimensão mas mais homogéneos nessas preferências. Esta identificação pode levar as empresas a desenvolver novos produtos ou serviços, e adequar a comunicação a ser efectuada de forma a promover os novos produtos criados.

A presente tese pretende focalizar-se nas estruturas de preferência dos produtos financeiros adquiridos pelos clientes de uma determinada instituição financeira. Tal análise irá contribuir para o desenho de acções de *cross-selling*, adequando melhor a oferta de produtos de forma a antecipar as necessidades dos clientes, aumentando assim, o seu respectivo valor para a empresa. Com efeito, a heterogeneidade entre os clientes respectivamente aos produtos financeiros adquiridos será evidenciada com a identificação de segmentos de clientes com

comportamentos diferentes. Cada um dos segmentos irá reunir um conjunto de clientes com necessidades financeiras bem específicas, exigindo assim uma abordagem de marketing diferenciada. O perfil dos clientes que compõem cada um dos segmentos será analisado com base em variáveis demográficas. Com este conhecimento do cliente “padrão” de cada segmento, os gestores de clientes poderão adequar a mensagem de marketing a ser veiculada pelos canais de comunicação. Por último, será avaliado o impacto das acções de marketing realizadas nos segmentos obtidos, de forma a perceber em que medida estas últimas contribuem para a heterogeneidade das necessidades financeiras dos clientes.

Desta forma, é proposto a utilização de modelos de clustering probabilísticos – modelos de mistura finita para a segmentação dos produtos bancários adquiridos de clientes recentes de uma Instituição Financeira. A ambição desta proposta, é evidenciar as mais valias que os referidos modelos podem ter na área da banca. Pretende-se mostrar que a utilização destes modelos pode tornar mais célere o processo de segmentação dos clientes desde a obtenção dos segmentos de clientes, das relações explicativas entre o comportamento de compra e variáveis de marketing, dos perfis dos mesmos com base em variáveis observáveis e classificação de novos clientes. Todas as referidas etapas que são muitas das vezes alcançadas na área da banca através da aplicação de K-means, regressão, análise discriminante, análise de variância simples, e outros métodos podem ser compilados num só passo através dos modelos de mistura.

Com efeito, a utilização dos modelos de mistura finita terá como objectivo o de identificar segmentos homogêneos de clientes relativamente aos produtos detidos, permitindo uma melhor adequação à heterogeneidade das necessidades e preferências dos clientes. Numa primeira fase será aplicado modelo de mistura sem variáveis concomitantes e numa segunda

fase serão introduzidas as variáveis concomitantes para permitir classificar novos clientes com base em informação mais acessível. Em ambos os casos a classificação de novos clientes no segmento mais adequado irá estipular a melhor oferta a propor bem como da forma mais adequada de contactar o cliente.

Será aplicado o modelo de mistura de regressões, com o objectivo de explicar o comportamento dos clientes face aos produtos detidos, com base em variáveis de marketing. Com efeito, será obtido um conjunto de estimativas explicando o impacto das variáveis de regressão nos produtos detidos pelos clientes para cada segmento identificado. Será incluída a componente concomitante do modelo – regressão do grupo de pertença do cliente em ordem a variáveis de caracterização do cliente – que permitem a classificação de futuros clientes nos segmentos previamente identificados. As estimativas serão obtidas pelos algoritmos EM (*Expectation Maximization*) e *Newton- Raphson*, e o número de segmentos será determinado por critérios de informação.

Os clientes alvo do estudo provêm de uma amostra cordialmente disponibilizada por uma Instituição Financeira que por razões de confidencialidade não será revelada a sua identidade. As variáveis base utilizadas para a segmentação correspondem à informação de posse de 8 produtos financeiros.

Para além deste capítulo de introdução, esta tese está organizada em 5 capítulos.

O Capítulo 2 revê o conceito de CRM – *Customer Relationship Management* e a necessidade de ferramentas estatísticas avançadas que o operacionalizem. O mesmo Capítulo aborda ainda

o conceito de segmentação, um aspecto fundamental em qualquer estratégia de marketing e, por definição, crítico na implementação de uma estratégia de CRM.

O Capítulo 3 introduz a ferramenta estatística – modelos de mistura – que aqui se propõe como ferramenta de segmentação na implementação de uma estratégia CRM.

O Capítulo 4 ilustra a metodologia proposta com uma aplicação a uma amostra de clientes de uma Instituição Financeira. Serão identificados os grandes grupos de necessidades financeiras dos clientes durante os primeiros meses de relacionamento com a Instituição Fénix (nome fictício). A simplicidade na identificação das estruturas relacionais, assim como a capacidade preditiva do modelo em identificar o segmento adequado, com base em variáveis demográficas, irá permitir diagnosticar quais as medidas de *cross-selling* futuras a serem tomadas. Será também avaliado o impacto da estratégia de marketing estabelecida, efectuando uma regressão dos indicadores de posse com base nos indicadores de marketing, ao mesmo tempo que os segmentos são obtidos. Este processo irá permitir fornecer informação sobre as condições de continuidade ou de melhoria da estratégia de marketing, e avaliar a necessidade de a diferenciar consoante os segmentos obtidos.

Por último, no Capítulo 5 efectua-se uma síntese dos principais resultados bem como propostas de análises futuras decorrentes dos resultados obtidos na presente tese.

## 2 Os conceitos de CRM e Segmentação de Mercado

### 2.1 CRM - *Customer Relationship Management*

O CRM - *Customer Relationship Management* é muitas das vezes considerado um sistema informático que permite planear e analisar as acções de marketing, identificar os alvos propensos à compra de produtos e gerir os contactos efectuados com os clientes através dos canais existentes, como por exemplo o *telemarketing* (Richardson e Weill, 1999), *mailing* (Turner, 2002; Bult e Wansbeek, 1995), e *banner*<sup>1</sup> (Manchanda *et al.*, 2006). Na área da Banca, à semelhança de outras actividades económicas, o CRM tem recebido especial atenção devido à necessidade de assegurar a satisfação e lealdade dos clientes, de modo a garantir a permanência dos mesmos e assegurar assim o retorno do investimento efectuado pelos accionistas face aos existentes concorrentes (Chen e Li, 2006).

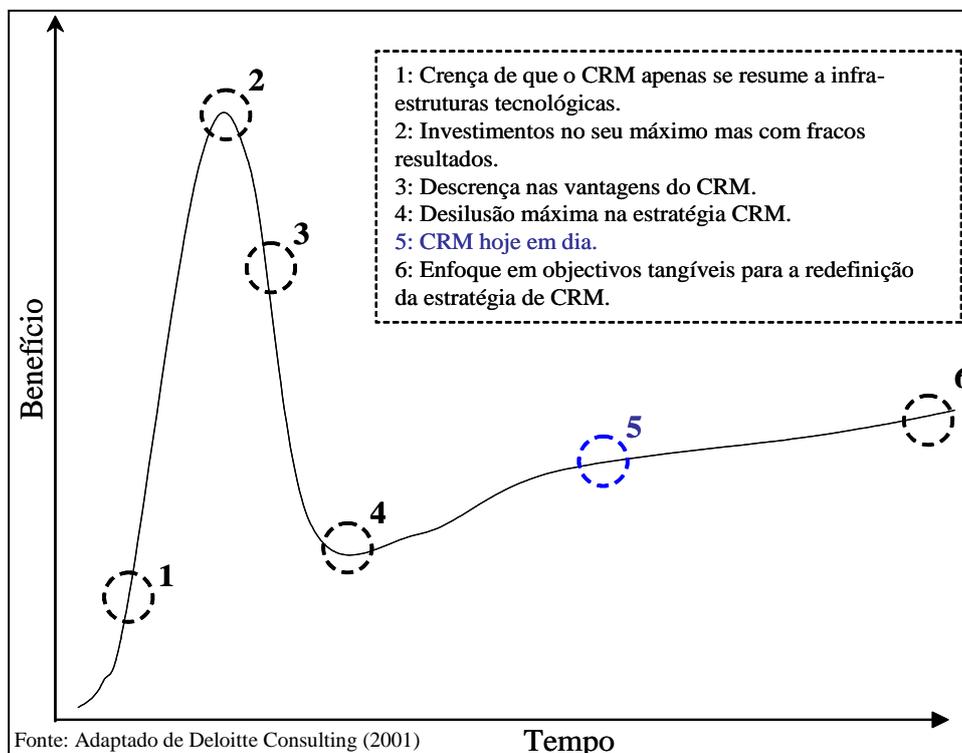
Graças ao desenvolvimento computacional, a partir de 1980 deu-se uma explosão na capacidade de armazenar dados sobre o cliente, criando a necessidade de os organizar para efeitos analíticos dado o potencial da utilização dessa informação. Antecipando-se à aparição de soluções informáticas para contornar o problema de análise de grandes volumes de dados, Peppers *et al.* (1999) introduziram o conceito do Marketing *One-to-One*, e Pine *et al.* (1993) o conceito de *Mass Customization*, que corresponde à definição dos produtos (ou serviços) de acordo com a especificação única de cada cliente. Com base nos conceitos anteriormente referidos, surgiram as soluções de *Software* e de *Hardware*, onde o CRM correspondia ao interface de ligação entre os diversos conjuntos de dados existentes sobre o cliente e a gestão

---

<sup>1</sup> O *banner* é uma publicidade de um produto ou serviço, introduzido numa secção de uma página de internet cujo conteúdo é composto por uma combinação de gráficos e textos. O *banner* contém uma ligação que permite aceder directamente à página de internet da empresa fornecedora do respectivo produto ou serviço (Manchanda *et al.*, 2006).

das actividades de marketing da empresa (Boulding *et al.*, 2005). Estas soluções de CRM tinham como objectivo implementar as infra-estruturas informáticas (por exemplo, um *data warehouse*<sup>1</sup> ou um *datamart*<sup>2</sup>), e sobre as mesmas analisar o comportamento do cliente desencadeando acções de marketing em conformidade com os resultados obtidos das análises efectuadas. Estas acções incidiam na aquisição de novos clientes, melhoria das relações já existentes e retenção dos clientes actuais. O *boom* do CRM deu-se por volta de 1990 quando elevados montantes foram investidos pelas empresas em *softwares* de CRM (Mendoza *et al.*, 2007).

Figura 1 – Fases de maturidade do CRM



A Deloitte Consulting (2001) analisou as diferentes fases (apresentadas por Gartner Research) de crescimento do CRM numa empresa (Figura 1). Este verifica que logo após o início da

<sup>1</sup> Um *data warehouse* é um armazém de dados, ou uma colecção de dados estruturados, consolidando a informação proveniente de vários sistemas operacionais e dedicado à ajuda na tomada de decisão (Berry e Linoff, 2000).

<sup>2</sup> Um *datamart* é um subconjunto dos dados armazenados num *data warehouse* respeitante a uma actividade específica. Contempla um universo de clientes bem delimitado e apenas serão considerados agregações de dados ou atributos considerados pertinentes para a actividade em questão (Lefébure e Venturi, 2005).

implementação de um projecto de CRM numa empresa, este passa por um ciclo de euforia, seguido de um período de desilusão derivado das consecutivas lacunas na definição do seu âmbito, objectivos e metas não concretizadas face ao investimento inicialmente efectuado.

Nesta última década, o CRM passou a ser uma das grandes preocupações nos mercados da Banca e dos Seguros pelo facto de não gerar o retorno esperado face ao investimento inicialmente efectuado (Ebner *et al.*, 2002). As principais causas apontadas para o insucesso dos primeiros projectos de CRM são a definição de objectivos ambiciosos que tornam irreais os prazos estipulados e a pouca adesão verificada entre os requisitos de negócio definidos e as necessidades dos utilizadores finais (Ebner *et al.*, 2002). Para uma integração contínua, e de forma a garantir que os orçamentos inicialmente previstos são respeitados aquando da implementação de uma estratégia de CRM numa organização, é primordial que haja uma definição clara das fronteiras do alcance do CRM. Assim, das inúmeras iniciativas de CRM possíveis, é necessário que a empresa se focalize nas que são verdadeiramente necessárias e não caia no erro de implementar tudo ao mesmo tempo (Peppers *et al.*, 1999; Mendoza *et al.*, 2007). Hoje em dia exige-se uma análise rigorosa e pormenorizada das implicações na implementação de uma estratégia de CRM, de forma a minimizar o risco de insucesso quando se pretende optar por uma visão centrada no cliente (Maklan *et al.*, 2005). Na realidade, o CRM é mais do que uma solução informática; não só permite estabelecer relações comportamentais usando sistemas informáticos para recolher e analisar os dados, mas também pretende integrar todos os departamentos da empresa para que todos colaborem na geração de valor para o cliente, criando ao mesmo tempo valor para o accionista (Boulding *et al.*, 2005).

O CRM, ou gestão da relação com o cliente, pode ser visto como uma *etapa organizacional*, que tem por objectivo *conhecer e satisfazer melhor os clientes identificados pelo seu potencial de actividade com a empresa e sua rendibilidade*, utilizando para tal a *pluralidade dos canais de contacto* existentes, num contexto de uma *relação de longo prazo* com o

cliente, de forma a *incrementar o retorno sobre as vendas e rentabilidade* da empresa (Alard e Guggémos, 2005). De acordo com os mesmos autores, o CRM é orientado por 6 princípios que devem em permanência ser considerados como um “todo”, mas devem também ser analisados individualmente de forma a avaliar as estratégias mais adequadas no desenvolvimento de cada um deles, de forma a garantir a simbiose entre as várias componentes:

1. *Etapa organizacional* - A relação com o cliente não é um procedimento tecnológico ou de departamento, mas sim um processo organizacional;
2. *Conhecer melhor e satisfazer melhor* - Estes dois enfoques são centrais ao CRM, um melhor conhecimento do cliente permite uma melhor resposta às suas necessidades;
3. *Identificação do potencial de actividade e rentabilidade* - Não se pode efectuar uma gestão eficaz da relação com o cliente sem a identificação do retorno do investimento efectuado. Estes dois aspectos permitem adequar as estratégias de interacção com o cliente de acordo com o seu potencial de acordo com o orçamento definido pela empresa. É importante conseguir distinguir o valor actual do cliente do seu valor potencial ou futuro. Por exemplo, um cliente que possui o seu património distribuído por diversos bancos, pode ter uma rentabilidade baixa num determinado banco, e ser considerado pelo mesmo como sendo um cliente com elevado potencial de rentabilidade. O “potencial futuro” de um cliente é um tema complexo, que passa pela identificação de um conjunto de indicadores que permitem inferir sobre o seu potencial futuro;
4. *Pluralidade dos canais de contacto* - A gestão da relação com o cliente passa por um contexto multicanal. É necessário compreender as reacções do cliente derivadas das iniciativas de contacto com origem na empresa e efectuadas através dos canais utilizados para o efeito (e.g., canal *Mailing*, canal *Telemarketing*, etc.). Por outro lado, é necessário

compreender as interacções do cliente com a empresa, e se as razões do contacto são diferentes consoante o canal utilizado (e.g., canal Balcão, canal internet, etc.);

5. *Relação de longo prazo* - Ao contrário de uma abordagem transaccional, a relação com o cliente passa por uma abordagem composta por uma série de transacções contextualizadas na relação existente com o cliente;
6. *Incremento de lucros e rendibilidade* - A gestão da relação com o cliente não pode ter como única finalidade a satisfação deste último, pelo facto do objectivo desta política ser o incremento das vendas e do retorno do investimento derivado das mesmas. Deve ser, assim, colocado em paralelo, o custo dos programas de “CRM” e o retorno esperado do investimento e, em última análise, estimar a rendibilidade consequente dos referidos investimentos.

Contrariamente à visão anterior, de acordo com Peppers *et al.* (1999), a implementação de uma estratégia de CRM passa por quatro etapas: a identificação dos clientes, a diferenciação entre eles, a interacção com os clientes, e a *customização* dos produtos e/ou serviços. A combinação destes quatro passos define os caminhos possíveis para a implementação de um marketing relacional. Consoante as necessidades, a empresa deverá iniciar ou focalizar os seus esforços na implementação da sua estratégia de CRM.

A perspectiva de Peppers *et al.* (1999) pode resumir-se na definição de um ecossistema constituído por três sistemas de actuação: o CRM Operacional, o CRM Colaborativo e o CRM Analítico. O CRM Operacional consiste na implementação da infra-estrutura de suporte ao cliente em cada interacção com a organização (SAS, 2002). O CRM Colaborativo são os pontos efectivos de contacto com os clientes, onde ocorrem as interacções entre eles e a empresa. Através dessas interacções a empresa obtém dados importantes de cada cliente, que

posteriormente serão utilizados na identificação de vantagens competitivas, através da personalização de produtos e serviços (Lemmi, 2007). O CRM Analítico ou *Customer Insight* tem como objectivo compreender as necessidades dos clientes de forma a adequar os produtos e as estratégias de marketing. Para tal, deve reunir todos os dados existentes sobre os clientes e integrá-los numa única plataforma informática, sobre a qual são efectuadas as necessárias análises estatísticas. Para Kamakura *et al.* (2005), o CRM Analítico assenta em três etapas: a aquisição, o desenvolvimento e a retenção dos clientes. A fase de aquisição corresponde à captação de novos clientes para a empresa. A segunda etapa tem como objectivo aumentar a rendibilidade dos clientes já existentes. Por último, a terceira fase do CRM Analítico, consiste em reter os clientes, evitando que estes efectuem negócio com empresas concorrentes. Os conceitos de CRM aqui apresentados, definem uma visão centrada no cliente, de modo a alcançar um maior conhecimento sobre o mesmo. Este conhecimento fornece à empresa a informação necessária para que esta possa decidir sobre as formas mais adequadas de interacção com os seus clientes, garantindo a satisfação e o aumento do valor para os mesmos, assim como do incremento dos seus lucros.

No mercado Português verifica-se que são poucas as empresas que conseguiram implementar com sucesso as soluções de CRM, isto é, que tenham atingido o retorno sobre o investimento efectuado inicialmente (Beirão, 2004). A mudança de agulhas de um Marketing tradicional (orientado para o produto) para um Marketing relacional (orientado para o cliente) não se concretizou dado os últimos anos de crise vividos em Portugal. Com efeito, Beirão (2004: 5) afirma que “... não vale a pena pensar em CRM se o modelo tradicional de negócio se mantiver, isto é, se a organização interna continuar centrada nos produtos ou em mercados geográficos e não organizada por segmentos de clientes”. Desta forma, a empresa deve passar de uma visão centrada no produto, para uma visão centrada no cliente. A mesma autora

aponta para a existência de “quintinhas” e disputas internas nas empresas que não favorecem a implementação do CRM.

No entanto, as consultoras apontam para alguns casos de sucesso. A Sociedade Central de Cerveja e bebidas (SCC) (SAP, 2006) desenvolveu um sistema informático integrado e centralizado no cliente, fazendo com que ao contactar o *Call Center*, o cliente tenha consciência de que a SCC sabe quem ele é, quais foram as suas necessidades passadas, dando assim à SCC as condições necessárias para prestar um serviço à medida das suas necessidades presentes. A Allianz Portugal (SAS, 2006), companhia de seguros, consegue identificar a melhor oferta a propor ao cliente no momento certo. O Banco Santander Totta (SAS, 2007) aumentou a sua taxa de sucesso nas acções de aquisição do cartão de crédito, e diminuiu a sua taxa de abandono antecipando as necessidades financeiras do cliente.

Na Banca, o CRM Analítico é utilizado para maximizar a rendibilidade de um cliente. Ebner *et al.* (2002) explica que uma forma de atingir esta meta é efectuando vendas cruzadas, isto é, realizando acções de *cross-selling*<sup>1</sup>. O relatório anual 2003 de Wells Fargo (2003: 16) diz que “80% do nosso crescimento advém da venda de mais produtos aos clientes existentes”. Para além de aumentar a rendibilidade, o *cross-selling* permite fortalecer a relação existente entre o cliente e a empresa. A dinâmica do *cross-selling* representa uma actividade importante na área da banca. Com efeito, a oferta de um pacote *bundled* (oferta constituída por mais do que um produto) garante uma maior fidelização do cliente. O *cross-selling* é função do ciclo de vida do cliente na empresa, sendo esta uma oportunidade de negócio quando a proposta de valor apresentada responder às necessidades do cliente. Uma forma de a banca definir o *cross-*

---

<sup>1</sup> O *cross-selling* visa incentivar o cliente actual a comprar produtos diferentes.

*selling* a ser proposto a cada cliente é com o recurso ao processo de segmentação (Speed e Smith, 1992).

## **2.2 Segmentação de mercado**

Existem tantas definições de segmentação quantos os investigadores na área do marketing. Uma das primeiras referências sobre a segmentação de mercados é da autoria de Wendell Smith em 1956. Este autor define a segmentação como sendo uma visão desagregada dos efeitos derivados da existência de produtos e estratégias de marketing diferentes, que por sua vez tendem a identificar diferentes comportamentos na procura, quando apenas um tipo de comportamento era anteriormente assumido. A segmentação consiste em descrever a heterogeneidade existente no mercado, conseqüente dos diferentes comportamentos e preferências observados, identificando um número de mercados com menor dimensão mas mais homogêneos nessas preferências. Esta identificação pode levar as empresas a desenvolver novos produtos ou serviços, e adequar a comunicação a ser efectuada de forma a promover os novos produtos criados. Nos dias de hoje, esta definição continua a responder às preocupações das empresas, pois esta definição apresenta o processo de segmentação como sendo um modelo conceptual que descreve a forma como o gestor observa o seu mercado, de forma a melhor conhecer o potencial de rendibilidade dos clientes existentes (Sarvary, 2005). Os segmentos obtidos não correspondem a agrupamentos naturais ou *a priori* dos indivíduos, mas são determinados com base em factores relevantes para a compreensão da estrutura existente no mercado (Smith, 1956). O objectivo de uma segmentação é encontrar grupos de indivíduos que partilham das mesmas preferências, assegurando que os grupos criados sejam suficientemente diferentes uns dos outros, e onde cada segmento pode ser alvo de uma mesma abordagem de marketing pelo facto de possuírem as mesmas necessidades ou preferências.

Wedel e Kamakura (1999) apontam para a existência de diversos tipos de segmentação consoante o objectivo a ser alcançado. Por exemplo, uma segmentação pode ser efectuada para evidenciar os diferentes comportamentos nas necessidades existentes nos indivíduos, ou pode ser efectuada com a finalidade de agrupar os clientes com base nas suas características demográficas. Apesar da primeira ter como resultado evidenciar os segmentos potencialmente mais lucrativos, esta depende frequentemente de informação dificilmente observável, sendo disponível, muitas das vezes, mediante a realização de inquéritos; a segunda permite identificar grupos de clientes com características semelhantes e que tornam mais simples o processo de selecção dos clientes para o envio de mensagens promocionais. Esta última será vantajosa caso esteja correlacionada com a estrutura de segmentos. Assim, verifica-se que para um mesmo universo de clientes diversas segmentações podem ser obtidas com diferentes finalidades. Apesar de raramente efectuado na prática, há toda a vantagem em que este processo seja simultâneo.

O CRM advoga a existência de heterogeneidade nos clientes no que respeita às suas necessidades de produtos ou serviços. Como tal, a segmentação tem vindo a tornar-se uma prática recorrente em muitas das organizações, permitindo o conhecimento das características que discriminam cada um dos segmentos de forma a providenciar a “*customização*” de produtos ou serviços conforme os segmentos de clientes obtidos. De forma a perceber essas diferenças, é necessário particionar a base de clientes em segmentos de menor dimensão, de forma a dotar os técnicos de marketing, do conhecimento necessário para a implementação de uma adequada estratégia de contacto e um correcto posicionamento dos produtos em cada segmento. Mendonza *et al.* (2007) considera que a segmentação dos clientes é um factor crítico de sucesso para a implementação de uma estratégia de CRM, pelo facto de representar um dos caminhos para o conhecimento do cliente.

Usualmente, as empresas utilizam os modelos de propensão com base em dados históricos do cliente para estimar a probabilidade de um cliente comprar um novo produto e o retorno esperado desta nova aquisição (Cohen, 2004). Mas esta forma de actuação não permite otimizar uma acção de Marketing. Pelo facto de ser efectuada uma única oferta, não é maximizado o retorno sobre o investimento efectuado no contacto, não propondo vários produtos em simultâneo. Assim, o conhecimento da probabilidade de um cliente adquirir apenas um produto não é suficiente quando uma empresa possui uma carteira de vários produtos a oferecer. Os departamentos de Marketing enfrentam um novo paradigma, que é o de conhecer quais os produtos a oferecer aos seus clientes e não quais os clientes a serem alvos de uma acção de marketing para a aquisição de um só produto (Cohen, 2004). Bijmolt *et al.* (2004) referem que o conhecimento sobre os produtos adquiridos pelos clientes, representa uma informação relevante para uma tomada de decisão no que diz respeito ao desenvolvimento e introdução de novos produtos no mercado. Pelo facto da segmentação permitir a obtenção de um maior conhecimento dos seus clientes, antecipando assim algumas das suas necessidades, esta permite gerar uma relação de proximidade com o cliente, perceber as diferentes relações estruturais existentes no mercado e assim avançar com estratégias de *cross-selling*.

Na área da banca, as técnicas de segmentação não são apenas utilizadas para o estudo da heterogeneidade dos clientes face aos produtos detidos. Kaynak e Harcar (2005) analisaram quais eram os factores relevantes que motivavam os clientes a escolher um banco rural (neste caso um dos bancos locais existentes na Pensilvânia, Estados Unidos da América) em detrimento de um banco nacional ou vice-versa. Este estudo visava identificar os atributos mais relevantes na escolha entre os dois tipos de banco e determinar a utilidade que detinham

determinados serviços prestados pelos bancos locais. Tal conhecimento, permitiu afinar a estratégia de captação de novos clientes na referida zona geográfica. A segmentação foi efectuada através de uma análise factorial. Foram identificados 6 factores para a discriminação do comportamento entre os clientes de bancos locais e os clientes de bancos nacionais. As diferenças estatisticamente significativas revelaram quais os serviços percebidos de formas diferente entre os dois grupos de clientes. Para além do conhecimento dos atributos diferenciáveis, esta análise permitiu ao gestor de marketing afinar o portfolio de produtos ou serviços a oferecer consoante o perfil dos clientes dos bancos locais e bancos nacionais. Esta perfilagem foi realizada utilizando o teste de independência qui-quadrado aplicado entre o indicador do tipo de banco a que pertencia o cliente e cada uma das variáveis descritivas.

No estudo acima apresentado (Kaynak e Harcar, 2005), os processos de segmentação e perfilagem efectuaram-se em 3 etapas:

Etapa 1: Identificação dos construtos com a realização de uma análise factorial sobre um conjunto de medidas da utilidade dos serviços prestados pelos bancos locais e nacionais. Dessa análise resultaram a identificação de 6 factores.

Etapa 2: Análise da igualdade da importância de cada um dos 6 factores entre os clientes de bancos locais e clientes de bancos nacionais. Essa análise permitiu identificar que factores eram percebidos de forma diferente entre os clientes de bancos locais e os clientes dos bancos nacionais.

Etapa 3: Efectuou-se uma perfilagem dos dois grupos de clientes, com recurso ao teste de independência do qui-quadrado. Analisou-se para cada uma das variáveis sócio-demográficas, a existência de independência nos dois grupos de clientes.

Lawson e Todd (2003) apresentaram um estudo com o objectivo de evidenciar a importância de uma segmentação efectuada com base em dados de preferência face a uma segmentação efectuada com base em variáveis demográficas para explicar a heterogeneidade dos clientes nas preferências dos métodos de pagamentos utilizados. Para a identificação dos segmentos foi utilizado o método *K-Means* sobre um conjunto de variáveis de preferência. Foram analisadas soluções com 2 a 5 segmentos, sendo escolhida a solução com 3 segmentos. A decisão do número de segmentos finais foi efectuada conjugando 3 critérios de avaliação: o critério da bondade do ajustamento “invertido” mediante uma análise discriminante, o critério de avaliação da relação de dependência entre as variáveis de segmentação e as diversas soluções; e por fim o critério de interpretação da solução final. O primeiro critério consiste em efectuar uma análise discriminante para cada conjunto de segmentos considerados, e com base numa amostra de dados perceber em qual dos conjuntos de segmento, a análise discriminante possui a taxa de má classificação mais baixa. O segundo critério pretende escolher o número de segmentos a ser considerado conforme as relações de dependência existentes com as variáveis de base de segmentação. O terceiro critério é uma medida de segurança no sentido em que não vale a pena apresentar uma proposta de segmentação se esta não pode ser interpretada, inviabilizando a tomada de medidas de actuação. A perfilagem dos 3 segmentos foi efectuada com base em ANOVA (Análise de variância simples) e tabelas de contingência. Por fim, foi utilizada uma análise discriminante para avaliar o poder preditivo entre as variáveis demográficas e os 3 segmentos.

No estudo desenvolvido por Kaynak e Harcar (2005), os processos de segmentação e perfilagem poderiam ter sido efectuados num só passo mediante a utilização do modelo de mistura de regressões com variáveis concomitantes. No caso do estudo de Lawson e Todd (2003), a aplicação de um modelo de mistura finita com concomitantes teria evitado as

múltiplas análises na identificação dos segmentos e perfilagem dos mesmos. Com efeito, este modelo permite obter a probabilidade de pertença de um cliente num dado segmento com base exclusivamente em variáveis concomitantes, sendo neste caso as variáveis demográficas utilizadas na perfilagem dos três segmentos. Por outro lado, a utilização dos critérios de informação teria tornado mais célere o processo de identificação do número de segmentos.

Na presente tese, iremos apresentar uma técnica de segmentação - modelos de mistura (Wedel e Kamakura, 1999) - que permite simultaneamente efectuar uma segmentação, para melhor compreender a heterogeneidade existente nas necessidades financeiras dos clientes, e obter as probabilidades de pertença *a posteriori*, as quais irão permitir classificar o cliente no segmento mais adequado dado um conjunto de características facilmente observáveis.



### **3 Modelo de mistura finita**

#### **3.1 Introdução**

Os modelos de mistura têm ganho protagonismo na área de estudos de mercado, pelas suas vantagens face aos métodos tradicionais de agrupamento. Os modelos de mistura assumem que os clientes provêm de diversos segmentos homogéneos do mercado que foram misturados em proporções desconhecidas. Não é dado a conhecer o segmento a que pertence o cliente, ou quantos segmentos existiam inicialmente, sendo estes não observáveis ou usualmente designados por latentes. Assim, o objectivo destes modelos é o de desfazer essas misturas de forma a ser revelado o número de segmentos e a definição das respectivas estruturas que compõem a população em análise (Cohen e Ramaswamy, 1998).

Cohen e Ramaswamy (1998) e Wedel e Kamakura (1999) explicam que a grande diferença entre as técnicas de agrupamento e os modelos de mistura consiste no tipo de problemas para os quais estes últimos conseguem dar resposta. As técnicas tradicionais de agrupamento são métodos descritivos que não entram em consideração com relações de dependência, ou sequer estruturas de predição. Embora os modelos de mistura possam ser utilizados na perspectiva de uma segmentação de carácter meramente descritivo, a sua principal vantagem, assenta na sua capacidade de segmentação e predição. Em resumo, os modelos de mistura finita assumem que a população é heterogénea, sendo constituída por duas ou mais sub-populações em proporções desconhecidas (Wedel e Kamakura, 1999), sendo o seu objectivo, o de identificar as componentes de mistura ou sub-populações, de forma a recuperar a estrutura da população em causa. Assim, a grande popularidade dos modelos de misturas finitas na área do Marketing deve-se principalmente a dois factores. O primeiro diz respeito à sua capacidade na descrição das relações estruturais existentes num dado mercado, sendo estas obtidas pela decomposição da heterogeneidade existente em segmentos de clientes homogéneos. Desta forma, os modelos de mistura finita podem ser utilizados como técnica de segmentação, permitindo a identificação de grupos de clientes semelhantes nos seus comportamentos (Wedel e Kamakura, 1999).

O segundo factor prende-se com a simplicidade na obtenção da função de probabilidade marginal, sendo esta definida pela mistura das funções de probabilidade condicionadas aos segmentos obtidos ponderadas pelas seus pesos relativos. Na prática, apenas é observada a distribuição marginal, uma vez que o indicador do segmento ou classe latente de onde é proveniente cada observação, não é observado, e devendo por isso ser estimado.

Para cada cliente, os modelos de mistura estimam a probabilidade de pertença a cada segmento obtido, permitindo assim perceber as relações de pertença que cada cliente possui com os diversos grupos. Muitos algoritmos de segmentação não têm incorporado essa estrutura “Fuzzy” (Cardoso *et al.*, 2000), sendo este factor considerado uma melhoria a ser incorporada em determinados métodos de agrupamentos. Tsai e Chiu (2004) referem esse ponto na utilização de um algoritmo genético para a realização de uma segmentação com base nos comportamentos de compra dos clientes. Saunders (1994) advoga que os clientes podem pertencer a mais do que um segmento dependendo das circunstâncias. Em determinados contextos, é possível verificar a existência de clientes que vão pertencendo a vários segmentos.

Nos modelos de mistura finita podem ser utilizadas variáveis de caracterização do cliente e também de caracterização do produto na sua medida original, sem haver a necessidade de efectuar uma transformação de forma a reduzir a informação a um único espaço métrico. A utilização da informação na sua medida original potencia a eficiência da segmentação (Tsai e Chiu, 2004). Tal facto não acontece nos métodos de agrupamentos hierárquicos tais como o método do vizinho mais próximo, onde os clientes são agrupados consoante uma medida de semelhança resultante de uma transformação efectuada aos dados. A não utilização da informação na sua medida de origem pode derivar numa segmentação que não está orientada na determinação das relações estruturais existentes no que diz respeito aos comportamentos de preferências existentes (Cohen e Ramaswamy, 1998). Os modelos de mistura também permitem trabalhar em simultâneo com variáveis nominais, ordinais e contínuas, ao contrário

do método clássico *K-means* onde se recomenda a utilização de variáveis contínuas (Saunders, 1994).

### 3.2 Modelos de mistura finita

Os modelos de mistura finita são muito utilizados para a análise de agrupamentos ou *clustering*, sendo utilizados nas mais diversas áreas de Marketing e Ciências Sociais (Wedel e Kamakura, 1999). De forma a tornar mais transparente a exposição destes modelos em grau crescente de complexidade, segue-se a definição das parcelas elementares que os compõem.

Seja uma amostra de  $n$  clientes, sendo o  $i$ -ésimo cliente da amostra designado por  $i$  ( $i=1, \dots, n$ ). Cada cliente possui ou não  $J$  possíveis produtos, sendo  $y_{ij}$  a escolha do produto  $j$  observado para o cliente  $i$ , utilizando para tal uma variável *dummy* ou indicatriz. O modelo de mistura finita, composto por  $S$  segmentos ou classes latentes para  $y_i$ , é definido pela seguinte função de probabilidade (Wedel e Kamakura, 1999):

$$f(y_i; \varphi) = \sum_{s=1}^S \pi_s f_s(y_i; \theta_s), \quad (1)$$

onde  $f_s(y_i; \theta_s)$  corresponde à função de probabilidade de  $y_i = (y_{i1}, \dots, y_{iJ})$  condicionada ao segmento  $s$  com o vector de parâmetros  $\theta_s$ ,  $\pi_s$  corresponde ao peso do segmento  $s$  na população e  $\varphi = (\pi_1, \dots, \pi_{S-1}, \theta_1, \dots, \theta_S)$  define o conjunto dos parâmetros a serem estimados. Note que apenas  $S-1$  pesos são independentes e estimados, uma vez que estão sujeitos às restrições:

$$0 < \pi_s \leq 1 \text{ e } \sum_{s=1}^S \pi_s = 1. \quad (2)$$

O modelo de mistura finita assenta no pressuposto da independência local que admite que, condicionado ao seu segmento de pertença, as variáveis  $y_i$  são independentes (Boter e Wedel, 1999). Desta forma, o modelo de mistura pode ser definido pela seguinte expressão:

$$f(\mathbf{y}_i; \boldsymbol{\varphi}) = \sum_{s=1}^S \pi_s \prod_{j=1}^J f_s(y_{ij}; \boldsymbol{\theta}_{js}). \quad (3)$$

É usual assumir que as funções densidades  $f_s(\mathbf{y}_i; \boldsymbol{\theta}_s)$  pertencem à mesma família paramétrica, no entanto esta restrição não é uma regra. Qualquer função de probabilidade pode ser utilizada num modelo de mistura, desde que se garanta que o modelo esteja identificado. Por exemplo, Wedel *et al.* (1993) utilizam a função de distribuição de Poisson para identificar os tipos de comportamentos existentes (ou segmentos) relativamente à incidência de compra de livros verificada num dado intervalo de tempo.

Por outro lado, para realizar uma segmentação com base em indicadores de compra de determinados produtos de lazer, Boter e Wedel (1999) recorrem à distribuição multinomial, e neste caso o modelo toma a seguinte forma:

$$f(\mathbf{y}_i; \boldsymbol{\varphi}) = \sum_{s=1}^S \pi_s \prod_{j=1}^J p_{js}^{y_{ij}}, \quad (4)$$

onde  $p_{js}$  corresponde à probabilidade do cliente  $i$  adquirir o produto  $j$  condicionado ao segmento de pertença  $s$ , e assume-se que esta é igual para todos os clientes pertencentes ao segmento  $s$ . A probabilidade do cliente  $i$  adquirir o produto  $j$  pode ser parametrizada da seguinte forma (Vermunt e Magidson, 2005):

$$p_{js} = \frac{\exp(\beta_{0j} + \beta_{js} y_{ij})}{\sum_{j=1}^J \exp(\beta_{0j} + \beta_{js} y_{ij})}, \quad (5)$$

onde  $\beta_{0j}$  representa a constante e  $\beta_{js}$  é o efeito do segmento  $s$  na compra do produto  $j$ , com  $\beta_{J_s}=0$ . O vector  $y_i$  é constituído por  $J$  variáveis dummy's, em que cada uma delas indica se o cliente adquiriu o produto  $j$ .

Desta forma,  $f(y_i; \varphi)$  é a probabilidade de observar o comportamento de aquisição dos produtos para o cliente  $i$ . O modelo de mistura finita providencia ainda as probabilidades  $a$  *posteriori* de pertença de cada cliente aos seus segmentos:

$$a_{is} = \frac{\pi_s \prod_{j=1}^J p_{js}^{y_{ij}}}{\sum_{r=1}^S \pi_r \prod_{j=1}^J p_{jr}^{y_{ij}}}. \quad (6)$$

Se o modelo de classe latentes for aplicado a uma base de novos clientes, estas probabilidades *a posteriori* irão identificar para cada cliente, qual o segmento de pertença mais provável (Boter e Wedel, 1999).

### 3.3 Modelos de mistura de regressões

Nos modelos tradicionais de regressão, o objectivo é estimar uma equação de regressão para toda a população. Assim, assume-se a existência de uma só população homogénea em que todos os clientes possuem o mesmo comportamento. Desta forma, estes modelos podem levar a conclusões erradas no caso da população ser heterogénea no que diz respeito aos comportamentos observados. Caso sejam identificadas sub-populações, e sobre as mesmas forem aplicados os modelos de regressão, então os resultados obtidos seriam mais concordantes com a verdadeira estrutura da população (DeSarbo e Ramaswamy, 1994). As

duas etapas, a segmentação e a regressão, são realizadas em simultâneo no modelo de mistura de regressões. O comportamento observado passa a ser explicado pelo conjunto de coeficientes de regressão estimados para cada um dos segmentos considerados. Por segmento, teremos a explicação do comportamento dos clientes observados, obtido como função das variáveis independentes consideradas, também designadas por covariáveis. Tal como nos modelos de mistura finita, também são calculadas as probabilidades de pertença *a posteriori*, permitindo assim, classificar um novo cliente num dos segmentos obtidos.

No marketing, existe cada vez mais a necessidade de identificar segmentos de clientes heterogéneos no que diz respeito à aquisição de bens, produtos ou serviço, e perceber para cada cliente os motivos que explicam a diversidade destes comportamentos (Wedel e Kamakura, 1999). Green *et al.* (1976) aplicaram os modelos de mistura de regressões na área das telecomunicações, onde realizaram uma segmentação sobre a utilização ou não de um novo serviço de telecomunicações, utilizando para factores de regressão covariáveis demográficas (educação, idade, mobilidade, e vencimento) de forma a melhor orientar as estratégias de marketing. Kamakura e Russell (1989) realizaram uma segmentação sobre a escolha da marca para a compra de um mesmo “produto”, explicando em cada um dos segmentos, as relações existentes entre as preferências observadas e a respectiva sensibilidade, ou não, ao preço praticado na marca seleccionada. Estes modelos podem ser aplicados em outras áreas. Vermunt e Magidson (2003) aplicaram estes modelos nas áreas sociais ao efectuar uma segmentação sobre o nível de satisfação dos trabalhadores de uma empresa, explicando os comportamentos observados com base na etnia, sexo, idade e zona geográfica.

Seja  $\mathbf{x}_{ij}$  o vector de covariáveis associado ao cliente  $i$  para o produto  $j$  e  $\mathbf{x}_i = \langle \mathbf{x}_{i1}, \dots, \mathbf{x}_{iJ} \rangle$  todas as covariáveis associadas ao cliente  $i$ , a função de distribuição marginal é então definida pela seguinte expressão:

$$f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\varphi}) = \sum_{s=1}^S \pi_s f_s(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}_s). \quad (7)$$

Nestes modelos também é assumido o pressuposto da independência local. Assim assume-se que as observações  $\mathbf{y}_i = \langle y_{i1}, \dots, y_{iJ} \rangle$  são independentes entre si condicionadas ao segmento de pertença  $s$ . Com base neste pressuposto, a obtenção da função de distribuição marginal é simplificada, pois de uma função de distribuição multivariada  $f_s(\mathbf{y}_i; \boldsymbol{\theta}_s)$ , passa a ser considerada o produto de funções de distribuição univariada, isto é, a função conjunta  $f_s(\mathbf{y}_i; \boldsymbol{\theta}_s)$  passa a ser definida pela seguinte expressão:

$$f_s(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}_s) = \prod_{j=1}^J f_s(y_{ij} | \mathbf{x}_{ij}, \boldsymbol{\theta}_{js}). \quad (8)$$

Ficamos, assim, com a seguinte função de distribuição marginal:

$$f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\varphi}) = \sum_{s=1}^S \pi_s \prod_{j=1}^J f_s(y_{ij} | \mathbf{x}_{ij}, \boldsymbol{\theta}_{js}). \quad (9)$$

Cada mistura é, assim, definida pela sua função de distribuição univariada  $f_s(y_{ij} | \mathbf{x}_{ij}, \boldsymbol{\theta}_{js})$ .

Uma vez estimados o vector de parâmetros  $\boldsymbol{\varphi}$ , as probabilidades *a posteriori* de um cliente  $i$  pertencer ao segmento  $s$  podem ser obtidas recorrendo ao teorema de Bayes (Wedel e Kamakura, 1999):

$$\alpha_{is} = \frac{\pi_s \prod_{j=1}^J f_s(y_{ij} | \mathbf{x}_{ij}, \boldsymbol{\theta}_{js})}{\sum_{r=1}^S \pi_r \prod_{j=1}^J f_r(y_{ij} | \mathbf{x}_{ij}, \boldsymbol{\theta}_{jr})}. \quad (10)$$

As estimativas de  $\alpha_{is}$  providenciam a distribuição da alocação dos clientes nos vários segmentos. Como já referido anteriormente, a função  $f_s(y_{ij} | \mathbf{x}_{ij}, \boldsymbol{\theta}_{js})$  utilizada na Expressão (9) pode tomar diversas formas. No âmbito do presente estudo, será analisado com maior detalhe os modelos de mistura de regressões para dados binários.

### 3.4 Modelo de mistura de regressões para dados binários

O modelo de mistura de regressões para dados binários foi desenvolvido no contexto do marketing por Kamakura e Russell (1989). A variável dependente resulta da observação da marca escolhida pelo cliente para a compra do produto “molho de tomate” durante um período de tempo. Os autores aplicaram o modelo de mistura para efectuar a identificação de segmentos de clientes com comportamentos de escolha semelhantes, e perceber em cada segmento, a importância do factor preço nos comportamentos observados. Para o presente estudo, não serão observadas as escolhas efectuadas durante um determinado período de tempo, mas iremos observar os produtos adquiridos de entre  $J$  produtos possíveis, e serão utilizadas várias variáveis independentes. Por se tratar de dados discretos passamos a utilizar a designação  $P(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\varphi})$  em vez da expressão  $f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\varphi})$ .

A variável dependente  $y_i$  corresponde à aquisição ou não de  $J$  produtos, e assumindo o pressuposto da independência local, a função massa de probabilidade marginal é definida

pelas misturas de  $J$  distribuições Bernoulli independentes condicionadas ao segmento de pertença  $s$  :

$$P(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\varphi}) = \sum_{s=1}^S \pi_s \prod_{j=1}^J p_{js}(\mathbf{x}_{ij}; \boldsymbol{\beta}_{js})^{y_{ij}} \mathbf{1} - p_{js}(\mathbf{x}_{ij}; \boldsymbol{\beta}_{js})^{\bar{1}-y_{ij}}, \quad (11)$$

onde  $y_{ij}$  assume o valor 1 no caso do cliente  $i$  ter adquirido o produto  $j$ , e toma o valor 0 caso contrário. Para além disso,  $p_{js}(\mathbf{x}_{ij}; \boldsymbol{\beta}_{js})$  corresponde à probabilidade do cliente  $i$  adquirir o produto  $j$  condicionado ao segmento  $s$ . As variáveis  $\mathbf{x}_{ij}$  são as covariáveis utilizadas para efectuar a regressão de  $p_{js}(\mathbf{x}_{ij}; \boldsymbol{\beta}_{js})$  e  $\boldsymbol{\beta}_{js}$  é o vector de coeficientes de regressão a serem estimados para o produto  $j$  no segmento  $s$ .

Para a estimação de  $p_{js}(\mathbf{x}_{ij}; \boldsymbol{\beta}_{js})$ , Kamakura e Russel (1989) optaram pela função de ligação logit. Assim, a probabilidade do cliente  $i$  ter adquirido o produto  $j$  condicionada ao segmento  $s$ , passa a ser definida por:

$$p_{js}(\mathbf{x}_{ij}; \boldsymbol{\beta}_{js}) = \frac{\exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}{1 + \exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}. \quad (12)$$

Os pesos dos segmentos  $\pi_s$  são obtidos pela seguinte transformação multinomial logit:

$$\pi_s = \frac{\exp(\lambda_s)}{\sum_{r=1}^S \exp(\lambda_r)}. \quad (13)$$

Para que a Expressão (13) possa ser identificada, isto é, para garantir a existência de um único conjunto de soluções, é necessário fixar uma das categorias (categoria de referência), sobre a qual as restantes serão comparadas. Para tal, coloca-se um dos parâmetros igual a zero. Usualmente esta restrição é imposta ao parâmetro da última categoria ou segmento considerado. Assim, temos  $\lambda_S = 0$ .

Neste caso foi utilizada a transformação multinomial logit, mas qualquer outra função de ligação pode ser usada desde que verifique as seguintes restrições:

$$0 < \pi_s \leq 1 \text{ e } \sum_{s=1}^S \pi_s = 1. \quad (14)$$

Para além de  $\pi_s$  representar o peso relativo do segmento  $s$ , este parâmetro é interpretado como sendo a probabilidade de um cliente escolhido ao acaso pertencer ao segmento  $s$ . Alterando a Expressão (11), pela actualização das parcelas  $p_{js}(\mathbf{x}_{ij}; \boldsymbol{\beta}_s)$  definida na Expressão (12), obtém-se a seguinte função massa de probabilidade,

$$P(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\varphi}) = \sum_{s=1}^S \pi_s \prod_{j=1}^J \left[ \left( \frac{\exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}{1 + \exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})} \right)^{y_{ij}} \left( 1 - \frac{\exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}{1 + \exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})} \right)^{1-y_{ij}} \right],$$

onde  $\boldsymbol{\varphi} = \{\boldsymbol{\lambda}, \boldsymbol{\beta}\}$  corresponde ao conjunto de parâmetros do modelos a serem estimados e  $y_{ij}$  toma o valor 1 caso o cliente  $i$  tenha adquirido o produto  $j$ , e toma valor 0 no caso contrário.

Após a obtenção das estimativas de  $\boldsymbol{\lambda}$  e  $\boldsymbol{\beta}$ , as probabilidades *a posteriori* são fornecida pela seguinte expressão:

$$a_{is} = \frac{\pi_s \prod_{j=1}^J \left[ \frac{\exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}{1 + \exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})} \right]^{y_{ij}} \left[ 1 - \frac{\exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}{1 + \exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})} \right]^{1-y_{ij}}}{\sum_{r=1}^S \pi_r \prod_{j=1}^J \left[ \frac{\exp(\beta_{0r} + \boldsymbol{\beta}_{jr} \mathbf{x}_{ij})}{1 + \exp(\beta_{0r} + \boldsymbol{\beta}_{jr} \mathbf{x}_{ij})} \right]^{y_{ij}} \left[ 1 - \frac{\exp(\beta_{0r} + \boldsymbol{\beta}_{jr} \mathbf{x}_{ij})}{1 + \exp(\beta_{0r} + \boldsymbol{\beta}_{jr} \mathbf{x}_{ij})} \right]^{1-y_{ij}}}. \quad (15)$$

### 3.5 Modelo de mistura de regressões com variáveis concomitantes

Os modelos de mistura de regressões de per si não permitem identificar os segmentos de pertença para novos clientes. Este problema acontece por exemplo nas áreas de Demografia, Psicologia, ou de Marketing, onde os segmentos são derivados à custa de variáveis base da segmentação, cuja respectiva obtenção para novos clientes implica custos elevados e processos morosos (por exemplo quando as variáveis base de segmentação derivam da realização de um inquérito). Estes grupos ou segmentos necessitam de ser perfilados em simultâneo com variáveis concomitantes que pretendem ser menos dispendiosas na sua obtenção, ou facilmente disponíveis tais como as variáveis demográficas (Wedel, 2002). Uma vez identificados os grupos, os novos clientes são classificados usando apenas a informação de caracterização demográfica.

Gupta e Chintagunta (1994) propõem no contexto da segmentação de mercados uma nova abordagem, onde as probabilidades de pertença *a priori* passam a ser funções de determinadas características que se espera, relacionadas com os segmentos, de forma a possibilitar a identificação do segmento de pertença para um novo cliente. As características referidas anteriormente são designadas por variáveis concomitantes, e definem um sub-modelo em que o segmento de pertença do cliente é definido pela variável latente discreta  $Z$ , e esta, passa a ser função das variáveis concomitantes. No estudo apresentado em Gupta e Chintagunta (1994) as variáveis concomitantes explicam a classificação dos clientes em segmentos, pelo que não devem ser ignoradas no processo de estimação do modelo de mistura de regressões. Assim, esta metodologia permite, não só caracterizar os segmentos de forma qualitativa e

quantitativa em termos de resposta, mas também identificar o segmento de pertença para um novo cliente mediante a utilização das variáveis concomitantes.

O modelo definido por Gupta e Chintagunta (1994) é uma extensão directa do modelo de mistura de regressões proposto por Kamakura e Russel (1989) mas agora são incluídas as variáveis concomitantes para a estimação das probabilidades de pertença dos clientes a cada segmento. Os autores assumem um modelo de regressão logístico para definir a probabilidade do cliente  $i$  pertencer ao segmento  $s$  dado um conjunto de variáveis concomitantes, e por razões de identificação do modelo é imposta a condição  $\gamma_S = 0$ :

$$\pi_{is} = \frac{\exp(w_i^T \gamma_s)}{1 + \sum_{r=1}^{S-1} \exp(w_i^T \gamma_r)}, \quad (16)$$

onde  $w_i^T$  corresponde ao vector transposto das variáveis concomitantes observadas para o cliente  $i$  e  $\gamma_s = (\gamma_1, \dots, \gamma_{S-1})$  corresponde ao vector dos coeficientes de regressão a serem estimados e representa o impacto entre pertencer ao segmento  $s$  e pertencer ao segmento  $S$  provocado pelo efeito das variáveis concomitantes. Por razões de identificação do modelo, o coeficiente  $\gamma_S$  toma o valor zero. Com esta parametrização, o segmento  $S$  passa a ser o segmento base de comparação. Os  $\pi_{is}$ , com  $i = 1, \dots, n$  e  $s = 1, \dots, S$ , devem respeitar as seguintes condições:

$$\sum_{s=1}^S \pi_{is} = 1 \text{ e } \pi_{is} > 0. \quad (17)$$

A mais valia deste modelo é, sem dúvida, a circunscrição dos segmentos num espaço bem identificado mediante as variáveis concomitantes, passando assim a ser viável, graças às probabilidades estimadas de pertença ao segmento obtidas para os novos clientes, apenas desenvolver acções promocionais para os segmentos pretendidos. Retomando a Expressão

(11), e considerando os pesos relativos das misturas definidos pela Expressão (16), obtém-se a seguinte função massa de probabilidade marginal do cliente  $i$ :

$$P(y_i | \mathbf{x}_i, \mathbf{w}_i, \boldsymbol{\varphi}) = \sum_{s=1}^S \pi_{is} \mathbf{w}_i \boldsymbol{\gamma}_s \prod_{j=1}^J P_{js}(y_{ij} | \boldsymbol{\beta}_{js}, \mathbf{x}_{ij}), \quad (18)$$

onde  $\boldsymbol{\varphi} = \boldsymbol{\pi}, \boldsymbol{\beta}$  é o vector de parâmetros do modelo a serem estimados.

Desta forma, Gupta e Chintagunta (1994) retomaram os dados utilizados por Kamakura e Russell (1989), e realizaram uma nova segmentação, onde as probabilidades de pertença do cliente aos segmentos, passaram a ser função das variáveis *Rendimento*, *Tamanho do agregado familiar* e *Idade do representante do agregado*. No caso de ser considerado o modelo descrito na Secção 3.4, mas incorporando agora no modelo as variáveis concomitantes, a probabilidade marginal passa a ser definida pela seguinte expressão:

$$P(y_i | \mathbf{x}_i, \mathbf{w}_i, \boldsymbol{\varphi}) = \sum_{s=1}^S \pi_{is}(\mathbf{w}_i, \boldsymbol{\gamma}_s) \prod_{j=1}^J \left[ \frac{\exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}{1 + \exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})} \right]^{\bar{y}_{ij}} \left[ 1 - \frac{\exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}{1 + \exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})} \right]^{\bar{I} - \bar{y}_{ij}}, \quad (19)$$

onde  $p_{js}(\mathbf{x}_{ij}, \boldsymbol{\beta}_{js}) = \frac{\exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}{1 + \exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}$  é a probabilidade do cliente  $i$  ter adquirido o

produto  $j$  condicionado ao segmento  $s$ , e  $\pi_{is} \mathbf{w}_i \boldsymbol{\gamma}_s = \frac{\exp \mathbf{w}_i^T \boldsymbol{\gamma}_s}{1 + \sum_{r=1}^{S-1} \exp \mathbf{w}_i^T \boldsymbol{\gamma}_r}$  corresponde à

probabilidade *a priori* do cliente  $i$  pertencer ao segmento  $s$ , sendo a respectiva probabilidade de pertença *a posteriori* definida pela seguinte expressão:

$$\alpha_{is} = \frac{\pi_{is}(\mathbf{w}_i; \boldsymbol{\gamma}_s) \prod_{j=1}^J \left[ \frac{\exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}{1 + \exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})} \right]^{\bar{y}_{ij}} \left[ 1 - \frac{\exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})}{1 + \exp(\beta_{0s} + \boldsymbol{\beta}_{js} \mathbf{x}_{ij})} \right]^{\bar{I} - \bar{y}_{ij}}}{\sum_{r=1}^S \pi_{ir}(\mathbf{w}_i; \boldsymbol{\gamma}_r) \prod_{j=1}^J \left[ \frac{\exp(\beta_{0r} + \boldsymbol{\beta}_{jr} \mathbf{x}_{ij})}{1 + \exp(\beta_{0r} + \boldsymbol{\beta}_{jr} \mathbf{x}_{ij})} \right]^{\bar{y}_{ij}} \left[ 1 - \frac{\exp(\beta_{0r} + \boldsymbol{\beta}_{jr} \mathbf{x}_{ij})}{1 + \exp(\beta_{0r} + \boldsymbol{\beta}_{jr} \mathbf{x}_{ij})} \right]^{\bar{I} - \bar{y}_{ij}}}, \quad (20)$$

mais genericamente:

$$\alpha_{is} = \frac{\pi_{is}(\mathbf{w}_i; \boldsymbol{\gamma}_s) \prod_{j=1}^J P_{js}(y_{ij} | \mathbf{x}_{ij}; \boldsymbol{\beta}_{js})}{\sum_{r=1}^S \pi_{ir}(\mathbf{w}_i; \boldsymbol{\gamma}_r) \prod_{j=1}^J P_{jr}(y_{ij} | \mathbf{x}_{ij}; \boldsymbol{\beta}_{jr})}. \quad (21)$$

A perfilagem dos segmentos derivados da aplicação deste modelo é efectuada com base na análise das probabilidades do indivíduo pertencer à categoria  $b$  da variável concomitante  $w_k$  com  $k = 1, \dots, K$  condicionadas ao segmento  $s$ :

$$P(w_k = b | z = s) = \frac{\sum_{w_{ik}=b} P(z = s | \mathbf{y}_i, w_{ik})}{\sum_{i=1}^n P(z = s | \mathbf{y}_i, w_{ik})} = \frac{\sum_{w_{ik}=b} \alpha_{is}}{\sum_{i=1}^n \sum_{s=1}^S \alpha_{is}}. \quad (22)$$

Para cada variável concomitante  $w_k$ , a soma destas probabilidades condicionadas para todas as respectivas categorias é igual a um dentro do segmento  $s$ .

### 3.6 Identificação do modelo

A identificação de um modelo é um tema recorrente nos modelos de mistura finita. Uma função densidade ou massa de probabilidade diz-se identificada se para dois conjuntos distintos de parâmetros, o modelo de mistura finita não resultar na mesma função de distribuição. Assim, um modelo de mistura diz-se identificado se e só se (Dias, 2004):

$$\begin{aligned} \sum_{s=1}^S \pi_s f_s(\mathbf{y}_i; \boldsymbol{\varphi}_s) &= \sum_{r=1}^R \pi_r f_r(\mathbf{y}_i; \boldsymbol{\varphi}_r) \\ \Rightarrow S = R \wedge \forall_{s=1, \dots, S} \exists_{r=1, \dots, R} : \pi_s &= \pi_r \wedge \boldsymbol{\varphi}_s = \boldsymbol{\varphi}_r, \end{aligned} \quad (23)$$

onde  $\boldsymbol{\varphi}_s$  e  $\boldsymbol{\varphi}_r$  correspondem aos parâmetros do modelo de mistura definidos ao nível do segmento  $s$  e  $r$ . Nota-se que a identificação ou não de um modelo não depende do procedimento adoptado para a estimação dos parâmetros.

Dada a complexidade da verificação da identificação do modelo, conforme definida pela Expressão (23), uma condição necessária mas não suficiente, a ser respeitada para a identificação do mesmo pode ser deduzida a partir do total de parâmetros a estimar, que deverá ser maior que o total de combinações que as observações podem tomar menos 1 (Dias, 2004):

$$\prod_{j=1}^J L_j - 1 > d_{\varphi}, \text{ isto é, } df > 0 \quad (24)$$

em que os graus de liberdade são definidos por:

$$df = \prod_{j=1}^J L_j - 1 - d_{\varphi}, \quad (25)$$

onde  $L_j$  representa o número de categorias da variável  $j$  e  $d_{\varphi}$  indica o número de parâmetros do modelo a serem estimados. No caso do modelo de mistura finita com 8 variáveis binárias, e considerando uma solução óptima constituída por 4 segmentos, será necessário proceder à estimação de 35 parâmetros<sup>1</sup>. Este modelo terá 220 graus de liberdade,

$$\text{isto é, } df = \prod_{j=1}^8 2 - 1 - 35 = 2^8 - 1 - 35 = 220.$$

Wedel e Kamakura (1999) aconselham a análise da matriz Hessiana das 2<sup>as</sup> derivadas parciais da função log-verosimilhança e, no caso da matriz ser definida positiva, o modelo diz-se localmente identificado.

---

<sup>1</sup> Para cada segmento  $s=1, 2$  e  $3$ , são estimados 8 parâmetros,  $\beta_{1s}, \dots, \beta_{8s}$ , correspondente aos efeitos dos 8 indicadores binários. No segmento 4, os efeitos são igualados a zero por causa da identificação do modelo. É necessário estimar as 8 constantes,  $\beta_{0j}$  com  $j=1, \dots, 8$ , de cada uma das regressões logit. Pelo facto de considerar-se 4 segmentos, temos 3 parâmetros a estimar correspondentes aos pesos  $\lambda_1, \lambda_2$  e  $\lambda_3$  com  $\lambda_4=0$ . Ficamos com um total de  $(8 \times 3)+8+3 = 35$ .

### 3.7 Algoritmo EM

Verificou-se que o método tradicional de estimação da máxima verosimilhança não consegue solucionar as equações demasiado complexo impostas pelos modelos de mistura. Para estes modelos, a função de log-verosimilhança é definida pela seguinte expressão:

$$\ell(\boldsymbol{\varphi} | \mathbf{x}, \mathbf{y}, \mathbf{w}) = \sum_{i=1}^n \ln P(\mathbf{y}_i | \mathbf{x}_i, \mathbf{w}_i, \boldsymbol{\varphi}). \quad (26)$$

Para resolver este problema de optimização, foi sugerido a aplicação do algoritmo EM que permite maximizar as complexas funções de verosimilhança dos modelos de mistura de regressões, através da expansão do universo inicial considerado a dados não observados.

No âmbito dos modelos de mistura, estes dados não observados correspondem à indicação do segmento de pertença do cliente  $i$  e são introduzidos no modelo pela seguinte função indicatriz (DeSarbo *et al.*, 1995):

$$z_{is} = \begin{cases} 1 & \text{se e só se o cliente } i \text{ pertence ao segmento } s \\ 0 & \text{caso contrário} \end{cases}. \quad (27)$$

O vector  $\mathbf{z}_i$  é definido pelas componentes  $\{z_{i1}, \dots, z_{is}\}^T$  e assume-se que para cada observação  $\mathbf{y}_i$ , o vector  $\mathbf{z}_i$  tem uma distribuição independente e identicamente distribuída dada por uma distribuição multinomial com probabilidades  $\boldsymbol{\pi}$ , isto é:

$$P_i(\mathbf{z}_i; \boldsymbol{\pi}) = \prod_{s=1}^S \pi_s^{z_{is}}, \quad (28)$$

onde  $P_i(\mathbf{z}_i; \boldsymbol{\pi})$  corresponde à função probabilidade *a priori* de observar o cliente  $i$  no segmento  $s$ . Considera-se ainda que os dados observados  $\mathbf{y}_1, \dots, \mathbf{y}_n$  são independentes entre si dado o vector  $\mathbf{z}_1, \dots, \mathbf{z}_n$ . Para o modelo com variáveis concomitantes, a função de verosimilhança conjunta de  $\mathbf{y}_i$  com  $\mathbf{z}_i$  é dada por (DeSarbo *et al.*, 1995):

$$\begin{aligned} P_i(\boldsymbol{\varphi}, \mathbf{y}_i, \mathbf{z}_i | \mathbf{x}_i, \mathbf{w}_i) &= P(\mathbf{z}_i | \mathbf{w}_i, \boldsymbol{\gamma}) P(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\beta}) \\ &= \prod_{s=1}^S \pi_s(\mathbf{w}_i, \boldsymbol{\gamma}_s)^{z_{is}} P_{i/s}(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\beta}_s)^{z_{is}}. \end{aligned} \quad (29)$$

A função de verosimilhança para os dados completos é dada por:

$$P(\boldsymbol{\varphi}; \mathbf{y}, \mathbf{z}, \mathbf{x}, \mathbf{w}) = \prod_{i=1}^n P_i(\boldsymbol{\varphi}; \mathbf{y}_i, \mathbf{z}_i, \mathbf{x}_i, \mathbf{w}_i), \quad (30)$$

e aplicando o logaritmo à função de verosimilhança ficamos com a seguinte expressão:

$$\ln P(\boldsymbol{\varphi}; \mathbf{y}, \mathbf{z}, \mathbf{x}, \mathbf{w}) = \sum_{i=1}^n \sum_{s=1}^S z_{is} \ln P_s(\mathbf{y}_i; \mathbf{x}_i, \boldsymbol{\beta}_s) + z_{is} \ln \pi_s(\mathbf{w}_i, \boldsymbol{\gamma}_s). \quad (31)$$

O valor esperado da função de log-verosimilhança definida na Expressão (31) é então maximizada mediante a aplicação do algoritmo EM que calcula as estimativas do modelo, executando alternadamente as etapas *E* e *M* até que não seja verificada qualquer melhoria em termos de valor da log-verosimilhança, sendo esta melhoria previamente especificada como critério de convergência do algoritmo.

Na etapa *E-Step*, o valor esperado de  $\ln P(\boldsymbol{\varphi}; \mathbf{y}, \mathbf{z}, \mathbf{x}, \mathbf{w})$  é calculado em ordem aos dados não observados  $\mathbf{z}_1, \dots, \mathbf{z}_n$  condicional nas observações  $\mathbf{y}, \mathbf{x}$  e  $\mathbf{w}$ , e um conjunto provisório de estimativas para o vector  $\boldsymbol{\varphi}$ . Observa-se que o valor esperado de  $\ln P(\boldsymbol{\varphi}; \mathbf{y}, \mathbf{z}, \mathbf{x}, \mathbf{w})$  é obtido substituindo o vector  $\mathbf{z}_1, \dots, \mathbf{z}_n$  pelos seus valores esperados  $E(z_{is} | \mathbf{y}_i, \boldsymbol{\varphi})$ , sendo estes idênticos às probabilidades de pertença *a posteriori*  $a_{is}$  definidos na Expressão (21).

Na etapa *M-Step*, o valor esperado  $E \ln l(\boldsymbol{\mu}; \mathbf{y}, \mathbf{z}, \mathbf{x}, \boldsymbol{w})$  é maximizado em ordem ao vector de parâmetros  $\boldsymbol{\varphi} = \boldsymbol{\mu}, \boldsymbol{\beta}$ .

O algoritmo segue assim os seguintes passos (Wedel e Kamakura, 1999):

1. Na primeira iteração  $h=0$ , o procedimento é inicializado com a definição do número de segmentos a serem considerados, e gerando um conjunto inicial de probabilidades *a posteriori*  $\alpha_{is}^{(0)}$ . Este conjunto pode ser obtido de forma aleatória ou de forma determinística utilizando os resultados de uma análise prévia de segmentação, como por exemplo o algoritmo *K-Means*;
2. Dado o conjunto  $\alpha_{is}^{(h)}$ , são obtidas as estimativas de  $\boldsymbol{\varphi} = \boldsymbol{\mu}, \boldsymbol{\beta}$ . A obtenção das mesmas só é possível por métodos numéricos, em geral o algoritmo Newton-Raphson ou uma variante deste;
3. Avaliação da convergência: O algoritmo termina no caso da diferença de valor das funções de log-verosimilhança calculadas em iterações sucessivas ser suficientemente pequena; caso contrário:  $h \leftarrow h+1$ , calcular as probabilidades *a posteriori*  $\alpha_{is}^{(h)}$  usando a Expressão (21), isto é,

$$\alpha_{is} = \frac{\pi_{is}(\boldsymbol{w}_i, \boldsymbol{\gamma}_s) \prod_{j=1}^J P_{js} \boldsymbol{y}_{ij} | \boldsymbol{x}_{ij}; \boldsymbol{\beta}_{js}}{\sum_{r=1}^S \pi_{ir}(\boldsymbol{w}_i, \boldsymbol{\gamma}_r) \prod_{j=1}^J P_{jr} \boldsymbol{y}_{ij} | \boldsymbol{x}_{ij}; \boldsymbol{\beta}_{jr}}$$
 e voltar à etapa 2.

Uma das vantagens do algoritmo EM, apontada por Wedel e Kamakura (1999), é a convergência monótona positiva do referido método para um máximo. Desta forma, o algoritmo EM, no mínimo, converge para um máximo local. Dada a possibilidade da função

de verosimilhança possuir vários máximos locais, a selecção dessas estimativas iniciais é crucial, para garantir a obtenção de um bom candidato a máximo global. Bandeen-Roche *et al.* (1997) sugerem vários conjuntos de valores iniciais de forma a certificar que o máximo global foi efectivamente atingido.

Algumas tentativas de alteração do algoritmo original EM foram ensaiadas com o objectivo de aumentar a velocidade de convergência desse algoritmo. Os algoritmos de EM modificados tentam garantir as características de estabilidade e simplicidade do algoritmo EM (Dias, 2004). Na presente tese será utilizado o algoritmo EM para o cálculo das estimativas para um número definido de iterações, passando posteriormente ao método numérico de Newton-Raphson no processo de estimação dos parâmetros até a convergência do modelo ser atingida.

### 3.8 Determinação do número de segmentos

O processo de estimação de um modelo de mistura de regressões só pode ser efectuado definindo *a priori* o número de segmentos  $S$ . A abordagem tradicional à escolha do número de segmentos é a utilização do teste de razão de verosimilhança. Contudo, para os modelos de mistura finita a sua utilização é problemática uma vez que a hipótese nula a ser testada está definida na fronteira do universo dos parâmetros, o que origina a não verificação da condição de regularidade de Cramér que define uma distribuição assintótica qui-quadrada para a respectiva estatística de teste<sup>1</sup>.

---

<sup>1</sup> Por exemplo, para três segmentos, com os respectivos pesos  $\pi_1$ ,  $\pi_2$  e  $\pi_3$ , pretendemos saber se o modelo com três segmentos é melhor do que o modelo obtido com dois segmentos. Se efectuarmos a respectiva análise, recorrendo ao teste da razão de verosimilhança, é avaliada a função de verosimilhança do modelo restrito com dois segmentos, face à função de verosimilhança do modelo com três segmentos. Temos, assim, o seguinte ensaio de hipóteses:

$$H_0: \pi_1 \neq 0, \pi_2 \neq 0, \pi_3 = 0 \text{ vs. } H_a: \pi_1 \neq 0, \pi_2 \neq 0, \pi_3 \neq 0 \text{ com } \pi_s \in [0, 1].$$

Verifica-se que o valor a ser testado na hipótese nula pertence à fronteira do domínio de  $\pi_s$  e consequentemente a condição de regularidade de Cramér não se verifica, fazendo com que a distribuição assintótica Qui-Quadrado não seja válida para a avaliação dos resultados do teste da razão de verosimilhança.

Desta forma, uma abordagem assente nos critérios de informação tem sido privilegiada na determinação do número de segmentos em modelos de mistura finita (Wedel e Kamakura, 1999). Estas medidas tentam estabelecer um equilíbrio entre o aumento na bondade do ajustamento verificado, face ao aumento do número de parâmetros utilizados. Estes critérios acabam por imputar um “custo” que depende directamente do total de parâmetros do modelo. Assim, quanto maior for o número de parâmetros, maior será a penalização introduzida na referida métrica, fazendo com que o aumento verificado na bondade do ajustamento não seja suficiente para superar o referido custo. Deste modo é seleccionado  $S$  que minimiza o seguinte critério (Dias, 2004):

$$C_S = -2l_S(\hat{\phi}; \mathbf{y}, \mathbf{x}, \mathbf{w}) + d_\phi a, \quad (32)$$

onde  $l_S(\hat{\phi}; \mathbf{y}, \mathbf{x}, \mathbf{w})$  é o logaritmo da função de verosimilhança para um modelo de  $S$  segmentos, calculado no ponto  $\hat{\phi} = \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\beta}}$ ,  $d_\phi$  é o número de parâmetros livres a serem estimados, e a constante  $a$  corresponde ao factor de penalização. Na Tabela 1 podem ser observados os vários critérios de informação utilizados na determinação do número de segmentos<sup>1</sup>.

**Tabela 1 – Critérios de Informação**

Critério de Informação	$a$
Critério de Informação Akaike (AIC) (Akaike, 1974)	2
Critério de Informação Bayesiano (BIC) (Schwarz, 1978)	$\ln n$
Critério de Informação Consistente Akaike (CAIC) (Bozdogan, 1987)	$\ln(n+1)$
Critério de Informação Akaike Modificado (AIC3) (Bozdogan, 1993)	3

Nota:  $n$  representa o total de indivíduos considerados no processo de estimação do modelo.

<sup>1</sup> Para uma lista mais vasta de possibilidades, vide e.g Brochado e Martins (2003).

O número de segmentos a ser considerado no modelo é uma decisão crucial no processo de segmentação. Com efeito, este número tem implicações na determinação das estruturas relacionais existentes no mercado e, pode influenciar as decisões a serem tomadas no desenvolvimento das estratégias de marketing.

Andrews e Currim (2003) argumentam que o critério de informação AIC3 definido por uma constante de penalização igual a 3 em vez do tradicional valor 2, apresenta ser um bom critério para a determinação do número de segmentos. No entanto, para uma amostra de grande dimensão, os critérios BIC e CAIC revelaram-se consistentes, sendo considerados também bons indicadores para a escolha do número de segmentos a reter.

De seguida, apresenta-se uma aplicação dos modelos apresentados, onde os coeficientes serão estimados pelo algoritmo EM com a variante Newton-Raphson. A decisão do número de segmentos a serem considerados para cada um dos modelos será efectuada com base na análise dos critérios de informação.

## 4 Uma aplicação à Banca

### 4.1 Introdução

O problema de marketing que apresentamos no presente estudo é o de compreender a heterogeneidade existente nos clientes de uma instituição financeira relativamente às suas necessidades financeiras e os eventuais impactos derivados das acções de marketing no portfolio dos produtos detidos pelos clientes.

O desenvolvimento de uma estratégia de CRM na sua vertente analítica pode ajudar na identificação das estruturas de comportamento e impactos das acções de venda dirigida, guiando o gestor de marketing no delineamento de abordagens de marketing mais adequadas aos clientes. Se não for conhecido o comportamento do cliente relativamente aos produtos detidos na sequência das acções realizadas, não será possível elaborar acções de *cross-selling* para os clientes a quem faça sentido os contactar com a respectiva oferta do produto.

Assim, a definição de acções de *cross-selling* apenas são viáveis se forem identificados os clientes para os quais faz sentido definir uma abordagem mais dirigida. Este estudo passará pela elaboração de um processo de segmentação conduzido em 3 etapas:

- Etapa 1: Os clientes vão ser segmentados com base nos produtos detidos. Esta segmentação irá permitir avaliar a existência de comportamentos heterogéneos no que diz respeito ao portfolio dos produtos detidos. Após a obtenção dos segmentos será efectuado o estudo do perfil com base em algumas variáveis demográficas. Iremos tentar perceber se de facto para cada um dos segmentos obtidos, os clientes são de natureza diferente.

- Etapa 2: Os clientes vão ser novamente segmentados com base nos indicadores de posse, mas os segmentos obtidos, serão em simultâneo, perfilados pelas variáveis

demográficas utilizadas na etapa 1. Os perfis dos segmentos obtidos na etapa 1 e na etapa 2 serão comparados de forma a identificar as diferenças entre as duas soluções.

- Etapa 3: Os clientes vão ser segmentados com base nos indicadores de posse, mas, em cada segmento será efectuada uma regressão destes indicadores com base em variáveis de marketing. Tal como descrito na etapa 2, os segmentos obtidos, serão perfilados pelas variáveis demográficas. Desta forma será possível avaliar o impacto das acções de marketing nas estruturas de produtos detidos pelos clientes.

## **4.2 Os dados**

Os dados foram cordialmente disponibilizados por uma instituição financeira, que por razões de confidencialidade, será designada por Instituição Financeira Fénix. Os dados foram trabalhados de forma a obter a informação agregada por cliente necessária à aplicação dos modelos anteriormente referidos. Para além de terem sido construídos os indicadores de posse para cada um dos produtos para os quais a Instituição forneceu a respectiva informação, foram também aplicados um conjunto de filtros gerais de exclusão de forma a serem contemplados apenas clientes que possam ser alvo de acção comercial. Foram, assim, excluídos clientes com idades inferiores a 18 anos e superiores a 65 anos. Todos os clientes com um comportamento não desejado do ponto de vista financeiro também não foram considerados. Desta forma, foram excluídos clientes com negativos injustificados, crédito vencido e uso fraudulento de cartões. Além disso, foi assegurado que os clientes considerados são novos clientes, no sentido que não foram clientes da Instituição Fénix no passado, tendo aberto a sua primeira conta à ordem em Janeiro de 2007. Apenas foram considerados clientes particulares, tendo sido excluídos, clientes empresas e clientes ENI's (empresa em nome individual). Desta

forma, no final foram considerados 5.558 clientes para o presente estudo. Para cada cliente sabe-se a idade, o sexo, a zona Markttest, os indicadores de posse para os produtos contemplados, e as acções de marketing realizadas de Janeiro a Outubro 2007.

A escolha dos clientes com uma relação bancária recente foi propositada. Por um lado, o objectivo é compreender as necessidades financeiras dos clientes no início da sua relação com a Instituição Financeira. Por outro lado, as estratégias de marketing existentes na Instituição são elaboradas para clientes com uma antiguidade mais elevada, e desta forma é necessário perceber se a referida estratégia é adequada para os clientes recentes. Poderia ter-se optado por clientes com uma data de abertura em Outubro de 2007, mas é frequente que as acções comerciais nos últimos meses sejam referentes aos produtos de Investimento que permitem aceder aos benefícios fiscais. Desta forma, para não correr o risco de capturar este efeito, optou-se por seleccionar o mês de Janeiro.

#### **4.2.1 Indicadores de posse**

Os indicadores de posse dizem respeito aos produtos adquiridos pelos clientes até Outubro de 2007. Estes indicam quais os produtos detidos pelo cliente, e tomam o valor 0 no caso do cliente não deter o produto e 1 no caso contrário.

Os produtos considerados no presente estudo são:

- *Cartão de Débito*: É um dos principais tipos de movimentação imediata a débito na conta à ordem, podendo ser utilizado para efectuar levantamentos diários nas caixas automáticas, e para efectuar compras nos estabelecimentos munidos de terminais de pagamento automático.

- *Cartão de Crédito*: Reúne todas as características verificadas num cartão de débito, e permite um meio de poder gerir o saldo da conta à ordem no dia-a-dia, pelo facto de dispor de um *plafond* de crédito mensal, com possibilidade de liquidar o montante devido de forma faseada ou pontual à data previamente acordada entre o cliente e a instituição. O pagamento faseado implica a aplicação de uma taxa sobre o montante não liquidado, e que permanece em dívida para o mês seguinte.

- *Serviço Financeiro*: Permite aceder a um conjunto de vantagens na adesão do serviço, mediante o pagamento de uma anuidade. Estas vantagens podem estar relacionadas com anuidades gratuitas nos cartões de débito e cartões de crédito, bonificações no *spread* do crédito ao consumo e crédito habitação, taxas mais rentáveis no crédito automóvel ou aconselhamento financeiro personalizado.

- *Depósitos a Prazo*: É uma aplicação segura, com capital e taxa de juro garantidos, onde se escolhe o prazo e o montante a ser aplicado.

- *Produto de Investimento*: É um produto de investimento com vários níveis de riscos, onde o cliente escolhe o mais adaptado ao seu perfil. Estão incluídos nesta categoria os fundos, os produtos estruturados, os certificado, as acções, e as obrigações.

- *Seguro de Capitalização*: Reúne os Planos de Poupança Reforma (PPR) e os *Unit Linked* (UL). Um PPR é plano de poupança vocacionado para fomentar a poupança para a reforma, e é comercializado na forma de um seguro de vida. Um UL é um produto de poupança de médio/longo prazo sob a forma de um seguro de vida, cujo

patrimônio é aplicado em fundos de investimento. Não há garantia de capital, nem qualquer taxa de rendimento mínima, não limitando por isso a rentabilidade do produto.

- *Crédito ao Consumo*: É um financiamento concedido pela instituição que pode atingir 100% do valor do bem ou serviço a financiar. O financiamento possui um limite mínimo e máximo, e o cliente pode optar por um prazo de pagamento calculado em número de meses sendo os limites mínimo e o máximo de pagamento definidos pela instituição.

- *Crédito Imobiliário*: Reúne o crédito à habitação, o crédito para obras, e o crédito sinal. O crédito à habitação destina-se à aquisição de habitação própria. O crédito para obras destina-se a financiar obras de beneficiação ou recuperação de habitação própria. O crédito sinal destina-se ao financiamento do sinal para aquisição de uma habitação.

**Tabela 2 – Taxas de Posse de Produto**

Produtos	Taxa de Posse
Cartão de Débito	89.51%
Cartão de Crédito	25.50%
Serviço Financeiro	16.43%
Depósito a prazo	8.70%
Crédito Imobiliário	8.60%
Crédito ao consumo	5.43%
Produto de Investimento	3.18%
Seguro de Capitalização	2.43%

Como se pode observar na Tabela 2, para a amostra considerada, o *Cartão de Débito* destaca-se dos restantes produtos, pois 89.51% dos clientes na amostra detêm este produto. Seguem

depois os produtos *Cartão de Crédito* e *Serviço Financeiro* com uma taxa de posse de 25.50% e 16.43%, respectivamente. O produto *Seguro de Capitalização* possui a menor taxa de posse alcançando os 2.43%.

#### 4.2.2 Variáveis demográficas

As variáveis demográficas utilizadas no presente trabalho são a *Idade*, o *Sexo* do cliente, e a sua *Zona Markttest*<sup>1</sup>. Verifica-se que 50% dos clientes da amostra têm uma idade inferior ou igual a 28 anos. A média situa-se nos 32 anos, indicando uma assimetria positiva. Apenas foram considerados clientes com idade compreendida entre os 18 e os 65 anos.

**Tabela 3 - Distribuição da variável *Sexo***

Variável <i>Sexo</i>	% de clientes
Feminino	43.72
Masculino	56.28

Para a amostra considerada, podemos ver na Tabela 3 existem mais clientes do sexo masculino (56.28%) do que clientes do sexo feminino (43.72%).

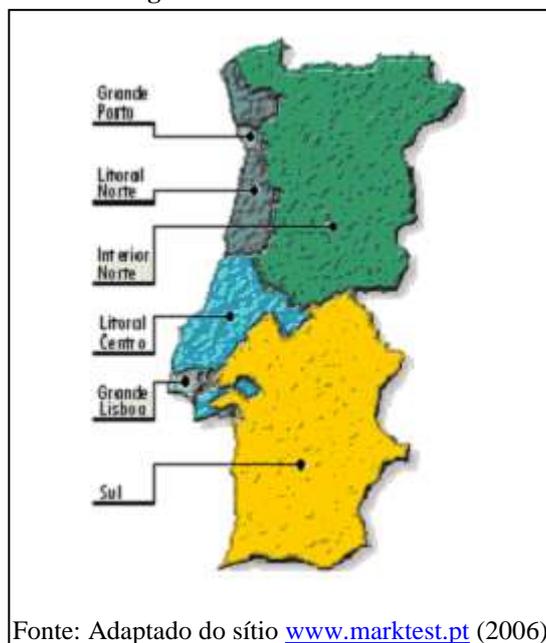
A zona ou região *Markttest* foi definida com base no distrito e no concelho do balcão com o qual o cliente tem um maior envolvimento. Assim, mesmo que o cliente tenha contas à ordem abertas em mais do que um balcão, apenas lhe será afecto uma *Zona Markttest*. Com efeito, no

---

<sup>1</sup> Para mais informação sobre as zonas Markttest consultar o sítio da Markttest [http://www.markttest.pt/Notas\\_Tecnicas/default.asp?strUrl=/notas\\_tecnicas/info/conteudos/documentos/regioes.asp](http://www.markttest.pt/Notas_Tecnicas/default.asp?strUrl=/notas_tecnicas/info/conteudos/documentos/regioes.asp), através do qual é disponibilizada uma base de dados com a classificação dos distritos e concelhos nas respectivas *Zonas Markttest*. A informação das zonas foi extraída do sítio em 13 de Setembro de 2006.

caso do cliente possuir mais do que uma conta em balcões diferentes, a zona Marktest do cliente corresponde à *Zona Marktest* do balcão com maior saldo contabilístico.

**Figura 2 – Zonas Marktest**



A Figura 2 apresenta a repartição do território português continental nas 6 zonas Marktest.

A amostra é constituída por 22.40% de clientes da *Grande Lisboa*, 19.54% do *Litoral Norte* e 18.37% do *Interior Norte* (Tabela 4).

**Tabela 4 – Distribuição dos clientes pelas Zonas Marktest**

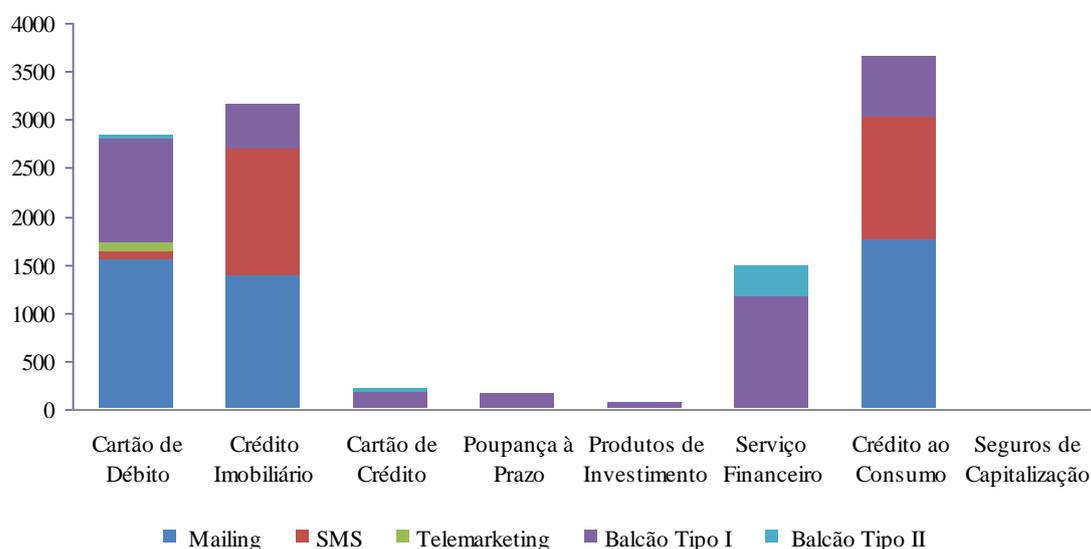
<u>Zona Marktest</u>	<u>% de clientes</u>
Grande Lisboa	22.40
Grande Porto	10.43
Interior Norte	18.37
Litoral Centro	14.82
Litoral Norte	19.54
Sul	9.52
Ilhas	4.91

A Zona Marktest com menor número de clientes é a zona *Ilhas* reunindo 4.91% dos clientes da amostra (Tabela 4).

### 4.2.3 Variáveis de marketing

As variáveis de marketing utilizadas no presente estudo correspondem à indicação da realização ou não, de uma campanha de aquisição para um dado produto financeiro. Foi recolhida a informação do canal utilizado para contactar o cliente para acções de aquisição. Os canais contemplados são o *mailing*, o SMS<sup>1</sup>, o *telemarketing*, o *balcão* tipo I, e o *balcão* tipo II. A diferença entre os canais balcão tipo I e balcão tipo II reside apenas na selecção dos clientes condicionando o tipo de abordagem a ser efectuado pelo agente comercial.

Figura 3 – Número de contactos efectuados por produto e canal de contacto



Observa-se da Figura 3, que para os produtos *de Poupança a Prazo*, *Produto de Investimento* e *Cartão de Crédito*, o número de contactos efectuados é muito inferior ao número de contactos efectuados para os restantes produtos. Para os clientes da amostra considerada, o

<sup>1</sup> Mensagem enviada directamente para o telemóvel do cliente.

*Mailing* é fortemente utilizado nos produtos de *Crédito Imobiliário*, *Crédito ao Consumo* e *Cartão de Débito*. De uma forma resumida de Janeiro a Outubro 2007, foram contactados para a aquisição do *Cartão de Débito* 1546 clientes por *Mailing*, 68 por SMS, 98 por *Telemarketing*, 1097 por Balcão tipo I e 29 por Balcão tipo II. Para o *Crédito Imobiliário*, foram contactados para acções de aquisição, 1382 clientes por *Mailing*, 1304 clientes por SMS e 473 clientes por Balcão tipo I. Para o *Cartão de Crédito*, foram contactados 174 por Balcão tipo I e 22 por Balcão tipo II. Para os produtos de *Depósito a Prazo*, foram contactados 141 clientes por *Mailing*. Para os *Produtos de Investimento*, foram apenas contactados 49 clientes. Para o *Serviço Financeiro*, foram contactados 1153 clientes por Balcão tipo I e 330 clientes por Balcão tipo II. O produto *Seguros de Capitalização* não foi alvo de qualquer acção de venda dirigida.

### **4.3 Segmentação baseada nos indicadores de posse de produtos**

#### **4.3.1 Modelo de mistura finita**

Como já referido anteriormente vamos, numa primeira fase, segmentar os clientes utilizando os indicadores de posse para as variáveis base de segmentação. Pretendemos com esta primeira etapa avaliar a existência de grupos de clientes com diferentes tipologias de portfolios de produtos detidos. Esta primeira segmentação é necessária para que seja avaliada a homogeneidade dos clientes no que concerne aos produtos que foram sendo adquiridos desde Janeiro a Outubro de 2007. Esta análise é primordial, pois só após a identificação de comportamentos diferenciados é que fará sentido pensar em delinear abordagens de marketing diferenciadas de forma a antecipar as necessidades dos clientes. Esta primeira iteração da segmentação, irá permitir perceber quais foram as necessidades dos clientes durante os primeiros meses de relacionamento com a Instituição Fénix, e perceber as estruturas

relacionais existentes entre os produtos para melhor compreender as medidas de *cross-selling* a serem tomadas. A identificação de vários segmentos permitirá aceder a uma visão mais centrada no cliente. Com efeito, cada segmento irá reunir clientes semelhantes relativamente aos produtos adquiridos durante os 10 primeiros meses de relacionamento com a instituição financeira. Esta compreensão das necessidades financeiras possibilita uma actuação antecipada da instituição financeira em termos de novos produtos a criar, ou acções comerciais a realizar, de forma a satisfazer necessidades futuras dos clientes, aumentando assim o seu respectivo valor (Ebner *et al.*, 2002).

Com base na amostra composta por 5.558 clientes, irá realizar-se uma segmentação sobre os 8 indicadores de posse, recorrendo ao modelo de misturas finitas descrito na Secção 3.2. Para a estimação dos modelos foi utilizado o Latent Gold 4.0. Foram estimados 6 modelos, variando o número de segmentos de 1 até 6, e a solução óptima apontada pelos critérios de informação BIC, AIC e AIC3 é o modelo constituído por 4 segmentos.

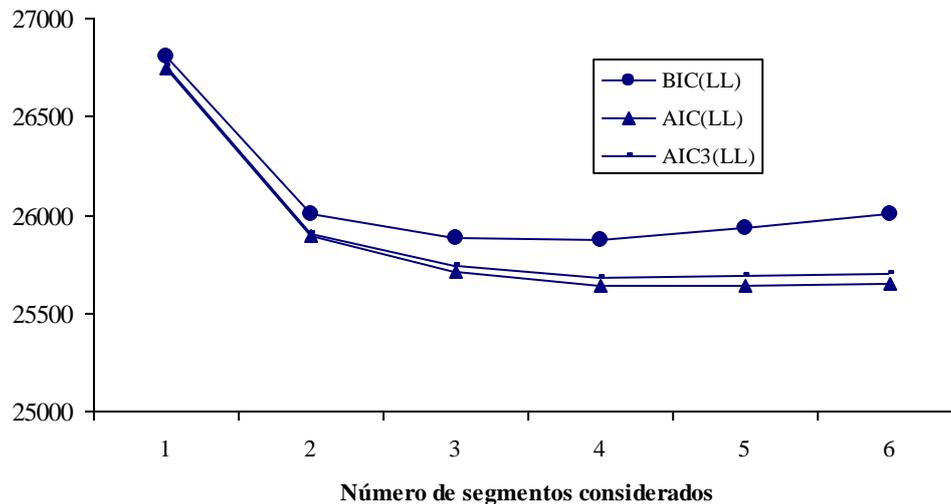
Repetiu-se várias vezes o processo de estimação utilizando diferentes valores iniciais para a estimação dos parâmetros do modelo, e todos os resultados apontam sempre para uma solução óptima com 4 segmentos. Também verificou-se que as estimativas obtidas são iguais para as várias soluções óptimas. O modelo foi estimado com recurso ao método de estimação EM para um máximo de 250 iterações, passando a estimar os parâmetros pelo método de Newton-Raphson com um limite máximo de 50 iterações. O modelo converge quando a soma relativa das diferenças dos valores das estimativas obtidos em iterações consecutivas for menor que

$$10^{-8}, \text{ ou seja } \sum_{p=1}^P \left| \frac{\hat{\phi}_p^{k} - \phi_p^{k-1}}{\hat{\phi}_p^{k-1}} \right| < 10^{-8}, \text{ sendo } P \text{ o número total de parâmetros a serem}$$

estimados.

Na Figura 4 observa-se que os critérios de informação penalizam os modelos constituídos por 5 ou mais segmentos. Desta forma, o modelo composto por 4 segmentos é considerado a solução óptima.

Figura 4 – Critérios de informação



O modelo assim estimado é composto por 4 classes latentes e possui a seguinte função massa de probabilidade:

$$P(y_i; \beta, \lambda) = \sum_{s=1}^4 \pi_s \prod_{j=1}^8 p_{js}^{y_{ij}} (1 - p_{js})^{1-y_{ij}}$$

onde

$$y_{is} = \begin{cases} 1 & \text{se e só se o cliente } i \text{ detem o produto } j \\ 0 & \text{caso contrário} \end{cases},$$

com j=1 para o *Cartão de Débito*, j=2 para o *Crédito Imobiliário*, j=3 para *Cartão de Crédito*, j=4 para *Depósito a Prazo*, j=5 para *Produto de Investimento*, j=6 para *Seguro de Capitalização*, j=7 para *Serviço Financeiro* e j=8 para *Crédito ao Consumo*. Tem-se ainda que  $p_{js}$  corresponde à probabilidade de aquisição do produto  $j$  condicionado ao segmento de pertença  $s$ , sendo esta probabilidade obtida mediante a estimação das constantes  $\beta_{0j}$  e

efeitos  $\beta_{js}$ , com  $j = 1, \dots, 8$  e  $s = 1, 2$  e  $3$  com  $\beta_{j4} = 0$  pelo facto do segmento 4 ser o segmento de referência no processo de estimação. Os valores  $\exp(\beta_{js})$  indicam quantas vezes mais é provável verificar a ausência do produto  $j$  no segmento  $s$  face à não ausência do mesmo no segmento 4.

As probabilidades de pertença  $\pi_s$  são obtidas mediante a estimação das constantes  $\lambda_s$  com  $s = 1, 2$  e  $3$  e  $\lambda_4 = 0$  para o garante da identificação do modelo.

Com base na Expressão (25), verifica-se que o modelo possui 220 graus de liberdade  $df$ , sendo assim respeitada a condição necessária para que o modelo seja identificado (Dias, 2004):

$$df = \prod_{j=1}^J L_j - 1 - d_{\beta, \lambda} = 2^8 - 1 - 35 = 220.$$

#### 4.3.2 Estimação dos parâmetros do modelo

Pelo algoritmo EM obtêm-se as estimativas das 8 constantes e dos 8 efeitos para cada um dos 8 indicadores binários para cada segmento, à excepção do último onde os efeitos são igualados a zero. Também serão estimadas as constantes  $\lambda_1$ ,  $\lambda_2$  e  $\lambda_3$  para a obtenção das estimativas dos pesos relativos dos segmentos, isto é,  $\hat{\pi}_1$ ,  $\hat{\pi}_2$ ,  $\hat{\pi}_3$  e  $\hat{\pi}_4$ . Ficamos com um total de  $8 + 8 \times 3 + 3 = 35$  parâmetros.

A avaliação das estimativas obtidas foi efectuada utilizando a estatística de Wald que permite testar a restrição do(s) parâmetro(s) ter(em) um valor nulo. No caso da estatística de teste não ser “estaticamente significativa”, podemos optar por não efectuar a estimação do parâmetro e

assim conseguir aumentar os graus de liberdade do modelo. A possibilidade de colocar determinados parâmetros com o valor zero pode ser extremamente benéfico quando o modelo implica a estimação de um número elevado de parâmetros, colocando em risco a condição necessária de identificação do modelo (ver Expressão (24)).

**Tabela 5 – Estimativas das constantes do modelo sem concomitantes**

Estimativa do vector de constantes $\beta_0$		Wald	Valor-p
Cartão de Débito -	$\hat{\beta}_{01}$	40.57	< 0.001
Crédito Imobiliário -	$\hat{\beta}_{02}$	0.02	0.9
Cartão de Crédito -	$\hat{\beta}_{03}$	2.70	0.1
Depósito a Prazo -	$\hat{\beta}_{04}$	42.02	< 0.001
Produto de Investimento -	$\hat{\beta}_{05}$	44.35	< 0.001
Seguro de Capitalização -	$\hat{\beta}_{06}$	40.84	< 0.001
Serviço Financeiro -	$\hat{\beta}_{07}$	8.97	0.0028
Crédito ao Consumo -	$\hat{\beta}_{08}$	34.42	< 0.001

Para um nível de significância igual a 0.05, as constantes de intersecção dos indicadores de *Crédito Imobiliário* e *Cartão de Crédito* não são estatisticamente significativas (Tabela 5).

**Tabela 6 – Estimativas dos efeitos do modelo sem concomitantes**

Estimativa do vector de efeitos $\beta_j$	Solução com 4 Segmentos				Wald (valor-p)	
	1	2	3	4		
Cartão de Débito -	$\hat{\beta}_{1s}$	-0.3565	1.956	-1.0011	0	< 0.001
Crédito Imobiliário -	$\hat{\beta}_{2s}$	2.6001	3.2176	2.5135	0	< 0.001
Cartão de Crédito -	$\hat{\beta}_{3s}$	3.2369	4.1492	1.6163	0	< 0.001
Depósito a Prazo -	$\hat{\beta}_{4s}$	2.8528	-1.2547	1.0398	0	< 0.001
Produto de Investimento -	$\hat{\beta}_{5s}$	2.3241	-1.6783	1.5993	0	< 0.001
Seguro de Capitalização -	$\hat{\beta}_{6s}$	2.6542	0.7075	0.9177	0	0.001
Serviço Financeiro -	$\hat{\beta}_{7s}$	2.1105	7.5895	-2.3186	0	< 0.001
Crédito ao Consumo -	$\hat{\beta}_{8s}$	1.3543	1.5756	-1.5069	0	< 0.001

Para um nível de significância igual a 0.05, todos os parâmetros coeficientes são estatisticamente significativos, isto é, diferentes de zero (Tabela 6).

**Tabela 7 – Constantes de regressão dos pesos relativos do modelo sem concomitantes**

Estimativa do vector de constantes $\lambda$	Solução com 4 segmentos				Wald Valor-p
	1	2	3	4	
	2.16	1.25	0.86	0	< 0.001

Para um nível de significância de 0.05, o conjunto dos parâmetros é estatisticamente diferente de zero (Tabela 7).

A interpretação das probabilidades condicionadas ao segmento dos indicadores de posse só pode ser correctamente efectuada se for respeitado o pressuposto da independência condicional referida na Secção 3.2. A avaliação da violação do pressuposto é efectuada com base na análise dos resíduos bivariados apresentados na Tabela 8. Assumindo que a amostra inicial é o resultado da mistura de quatro subamostras, espera-se que o modelo de mistura retenha as relações estruturais existentes, verificando-se dentro de cada segmento, a independência entre os indicadores de posse. No caso de existir independência entre os Indicadores de Posse em cada um dos segmentos estimados, os resíduos bivariados seguem uma distribuição *Qui-quadrado* com um grau de liberdade. Assim para valores dos resíduos bivariados superiores a 3.84, estes são estatisticamente significativos, e é rejeitada a hipótese da independência local. No caso de se verificar, para um par de indicadores de posse, um resíduo bivariado superior a 3.84, significa que existe um efeito conjunto entre os respectivos indicadores que não foi considerado, pelo que este efeito deve ser incluído no modelo.

Todos os resíduos bivariados são inferiores a 3.84, possibilitando a interpretação correcta das probabilidades de posse condicionadas aos segmentos. Passamos agora a descrever os segmentos obtidos, percebendo assim as relações estruturais no que concerne aos produtos adquiridos pelos clientes para a amostra considerada.

**Tabela 8 – Resíduos Bivariados do modelo sem concomitantes**

Indicadores de Posse	Cartão de Débito	Crédito Imobiliário	Cartão de Crédito	Depósito a Prazo	Produto de Investimento	Seguro de Capitalização	Serviço Financeiro	Crédito ao Consumo
Cartão de Débito	.							
Crédito Imobiliário	0.60	.						
Cartão de Crédito	0.05	0.01	.					
Depósito a Prazo	0.06	0.56	0.07	.				
Produto de Investimento	0.13	0.35	0.007	1.52	.			
Seguro de Capitalização	3.76	0.06	0.59	0.11	2.6	.		
Serviço Financeiro	0.19	0.08	0.01	0.13	0.05	0.08	.	
Crédito ao Consumo	1.00	0.65	0.68	2.00	0.002	3.31	0.001	.

### 4.3.3 Relações estruturais existentes nos segmentos

Como se pode observar na Tabela 9, o modelo composto por 4 segmentos, é constituído por um primeiro segmento com um peso muito significativo pois reúne mais de metade dos clientes da amostra ( $\hat{\pi}_1 = 0.56$ ), tendo o último segmento uma dimensão bastante reduzida com um peso relativo  $\hat{\pi}_4 = 0.06$

**Tabela 9 – Dimensão dos segmentos do modelo sem concomitantes**

Segmento	1	2	3	4
$\hat{\pi}_s$	0.56	0.23	0.15	0.06

As probabilidades condicionadas permitem perceber a estrutura de cada segmento no que diz respeito aos produtos detidos. Para a Expressão (5), com base na amostra Fénix obtiveram-se as estimativas apresentadas na Tabela 10.

Estas probabilidades condicionadas, quando comparadas com o modelo agregado, isto é, o modelo que apenas considera um segmento, evidencia os comportamentos diferenciados na detenção dos produtos.

**Tabela 10 – Probabilidades condicionadas de posse do modelo sem concomitantes**

Estimativas das probabilidades condicionadas $\hat{p}_{js}$	Solução com 4 segmentos				Agregado	
	1	2	3	4		
<i>Cartão de Débito</i>	$\hat{p}_{1s}$	0.96	0.68	0.98	0.94	0.90
<i>Crédito Imobiliário</i>	$\hat{p}_{2s}$	0.06	0.04	0.07	0.48	0.09
<i>Cartão de Crédito</i>	$\hat{p}_{3s}$	0.18	0.08	0.53	0.85	0.25
<i>Depósito a Prazo</i>	$\hat{p}_{4s}$	0.01	0.31	0.04	0.11	0.09
<i>Produto de Investimento</i>	$\hat{p}_{5s}$	0.003	0.12	0.01	0.03	0.03
<i>Seguro de Capitalização</i>	$\hat{p}_{6s}$	0.007	0.04	0.04	0.08	0.02
<i>Serviço Financeiro</i>	$\hat{p}_{7s}$	0.045	0.0002	0.80	0.28	0.16
<i>Crédito ao Consumo</i>	$\hat{p}_{8s}$	0.02	0.01	0.24	0.07	0.05

No segmento 1 (56%) a grande maioria dos clientes possuem o *Cartão de Débito* (0.96), sendo este o produto menos complexo dos produtos em análise. Os restantes produtos possuem probabilidades condicionadas baixas, sendo o *Cartão de Crédito* o segundo produto com maior probabilidade condicionada de posse com 0.18. Este produto reúne todas as características verificadas num *Cartão de Débito*, mas permite ainda um meio de poder gerir o saldo da conta à ordem no dia-a-dia, pelo facto de dispor de um crédito mensal, com possibilidade de liquidar o montante devido de forma faseada ou pontual. Este segmento reúne clientes com necessidades muito simples, que é a de possuir um cartão da Instituição Fénix para efectuar levantamentos e pagamentos. Os respectivos clientes ainda não evidenciam nenhum comportamento devedor ou aforrador, não tendo ainda assim nenhuma responsabilidade ou necessidade financeira. Este facto leva-nos a crer que estes clientes ainda estão numa fase muito inicial das suas vidas activas ou ainda, depender de outros de forma

parcial e de um ponto de vista financeiro. Como tal, este segmento será designado por *Recém-Chegados*.

No segmento 2 (23%) é onde se verifica a probabilidade condicionada de posse do *Cartão de Débito* mais baixa (0.68) face aos segmentos 1, 3 e 4 onde a mesma é no mínimo de 0.94. No entanto, é neste segmento que se verificam as probabilidades condicionadas de posse mais altas do *Depósito a Prazo* e *Produto de Investimento* com 0.31 e 0.12 respectivamente face aos restantes segmentos. Este segmento reúne assim clientes *Aforradores*.

No segmento 3 com uma dimensão de 15%, podem ser observadas as probabilidades condicionadas de posse do *Cartão de Débito*, *Cartão de Crédito*, *Serviço Financeiro e Crédito ao Consumo* mais elevadas dos 4 segmentos com probabilidades de 0.98, 0.57, 0.80 e 0.24, respectivamente. Repara-se que comparativamente ao modelo agregado, as probabilidades condicionadas de posse do *Serviço Financeiro e Crédito ao Consumo* são 5 vezes maiores do que as respectivas homólogas. Este segmento reúne clientes com um elevado índice de *cross-selling*, e portanto com um nível de fidelização significativo com a instituição Fénix. Para além disso, pode observa-se um comportamento devedor de médio prazo, podendo este dever-se a uma situação estável na sua vida activa, permitindo aos mesmos recorrer ao crédito para a aquisição de bens que não seria possível com o património actual detido. Este segmento parece ser caracterizado por clientes que estão numa fase mais avançada no seu ciclo de vida, e estão mais comprometidos com a instituição Fénix quando comparado com o segmento 1. Este segmento já indica alguma fidelização do cliente ao banco, pois compromete o cliente a possuir um saldo à ordem suficiente para cobrir as despesas contraídas com o respectivo cartão ou as prestações do *Crédito ao Consumo*. Este segmento reúne assim os clientes *Fidelizados q.b.*

No segmento 4, com a dimensão mais pequena de 6%, é provavelmente o segmento mais rendível. Com efeito, é neste segmento que se verificam as probabilidades condicionadas mais elevadas de posse do *Crédito Imobiliário* (0.48) e *Cartão de Crédito* (0.85). A probabilidade condicionada ao segmento do *Crédito Imobiliário* é 5 vezes maior do que a do modelo agregado. Pelo facto de registar a probabilidade mais alta de posse do produto *Crédito Imobiliário*, este segmento reúne os clientes *Fiéis* à Instituição Fénix, devendo esta ser o respectivo banco principal. Com efeito, a posse do respectivo produto implica uma relação de médio e longo prazo com a instituição. Este segmento reúne os clientes *Fiéis*.

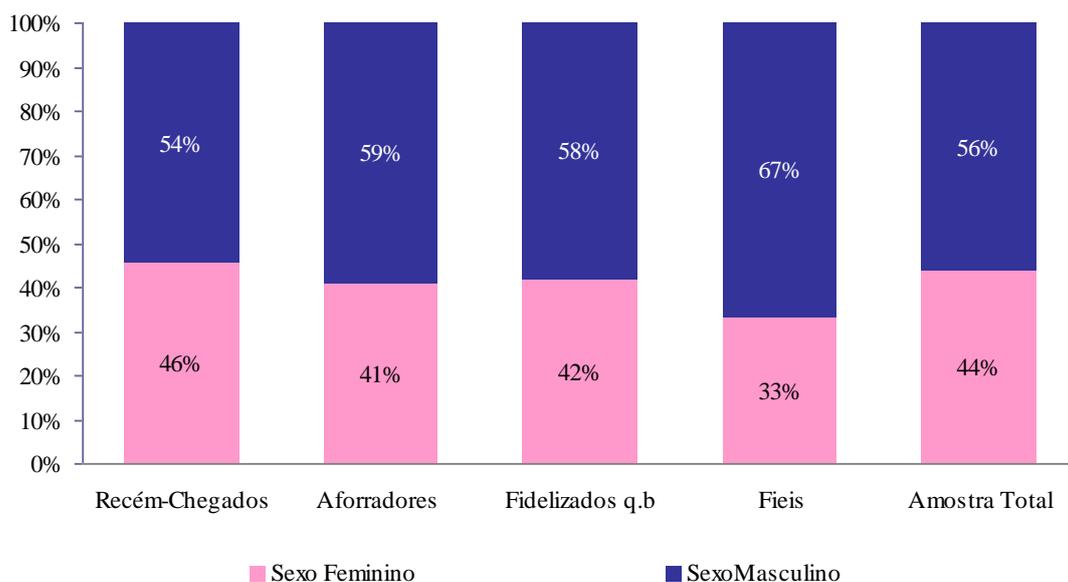
Este primeiro modelo permite-nos perceber como a heterogeneidade relativamente à posse dos produtos se distribui de forma homogénea na solução obtida com 4 segmentos. Agora que as relações dos comportamentos de posse encontram-se mais clarificadas, é preciso perceber quem são os clientes que compõem os segmentos obtidos. Uma análise cruzada entre os segmentos de clientes e variáveis de caracterização do cliente irão permitir alcançar este objectivo.

#### **4.3.4 Perfis dos segmentos**

Com base nas probabilidades *a posteriori* obtidas com a Expressão (6), os clientes são classificados em cada um dos 4 segmentos. Por exemplo, o cliente  $i$  será classificado no segmento 4 se  $\hat{\alpha}_{i4} > \hat{\alpha}_{is}$ ,  $s=1, 2$  e  $3$ .

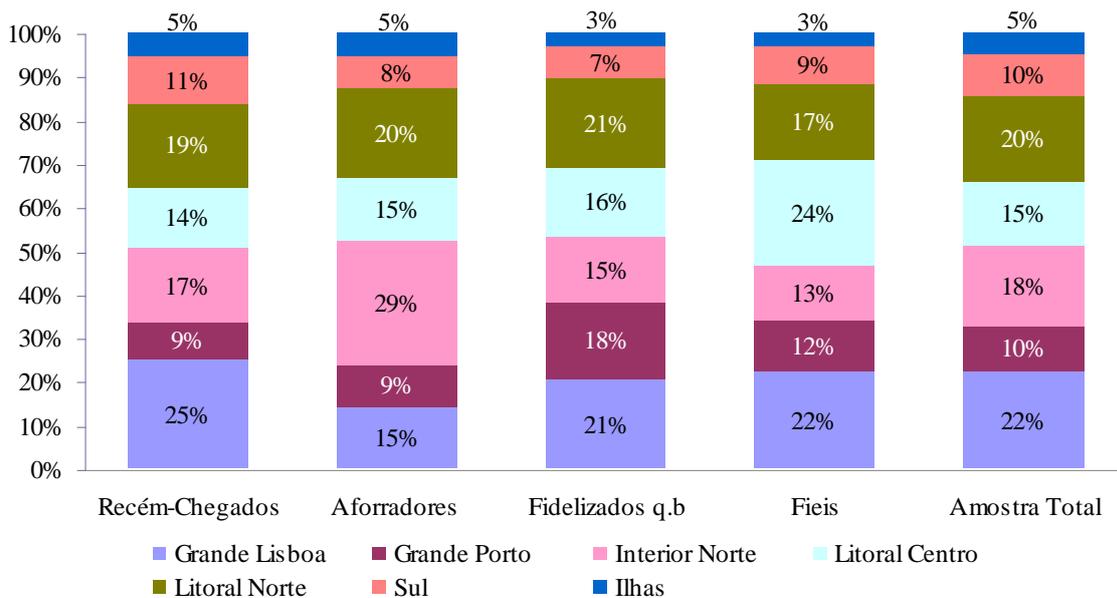
Ao efectuar uma análise das características dos clientes que constituem cada um dos segmentos analisados no ponto anterior, estaremos em melhores condições para perceber as motivações que os levaram a tais comportamentos no que concerne aos produtos detidos.

**Figura 5 – Distribuição da variável Sexo nos segmentos**



Verifica-se que nos segmentos dos *Recém-Chegados*, *Aforradores* e *Fidelizados q.b.*, a variável *Sexo* tem uma distribuição semelhante à verificada na amostra total. No segmento dos *Fiéis*, o segmento 4, pode observar-se a predominância dos clientes do sexo masculino com 67% face aos clientes do sexo feminino com 33% (Figura 5).

**Figura 6 – Distribuição da variável Zona Marktest nos segmentos**



No segmento dos *Recém-Chegados*, a distribuição dos clientes pelas Zonas Markttest é em muito semelhante ao verificado na amostra total, destacando-se no entanto a grande presença da *Grande Lisboa*, reunindo 25% dos clientes (Figura 6).

No segmento dos *Aforradores*, a zona do Interior Norte e Litoral Norte têm uma forte presença, com uma percentagem de 29% e 20%, respectivamente. Verifica-se uma menor percentagem de clientes da Zona da *Grande Lisboa* (15%) comparada com os restantes segmentos (Figura 6).

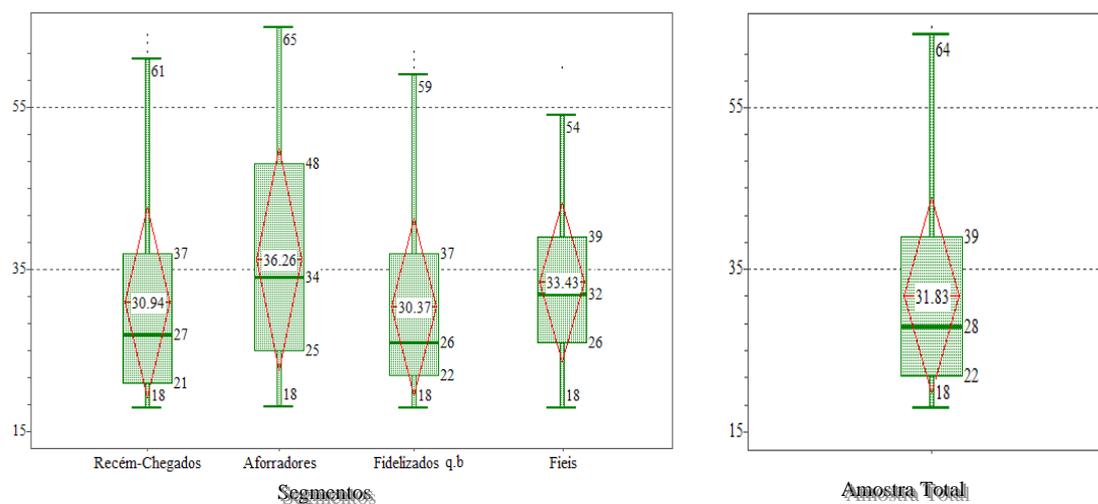
No segmento dos *Fidelizados q.b.*, o *Grande Porto* (18%) tem uma forte presença quando comparada com a da amostra total (10%), e verifica-se que as zonas mais predominantes são a *Grande Lisboa* e o *Litoral Norte* reunindo cada uma 21% dos clientes da amostra considerada (Figura 6).

No segmento dos *Fiéis*, a *Grande Lisboa* e o *Litoral Centro* são as zonas mais predominantes com 22% e 24% dos clientes. É neste segmento que se verifica a maior presença de clientes no *Litoral Centro* quando comparado com os restantes segmentos (Figura 6).

Os segmentos *Recém-Chegados* e *Fidelizados q.b.* possuem uma distribuição da idade semelhantes: 50% dos clientes têm idade menor ou igual a 27 anos, com 75% a não ultrapassar os 37 anos. São os segmentos mais jovens com uma média aproximada de 31 anos (Figura 7).

O segmento dos *Aforradores* é o segmento com clientes de idades mais elevadas, com uma média de 36 anos, pois 50% da população possuem uma idade superior ou igual a 34 anos.

**Figura 7 – Distribuição da variável *Idade* nos segmentos**



O segmento dos *Fiéis* possui uma mediana próxima da média com 32 e 33 anos respectivamente e 50% dos clientes possuem idades entre os 26 e os 39 anos. É um segmento onde se verifica uma média de idade superior aos segmentos dos *Recém-Chegados* e *Fidelizados q.b.* e inferior à do segmento dos *Aforradores*.

Este conjunto de informação permite perceber que a oferta de produtos/serviços pode ser adaptado para cada um dos segmentos, visto que os mesmos reúnem clientes com diferentes necessidades financeiras. Por outro lado, a simples análise descritiva efectuada aos segmentos obtidos, mostra que as estratégias de marketing devem ser adaptadas às características demográficas dos clientes em cada um dos segmentos. Estas duas resoluções - adequar a oferta e adaptar a estratégia de marketing - deve potenciar o aumento do *cross-selling*.

No entanto, esta segmentação é meramente explicativa, não fornecendo uma forma de classificar novos clientes nos segmentos definidos. A classificação dos novos clientes é primordial para que os mesmos possam ser identificados e conseqüentemente contactados (Gupta e Chintagunta, 1994). Efectuar a classificação dos novos clientes mediante uma

análise descritiva pode levar a resultados errados. Na caracterização dos segmentos efectuada na presente Secção, verificamos que a distribuição da variável *Idade* é muito semelhante nos segmentos dos clientes *Recém-Chegados* e *Fidelizados q.b.* Como tal, a diferenciação dos clientes destes dois segmentos não pode ser baseada apenas na variável idade. É necessário conjugar esta última com as variáveis *Sexo* e *Zona Marktest*, para distingui-los. Este processo é pouco claro e muito pouco fiável.

Desta forma, vamos estimar o modelo em que as variáveis *Sexo*, *Zona Marktest* e *Idade* serão incluídas na estimação das probabilidades de pertença.

#### **4.4 Segmentação baseada nos indicadores de posse e utilização das variáveis demográficas**

##### **4.4.1 Modelo de mistura finita com variáveis concomitantes**

Por um lado, já foram identificados os diferentes tipos de portfolios de produtos detidos pelos clientes originando 4 segmentos, sendo estes os clientes *Recém-Chegados*, *Aforradores*, *Fidelizados q.b.* e *Fiéis*; por outro lado, na análise efectuada à perfilagem dos segmentos, verificou-se que a distribuição das variáveis *Sexo*, *Zona Marktest* e *Idade* não é igual para todos os segmentos. Por esta razão torna-se importante obter um método de classificação que, com base nestas últimas variáveis, permita efectuar a afectação do cliente no segmento correcto. Esta possibilidade de classificar um novo cliente com base nas suas características demográficas, permite identificar rapidamente o seu segmento de pertença e em conformidade, actuar sobre o mesmo com um marketing dirigido e ajustado às suas necessidades (Gupta e Chintagunta, 1994). A mais valia deste modelo é, sem dúvida, a circunscrição dos segmentos num espaço bem identificado mediante a utilização das variáveis

concomitantes. Com efeito, este modelo possibilita identificar o segmento mais provável de pertença do cliente através das respectivas probabilidades de pertença. Desta forma, possibilita direccionar as acções apenas para os clientes do segmento para o qual se destina a mensagem de marketing.

O modelo estimado neste capítulo é muito semelhante ao estimado no capítulo anterior, com o acréscimo do facto das probabilidades de pertença *a priori*  $\pi_s$  serem agora função das variáveis concomitantes *Idade*, *Sexo*, e *Zona Marktest*, isto é,  $\pi_{is} = \lambda_i \gamma_s$  com  $w$  o vector das variáveis concomitantes e  $\gamma$  o vector de parâmetros a estimar. Com efeito, já não é estimado o vector de constantes  $\lambda = \lambda_1, \dots, \lambda_{S-1}$ , e passa a ser estimado o vector  $\gamma = \gamma_{01}, \dots, \gamma_{0, S-1}, \gamma_{11}, \dots, \gamma_{1, S-1}, \dots, \gamma_{K1}, \dots, \gamma_{K, S-1}$  definido previamente na Expressão (16), sendo  $S$  o número de segmentos e  $K$  o número total de parâmetros a serem estimados em cada segmento, correspondentes aos efeitos das variáveis concomitantes em cada segmento. Para a variável concomitante *Idade* é estimado um parâmetro  $\gamma_{1s}$ . Para a variável *Sexo*, por ser uma variável binária, apenas é estimado um parâmetro ( $\gamma_{2s}$ ) para cada segmento, sendo a categoria *masculino* a categoria de referência. Pelo facto da variável *Zona Marktest*, ser uma variável nominal com 7 categorias, o modelo irá construir 6 variáveis *dummy*, procedendo assim à estimação, em cada segmento, de 6 parâmetros  $\gamma_{3s}, \dots, \gamma_{8s}$ , ficando a sétima categoria, a zona *Ilhas*, a categoria de referência. O vector  $\gamma_s$  indica, assim, o impacto provocado pelo efeito das variáveis concomitantes, entre pertencer ao segmento  $s$ , e pertencer ao segmento  $S$ . O último segmento  $S$  é novamente considerado o segmento base de comparação.

#### 4.4.2 Estimação dos parâmetros do modelo com variáveis concomitantes

Procedeu-se à estimação do modelo, variando o número de segmentos de 1 a 6, e a solução óptima apontada pelo critério BIC foi uma solução com 4 segmentos. Repetiu-se várias vezes o processo de estimação utilizando diferentes soluções iniciais na estimação iterativa dos parâmetros do modelo, e a solução com 4 segmentos foi sempre identificada como sendo a solução óptima.

No que diz respeito à estimação do vector de constantes  $\beta_0$ , para um nível de significância igual a 0.05, apenas a constante de intersecção do indicador de *Cartão de Débito* não é estatisticamente significativa (ver apêndice A para mais detalhe).

Quando considerado um nível de significância de 0.05 todos os parâmetros do vector dos efeitos dos indicadores de posse  $\beta_j$ , são estatisticamente significativos, isto é, diferentes de zero (ver apêndice A para mais detalhe). No entanto, para um nível de significância de 0.01, as estimativas dos efeitos para o indicador de *Seguro de Capitalização* não são estatisticamente diferentes de zero.

Da análise da Tabela 11, verifica-se que o indicador de posse de *Seguro de Capitalização* possui resíduos bivariados elevados com o *Produto de Investimento*, *Cartão de Débito* e *Cartão de Crédito*.

**Tabela 11 – Resíduos Bivariados do modelo com concomitantes**

Indicadores de Posse	Cartão de Débito	Crédito Imobiliário	Cartão de Crédito	Depósito a Prazo	Produto de Investimento	Seguro de Capitalização	Serviço Financeiro	Crédito ao Consumo
Cartão de Débito	.							
Crédito Imobiliário	<b>5.35</b>	.						
Cartão de Crédito	0.14	<b>28.14</b>	.					
Depósito a Prazo	0.09	1.74	0.79	.				
Produto de Investimento	1.36	0.02	0.02	0.16	.			
Seguro de Capitalização	<b>4.75</b>	3.50	<b>5.08</b>	3.29	<b>4.19</b>	.		
Serviço Financeiro	0.23	1.87	0.25	1.67	0.04	0.45	.	
Crédito ao Consumo	0.62	<b>3.85</b>	0.54	2.33	0.23	2.85	0.75	.

Relembra-se que este indicador, no modelo de mistura finita sem concomitantes, não tinha sido relevante para a descrição das relações de posse dos produtos nos segmentos obtidos. Por outro lado, este produto possui a taxa mais baixa de posse na amostra considerada (2.43%). Assim, decidiu-se re-estimar o modelo não considerando o indicador de posse do produto de *Seguro de Capitalização* no conjunto das variáveis base de segmentação.

#### **4.4.3 Estimação dos parâmetros do modelo não considerando o indicador de posse Seguro de Capitalização**

Um novo modelo foi estimado tendo para variáveis base de segmentação os indicadores de posse de *Cartão de Débito*, *Crédito Imobiliário*, *Cartão de Crédito*, *Poupança a Prazo*, *Produto de Investimento*, *Serviço Financeiro*, e *Crédito ao Consumo*. As variáveis concomitantes são as mesmas que as referenciadas na Secção anterior, isto é, as variáveis

*Idade, Sexo e Zona Marktest*. Para este novo modelo, a solução óptima apontada pelo critério de informação BIC foi a composta por 5 segmentos (ver Apêndice B para mais detalhe).

Assim, para uma solução com 5 segmentos foram agora estimados 35 parâmetros do vector  $\beta$  para a obtenção das probabilidades condicionadas  $\hat{p}_{js}$ , e foram estimados 36 parâmetros do vector  $\gamma$  para a obtenção das probabilidades de pertença  $\hat{\pi}_j, \gamma_j$ . Relativamente ao primeiro grupo de 35 parâmetros, é necessário estimar:

- 7 constantes, uma por cada indicador, sendo estas  $\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{05}, \beta_{06}$  e  $\beta_{07}$ , respectivamente;
- 28 efeitos correspondentes ao efeito de cada indicador para cada um dos segmentos obtidos, à excepção do 5º segmento onde os parâmetros são igualados a zero pelo facto deste último ser o segmento de referência. Os parâmetros  $\beta_{15}, \beta_{25}, \beta_{35}, \beta_{45}, \beta_{55}, \beta_{65}$  e  $\beta_{75}$  são então igualados a zero.

Relativamente ao segundo conjunto de 36 parâmetros, é necessário estimar:

- 4 constantes,  $\gamma_{01}, \gamma_{02}, \gamma_{03}$  e  $\gamma_{04}$ ;
- 32 efeitos correspondentes ao efeito de cada categoria das variáveis concomitantes para as variáveis discretas ou ao efeito da variável no seu todo para as variáveis contínuas:
  - Para a variável contínua *Idade* são estimados 4 parâmetros,  $\gamma_{11}, \gamma_{12}, \gamma_{13}$  e  $\gamma_{14}$  respeitante aos segmentos 1, 2, 3 e 4, respectivamente.
  - Para a variável binária *Sexo*, a categoria *Masculino* é a categoria de referência e como tal apenas são estimados 4 parâmetros para a categoria *Feminino*,  $\gamma_{21}, \gamma_{22}, \gamma_{23}$  e  $\gamma_{24}$ , respeitante aos segmentos 1, 2, 3 e 4, respectivamente;
  - Para a variável nominal *Zona Marktest*, a categoria *Ilhas* é a categoria de referência e desta forma, são estimados os parâmetros para as restantes 6 zonas

para os segmentos 1, 2, 3 e 4, perfazendo um total de 24 parâmetros,  $\gamma_{ks}$  com  $k=1, 2, 3, 4, 5$  e  $s=1, 2, 3$  e  $4$ .

- O 5º segmento é o segmento de referência, ficando os efeitos das respectivas variáveis concomitantes igualados a zero, isto é,  $\gamma_{15}, \gamma_{25}, \gamma_{35}, \gamma_{45}, \gamma_{55}, \gamma_{65}, \gamma_{75}$  e  $\gamma_{85}$  são nulos.

Desta forma, o modelo ficou com 5487 graus de liberdade e 71 parâmetros a serem estimados. Verifica-se, assim, a condição necessária para identificação do modelo. Os parâmetros dos vectores  $\beta_0$  e  $\beta_j$  revelaram-se estatisticamente significativos com excepção da constante e respectivo efeito do indicador de posse *Crédito Imobiliário* com um valor-p da estatística de Wald de 0.81 e 0.12 respectivamente (ver apêndice B para mais detalhe).

Para a avaliação da significância dos parâmetros  $\gamma$  foi novamente utilizada a estatística de Wald.

**Tabela 12 – Constantes de regressão dos pesos relativos do modelo com concomitantes**

Estimativa do vector de constantes $\gamma_0$	Solução com 5 segmentos					Wald (valor-p)
	1	2	3	4	5	
	1.44	11.6	-1.52	0.60	0	<0.001

Para um nível de significância de 0.05, o conjunto dos parâmetros do vector de constantes é estatisticamente diferente de zero (Tabela 12).

Da observação da Tabela 13, verificamos que para um nível de significância de 0.05, todos os parâmetros do vector  $\gamma$  são estatisticamente diferentes de zero. Todas as condições encontram-se reunidas para passarmos à descrição dos segmentos obtidos, e efectuar a perfilagem dos mesmos mediante as probabilidades condicionadas ao segmento.

**Tabela 13 – Efeitos de regressão dos pesos relativos do modelo com concomitantes**

Estimativa dos efeitos $\gamma_{ks}$		Solução com 5 segmentos					Wald (valor-p)
<i>Idade</i>	$\gamma_{1s}$	0.003	-0.445	0.056	-0.020	0	< 0.001
<i>Sexo Feminino</i>	$\gamma_{2s}$	1.2474	0.243	0.644	0.695	0	< 0.001
<i>Grande Lisboa</i>	$\gamma_{3s}$	-0.794	0.240	-1.585	-0.272	0	
<i>Grande Porto</i>	$\gamma_{4s}$	-1.650	-0.681	-0.996	0.337	0	
<i>Interior Norte</i>	$\gamma_{5s}$	-1.455	-0.057	0.166	-0.309	0	
<i>Litoral Centro</i>	$\gamma_{6s}$	-1.423	-0.754	-0.954	-0.257	0	< 0.001
<i>Litoral Norte</i>	$\gamma_{7s}$	-0.623	-0.535	-0.304	0.240	0	
<i>Sul</i>	$\gamma_{8s}$	-1.019	1.447	-1.467	-0.709	0	

#### 4.4.4 Relações estruturais existentes nos segmentos

A distribuição dos clientes pelos segmentos não está tão concentrada no primeiro segmento ( $\hat{\pi}_1=0.35$ ) como verificado no modelo anterior ( $\hat{\pi}_1=0.56$ ). Todos os segmentos possuem pesos substanciais, o que permite efectuar, caso seja necessário, análises específicas para cada segmento (Tabela 14). Lembra-se que no modelo anterior, o segmento com menor dimensão detinha um peso relativo de apenas  $\hat{\pi}_4=0.06$ .

**Tabela 14 – Dimensão dos segmentos do modelo com concomitantes**

Segmento	1	2	3	4	5
$\hat{\pi}_s$	0.35	0.23	0.15	0.14	0.13

Com base nas probabilidades condicionadas, passamos agora à descrição das relações estruturais existentes, com base nas suas respectivas estimativas disponibilizadas na Tabela 15.

**Tabela 15 – Probabilidades condicionadas de posse do modelo com concomitantes**

Estimativas das probabilidades condicionadas $\hat{p}_{js}$		Solução com 5 segmentos					Agregado
		1	2	3	4	5	
<i>Cartão de Débito</i>	$\hat{p}_{1s}$	0.99	0.94	0.51	0.98	0.90	0.90
<i>Crédito Imobiliário -</i>	$\hat{p}_{2s}$	0.0003	0	0.0002	0.11	0.56	0.09
<i>Cartão de Crédito -</i>	$\hat{p}_{3s}$	0.20	0.17	0.08	0.60	0.40	0.25
<i>Depósito a prazo -</i>	$\hat{p}_{4s}$	0.05	0.06	0.28	0.04	0.07	0.09
<i>Produto de Investimento</i>	$\hat{p}_{5s}$	0.01	0.01	0.15	0.002	0.02	0.03
<i>Serviço Financeiro -</i>	$\hat{p}_{6s}$	0.03	0.11	0.006	0.86	0.09	0.16
<i>Crédito ao Consumo -</i>	$\hat{p}_{7s}$	0.01	0.05	0.01	0.24	0.03	0.05

No segmento 1 ( $\hat{\pi}_1=0.35$ ) verifica-se que apenas 1% dos clientes não possuem o *Cartão de Débito* ( $\hat{p}_{11}=0.99$ ). A probabilidade condicionada de posse do *Cartão de Crédito* é significativa com  $\hat{p}_{31}=0.20$ . As restantes probabilidades condicionadas são extremamente baixas, não ultrapassando o valor de 0.05. Este segmento reúne clientes com um comportamento de posse de produtos muito semelhante com o segmento *Recém-Chegados* na solução de 4 segmentos obtida no modelo descrito na Secção 4.3.3.

No segmento 2, correspondente a 25% dos clientes ( $\hat{\pi}_2=0.25$ ), verifica-se que com probabilidade  $\hat{p}_{12}=0.94$ , os clientes possuem o *Cartão de Débito*. Verifica-se também um peso significativo do *Cartão de Crédito* e do *Serviço Financeiro* com  $\hat{p}_{32}=0.17$  e  $\hat{p}_{62}=0.11$ , respectivamente. Este segmento possui um comportamento muito semelhante ao anterior pelo facto de reunir clientes com produtos financeiros simples. No entanto, o peso da probabilidade condicionada do indicador de posse do produto *Serviço Financeiro* pode evidenciar um segmento de clientes dispostos a construir uma relação duradoura com a Instituição Fénix. Com efeito, a aquisição desse produto permite aceder a um conjunto de benefícios para além da anuidade gratuita para determinados cartões. Este permite, em determinadas condições, um *spread* mais reduzido nos produtos de *Crédito ao Consumo* e de *Crédito Imobiliário*.

Com base nesta análise, os segmentos 1 e 2 parecem ser semelhantes nas suas estruturas de produtos adquiridos. Não é clara neste momento a razão pela qual eles não foram agregados num só segmento. A estimação de um modelo de mistura finita com variáveis concomitantes poderá ajudar-nos a compreender quais os factores que levaram à necessidade de os diferenciar criando por isso dois segmentos. Desta forma, estes dois segmentos irão manter o mesmo nome, os *Recém-Chegados*.

No segmento 3 ( $\hat{\pi}_3=0.15$ ) a probabilidade condicionada de posse de *Cartão de Débito* é a mais baixa com  $\hat{p}_{13} = 0.51$  quando comparada com os restantes segmentos, onde a mesma é no mínimo de 0.90. Por outro lado, é neste segmento que se verificam as probabilidades condicionadas de posse mais elevadas para os produtos *Depósito a Prazo* e *Produto de Investimento* com  $\hat{p}_{43}=0.28$  e  $\hat{p}_{53}=0.15$  comparativamente aos restantes segmentos. Este segmento é muito semelhante ao segmento dos *Aforradores* identificado na Secção 4.3.3. Desta forma, este segmento será também baptizado de segmento *Aforrador*.

No segmento 4 ( $\hat{\pi}_4=0.14$ ), verifica-se a 2ª probabilidade condicionada de posse do *Cartão de Débito* mais elevada com  $\hat{p}_{14}=0.98$ . As probabilidades condicionadas de posse dos produtos *Cartão de Crédito*, *Serviço Financeiro* e *Crédito ao Consumo* são as mais elevadas dos 5 segmentos com probabilidades de 0.60, 0.86, e 0.24 respectivamente. Comparativamente com o segmento 3 do modelo apresentado na Secção 4.3.3, as estruturas entre os dois segmentos são muito semelhantes. Como já referido anteriormente, este segmento já indica alguma fidelização do cliente ao banco, pois compromete o cliente a possuir um saldo à ordem suficiente para cobrir as despesas contraídas com o respectivo cartão ou as prestações do

*Crédito ao Consumo*. Neste sentido, este segmento também vai ser designado de *Fidelizados q.b.*

No segmento 5 ( $\hat{\pi}_3=0.13$ ), encontramos o *filet mignon* dos clientes mais fiéis. Este segmento apresenta a probabilidade condicionada mais elevada de posse do *Crédito Imobiliário* com  $\hat{p}_{25}=0.56$ . Comparativamente ao segmento dos *Fiéis* (ver Secção 4.3.3), este já não possui a probabilidade condicionada mais elevada de posse de *Cartão de Crédito*, mas mesmo assim encontra-se em 2º lugar com  $\hat{p}_{35}=0.40$ . Pelo facto deste segmento reunir os clientes que na sua grande maioria são detentores de um *Crédito Imobiliário*, vai ser designado de segmento *Fiel*.

Este modelo possui dois segmentos com estruturas semelhantes às obtidas no modelo de mistura finita simples. São estes os *Aforradores* e os *Fidelizados q.b.* O segmento dos *Recém-Chegados* do modelo de mistura finita simples (ver Secção 4.3.3) parece ter sido dividido em dois sub-segmentos, os segmentos 1 e 2. O segmento de menor dimensão é em ambos os modelos - mistura finita simples e mistura finita com concomitantes - o que reúne os clientes com predominância em deter o produto *Crédito Imobiliário*. A análise das probabilidades condicionadas ao segmento poderá elucidar-nos sobre o papel das variáveis concomitantes na formação dos segmentos. Em particular, pretende compreender-se em que medida os segmentos 1 e 2 - os *Recém-Chegados* - são diferentes, percebendo assim a razão pela qual os mesmos foram subdivididos em dois sub-segmentos.

#### **4.4.5 Perfis dos segmentos**

A grande vantagem do modelo de mistura com variáveis concomitantes é o facto de os segmentos serem perfilados em simultâneo com o processo de estimação das probabilidades condicionadas aos segmentos. Com a perfilagem efectuada desta forma, permite, por um lado

perceber quais as características dos clientes pertencentes a cada segmento, e por outro, a classificação de novos clientes nos seus respectivos segmentos usando apenas a informação de caracterização demográfica e geodemográfica (Gupta e Chintagunta, 1994).

**Tabela 16 – Probabilidades condicionadas ao segmento**

Estimativas das probabilidades condicionada ao segmento			Solução com 5 segmentos					Agregado
			1	2	3	4	5	
<i>Idade</i>	dos 18 aos 21 anos	$\hat{p}_{\psi_1} = 1; z = s$	0.09	0.68	0.04	0.15	0.05	0.22
	dos 22 aos 25 anos	$\hat{p}_{\psi_1} = 2; z = s$	0.15	0.247	0.08	0.21	0.12	0.17
	dos 26 aos 32 anos	$\hat{p}_{\psi_1} = 3; z = s$	0.26	0.07	0.17	0.25	0.32	0.21
	dos 33 aos 43 anos	$\hat{p}_{\psi_1} = 4; z = s$	0.27	0.003	0.25	0.23	0.32	0.20
	dos 44 aos 65 anos	$\hat{p}_{\psi_1} = 5; z = s$	0.23	0.000	0.46	0.16	0.19	0.20
	Média		35	21	42	32	35	32
<i>Sexo</i>	Feminino	$\hat{p}_{\psi_2} = 0; z = s$	0.56	0.41	0.37	0.42	0.26	0.44
	Masculino	$\hat{p}_{\psi_2} = 1; z = s$	0.44	0.59	0.63	0.58	0.74	0.56
<i>Zona</i>	Grande Lisboa	$\hat{p}_{\psi_3} = 1; z = s$	0.28	0.25	0.09	0.19	0.21	0.22
	Grande Porto	$\hat{p}_{\psi_3} = 2; z = s$	0.07	0.10	0.08	0.20	0.12	0.10
	Interior Norte	$\hat{p}_{\psi_3} = 3; z = s$	0.11	0.21	0.37	0.14	0.18	0.18
	Litoral Centro	$\hat{p}_{\psi_3} = 4; z = s$	0.14	0.12	0.15	0.17	0.20	0.15
	Litoral Norte	$\hat{p}_{\psi_3} = 5; z = s$	0.22	0.17	0.19	0.22	0.15	0.20
	Sul	$\hat{p}_{\psi_3} = 6; z = s$	0.11	0.11	0.07	0.05	0.11	0.10
	Ilhas	$\hat{p}_{\psi_3} = 7; z = s$	0.07	0.04	0.05	0.03	0.03	0.05

No segmento 1, comparativamente aos resultados obtidos com o modelo de mistura finita simples (ver Secção 4.3.4), verificam-se grandes diferenças nas variáveis *Sexo* e *Idade*. Neste segmento, os pesos das categorias invertem-se face ao seu homólogo (ver Secção 4.3.4), onde se observa a probabilidade condicionada mais alta para a categoria do sexo *Feminino* com  $\hat{p}_{\psi_2} = 0; z = 1 = 0.56$  (Tabela 16). No que diz respeito à variável *Idade*, a probabilidade de pertença é mais elevada para idades superiores a 26 anos. Este facto contraria o comportamento do seu homólogo, visto este último ser um dos segmentos mais jovens com 50% dos indivíduos concentrados nas idades compreendidas entre os 18 e os 27 anos. Para este segmento a média de idade é igual a 35 anos, não sendo assim o segmento mais jovem,

ao contrário do observado no modelo anterior. Para o actual segmento 1, a designação *Recém-Chegados* não se coaduna com as probabilidades condicionadas obtidas para o seu homólogo relativamente à variável *Idade*. Este segmento tem uma média de Idade de 35 anos, não podendo reunir clientes numa fase inicial do seu ciclo de vida financeira. Estes clientes devem possuir uma relação financeira com outra instituição onde terão certamente produtos de médio e longo prazo, mas com a necessidade de possuir outra conta numa instituição diferente da primeira de forma a deter alguma independência. Possivelmente estes clientes possuem a sua primeira relação bancária noutra instituição, sendo a Instituição Fénix o seu segundo banco. Por outro lado, pelo facto da probabilidade condicionada ao segmento 1 ser mais alta para a categoria do sexo *Feminino* pode corroborar esta hipótese. Com efeito, as mulheres conseguem cada vez mais a sua autonomia financeira, exigindo uma certa liberdade e independência na forma como pretendem gerir o seu capital. Assim, este segmento passa a ser designado por *Independentes*.

O segmento 2 é, sem dúvida, o mais jovem, pois possui uma probabilidade condicionada ao segmento de 0.68 para as idades compreendidas entre os 18 e os 21 anos. A probabilidade condicionada na categoria do sexo *Masculino* é superior,  $\hat{p}_{\text{♣}_2} = 1; z = 2_{\text{--}} = 0.59$ , à da categoria do Sexo *Feminino* com  $\hat{p}_{\text{♣}_2} = 0; z = 2_{\text{--}} = 0.41$  (Tabela 16). Verificam-se as probabilidades mais elevadas nas zonas da *Grande Lisboa* e do *Interior Norte* com  $\hat{p}_{\text{♣}_3} = 1; z = 2_{\text{--}} = 0.25$  e  $\hat{p}_{\text{♣}_3} = 3; z = 2_{\text{--}} = 0.21$ , respectivamente. Mesmo assim, a zona do *Litoral Norte* possui uma probabilidade condicionada de pertença significativa de 0.17. Para este segmento, já recuperámos as características identificadas no segmento 1 na Secção 4.3.4. Este segmento continuará a ser designado por *Recém-Chegados*.

O segmento 3, os *Aforradores*, possui um comportamento semelhante ao seu homólogo do modelo de mistura finita sem concomitantes. Com efeito, este segmento permanece o mais “velho” com uma probabilidade condicionada ao segmento máxima para as idades superiores aos 44 anos igual a  $\hat{p}_{\psi_1=5; z=3}^- = 0.46$ . Verifica-se uma probabilidade condicionada para a categoria Sexo *Masculino* superior ao da categoria do Sexo *Feminino* com 0.63 e 0.37, respectivamente. Continua a ser a zona do *Interior Norte* a predominar com uma probabilidade condicionada ao segmento 3 de  $\hat{p}_{\psi_3=3; z=3}^- = 0.37$ . A zona *Litoral Norte* permanece a segunda mais influente com a uma probabilidade de  $\hat{p}_{\psi_3=5; z=3}^- = 0.19$  (Tabela 16).

O segmento 4, os *Fidelizados q.b.*, também reúne clientes com características semelhantes ao seu homólogo. No que diz respeito à *Idade*, as probabilidades tomam valores médios mais altos para as idades compreendidas entre os 26 e os 32 anos e entre os 33 e 43 anos com  $\hat{p}_{\psi_1=3; z=4}^- = 0.25$  e  $\hat{p}_{\psi_1=4; z=4}^- = 0.23$ , respectivamente. Este segmento é o segundo mais jovem a seguir ao segmento 2. Verifica-se uma probabilidade condicionada de pertença superior nos clientes do Sexo *Masculino* com  $\hat{p}_{\psi_2=1; z=4}^- = 0.58$  (Tabela 16). As zonas *Marktest* com maiores probabilidades condicionadas continuam a ser o *Litoral Norte* e a *Grande Lisboa*, com  $\hat{p}_{\psi_3=5; z=4}^- = 0.22$  e  $\hat{p}_{\psi_3=1; z=4}^- = 0.19$ , respectivamente.

O segmento 5, os *Fiéis*, possui características semelhantes ao seu homólogo. É neste segmento que também se verifica a probabilidade condicionada mais alta para a categoria do sexo *Masculino* com  $\hat{p}_{\psi_2=1; z=5}^- = 0.74$ . No que diz respeito à *Idade*, as probabilidades tomam valores médios mais altos para as idades compreendidas entre os 26 e os 32 anos e entre os 33 e 43 anos com  $\hat{p}_{\psi_1=3; z=5}^- = 0.32$  e  $\hat{p}_{\psi_1=4; z=5}^- = 0.32$ , respectivamente. Permanece o 2º segmento mais velho a seguir ao segmento *Aforradores*. As zonas da *Grande*

*Lisboa* e do *Litoral Centro* são as zonas com probabilidade condicionada mais elevada com  $\hat{p}_{\kappa_3} = 1; z = 5_{-} = 0.21$  e  $\hat{p}_{\kappa_3} = 4; z = 5_{-} = 0.20$ .

Verifica-se, assim, que a aplicação do modelo de mistura com variáveis concomitantes permitiu identificar dois sub-segmentos dentro do segmento *Recém-Chegados* identificado no modelo de mistura sem concomitantes. Estes dois sub-segmentos, o segmento 1 e o segmento 2 apesar de possuírem comportamentos semelhantes no que diz respeito aos produtos detidos, as variáveis concomitantes mostram que os mesmos reúnem clientes com características muito diferentes. Tal facto pode ter implicações ao nível do marketing, podendo ser necessário uma estratégia diferenciada consoante o segmento.

Em termos de perfis, os quatro segmentos, os *Recém-Chegados*, *Aforradores*, *Fidelizados q.b.* e *Fiéis*, são semelhantes aos seus homólogos identificados no modelo de mistura finita sem concomitantes (Tabela 17).

**Tabela 17 – Resumo dos perfis dos segmentos obtidos**

Variáveis Concomitantes	Segmentos				
	Independentes	Recém-Chegados	Aforradores	Fidelizados q.b.	Fiéis
Idade	26 aos 43 anos	18 aos 21 anos	44 aos 65 anos	26 aos 43 anos	26 aos 43 anos
Sexo	Predominância das Mulheres	Predominância dos Homens	Predominância dos Homens	Predominância dos Homens	Predominância acentuada dos Homens
Zona Marktest	Grande Lisboa e Litoral Norte	Grande Lisboa e Interior Norte	Interior Norte e Litoral Norte	Litoral Norte, Grande Porto e Grande Lisboa	Grande Lisboa e Litoral Centro

#### 4.4.6 Classificação dos clientes

Um dos benefícios resultante da utilização do modelo de mistura consiste na obtenção das probabilidades *a posteriori* de pertença para novos clientes. Estas últimas podem ser obtidas com base nas variáveis de base da segmentação. Muitas das vezes, esta informação é disponível apenas para o processo de segmentação, não estando habitualmente disponível para os clientes que não participaram no estudo. Tal facto, impossibilita uma classificação efectiva de novos clientes num dos segmentos obtidos, podendo transformar toda a segmentação num estudo meramente descritivo. Uma forma de contornar este problema passa pela utilização de variáveis disponíveis quer durante o processo de segmentação quer posteriormente quando os segmentos já foram identificados. A utilização de tal informação, isto é, o recurso às variáveis concomitantes, só é útil se estas possuírem uma capacidade explicativa das relações de pertença do cliente ao segmento  $Z$ . Com efeito, foi verificado na Secção 4.4.3 que as variáveis *Idade*, *Sexo* e *Zona Markttest* possuem esta capacidade explicativa. O *Software Latent Gold 4.0* estima a probabilidade *a posteriori*,  $\hat{\alpha}_{is}$ , com  $s=1, 2, 3, 4$  e  $5$ , dos clientes pertencerem aos segmentos. Esta estrutura “fuzzy” de pertença obriga à definição de uma regra no caso de se pretender alocar o cliente a um único segmento. Neste caso, o cliente é classificado no segmento com a probabilidade de pertença mais elevada. Dada a variável concomitante, é possível analisar o impacto da mesma nos vários segmentos, com base no cálculo das respectivas médias das probabilidades de pertença aos segmentos. Por outras palavras, para uma determinada categoria da variável concomitante, podemos observar a que segmento esta última tem maior probabilidade de pertencer, invertendo a relação de Bayes. A diferença entre a análise efectuada na Secção anterior - Perfis dos Segmentos - e a presente análise, consiste na perspectiva de análise dos segmentos. No primeiro caso, é fixado o segmento e efectuada uma caracterização do mesmo com base nas variáveis concomitantes.

Este processo é repetido para todos os segmentos de forma a estabelecer os perfis de cada um deles. Nota-se que neste caso as probabilidades condicionadas ao segmento somam um para cada uma das variáveis concomitantes. No segundo caso, é fixada a categoria da variável concomitante e é efectuada uma análise da importância da mesma ao longo dos segmentos. Para uma determinada categoria da variável concomitante, serão analisadas as probabilidades de pertença para todos os segmentos, sendo a soma das mesmas igual a um.

**Tabela 18 – Probabilidades condicionadas de pertença**

Estimativas das probabilidades condicionada ao segmento	Segmentos					
	<i>Independentes</i>	<i>Recém-Chegados</i>	<i>Aforradores</i>	<i>Fidelizados q.b.</i>	<i>Fiéis</i>	
<i>Idade</i>	dos 18 aos 21 anos	0.139	0.714	0.024	0.077	0.046
	dos 22 aos 25 anos	0.314	0.349	0.068	0.157	0.112
	dos 26 aos 32 anos	0.440	0.079	0.122	0.193	0.166
	dos 33 aos 43 anos	0.469	0.003	0.191	0.164	0.172
	dos 44 aos 65 anos	0.405	0.000	0.352	0.100	0.143
	<i>Sexo</i>	Feminino	0.446	0.219	0.126	0.133
Masculino		0.277	0.248	0.169	0.140	0.166
<i>Zona Marktest</i>	Grande Lisboa	0.442	0.265	0.265	0.112	0.121
	Grande Porto	0.232	0.232	0.232	0.265	0.150
	Interior Norte	0.201	0.270	0.270	0.105	0.124
	Litoral Centro	0.323	0.196	0.196	0.159	0.169
	Litoral Norte	0.400	0.198	0.198	0.155	0.099
	Sul	0.413	0.262	0.262	0.075	0.143
	Ilhas	0.516	0.191	0.191	0.082	0.067

No que diz respeito à variável *Idade*, verificou-se que o segmento dos *Independentes* é caracterizado por clientes com idades compreendidas entre os 26 e os 43 anos. Quando analisadas as respectivas faixas etárias na Tabela 18, observa-se que são nestas que este

segmento possui as probabilidades condicionadas de pertença mais elevadas, com um valor de 0.440 para as idades compreendidas entre os 26 anos e os 32 anos e um valor de 0.469 para as idades compreendidas entre os 33 anos e os 43 anos. Verificou-se também que a categoria do *Sexo Feminino* tinha uma forte predominância face ao *Sexo Masculino* no segmento dos *Independentes*. É novamente neste segmento que a probabilidade de pertença é a mais elevada quando condicionada à categoria do *Sexo Feminino* alcançando um valor de 0.446. No que diz respeito à variável *Zona Markttest*, as zonas da *Grande Lisboa* e do *Litoral Norte* são as mais predominantes neste segmento. Mais uma vez verifica-se que o segmento dos *Independentes*, para as referidas zonas, possui as probabilidades condicionadas de pertença mais elevadas. No que concerne as zonas *Sul* e *Ilhas* e analisando novamente a Tabela 18, observa-se que neste segmento, estas zonas possuem os pesos mais elevados quando comparados com os pesos obtidos nos demais segmentos. Relembra-se que em geral as probabilidades de pertencer às zonas *Ilhas* e *Sul* dado o segmento eram baixas em todos eles atingindo um máximo de 0.11 e 0.07, respectivamente. Com esta análise verifica-se que dados as referidas zonas, as probabilidades de pertença são mais elevadas no segmento dos *Independentes*.

A grande característica do segmento *Recém-Chegados* é o de reunir os clientes mais jovens. Tal facto se verifica na probabilidade condicionada de pertença que toma o valor de 0.714 para a faixa etária dos 18 aos 21 anos. Este segmento possui ainda a probabilidade mais elevada para as idades compreendidas entre os 22 e os 25 anos. Este segmento ainda possui as segundas probabilidades de pertença de maior valor quando condicionadas às categorias do *Sexo Feminino* e *Sexo Masculino*. No que diz respeito à variável *Zona Markttest*, a probabilidade de pertença condicionada ao *Interior Norte* é a mais elevada nos *Recém-Chegados* sendo esta zona uma das mais importantes aquando da perfilagem do referido

segmento. A *Grande Lisboa*, a zona com maior peso no segmento, possui a segunda probabilidade de pertença mais elevada neste segmento.

A principal característica dos *Aforradores* é o facto da respectiva distribuição da variável *Idade* tomar os valores mais elevados nas faixas etárias de idades mais elevadas. Para as idades compreendidas entre os 44 e os 65 anos, as probabilidades de pertença condicionadas são mais elevadas no segmento dos *Independentes*. No entanto, é nos *Aforradores* que se verificam as segundas probabilidades de pertença mais elevadas. Por outro lado, é nas faixas etárias dos 18 aos 21 anos e 22 aos 25 anos que este segmento possui as probabilidades de pertença mais baixas.

Para os *Fidelizados q.b.*, a principal característica é o peso que possui o *Grande Porto* quando comparado com os pesos homólogos nos restantes segmentos. É neste segmento que é verificada a sua probabilidade de pertença mais elevada.

A variável *Sexo* foi uma das variáveis mais importantes para a descrição do perfil do segmento dos *Fiéis*. Com efeito, é neste segmento que se verifica a maior disparidade entre os clientes do *Sexo Feminino* (0.26) e o *Sexo Masculino* (0.74). Este comportamento tem reflexo nas probabilidades de pertença pelo facto de quando condicionada ao *Sexo Feminino*, é neste segmento que se verifica a probabilidade de pertença mais baixa.

Esta análise permite completar e tornar mais rica a análise efectuada sobre os perfis dos segmentos. Como exemplo, o segmento dos *Aforradores* caracterizado por clientes com Idades mais elevadas, possui a probabilidade de pertença mais baixa para as idades mais

novas, reforçando assim o que já tinha sido verificado na análise de perfis, mas agora analisado de um ângulo diferente.

#### **4.5 Segmentação e regressão em simultâneo com recurso a variáveis de marketing**

##### **4.5.1 Modelo de mistura de regressões**

A aplicação do modelo de mistura finita permitiu perceber a heterogeneidade das necessidades financeiras dos clientes ao longo dos 10 meses de relacionamento com a Instituição Fénix. Determinados produtos poderão ter sido a razão pela qual o cliente iniciou a sua relação com esta Instituição, sendo os restantes adquiridos posteriormente ao longo destes 10 meses de relacionamento. A utilização das variáveis concomitantes permitiu ainda perceber os diferentes tipos de clientes relacionados com os diferentes comportamentos observados no que diz respeito aos produtos detidos. Não obstante os benefícios verificados com a aplicação dos modelos referidos anteriormente, o CRM não visa apenas a caracterização dos diferentes comportamentos dos clientes face aos produtos adquiridos. É também necessário perceber qual foi o retorno do investimento das estratégias de marketing, de forma a avaliar as condições de continuidade ou de melhoria das mesmas. Assim, pretende perceber-se se as acções realizadas pela Instituição Financeira no período de Janeiro a Outubro de 2007 tiveram impacto nas relações estruturais no que diz respeito aos produtos detidos pelos clientes.

Na presente secção, procuramos compreender se as acções executadas ao longo dos referidos meses, tiveram impacto no comportamento dos clientes no que concerne aos produtos detidos pelos mesmos. Para o efeito é novamente considerada a amostra utilizada na Secção 4.4 para a estimação dos efeitos das acções de marketing para cada segmento e para tal será aplicado o

modelo de mistura de regressões com variáveis concomitantes. Tal como Sarvary (2005) aponta, pretendemos obter uma segmentação que descreva a forma como o gestor pretende observar o seu mercado, identificando os segmentos cujo comportamento dos produtos detidos pode ser devido às acções de marketing, e assim adequar as abordagens de Marketing dirigidas de forma a aumentar o índice de *cross-selling*.

Para a estimação destes modelos foi novamente utilizado o *Software Latent Gold 4.0*. Os indicadores de posse considerados são os mesmos que os utilizados na Secção 4.4, isto é, os indicadores de posse de *Cartão de Débito*, *Crédito Imobiliário*, *Cartão de Crédito*, *Depósito a Prazo*, *Produto de Investimento*, *Serviço Financeiro* e *Crédito ao Consumo*. Utilizaram-se as mesmas variáveis concomitantes, isto é, as variáveis *Idade*, *Sexo* e *Zona Marktest*.

#### **4.5.2 Análises bivariadas**

Antes de iniciar o processo de estimação dos vários modelos, vai ser efectuada uma análise bivariada entre os indicadores de posse e os indicadores de *marketing* para averiguar a existência de pelo menos uma relação de dependência. Se não forem observadas associações entre os indicadores de posse e os indicadores de marketing, é de esperar que a aplicação do modelo de mistura de regressões seja de pouca utilidade. Para tal, foi avaliada a independência entre os dois tipos de indicadores recorrendo ao teste do Qui-quadrado para tabelas de contingência de dimensão  $2 \times 2$  (ver Apêndice C).

Para tabelas de contingência de dimensão  $2 \times 2$ , a rejeição da Hipótese Nula da independência entre os dois indicadores, dá-se quando o valor da estatística de teste é superior ao valor 3.84,

que corresponde ao quantil de probabilidade 0.95 de uma distribuição Qui-quadrado com um grau de liberdade.

**Tabela 19 – Valor da estatística de teste do qui-quadrado**

Indicador de Posse	Indicador de contacto efectuado pelo Canal				
	Mailing	SMS	Telemarketing	Balcão tipo I	Balcão tipo II
Cartão de Débito	67,61	0,20	5,86	1,76	0,00
Crédito Imobiliário	137,28	117,83	<b>n.a</b>	43,98	<b>n.a</b>
Cartão de Crédito	<b>n.a</b>	<b>n.a</b>	<b>n.a</b>	0,17	25,94
Poupança à Prazo	<b>n.a</b>	<b>n.a</b>	<b>n.a</b>	0,05	<b>n.a</b>
Produtos de Investimento	<b>n.a</b>	<b>n.a</b>	<b>n.a</b>	87,39	<b>n.a</b>
Serviço Financeiro	<b>n.a</b>	<b>n.a</b>	<b>n.a</b>	192,50	14,91
Crédito ao Consumo	2,74	0,45	<b>n.a</b>	0,03	<b>n.a</b>

Nota: O termo **n.a** indica que nenhum cliente detentor do produto foi alvo de acção de aquisição pelo canal.

Verifica-se que para os produtos de *Poupança a Prazo* e *Crédito ao Consumo*, os valores da estatística de teste (quando aplicável) não fornecem evidências suficientes para rejeitar a hipótese de independência entre os respectivos indicadores de posse e qualquer um dos indicadores de contacto efectuado pelos canais (Tabela 19). O indicador de *Mailing* é estatisticamente significativo para os indicadores de posse dos produtos *Cartão de Débito* e *Crédito Imobiliário*. É rejeitada a hipótese de independência entre o indicador de SMS e o indicador *Crédito Imobiliário*. Para o indicador de *Telemarketing*, apenas existe uma relação de dependência com o indicador de posse de *Cartão de Débito*. No entanto, para um nível de significância de 0.01, a hipótese de independência não é rejeitada. É rejeitada a hipótese de independência entre o indicador de *Balcão tipo I* e os indicadores de posse de *Serviço Financeiro*, *Produto de Investimento*, e *Crédito Imobiliário*. Para o indicador de *Balcão tipo II*, os valores da estatística de teste verificados dos indicadores de posse de *Cartão de Crédito* e *Serviço Financeiro*, evidenciam a existência de associação. Para cada indicador de contacto

efectuado em acções de venda dirigida, existe evidência estatística de associação com pelo menos um indicador de posse.

Para além de analisar o grau de associação entre os indicadores de posse e as variáveis de marketing, realizou-se também regressões logísticas para cada um dos indicadores de posse.

**Tabela 20 – Resultado das regressões logísticas realizadas para todos os indicadores de posse**

Variáveis	Cartão de Débito	Crédito Imobiliário	Cartão de Crédito	Poupança a Prazo	Investimento e Poupanças	Crédito ao Consumo	Serviço Financeiro
Constante	2.53 (<.0001)	-4.0883 (<.0001)	-0.1036 (0.6775)	-2.3204 (<.0001)	-2.2514 (<.0001)	-2.8238 (<.0001)	-2.6439 (<.0001)
Mailing	0.50 (<.0001)	-0.9781 (<.0001)	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>	-0.2064 (0.0642)	<b>n.a</b> <b>n.a</b>
SMS	-0.31 (0.166)	-0.1998 (0.3125)	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>	0.1074 (0.3713)	<b>n.a</b> <b>n.a</b>
Telemarketing	0.52 (0.0771)	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>
Balcão tipo I	0.05 (0.3545)	0.8485 (0.0194)	-0.0717 (0.61)	0.0334 (0.8211)	1.2326 (<.0001)	0.0761 (0.4509)	-1.0381 (<.0001)
Balcão tipo II	-0.16 (0.6165)	<b>n.a</b> <b>n.a</b>	1.0462 (<.0001)	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>	<b>n.a</b> <b>n.a</b>	-1.0381 (0.0205)

Nota: valor-p entre parêntesis

Os resultados obtidos das regressões (Tabela 20) não são totalmente concordantes com as associações verificadas na Tabela 19. Com efeito, existem diferenças no que diz respeito à importância dos indicadores de *Telemarketing* e *SMS*. Os efeitos dos indicadores de *SMS* e *Telemarketing* não são estatisticamente significativos nas regressões dos indicadores de posse do *Crédito Imobiliário* e *Cartão de Débito*, respectivamente. Verifica-se que todos os indicadores de posse possuem pelo menos um indicador de marketing, *Mailing*, *Balcão tipo I*

ou *Balcão tipo II*, com um efeito estatisticamente significativo. Os indicadores de *Telemarketing* e *SMS* não possuem qualquer relação com os indicadores de posse.

### 4.5.3 Preparação dos dados para aplicação do modelo

O modelo de mistura de regressões irá efectuar uma regressão dos indicadores de posse sobre os indicadores de contacto em acções de venda dirigida para todos os segmentos. Por outras palavras, pretende explicar-se as relações estruturais dos produtos detidos com base nos indicadores de marketing por cada um dos segmentos obtidos. Tem-se, assim, 7 variáveis dependentes correspondentes aos indicadores de posse, e 5 variáveis independentes correspondentes às variáveis de marketing. Estas últimas são também designadas por covariáveis.

**Figura 8 – Variável dependente para o modelo de mistura de regressões**

Cliente	Indicador de Posse	Posse do Produto	Posição na sequência
A	Cartão de Débito	1	1
A	Crédito Imobiliário	0	2
A	Cartão de Crédito	0	3
A	Poupança à Prazo	0	4
A	Produto de Investimento	1	5
A	Serviço Financeiro	0	6
A	Crédito ao Consumo	1	7



Sequência	1	2	3	4	5	6	7
Cliente A	1	0	0	0	1	0	1

No entanto, o *Latent Gold 4.0* com a qual foram estimados os modelos das Secções 4.3 e 4.4, efectua a estimação do modelo de mistura de regressões com a definição de uma única

variável dependente. Pelo facto de trabalhar-se não com uma, mas sim com 7 variáveis dependentes, existe a necessidade de contornar esse problema construindo uma nova variável que passa a conter a informação dos 7 indicadores de posse. Como se pode observar na Figura 8, esta nova variável é constituída por uma sequência de 7 caracteres, onde cada posição pode tomar o valor 0 ou 1 consoante a posse ou não do produto, sendo que a posição na respectiva sequência diz sempre respeito ao mesmo produto.

#### **4.5.4 Modelo de mistura de regressões com concomitantes**

As concomitantes utilizadas para a estimação do presente modelo são novamente as variáveis *Idade*, *Sexo* e *Zona Market*. Para os 5.558 clientes da amostra considerada para a estimação dos modelos anteriores, foram obtidos diferentes modelos com 1, 2, 3, 4, 5 e 6 segmentos. Foram utilizados diferentes valores iniciais para o processo de estimação do modelo. Para além da não concordância na solução óptima, as estatísticas de Wald apresentaram sempre valores-p superiores ao tradicional nível de significância de 0.05 para a avaliação da significância de variáveis. Por outras palavras, os valores-p obtidos resultantes da estatística de Wald revelam que as variáveis de marketing não são estatisticamente significativas, sendo as estimativas obtidas para os respectivos efeitos iguais a zero.

Analisou-se a distribuição da variável dependente (ver Apêndice C), e verificou-se que determinados valores da variável tinham uma frequência relativa inferior a 5%. De forma a diminuir o número de valores distintos que a variável dependente pode tomar, extraiu-se uma sub-amostra retirando as categorias com uma frequência relativa inferior a 5%. O resultado dessa iniciativa originou numa nova amostra composta por 4.180 clientes, na qual a variável dependente assume a distribuição apresentada na Tabela 21. Procedeu-se à estimação do modelo, considerando a amostra constituída por 4.180 clientes. Mesmo assim os resultados

obtidos não foram estatisticamente significativos. Com efeito, a solução óptima composta por 5 segmentos teve os mesmos resultados que os observados com a amostra original. Verificou-se que os valores-p obtidos da estatística de Wald revelam que as variáveis de marketing não são estatisticamente significativas, isto é não são explicativas do comportamento dos produtos detidos nos segmentos (ver Apêndice D).

**Tabela 21 – Distribuição da variável dependente na nova amostra**

Variável Dependente	N.º de Observações	%
0000000	293	7.0
1000000	2636	63.1
1000010	303	7.2
1010000	648	15.5
1010010	300	7.2
Total	4180	100

Apesar de ter-se verificado algumas relações de associação, confirmadas depois com a realização de regressões logísticas para cada um dos indicadores de posse com base nas variáveis de marketing, as respectivas relações revelaram-se sem efeito quando aplicado o modelo de mistura de regressões. Tudo indica que não existe evidência da capacidade dos indicadores de marketing em explicar os diferentes tipos de comportamentos no que concerne os produtos detidos pelos clientes em cada um dos segmentos.

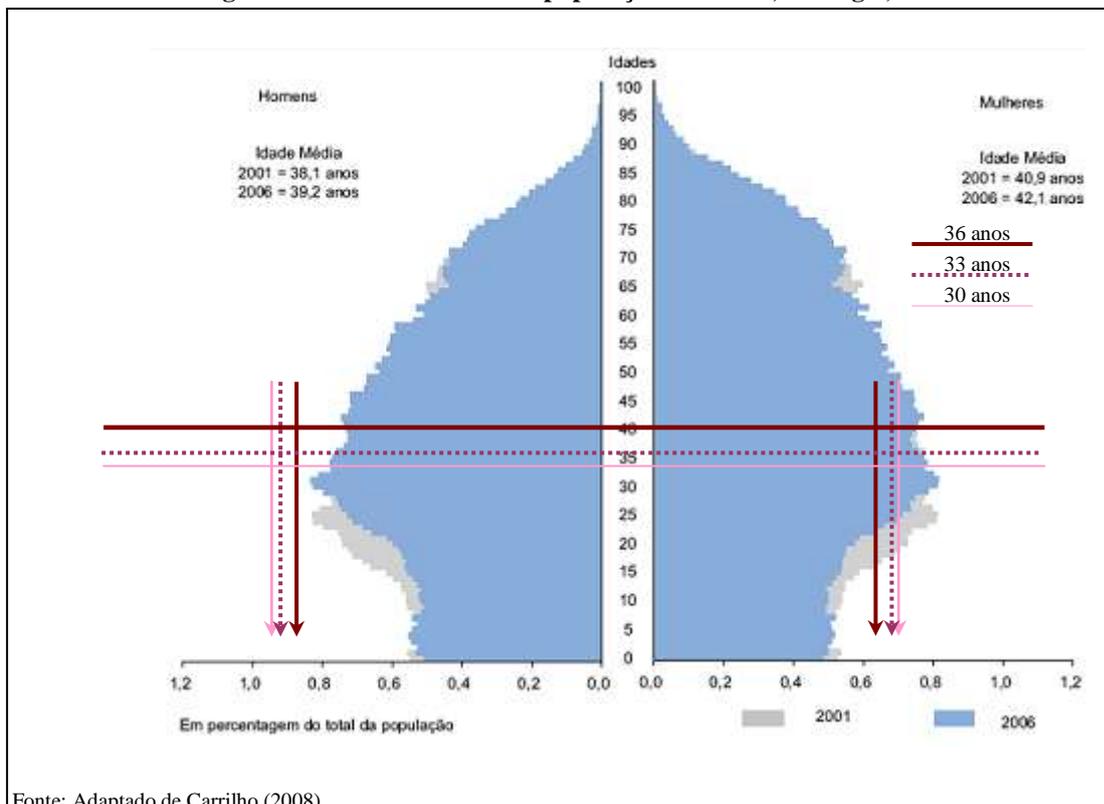
#### **4.6 Discussão**

A segmentação resultante da aplicação do modelo de mistura finita sem concomitantes, possibilitou a identificação de quatro segmentos.

Relativamente à composição dos segmentos no que diz respeito à variável *Sexo*, verifica-se que esta acaba por seguir a distribuição verificada na pirâmide etária da população residente

em Portugal. Na Figura 9, a pirâmide mostra que os homens são predominantes face às mulheres para as médias de idade obtidas nos segmentos. No entanto, as frequências relativas não se distanciam muito entre os dois géneros, esperando verificar-se a mesma ordem de importância na distribuição da variável *Sexo* ao longo dos segmentos. Para o segmento dos *Fieis*, esta ordem de importância não é verificada. Neste segmento as mulheres atingem a sua frequência mínima com 33%. Esta predominância considerável dos homens face às mulheres poderá ser explicada por diversos motivos. Um deles, de carácter mais organizacional, diz respeito aos processos operacionais da Instituição Financeira. Na abertura de uma conta à ordem, o primeiro titular da mesma é muitas das vezes a identidade representativa dos produtos posteriormente adquiridos. Desta forma, se o primeiro titular for o homem na aquisição de um crédito à habitação, este passará a ser a figura representativa do produto adquirido. Um segundo motivo, de carácter mais sociológico, diz respeito ao papel dos géneros na vida familiar de um casal.

**Figura 9 – Pirâmide etária da população residente, Portugal, 2001 e 2006**



A sociedade portuguesa ainda é uma sociedade patriarcal onde o representante do agregado familiar é predominantemente o homem. Neste sentido, a discrepância existente neste segmento pode dever-se ao facto do homem se dar como primeiro proponente na conta de crédito à habitação, e neste sentido apesar da aquisição ser efectuada pelo casal, o homem acaba por ser o representante dessa aquisição.

Assiste-se também ao fenómeno dos jovens viverem cada vez mais a sós ou em agregados sem relação parental (Aboim, 2003). Este facto deriva de uma maior individualização nos jovens entre os 20 e 29 anos, que optam por viver separados dos pais, quer por razões educacionais, que por razões de independência económica derivada à sua entrada na vida activa (Aboim, 2003). Quando derivado de motivos educacionais, podemos associar este comportamento de autonomia individual ao segmento dos *Recém-Chegados*, e quando derivado pela entrada na vida activa ao segmento dos *Fidelizados q.b.* Tais comportamentos permitem compreender melhor o portfolio dos clientes em um dos segmentos. Os clientes *Recém-chegados* apenas necessitam de movimentar a sua conta com depósitos e levantamentos de dinheiro, enquanto que os que iniciaram a sua vida activa podem comprar o seu primeiro carro, efectuar a sua primeira viagem fora de Portugal, ou simplesmente investir na compra de um computador. Uma análise mais detalhada da finalidade do *Crédito ao Consumo* contraído pode ajudar-nos a compreender os tipos de necessidades de crédito deste último segmento. Nota-se que nas condições anteriormente referidas, existe uma predominância clara do referido comportamento no sexo masculino (Guerreiro, 2003). Tal facto também se verifica na distribuição da variável *Sexo* nos dois segmentos referidos. É importante realçar que a entrada dos jovens na vida activa é cada vez mais tardia devido a uma maior escolarização, e conseqüente dependência económica dos mesmos dos pais (Carrilho, 2008). Este fenómeno pode explicar a incidência no segmento dos *Recém-*

*Chegados* nas zonas *Marktest Grande Lisboa, Litoral Norte e Interior Norte* e no segmento dos *Fidelizados q.b.* nas zonas de *Grande Lisboa, Litoral Norte e Grande Porto* pelo facto das mesmas apresentarem melhores condições de ensino e de trabalho.

No segmento dos *Fieis*, de um ponto de vista de negócio, é natural que se verifique a concentração dos clientes em torno dos 32 anos de idade. A concessão de crédito à habitação pode ser efectuada até um prazo máximo de 50 anos (com idade máxima a rondar os 75 anos de idade), o que possibilita diluir a dívida ao longo do tempo, conseguindo desta forma uma prestação mais baixa. Mas por outro lado, quanto maior for a idade do cliente ao contrair o empréstimo, mais elevada será o prémio de seguro de vida a ser pago pelo mesmo. Pelas razões já enunciadas anteriormente (Aboim, 2003), o facto dos jovens retardarem a sua entrada na vida activa, explica também o facto dos mesmos, só a partir dos 30 anos, começarem a reunir as condições para garantir uma estabilidade económica e suportar as prestações de um crédito à habitação.

Nos últimos anos, têm sido cada vez mais os alertas dados para a importância de poupar como benefícios para os clientes, mas também para a economia em geral. Depois do Estado Português ter acabado com os benefícios fiscais nas contas poupança habitação em 2005, o Orçamento de Estado de 2006, trouxe de volta os benefícios fiscais, permitindo uma dedução à colecta do IRS dos valores aplicados em planos de poupança reforma. Tal facto deve-se à capacidade cada vez mais reduzida do Estado em corrigir os desequilíbrios financeiros na segurança social ao longo dos próximos anos. Torna-se claro a necessidade de poupar para conseguir uma melhor reforma. É natural que a consciencialização deste problema social atinja os indivíduos que estão no activo. No entanto, dado o sobre-endividamento das famílias portuguesas na aquisição de automóvel, casa própria, mobiliário e electrodomésticos, a

possibilidade de poupar não se encontra ao alcance de todos os agregados familiares. Isto significa que apenas quando forem amortizadas algumas dessas dívidas se tornará viável aplicar valores em produtos de poupança. Esta realidade portuguesa pode explicar o facto do segmento dos *Aforradores* ser o de idade mais elevada comparativamente aos restantes. Pois, só com a existência de uma folga financeira verificada em idades mais avançadas, é que os clientes poderão ter capacidade para poupar. Ao observar a distribuição dos clientes pertencentes ao respectivo segmento, verifica-se que esta preocupação é predominante no Norte de Portugal, com 29% dos clientes pertencentes à zona *Interior Norte* e 20% pertencentes à zona *Litoral Norte*.

A segmentação obtida do modelo de mistura finita com variáveis concomitantes, respeita os grupos obtidos na segmentação anterior, contudo subdividiu o segmento anteriormente designado por *Recém-Chegados* em dois segmentos, os *Independentes* e os *Recém-Chegados* novamente. A identificação do segmento dos *Independentes* pode ser explicada pelo facto de se verificar um aumento da taxa de actividade feminina que passou de 53% em 1991 para 65% em 2001 (Aboim, 2003).

A utilização dos modelos de mistura e de mistura de regressões são ferramentas de ajuda na implementação de uma estratégia de CRM. Os primeiros não podem ser descuidados pois permitem perceber as relações estruturais existentes na população no que diz respeito aos primeiros produtos adquiridos pelos clientes. Esta análise permite perceber as necessidades dos clientes e, desta forma, delinear um plano estratégico de contacto. Os segundos permitem avaliar o impacto da estratégia adoptada e desta forma avaliar a sua continuidade, melhoria ou simplesmente o seu término. A aplicação do modelo de mistura com variáveis concomitantes, no que diz respeito aos segmentos obtidos, permitiu sugerir as seguintes medidas de actuação:

- no caso do segmento dos *Independentes*, conseguir por exemplo, acordos com Institutos de Beleza credenciados e com notoriedade pública com estabelecimentos na *Grande Lisboa* e no *Litoral Norte*, de forma a oferecer um financiamento com condições vantajosas em tratamentos de beleza. Este tipo de iniciativa pode derivar na concessão de um produto inovador cujo o tema central é o bem estar da mulher;
- nos *Recém-chegados*, e dado ser o segmento mais jovem (média de idade de 21 anos), o potencial pode estar na componente educacional. Uma sugestão pode ser o desenvolvimento de estratégias de *cross-selling* que promovam o patrocínio destes jovens nos seus estudos através de produtos de *Crédito ao Consumo* com condições vantajosas dada a idade do cliente;
- nos *Fidelizados q.b.*, o gestor de marketing pode realizar um conjunto de acções meramente informativas de forma a dar conhecimento das vantagens associadas ao produto *Serviço Financeiro* para aumentar o *cross-selling* nos produtos de *Cartão de Crédito*, *Crédito ao Consumo* e *Crédito Imobiliário*. Os contactos poderão incidir nos clientes residentes nas zonas do *Litoral*, *Grande Lisboa* e *Grande Porto*. Deve perceber-se quais os factores que distinguem a zona *Sul* das enunciadas anteriormente, para aumentar o índice de *cross-selling* nesta região;
- no segmento dos *Fiéis*, as acções devem ser dirigidas a clientes do sexo *Masculino*, com idade entre os 33 e 43 anos, com o propósito de oferecer o *Serviço Financeiro* que traz vantagens nas anuidades de cartões e nos *spreads* do crédito à habitação. Da mesma forma, estas acções devem ser dirigidas a clientes localizados em zonas Marktest do *Litoral*, *Grande Lisboa* e *Grande Porto*;
- nos *Aforradores*, deve seleccionar-se os clientes com idades superiores a 44 anos das zonas do *Interior Norte* e *Litoral Norte* e efectuar um acompanhamento em termos de divulgação de novos produtos de *Poupança a Prazo* ou de *Investimento*.

A aplicação do modelo de mistura de regressões não conseguiu mostrar evidências de que as acções comerciais de venda dirigida têm poder explicativo sobre os produtos adquiridos pelos clientes ao nível dos segmentos. É importante ressaltar que a única acção concertada para novos clientes, para além de confirmar se o cliente está satisfeito com o serviço prestado pelo agente comercial aquando da abertura de conta, também visa a colocação do *Cartão de Débito*. Este último é quase sempre adquirido aquando da abertura de conta, transformando a acção de venda numa acção de controlo da qualidade do serviço prestado. Para além das acções de venda dirigida, a Instituição Fénix possui uma estratégia de comunicação externa agressiva que utiliza para o efeito os meios de comunicação tais como a televisão, a rádio e a imprensa. A realização de campanhas com a utilização dos referidos recursos poderão ter anulado o efeito das acções dirigidas, sendo as mesmas percepcionadas como um reforço da mensagem comercial divulgada nos meios de comunicação anteriormente referidos. Esta análise pode indiciar uma abordagem de canal inadequada, uma estratégia de oferta discordante com as necessidades dos clientes, ou ainda a ineficiência das acções de venda dirigida para clientes recentes quando estão no “ar” campanhas publicitárias dos mesmos produtos.

A aplicação dos modelos de mistura com concomitantes permitiu-nos perceber as estruturas existentes nos segmentos obtidos e a forma como os mesmos se relacionam com as variáveis demográficas. Assim, conseguiu-se perceber que tipo de indivíduos é composto cada segmento, adequando melhor a estratégia de marketing para futuros clientes. É de salientar que a maior parte das iniciativas de marketing desta instituição, selecciona clientes com conta aberta há mais de 6 meses, pelo facto da informação de risco sobre o cliente ser efectuada com base num histórico de 6 meses.

Este estudo, baseado nos clientes com uma antiguidade de 10 meses, apenas efectuou a análise para os clientes que iniciaram a sua relação em Janeiro de 2007. A componente longitudinal não foi incorporada no presente trabalho. Tal abordagem teria permitido uma melhor compreensão dos segmentos, e desta forma melhor perceber as decisões de marketing a serem tomadas. Uma análise semelhante efectuada com base numa amostra de clientes com uma maior antiguidade de relacionamento com a Instituição Financeira poderia ter revelado outros resultados. É importante salientar que a escolha de clientes com uma relação bancária recente foi propositada. Por um lado, o objectivo foi o de compreender as necessidades financeiras dos clientes no início da sua relação com a Instituição Financeira com o Banco. Por outro lado, as estratégias de marketing existentes na Instituição são elaboradas para clientes com uma antiguidade mais elevada. A pretensão foi a de perceber se a referida estratégia era adequada para os clientes recentes.

Para uma investigação futura, a introdução da componente longitudinal no modelo, estando esta ausente no presente estudo, poderá inferir sobre a ordem pela qual os produtos são adquiridos em cada segmento e afinar ainda mais as futuras estratégias de marketing a serem personalizadas em cada segmento.

## 5 Conclusão

No ramo bancário, os produtos desenvolvidos são cada vez mais complexos e diversificados por forma a dar resposta às necessidades financeiras do cliente, garantindo assim a sua permanência na Instituição Financeira. Uma das formas utilizadas para fidelizar um cliente é a realização de acções de *cross-selling*, onde são propostos novos produtos aos clientes já existentes. Para que as ofertas sejam bem recebidas é preciso “acertar” no conjunto de produtos a oferecer. O objectivo único do CRM Analítico (Kamakura *et al.*, 2005) é conhecer o cliente para melhor o “servir”. Para tal, uma das suas metas é perceber quais as acções de *cross-selling* a serem desenvolvidas de forma a antecipar as necessidades futuras dos clientes. Na tentativa de solucionar este problema de marketing, recorreu-se a um dos pilares do CRM Analítico, a utilização de um processo de segmentação mediante a aplicação dos modelos de misturas sem e com recurso às variáveis concomitantes. Esta dissertação abordou duas vertentes das muitas existentes dos modelos de mistura para escolha discreta, os modelos binários de mistura finita sem e com variáveis concomitantes, assim como os modelos de mistura de regressões multinomiais sem e com variáveis concomitantes.

Os modelos de mistura binários demonstraram ser uma técnica de segmentação eficiente para melhor perceber as relações estruturais existentes, no que diz respeito à posse de produtos financeiros. A aplicação dos modelos de mistura foi efectuada com base em indicadores de posse detidos pelos clientes em Outubro de 2007, os quais iniciaram a sua relação financeira com a Instituição Fénix em Janeiro de 2007. Posteriormente aplicou-se um modelo de mistura de regressões. Estes modelos são uma extensão dos modelos de mistura como resultado de incluírem covariáveis (Dias, 2004), transformando a natureza descritiva do modelo de mistura, num modelo explicativo. Com efeito, o modelo de mistura de regressões possibilita a

estimação de um modelo de regressão para cada um dos segmentos obtidos. Tanto os modelos de mistura finita como os modelos de mistura de regressões permitem a incorporação de variáveis concomitantes, permitindo a obtenção de probabilidades de pertença *a posteriori* de forma a efectuar a classificação de novos clientes nos segmentos mais adequados.

A segmentação resultante da aplicação dos modelos de mistura finita sem concomitantes identificou 4 segmentos de clientes: os *Recém-chegados*, *Fidelizados q.b.*, *Fiéis* e *Aforradores*. O segmento *Recém-Chegados* aparentou agrupar clientes que se encontravam numa mesma fase do seu ciclo de vida do seu relacionamento com a instituição. No entanto, a aplicação do modelo de mistura finita com concomitantes, subdividiu este último em dois sub-segmentos, ficando assim, uma solução composta por 5 segmentos: os *Independentes*, *Recém-chegados*, *Fidelizados q.b.*, *Fiéis* e *Aforradores*. A existência de heterogeneidade nas necessidades financeiras dos clientes e os diferentes perfis verificados, ajuda-nos a perceber que a realização de uma única estratégia de *cross-selling* não é viável. Para além do cliente de um dado segmento poder considerar a oferta completamente desajustada, pode deixar a impressão de que a Instituição não o conhece, sendo esta consequência a antítese do CRM.

Os resultados obtidos definem uma base de trabalho para o desenvolvimento de estratégias de *cross-selling* que terão por objectivo fidelizar o cliente e aumentar a rendibilidade do mesmo, respondendo assim a um dos objectivos do CRM (Ebner *et al.*, 2002). No entanto, existem neste trabalho, algumas limitações que convém realçar:

- construção da amostra: a amostra foi construída por conveniência. A recolha da amostra incidiu nos clientes recentes de forma a perceber as suas motivações financeiras para iniciarem uma relação bancária, e compreender o impacto da estratégia adoptada sobre os mesmos;

- disponibilidade das variáveis concomitantes: Apenas foram consideradas as variáveis *Sexo*, *Idade* e *Zona Markt*. A incorporação da informação de domiciliação de vencimentos e habilitação literária poderia ter sido uma mais valia para melhor caracterizar os perfis dos segmentos. Na literatura (e.g. Li *et al.*, 2005), estas últimas são consideradas variáveis importantes para a realização de um estudo de segmentação com o objectivo de fomentar estratégias de *cross-selling*. Por outro lado, a informação da remuneração domiciliada teria permitido analisar a existência de segmentos constituídos pelos clientes mais rendíveis e assim adequar as ofertas de crédito nesse sentido.

A segmentação realizada neste trabalho apenas constitui uma das peças do *puzzle* do CRM – Analítico. Para satisfazer as necessidades do cliente antecipadamente é preciso completar este *puzzle* e analisar os restantes ângulos de observação do cliente de forma a garantir uma visão centrada sobre o mesmo. Em termos de investigação futura, será importante:

- estimar um modelo que identifique o próximo produto a ser oferecido para cada segmento obtido no presente estudo (Li *et al.*, 2005; Paas *et al.*, 2006). A possibilidade da introdução de uma componente longitudinal no modelo, permitirá inferir a ordem pela qual os produtos são adquiridos em cada segmento, e perceber o espaçamento temporal a respeitar entre cada uma das ofertas;
- analisar a forma como os clientes utilizam os canais de contacto. O entendimento da função de cada canal, e a sua frequência de utilização permitem desenvolver uma abordagem de marketing de acordo com o perfil de utilização dos canais do cliente (Cohen, 2004);
- maximizar o valor do investimento efectuado nas acções de marketing, sabendo seleccionar os clientes mais propensos à compra (Cohen, 2004);

- perceber o valor futuro que o cliente poderá vir a ter. Por outras palavras não limitar os esforços de marketing nos clientes actualmente rendíveis, mas também nos que possuem uma probabilidade elevada de crescimento de rendibilidade.

A utilização dos modelos de mistura com concomitantes permitiu atingir um dos objectivos do CRM analítico, o de melhor perceber clientes com 10 meses de relacionamento com a Instituição no que diz respeito às suas necessidades financeiras. Este conhecimento irá permitir, com base nas probabilidades de pertença *a posteriori*, decidir sobre a forma de melhor contactar e quais os produtos adequados a oferecer aos futuros clientes com as mesmas condições que os analisados no presente estudo. Verifica-se também que os segmentos obtidos conseguiram reflectir a realidade comportamental existente em Portugal, e desta forma são concordantes com as transformações dos comportamentos sociais e demográficos sucedidos em Portugal na última década.

## 6 Referências

- Aboim, S. (2003), Evolução das estruturas domésticas, *Sociologia, Problemas e Práticas* 43, 13-30.
- Akaike, H. (1974), A new look at statistical model identification, *IEEE Transactions on Automatic Control*, AC-19(6), 716-723.
- Alard, P. e P-A. Guggémos (2005), *CRM – Les Clés de la Réussite*. Éditions d'Organisation, Paris.
- Andrews, R. L. e I. S. Currim (2003), A comparison of segment retention criteria for finite mixture logit models, *Journal of Marketing Research* XL, 235-243.
- Bandeem-Roche, K., D. L. Miglioretti, S.L. Zeger e P.J. Rathouz (1997), Latent variable regression for multiple discrete outcomes, *Journal of the American Statistical Association* 2(440), 1375-1386.
- Beirão, A. (2004), O cliente no centro das atenções, *Suplemento da revista Marketeer* 93, 8-21.
- Berry, M.J.A e G.S. Linoff (2000), *Mastering Data Mining: The Art and Science of Customer Relationship Management*. Wiley (2nd Edition), New York.
- Bijmolt, T.H.A., L.J. Paas e J. K. Vermunt (2004), Country and consumer segmentation: Multi-Level latent class analysis of financial product ownership, *International Journal of Research in Marketing* 21, 323-340.
- Boter, J. e M. Wedel (1999), Segmentation of hedonic consumption: An application of latent class analysis to consumer transaction databases, *Journal of Market Focused Management* 3, 295-311.
- Boulding, W., R. Staelin, M. Ehret e W.J. Johnson (2005), A customer relationship management Roadmap: What is known, potential pitfalls and where to go, *Journal of Marketing* 69, 155-166.

Bozdogan, H. (1987), Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, 52(3), 345-370.

Bozdogan, H. (1993), Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse –Fisher information matrix, in Opitz , O., B. Lausen, and R. Klar (eds.), *Information and Classification, Concepts, Methods and Applications*, Berlin: Springer-Verlag,40-54.

Brochado, A. O. e F.V. Martins (2003). Assessing the number of components in the mixture models: a review, *FEP Working Papers*, Research – Work in Progress, 194, 1-31.

Bult, J. R. e T. Wansbeek (1995), Optimal selection for direct mail, *Marketing Science* 14(4), 378-394.

Cardoso, M. G. M. S, I. H. Themido e F.M. Pires (2000), Estruturando modelos e processos de segmentação, *Sociedade Portuguesa de Estatística, Actas do VII Congresso da Sociedade Portuguesa de Estatística*, 259-279.

Carrilho, M. J., L. Patrício (2008), A situação demográfica recente em Portugal, *Revista de Estudos Demográficos*, 42, 109-155.

Chen, Y. e L. Li (2006), Deriving information from CRM for knowledge management - A note on a commercial bank, *Systems Research and Behavioural Science* 23, 141-146.

Cohen, M. D. (2004). Exploiting response models – optimizing cross-sell and up-sell opportunities in banking, *Information Systems* 29, 327-341.

Cohen, S. H. e V. Ramaswamy (1998), Latent segmentations models, *Marketing Research*, 10(2), 14-21.

Deloitte Consulting (2001), How to eat the CRM elephant. *Staight talk series, n.º 2. New York*.

Desarbo, W.S. e V. Ramaswamy (1994), CRISP - Customer response based iterative segmentation procedures for response modeling in direct marketing, *Journal of Direct Marketing* 8(3), 7-20.

Desarbo, W.S., V. Ramaswamy e R. Chatterjee (1995), Analysing constant-sum multiple criterion data: A segment-level approach, *Journal of Marketing Research*, XXXII, 222-232.

Dias, G.J. (2004), Finite mixture models: Review, applications, and computer-intensive methods. Ph.D. Thesis, University of Groningen, *Research School, Systems, Organization and Management*).

Ebner, M., A. Hu, D. Levitt e J. McCrory (2002), How to rescue CRM, *McKinsey Quarterly Special Edition, Technologie 4*, 49-57.

Green, P. E., F. J. Carmone e D.P. Wachspress (1976), Consumer segmentation via latent class analysis, *Journal of Consumer Research* 3, 170-174.

Guerreiro, M. D. (2003), Pessoas sós: múltiplas realidades, *Sociologia, Problemas e Práticas* 43, 31-49.

Gupta, S. e P. K. Chintagunta (1994), On using demographic variables to determine segment membership in logit mixture models, *Journal of Marketing Research* XXXI, 128-136.

Kamakura, W. A., C. F. Mela, A. Ansari, A. Bodapati, P. Fader, R. Iyengar, P. Naik, S. Neslin, P. Sun, P. Verhoef, M. Wedel e R. Wilcox (2005), Choice models and customer relationship management, *Marketing Letters* 16 (3/4), 279-291.

Kamakura, W. A. e G. J. Russell (1989), A probabilistic choice model for market segmentation and elasticity structure, *Journal of Marketing Research* XXVI, 379-390.

Kaynak, E. e T. D. Harcar (2005), American consumers' attitudes towards commercial banks: A comparison of local and national bank customers by use of geodemographic segmentation, *The International Journal of Bank Marketing*, 21(1), 73-89.

Lawson, R. e S. Todd (2003), Consumer preferences for payment methods: a segmentation analysis, *The International Journal of Bank Marketing*, 21(2/3), 72-79.

Lefébure, R. e G. Venturi (2005), *Gestion de la Relation Client. Solutions d'entreprise*: Eyrolles, Paris.

Lemmi, M., (2007), *Resolve! CRM - Customer Relationship Management*, texto não publicado, [http://www.resolvenet.com.br/cons\\_crm.asp](http://www.resolvenet.com.br/cons_crm.asp) [consultado em 18-12-2007].

Li, S., B. Sun e R.T. Wilcox (2005), Cross-selling sequentially ordered products: An application to consumer banking services, *Journal of Marketing Research* XLII, 233-239.

Maklan, S., S. Knox e L. Ryals (2005), Using real options to help build the business case for CRM investment, *Long Range planning journal* 38, 393-410.

Manchanda, P., J. -P. Dubé, K.Y. Goh e P.K. Chintagunta (2006), The effect of banner advertising on internet purchasing, *Journal of Marketing Research* XLIII, 98-108.

Markttest (2006), Notas técnicas: Regiões Markttest, texto não publicado, [http://www.markttest.pt/Notas\\_Tecnicas/default.asp?strUrl=/notas\\_tecnicas/info/conteudos/documentos/regioes.asp](http://www.markttest.pt/Notas_Tecnicas/default.asp?strUrl=/notas_tecnicas/info/conteudos/documentos/regioes.asp) [consultado em 13-09-2006].

Mendoza, L. E., A. Marius, M. Pérez e A.C. Grimán (2007), Critical success factors for a customer relationship management, *Information and Software Technology* 47, 913-945.

Paas, L. J., T. H. A. Bijmolt e J.K. Vermunt (2006), Acquisition patterns of financial products: A longitudinal investigation, *Journal of Economic Psychology* 28, 229-241.

Peppers, D., M. Rogers e B. Dorf (1999), Is your company ready for one-to-one marketing, *Harvard Business Review* 77(1), 151-160.

Pine II, B., V. Bart e B. V. Boynton (1993), Making mass customization work, *Harvard Business Review* 71(5), 108-118.

Richardson, P. e P. Weill (1999), Telstra's national telemarketing centre, *Journal of Information Technology* 14, 217-234.

Sarvary, M. (2005), Market segmentation, target market selection and product positioning, *Harvard Business School* 15, 1-6.

SAP (2006), mySAP™ CRM customer success stories: A comprehensive reference guide, texto não publicado, [http://www.sap.com/solutions/business-suite/crm/pdf/CS\\_2006\\_Story\\_Book.pdf](http://www.sap.com/solutions/business-suite/crm/pdf/CS_2006_Story_Book.pdf) [consultado em 18-12-2007].

SAS (2002), CRM – Uma abordagem analítica, *Course Notes*.

SAS (2006), Who ensures the insurer?; Allianz Portugal enhances customer value with SAS®, texto não publicado, <http://www.sas.com/success/allianzportugal.html> [consultado em 18-12-2007].

SAS (2007), Bank's conversion rates jump to 30 percent; Portugal's Banco Santander Totta gets to know its customers even better with SAS®, texto não publicado, <http://www.sas.com/success/santandertotta.html> [Consultado em 18-12-2007].

Saunders, J. (1994), Cluster analysis, *Journal of Marketing Management* 10, 13-28.

Schwarz, G. (1978), Estimating the dimension of a model, *The Annals of Statistics*, 6(2), 461-464.

Smith, W. (1956), Product differentiation and market segmentation as alternative marketing strategies, *The Journal of Marketing* July, 1-8.

Speed, R. e G. Smith (1992), Retail financial services segmentation, *The Service Industries Journal*, 12(3), 368-383.

Tsai, C. -Y. e C. -C. Chiu (2004), A purchase-based market segmentation methodology, *Expert Systems with Applications* 27, 265-276.

Turner, J. (2002), Advertising credit products and loans, *ABA Bank Marketing* 8, 38.

Vermunt, J. K. e J. Magidson (2003), Latent class models for classification, *Computational Statistics & Data Analysis* 41, 531-537.

Vermunt, J. K. e J. Magidson (2005), *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Belmont Massachusetts: Statistical Innovations Inc.

Wedel, M. (2002), Concomitant variables in finite mixture models, *Statistica Neerlandica* 56(3), 362-375.

Wedel, M., W. S. Desarbo, J.R. Bult e V. Ramaswamy (1993), A latent class poisson regression model for heterogeneous count data, *Journal of applied Econometrics* 8, 397-411.

Wedel, M. e W. Kamakura (1999), *Market Segmentation Conceptual and Methodological Foundations*. International Series in Quantitative Marketing (2nd Edition), Springer, Dordrecht: Kluwer Academic.

Wells Fargo (2003), Measuring the next stage of success: The 15 most important measures of success in today's financial services industry, *Wells Fargo & Company Annual report 2003*, [https://www.wellsfargo.com/pdf/invest\\_relations/wf\\_2003annualreport](https://www.wellsfargo.com/pdf/invest_relations/wf_2003annualreport) [consultado em 18-12-2007].

## 7 Apêndices

**Apêndice A** - Modelo de mistura finita com concomitantes: Resultados detalhados do modelo de mistura finita com variáveis concomitantes considerando o Indicador de *Seguro de Capitalização*.

<b>Informação sobre os modelos estimados</b>			
Nº de Segmentos	Log-Verossimilhança	BIC	N.º de parâmetros
1	-13367.4	26803.79	8
2	-12876.2	25967.97	25
3	-12621.3	25604.84	42
4	-12522	25552.10	59
5	-12448.9	25553.07	76
6	-12416.8	25635.51	93

<b>Estimativas das constantes de regressão</b>			
Estimativa do vector de constantes $\beta_0$		Wald	Valor-p
Cartão de Débito -	$\hat{\beta}_{01}$	0.02	0.88
Crédito Imobiliário -	$\hat{\beta}_{02}$	32.44	< 0.001
Cartão de Crédito -	$\hat{\beta}_{03}$	111.22	< 0.001
Depósito a Prazo -	$\hat{\beta}_{04}$	30.72	< 0.001
Produto de Investimento -	$\hat{\beta}_{05}$	133.32	< 0.001
Seguro de Capitalização -	$\hat{\beta}_{06}$	184.24	< 0.001
Serviço Financeiro -	$\hat{\beta}_{07}$	20.60	< 0.001
Crédito ao Consumo -	$\hat{\beta}_{08}$	22.81	< 0.001

<b>Estimativas dos efeitos de regressão dos indicadores</b>						
Estimativa dos vector de efeitos $\beta_j$		Solução com 4 Segmentos				Wald (valor-p)
		1	2	3	4	
Cartão de Débito -	$\hat{\beta}_{1s}$	-2.92	-3.00	-3.40	0	< 0.001
Crédito Imobiliário -	$\hat{\beta}_{2s}$	-1.80	6.64	-2.08	0	< 0.001
Cartão de Crédito -	$\hat{\beta}_{3s}$	-1.38	-0.75	-2.88	0	< 0.001
Depósito a Prazo -	$\hat{\beta}_{4s}$	2.28	1.89	2.24	0	< 0.001
Produto de Investimento -	$\hat{\beta}_{5s}$	2.37	2.86	3.56	0	< 0.001
Seguro de Capitalização -	$\hat{\beta}_{6s}$	0.91	6.79	0.08	0	0.015
Serviço Financeiro -	$\hat{\beta}_{7s}$	0.41	-2.72	-5.90	0	< 0.001
Crédito ao Consumo -	$\hat{\beta}_{8s}$	-0.15	-1.58	-3.40	0	< 0.001

**Apêndice B** - Modelo de mistura finita com concomitantes não considerando o indicador de posse de *Seguros de Capitalização*: Resultados detalhados do modelo de mistura finita com variáveis concomitantes não considerando o Indicador de *Seguro de capitalização*.

**Informação sobre os modelos estimados**

Nº de Segmentos	Log-Verossimilhança	BIC	N.º de parâmetros
1	-12732.16	25524.69	7
2	-12243.16	24684.64	23
3	-12002.73	24341.76	39
4	-11903.90	24282.06	55
5	-11833.49	24279.21	71
6	-11777.82	24305.84	87

**Estimativas das constantes de regressão**

Estimativa do vector de constantes $\beta_0$	Wald	Valor-p
Cartão de Débito - $\hat{\beta}_{01}$	136.18	< 0.001
Crédito Imobiliário - $\hat{\beta}_{02}$	0.06	0.81
Cartão de Crédito - $\hat{\beta}_{03}$	8.97	0.0027
Depósito a Prazo - $\hat{\beta}_{04}$	132.56	< 0.001
Produto de Investimento - $\hat{\beta}_{05}$	105.15	< 0.001
Serviço Financeiro - $\hat{\beta}_{06}$	21.13	< 0.001
Crédito ao Consumo - $\hat{\beta}_{07}$	61.39	< 0.001

**Estimativas dos efeitos de regressão dos indicadores**

Estimativa dos vector de efeitos $\beta_j$		Solução com 5 Segmentos					Wald (valor-p)
		1	2	3	4	5	
Cartão de Débito - $\hat{\beta}_{1s}$	-3.19	-0.58	2.18	-1.54	0	< 0.001	
Crédito Imobiliário - $\hat{\beta}_{2s}$	8.25	10.85	8.75	2.35	0	0.12	
Cartão de Crédito - $\hat{\beta}_{3s}$	1.00	1.18	2.06	-0.80	0	< 0.001	
Depósito a Prazo - $\hat{\beta}_{4s}$	0.34	0.10	-1.72	0.40	0	< 0.001	
Produto de Investimento - $\hat{\beta}_{5s}$	1.1216	0.4746	-1.9173	2.6583	0	< 0.001	
Serviço Financeiro - $\hat{\beta}_{6s}$	1.3032	-0.1881	2.7276	-4.1423	0	< 0.001	
Crédito ao Consumo - $\hat{\beta}_{7s}$	0.7733	-0.4285	1.0362	-2.3202	0	< 0.001	

**Resíduos Bivariados**

Indicadores de Posse	Cartão de Débito	Crédito Imobiliário	Cartão de Crédito	Depósito a Prazo	Produto de Investimento	Serviço Financeiro	Crédito ao Consumo
Cartão de Débito	.						
Crédito Imobiliário	0.036	.					
Cartão de Crédito	0.002	0.731	.				
Depósito a Prazo	0.282	0.481	0.074	.			
Produto de Investimento	2.809	0.089	0	0.783	.		
Serviço Financeiro	0.046	0.014	1.281	1.073	0.0002	.	
Crédito ao Consumo	0.885	1.269	0.084	2.305	0.134	0.329	.

**Apêndice C - Modelo de mistura de regressões: Análise bivariada**

**Tabelas de Contingências entre indicadores de posse e indicadores de contacto efectuado por canal**

	Indicador de contacto efectuado por Mailing		
Indicador de posse de cartão de débito	0	1	Total
0	505	78	583
1	3507	1468	4975
Total	4012	1546	5558

	Indicador de contacto efectuado por SMS		
Indicador de posse de cartão de débito	0	1	Total
0	577	6	583
1	4913	62	4975
Total	5490	68	5558

	Indicador de contacto efectuado por Telemarketing		
Indicador de posse de cartão de débito	0	1	Total
0	580	3	583
1	4880	95	4975
Total	5460	98	5558

	Indicador de contacto efectuado por Balcão do tipo I		
Indicador de posse de cartão de débito	0	1	Total
0	480	103	583
1	3981	994	4975
Total	4461	1097	5558

	Indicador de contacto efectuado por Balcão Tipo II		
Indicador de posse de cartão de débito	0	1	Total
0	505	78	583
1	3507	1468	4975
Total	4012	1546	5558

	Indicador de contacto efectuado por Mailing		
Indicador de posse de Crédito Imobiliário	0	1	Total
0	3711	1369	5080
1	465	13	478
Total	4176	1382	5558

	Indicador de contacto efectuado por SMS		
Indicador de posse de Crédito Imobiliário	0	1	Total
0	3792	1288	5080
1	462	16	478
Total	4254	1304	5558

	Indicador de contacto efectuado por Telemarketing		
Indicador de posse de Crédito Imobiliário	0	1	Total
0	5080	0	5080
1	478	0	478
Total	5558	0	5558

	Indicador de contacto efectuado por Balcão do tipo I		
Indicador de posse de Crédito Imobiliário	0	1	Total
0	4609	471	5080
1	476	2	478
Total	5085	473	5558

	Indicador de contacto efectuado por Balcão Tipo II		
Indicador de posse de Crédito Imobiliário	0	1	Total
0	5080	0	5080
1	478	0	478
Total	5558	0	5558

	Indicador de contacto efectuado por Mailing		
Indicador de posse de Cartão de Crédito	0	1	Total
0	4141	0	4141
1	1417	0	1417
Total	5558	0	5558

	Indicador de contacto efectuado por SMS		
Indicador de posse de Cartão de Crédito	0	1	Total
0	4141	0	4141
1	1417	0	1417
Total	5558	0	5558

	Indicador de contacto efectuado por Telemarketing		
Indicador de posse de Cartão de Crédito	0	1	Total
0	4141	0	4141
1	1417	0	1417
Total	5558	0	5558

	Indicador de contacto efectuado por Balcão do tipo I		
Indicador de posse de Cartão de Crédito	0	1	Total
0	4009	132	4141
1	1375	42	1417
Total	5384	42	5558

	Indicador de contacto efectuado por Balcão Tipo II		
Indicador de posse de Cartão de Crédito	0	1	Total
0	4135	6	4141
1	1401	16	1417
Total	5536	22	5558

	Indicador de contacto efectuado por Mailing		
Indicador de posse de Poupança à Prazo	0	1	Total
0	5075	0	5075
1	483	0	483
Total	5558	0	5558

	Indicador de contacto efectuado por SMS		
Indicador de posse de Poupança à Prazo	0	1	Total
0	5075	0	5075
1	483	0	483
Total	5558	0	5558

	Indicador de contacto efectuado por Telemarketing		
Indicador de posse de Poupança à Prazo	0	1	Total
0	5075	0	5075
1	483	0	483
Total	5558	0	5558

	Indicador de contacto efectuado por Balcão do tipo I		
Indicador de posse de Poupança à Prazo	0	1	Total
0	4947	128	5075
1	470	13	483
Total	5417	141	5558

	Indicador de contacto efectuado por Balcão Tipo II		
Indicador de posse de Poupança à Prazo	0	1	Total
0	5075	0	5075
1	483	0	483
Total	5558	0	5558

	Indicador de contacto efectuado por Mailing		
Indicador de posse de Produto de Investimento	0	1	Total
0	5381	0	5381
1	177	0	177
Total	5558	0	5558

	Indicador de contacto efectuado por SMS		
Indicador de posse de Produto de Investimento	0	1	Total
0	5381	0	5381
1	177	0	177
Total	5558	0	5558

	Indicador de contacto efectuado por Telemarketing		
Indicador de posse de Produto de Investimento	0	1	Total
0	5381	0	5381
1	177	0	177
Total	5558	0	5558

	Indicador de contacto efectuado por Balcão do tipo I		
Indicador de posse de Produto de Investimento	0	1	Total
0	5345	36	5381
1	164	13	177
Total	5509	49	5558

	Indicador de contacto efectuado por Balcão Tipo II		
Indicador de posse de Produto de Investimento	0	1	Total
0	5381	0	5381
1	177	0	177
Total	5558	0	5558

	Indicador de contacto efectuado por Mailing		
Indicador de Posse de Serviço Financeiro	0	1	Total
0	4645	0	4645
1	913	0	913
Total	5558	0	5558

	Indicador de contacto efectuado por SMS		
Indicador de Posse de Serviço Financeiro	0	1	Total
0	4645	0	4645
1	913	0	913
Total	5558	0	5558

	Indicador de contacto efectuado por Telemarketing		
Indicador de Posse de Serviço Financeiro	0	1	Total
0	4645	0	4645
1	913	0	913
Total	5558	0	5558

	Indicador de contacto efectuado por Balcão do tipo I		
Indicador de Posse de Serviço Financeiro	0	1	Total
0	3526	1119	4645
1	879	34	913
Total	4405	1153	5558

	Indicador de contacto efectuado por Balcão Tipo II		
Indicador de Posse de Serviço Financeiro	0	1	Total
0	4344	301	4645
1	884	29	913
Total	5228	330	5558

	Indicador de contacto efectuado por Mailing		
Indicador de Posse de Crédito ao Consumo	0	1	Total
0	3590	1666	5256
1	220	82	302
Total	3810	1748	5558

	Indicador de contacto efectuado por SMS		
Indicador de Posse de Crédito ao Consumo	0	1	Total
0	4034	1219	5256
1	237	65	302
Total	4274	1284	5558

	Indicador de contacto efectuado por Telemarketing		
Indicador de Posse de Crédito ao Consumo	0	1	Total
0	5256	0	5256
1	302	0	302
Total	5558	0	5558

	Indicador de contacto efectuado por Balcão do tipo I		
Indicador de Posse de Crédito ao Consumo	0	1	Total
0	4665	591	5256
1	267	35	302
Total	4932	626	558

	Indicador de contacto efectuado por Balcão Tipo II		
Indicador de Posse de Crédito ao Consumo	0	1	Total
0	5256	0	5256
1	302	0	302
Total	5558	0	5558

#### Distribuição da variável dependente do modelo de mistura de regressões

			Continuação			Continuação		
Variável dependente	Freq. Absoluta	%	Variável dependente	Freq. Absoluta	%	Variável dependente	Freq. Absoluta	%
0000000	293	5.27	1000000	2636	47.43	1011001	4	0.07
0000001	5	0.09	1000001	65	1.17	1011010	12	0.22
0000010	12	0.22	1000010	303	5.45	1011011	7	0.13
0000100	41	0.74	1000011	75	1.35	1011100	3	0.05
0001000	110	1.98	1000100	68	1.22	1011110	1	0.02
0001100	8	0.14	1000101	2	0.04	1100000	177	3.18
0010000	50	0.90	1000110	1	0.02	1100001	2	0.04
0010001	2	0.04	1000111	2	0.04	1100010	29	0.52
0010010	9	0.16	1001000	214	3.85	1100011	6	0.11
0010011	2	0.04	1001001	7	0.13	1110000	122	2.20
0010100	2	0.04	1001010	9	0.16	1110001	8	0.14
0011000	7	0.13	1001011	3	0.05	1110010	50	0.90
0011100	2	0.04	1001100	26	0.47	1110011	7	0.13
0100000	24	0.43	1010000	648	11.66	1110100	3	0.05
0100010	3	0.05	1010001	33	0.59	1111000	9	0.16
0100011	1	0.02	1010010	300	5.40	1111010	5	0.09
0110000	5	0.09	1010011	69	1.24	1111110	1	0.02
0110010	4	0.07	1010100	11	0.20	Total	5558	100
0111000	2	0.04	1010110	1	0.02			
0111001	1	0.02	1011000	37	0.67			
			1100100	4	0.07			
			1101000	12	0.22			
			1101001	1	0.02			
			1101010	1	0.02			
			1101100	1	0.02			

**Apêndice D - Modelo de mistura de regressões com concomitantes:** Critérios de informação quando considerada a amostra composta por 4.180 clientes.

**Critérios de Informação com 5 covariáveis relacionadas com as acções *bellow the line***

N.º de Segmentos	Critério de Informação		
	BIC	AIC	AIC3
1	66873	66721	66745
2	34264	33934	33986
3	25473	24966	25046
4	22337	21652	21760
<b>5</b>	<b>15569</b>	<b>14707</b>	<b>14843</b>
6	15751	14712	14876

**Estimativas das constantes de regressão para a amostra composta por 4.180 clientes**

Estimativa do vector de constantes $\beta_0$	Segmentos					Desvio Padrão
	1	2	3	4	5	
0000000	-2.13	-1.92	-1.84	-1.87	3.97	1.54
1000000	7.66	0.30	0.38	0.35	0.46	3.53
1000010	-2.10	-1.89	4.31	-1.84	-1.73	1.64
1010000	-1.33	5.41	-1.04	-1.07	-0.96	2.42
1010010	-2.11	-1.90	-1.82	4.43	-1.74	1.67

**Estatísticas de Teste Wald aplicadas às constantes de regressão**

Estatística de teste	Valor da estatística	Valor-p
Wald	5.98	1

**Estimativas dos efeitos de regressão dos indicadores de marketing**

Estimativa dos vector de efeitos $\beta$		Segmentos					Desvio Padrão
Indicadores de contacto	Variável dependente	1	2	3	4	5	
<i>Balcão Tipo I</i>							
Não Contactado	0000000	-0.16	-0.15	-0.27	-0.22	0.91	-0.09
	1000000	0.63	-0.15	-0.27	-0.22	-0.23	0.32
	1000010	-0.16	-0.15	1.09	-0.22	-0.23	-0.08
	1010000	-0.16	0.60	-0.27	-0.22	-0.23	-0.06
	1010010	-0.16	-0.15	-0.27	0.88	-0.23	-0.09
Contactado	0000000	0.16	0.15	0.27	0.22	-0.91	0.09
	1000000	-0.63	0.15	0.27	0.22	0.23	-0.32
	1000010	0.16	0.15	-1.09	0.22	0.23	0.08
	1010000	0.16	-0.60	0.27	0.22	0.23	0.06
	1010010	0.16	0.15	0.27	-0.88	0.23	0.09

<i>Telemarketing</i>							
Não contactado	0000000	-0.50	-0.64	-0.52	-0.60	2.75	-0.30
	1000000	2.00	-0.64	-0.52	-0.60	-0.69	1.03
	1000010	-0.50	-0.64	2.09	-0.60	-0.69	-0.35
	1010000	-0.50	2.56	-0.52	-0.60	-0.69	-0.05
	1010010	-0.50	-0.64	-0.52	2.39	-0.69	-0.33
Contactado	0000000	0.50	0.64	0.52	0.60	-2.75	0.30
	1000000	-2.00	0.64	0.52	0.60	0.69	-1.03
	1000010	0.50	0.64	-2.09	0.60	0.69	0.35
	1010000	0.50	-2.56	0.52	0.60	0.69	0.05
	1010010	0.50	0.64	0.52	-2.39	0.69	0.33
<i>Balcão Tipo II</i>							
Não contactado	0000000	-0.33	-0.44	-0.65	-0.49	1.72	-0.24
	1000000	1.33	-0.44	-0.65	-0.49	-0.43	0.66
	1000010	-0.33	-0.44	2.60	-0.49	-0.43	-0.16
	1010000	-0.33	1.75	-0.65	-0.49	-0.43	-0.05
	1010010	-0.33	-0.44	-0.65	1.97	-0.43	-0.21
Contactado	0000000	0.33	0.44	0.65	0.49	-1.72	0.24
	1000000	-1.33	0.44	0.65	0.49	0.43	-0.66
	1000010	0.33	0.44	-2.60	0.49	0.43	0.16
	1010000	0.33	-1.75	0.65	0.49	0.43	0.05
	1010010	0.33	0.44	0.65	-1.97	0.43	0.21
<i>Mailing</i>							
Não contactado	0000000	-0.02	-0.12	-0.08	-0.15	0.53	-0.01
	1000000	0.06	-0.12	-0.08	-0.15	-0.13	0.00
	1000010	-0.02	-0.12	0.30	-0.15	-0.13	-0.03
	1010000	-0.02	0.47	-0.08	-0.15	-0.13	0.04
	1010010	-0.02	-0.12	-0.08	0.59	-0.13	0.00
Contactado	0000000	0.02	0.12	0.08	0.15	-0.53	0.01
	1000000	-0.06	0.12	0.08	0.15	0.13	0.00
	1000010	0.02	0.12	-0.30	0.15	0.13	0.03
	1010000	0.02	-0.47	0.08	0.15	0.13	-0.04
	1010010	0.02	0.12	0.08	-0.59	0.13	0.00
<i>SMS</i>							
Não contactado	0000000	-0.14	-0.17	-0.08	-0.11	0.80	-0.07
	1000000	0.56	-0.17	-0.08	-0.11	-0.20	0.30
	1000010	-0.14	-0.17	0.32	-0.11	-0.20	-0.11
	1010000	-0.14	0.67	-0.08	-0.11	-0.20	-0.01
	1010010	-0.14	-0.17	-0.08	0.44	-0.20	-0.10
Contactado	0000000	0.14	0.17	0.08	0.11	-0.80	0.07
	1000000	-0.56	0.17	0.08	0.11	0.20	-0.30
	1000010	0.14	0.17	-0.32	0.11	0.20	0.11
	1010000	0.14	-0.67	0.08	0.11	0.20	0.01
	1010010	0.14	0.17	0.08	-0.44	0.20	0.10

**Estadísticas de Teste Wald aplicadas aos efeitos de regressão dos indicadores de marketing**

	Wald	Valor-p
Balcão Tipo I	0.55	1
Telemarketing	2.70	1
Balcão Tipo II	2.24	1
Mailing	0.11	1
SMS	0.10	1



## MODELOS DE MISTURA EM CRM

Claudia da Encarnação Calado