

SEGMENTAÇÃO COM ÁRVORES DE CLASSIFICAÇÃO

Fernando Augusto de Sá Neves dos Santos
Elizabeth Reis

RESUMEN

Estabelecer perfis ou regras de segmentação pode tornar mais compreensível os desejos e as vontades dos públicos e contribuir para decisões de gestão mais fáceis. Modelos de discriminação lógicos fornecem resultados de fácil compreensão, dos quais as árvores de decisão surgem como os mais simplistas. Com base numa classificação de comportamentos de consumo, procuramos encontrar regras de segmentação da população jovem. Constatámos que ao contrário do que seria de esperar, não foram as variáveis demográficas que melhor diferenciam os jovens, mas sim as despesas de compra em artigos específicos. Apesar de as respostas terem sido obtidas numa grande diversidade de países (Portugal, Espanha, Alemanha, Grã-Bretanha, Macau, China, Moçambique e Quênia), são os comportamentos de compra que melhor segmentam os indivíduos, demonstrando claramente a tendência para que o consumo seja cada vez mais entendido como transnacional.

PALABRAS CLAVE: Segmentação, Árvores de Classificação, Consumo.

ABSTRACT

Establishing patterns or segmentation rules can provide information about needs and wishes and contribute to easier solutions to management decision. Discriminating logical models present results more understandable. Logical discriminant models provide results of easy comprehension, and decision trees are the simplistic. Using a young people consuming behavior classification we try to obtain rules that generate segment profiles. Contrary to expected the demographic variables weren't the best descriptors of segments but the attitudes and amounts spend on consuming. Even the fact that answers are obtain in different countries (Portugal, Spain, Germany, United Kingdom, Macau, China, Mozambique and Kenya), the consuming behavior represent the best reliable segment profile of young people, showing the tendency to transnational consumer behavior.

KEY WORDS: Segmentation, Classification Trees, Consumption.

1. INTRODUÇÃO

As vontades e/ou preferências dos indivíduos relativamente ao consumo reflectem as muitas influências que os indivíduos vão adquirindo desde que nascem e resultam também de condições do momento de compra. Para

além do factor educação, formação, experiências de vida que influenciam a determinadas atitudes perante o consumo, existem outras condicionantes que fazem actuar ou condicionam a decisão de compra e que podem resultar de outros factores mais específicos e relacionados directamente com as escolhas de cada indivíduo.

processo de globalização não transforma os indivíduos em consumidores globais (Kotler e Armstrong, 1987). A globalização permite reconhecer os padrões de comportamento, as necessidades e os desejos comuns entre os indivíduos de culturas diferentes (Steenkamp e Hofstede, 2002). A existência de segmentos de consumidores que transcendem as fronteiras dos países (Yavas, Verhage e Green, 1992), obrigam as empresas a identificá-los para melhor responder às suas necessidades, adaptando-se às suas exigências, ao nível, estúdio e estilo de vida, hábitos e, mesmo, aos valores que aprovam (Roberts, 1998).

Um dos desafios chave para as empresas é lidar efectivamente com uma estrutura heterogénea de necessidades e de desejos dos consumidores, quer a nível nacional quer a nível internacional. Para alguns produtos ou indústrias as fronteiras nacionais começam a ser menos importantes, tal como o princípio de organização por actividades internacionais ou estratégias multidomésticas (Steenkamp, Hofstede, 2002). O objectivo das empresas quando estudam e pretendem efectuar uma segmentação de mercado ou de consumidores é alcançar um maior lucro, pela crença de que o tratamento de forma diferenciada de consumidores ou mercados, em vez de tratamento igual para todos, poder contribuir para essa finalidade.

2. ESTUDO

Ao nível do processo de decisão de compra existem para Lendrevie *et al.* (1996) cinco aspectos principais do comportamento de compra: as motivações de consumidores, os critérios de escolha entre marcas, o grau de envolvimento em relação ao produto, o grau de premeditação da compra e as fontes de informação dos conselhos que os consumidores recebem. Comum e acima destes aspectos existem ainda muitos dos factores “físicos” que influenciarão a decisão final de compra. Considerando estes factos, o nosso estudo vai incidir sobre factores que influenciam a compra e pretende conhecer se existe um padrão comum entre jovens de diferentes países quando assumem determinado perfil de comportamento de consumo.

Ao longo da vida os indivíduos vão recebendo diferentes influências e conhecimentos que ajudam a “moldar” as suas atitudes, preferências e valores que subsistem no momento de compra. Quando numa idade mais adulta, com uma construção mental e psicológica definida, tomam decisões em função de outros factores como a apresentação e a proximidade da loja, as características dos produtos, a imagem da marca, as imagens positivas memorizadas da publicidade, etc. Neste sentido pretendemos encontrar ou identificar regras que traduzam estas influências no padrão ou nas escolhas de consumo.

Ao proceder a uma segmentação pretende-se distinguir da melhor forma cada grupo de indivíduos ou objectos através das características que melhor os diferenciam. As características psicológicas e físicas, sexo, idade, país de origem, preferências, etc. para os indivíduos e as formas de apresentação, cor, localização, etc. para os objectos. Encontrar ou definir perfis permite caracterizar e distinguir estes indivíduos ou objectos. A segmentação deve ter por base características objectivas por forma a torná-la clara para que se possa considerar como válida e para consolidar a estrutura de segmentação obtida. Para Cardoso (2003), conseguir identificar e integrar as características de diferenciação num modelo que permita classificar novos indivíduos em cada segmento, reforça a identidade de cada segmento. Neste processo de identificação existem diferenças entre consumidores ao nível geográfico, demográfico, firmográfico, comportamento, processo de decisão, forma de

comprar, opções de compra, factores de situação ou de influência, personalidade, estilo de vida, psico-gráficos, etc. (Bock e Uncles, 2002).

Podemos recorrer a diversas metodologias para encontrar ou definir perfis ou, então, identificar apenas algumas características que permitam diferenciar os indivíduos de forma clara quanto às suas preferências ou factores que determinam a sua conduta. Classificar os indivíduos é uma forma de encontrar regras de discriminação entre eles para encontrar características comuns. As formas tradicionais de o fazer passam por efectuar a diferenciação de características demográficas, como local de origem, país, raça, sexo, etc.

Contudo, numa sociedade cada vez mais global e transcultural devemos tentar encontrar as características comuns dos indivíduos para além das suas diferenças de naturalidade. Levitt (1983), já referenciava a existência de consumidores similares que transcendem as fronteiras de cada país, as suas características e preferências são partilhadas entre jovens que habitam nos pontos mais distantes do planeta, que pertencem a diferentes países, com cor de pele muito diferente, língua e até, nalguns casos, idades muito diferentes. Aliás, entre os jovens existe um conjunto de preocupações e de modos de estar que reflectem particularidades deste segmento da população independentemente de todas as restantes variáveis demográficas ou culturais (Graham, Hamdan, 1987; Moses, 2000). Sabemos que a idade reflecte comportamentos e ideais que diferenciam os indivíduos entre si, de forma clara, pelas vontades e pelas características psicológicas que reflectem preferências e objectivos diferenciados, com conceitos de tempo, de bem-estar e de gostos diferentes de outros indivíduos com idades diferentes.

O objectivo deste trabalho consiste em identificar regras que expliquem as diferenças de comportamentos, identificar as variáveis ou factores que se relacionam com o comportamento de compra dos jovens e conhecer os factores ou variáveis que determinam as escolhas efectivas dos jovens consumidores.

Assim, na nossa procura de pontos comuns entre os diferentes jovens intentamos proceder à análise de características partilhadas por jovens universitários de diferentes países quando tomam decisões de consumo. Com base em diferentes comportamentos procedemos à análise da implicação que o local de compra, a imagem e qualidade do produto, o tipo de compras que efectuem e, ainda, as variáveis demográficas (sexo e escalão etário) e culturais (o país de origem), têm ou se relacionam com esse comportamento. Que consciência têm de factores que influenciam na tomada das suas decisões, entre a imagem que possuem do produto, o local de compra ou marca, o conceito que têm de produto e local e de outros factores que condicionam a sua acção. Considerando, também, os montantes de despesa em diferentes tipos de artigos e nos diversos locais de compra.

As variáveis relacionadas com o consumo envolvem diferentes escalões de despesa (classe 1: <12,48 Euros; classe 2: 12,48 a 37,41 Euros; classe 3: >37,41 Euros) em várias rúbricas, nomeadamente, Vestuário, Produtos de Higiene, Alimentação, Calçado, artigos de Luxo, Diversão, Livros, Alimentos de conveniência, Vestuário para criança, artigos para o Lar, Combustíveis, Acessórios para Automóvel e Outros. Relativamente aos hábitos de consumo, estes referem-se ao conhecimento sobre se é costume (sim/não) efectuar as compras nos mesmos locais e se ocorrem na sua maioria em Centros Comerciais, no Comércio Local ou em Super/Hipermercados, se compram produtos de marca e se são fiéis a marcas ou a lojas específicas. Com base numa escala de Likert de 5 níveis (1: não importante a 5: extremamente importante) solicitou-se aos jovens que apontassem o nível de importância que atribuem a cada um dos aspectos que influenciam as suas compras (Preço, Conveniência do Produto, Conveniência do Local de Compra, Qualidade do Produto, Imagem do Produto, Imagem do Local de Compra, Características do Produto, Promoção Especial, Imagem de Marca, Publicidade ou Outro).

4.METODOLOGIA

Com a complexidade de informação que se justifica reunir para melhor compreender as opções dos consumidores e com a crescente inovação ao nível das novas tecnologias, as alternativas ou métodos para efectuar e definir uma estrutura de segmentos tem-se transformado numa tarefa mais fácil, mas nem sempre mais simples. As ciências da Computação geraram, contudo, alternativas que vêm de encontro a essa grande quantidade de informação e à necessidade de procedimentos simples. A Aprendizagem Automática é uma das áreas que tem originado modelos que vêm simplificar o trabalho dos analistas de mercado. A utilização de meios automáticos ou semiautomáticos para análise de dados, *Data Mining*, permitem compreender a grande quantidade de informação que as empresas estão a compilar. Entre as vantagens destes modelos contam-se a não exigência dos pressupostos habituais nos métodos paramétricos (distribuição normal para uma variável, por exemplo), que permitem reconhecer interações entre as variáveis, os dados omissos não comprometem os resultados e a interpretação dos resultados é simples (Cardoso, 2001).

As Árvores de Classificação são modelos discriminantes lógicos que pretendem prever uma variável nominal a partir de um conjunto de variáveis explicativas. Representam-se através de uma estrutura hierárquica que se assemelha a uma árvore invertida. A árvore de classificação representa uma série de questões e a resposta a cada pergunta determina a questão que se coloca em seguida. Se todas as perguntas forem bem escolhidas obtemos uma série diminuta mas clara, que permite classificar um novo dado. Segundo Berry e Linoff (1997) as árvores de decisão vão estabelecer regras padrão que permitem classificar cada sujeito ou objecto.

Árvores de Classificação fornecem modelos discriminantes com regras simples e lógicas para a segmentação tornando mais fácil a segmentação. Estas fornecem estruturas de segmentação com base em critérios binários de diferenciação. O perfil pode iniciar-se com base em critérios demográficos e incluir posteriormente variáveis de comportamento, hábitos ou opções. Aliás, as Árvores de Classificação, decisão ou regressão são uma forma de obviar aos problemas que os dados reflectem, o número elevado de não respostas, o não exigirem a normalidade dos dados associados aos métodos paramétricos, salientarem a não linearidade ou interacção que se possa verificar entre os dados, permitirem a análise de relações estruturais entre as diversas variáveis implicadas no modelo que se pretende definir (Berry e Linoff, 1997). Apresentam, ainda, como vantagem a capacidade do tratamento adequado de atributos qualitativos, bem como a fácil interpretação das soluções. Para lidar com muitos e diversos tipos de variáveis e para ultrapassar os problemas com *missing values*. Os algoritmos que melhor geram regras discriminatórias e interpretáveis consideram-se relevantes para o processo de definição de perfis. Os mais conhecidos são CHAID (Chi-Square Automatic Interaction Detection) e o CART (Classification and Regression Trees).

As Árvores de classificação estabelecem e dividem em caixas estanques, pontos, produtos, categorias, indivíduos ou objectos similares, que obtém a mesma classificação devido às regras que a árvore estabeleceu em contraste com a classificação obtida por métodos estatísticos, linear, logística, quadrática ou de discriminação que tentam partir os dados em classes, desenhando uma curva elíptica através do espaço e que nos permite intitular essa classificação como fraca (Berry e Lynoff, 1997). Embora, para Johnson (2001), a análise de clusters permita obter resultados semelhantes ou mesmo superiores aos que resultam de uma análise por árvores de decisão.

As Árvores de Classificação são uma forma alternativa de segmentação, classificação ou regressão que estabelecem regras simples. Estas regras definem perfis que caracterizem segmentos de consumidores, indivíduos ou objectos como o tipo de influências que assistem às decisões de compra, como é o nosso caso (Curry e Moutinho, 1994). Uma Árvore de Classificação representa uma série de questões hierarquizadas sobre a

categorização de um indivíduo que permite determinar a classe ou grupo a que esse indivíduo pertence. Separam os indivíduos em grupos distintos de resposta ou tipo de compra. Cada ramo ou folha da árvore, denominado de nó, representa uma série de questões ou padrões sobre a categoria do(s) objecto(s) ou indivíduo(s). O padrão representa as regras usadas para classificar os dados.

“Automatic interaction detection” deu origem ao algoritmo CHAID e este algoritmo pretende detectar relações estatísticas entre as variáveis. Este algoritmo pára o crescimento da árvore antes de ocorrer o *overfitting*, evitando a “poda” que pode impôr outros defeitos ou erros, está restrito a variáveis categóricas. Com recurso ao software Answer Tree disponível com o SPSS e através do algoritmo Chi-Squared Automatic Interaction Detector (CHAID) procedemos a análise primária de definição de regras que caracterizem e distingam os grupos de jovens estudantes, em função do comportamento de compra em determinado local.

4. ANÁLISE DE RESULTADOS

Com base numa classificação obtida através do procedimento *k-means*, aplicado aos hábitos de compra da juventude, obtiveram-se 3 segmentos ou grupos de indivíduos que representam comportamentos diferenciados de acordo com os hábitos de consumo e de compra que as variáveis atrás enunciadas reflectem. Assim, da segmentação obtida, resulta um primeiro grupo que reúne 25,5% (566) dos jovens com o hábito de compra no comércio local, um segundo grupo com 38,9% (865) dos jovens que compram em centros comerciais e um terceiro grupo a ser representado por 35,6% (792) dos jovens que compram em super/hipermercados.

Após várias análises exploratórias, estimando diferentes modelos que explicassem a classificação obtida, através da identificação e definição das regras ou factores que influenciam esse seu comportamento, verificámos a insustentabilidade de qualquer dos modelos simulados. Este resultado vem confirmar anteriores estudos, onde os valores ou os factores que precedem comportamentos nem sempre conseguem estabelecer uma relação coerente ou consistente. Como referem Helfer e Orsoni (1987) ou, como o comprovou o estudo de La Pierre (Lima, 1993) e de Hini, Gendall e Kendall (1995), não existe coerência entre as atitudes e os comportamentos. Pelo que, os resultados e os diferentes modelos tentados não explicam de forma satisfatória a variância existente entre os diferentes grupos.

Procedemos então a uma análise das restantes variáveis que melhor explicavam essas diferenças de comportamento.

5. ÁRVORE OBTIDA

Segundo Cardoso e Moutinho (2003), as variáveis demográficas são mais fiáveis para segmentar do que as atitudes ou características psicográficas. Porém, no nosso estudo, este princípio não foi observado. O facto de se tratar de um grupo de jovens, apesar de países muito diferentes, a verdade é que as variáveis que melhor estabelecem as diferenças de comportamentos relacionam-se com o consumo de diversos artigos considerados. O vestuário, a compra de livros e os gastos com alimentos de conveniência são, de facto, as variáveis que mais estabelecem a diferenciação dos grupos e só depois surge a explicação com base na variável demográficas (sexo) e a variável cultural (país). A classificação obtida reparte de forma indiferenciada os jovens pelos 3 grupos, como vemos pelo Quadro 1.

A construção da árvore de decisão através do algoritmo CHAID, começa sem solicitar a validação do modelo. Para um primeiro teste efectuámos várias tentativas de construção da árvore de decisão com regras de paragem muito “largas”, através de níveis elevados para a árvore de decisão e números mínimos de casos por nó, donde resultaram *outputs* de grande complexidade, com difícil leitura e interpretação (Anexo 1). Procedemos posteriormente à construção da árvore de classificação com um número mínimo de 50 observações no “nó pai” ou “nó raiz” (*parent node*) e um número mínimo de 20 casos no “nó folha” ou “nó filho” (*child node*) e um máximo de 6 níveis da árvore de decisão.

Apresenta-se em seguida a caracterização discriminante dos três segmentos definidos, realizada por meio do algoritmo CHAID. Este algoritmo considera, então, determinante na explicação da variância a variável *Despesas em vestuário* como a que proporciona a agregação mais homogénea dos jovens da amostra (2223 casos), das quais resultam associações possíveis e significativas para 43 pares de categorias. O nó raiz apresenta um valor de previsão de 2.1 (média) e um desvio padrão de 0.78 pelo que, pelo risco estimado de 0,155938 o modelo explica cerca de 74% da variância, que se poderia considerar um modelo bem ajustado. Porém, com o número elevado de “nós folha” e a pequena expressão de elementos na grande maioria dos “nós folha” (inferior a 2%, Quadro 2), podemos considerar o modelo demasiado exaustivo e complexo para interpretação.

Quadro 1: Distribuição dos clusters por sexo, escalões etários e país

		Clusters						Total	
		1		2		3			
		Resp	%	Resp	%	Resp	%	Resp	%
Sexo	Masculino	245	43,4%	344	39,8%	392	49,9%	981	44,3%
	Feminino	320	56,6%	520	60,2%	394	50,1%	1234	55,7%
grupo etário (3)	18-20 anos	205	36,5%	314	36,7%	305	39,2%	824	37,5%
	21-23 anos	246	43,9%	381	44,5%	324	41,6%	951	43,3%
	mais de 23 anos	110	19,6%	161	18,8%	150	19,3%	421	19,2%
Portugal		177	31,3%	241	27,9%	82	10,4%	500	22,5%
Grã-Bretanha		71	12,5%	88	10,2%	25	3,2%	184	8,3%
Espanha		67	11,8%	184	21,3%	79	10,0%	330	14,8%
Macau		37	6,5%	84	9,7%	74	9,3%	195	8,8%
China		1	,2%	21	2,4%	175	22,1%	197	8,9%
Alemanha		48	8,5%	102	11,8%	57	7,2%	207	9,3%
Moçambique		93	16,4%	47	5,4%	69	8,7%	209	9,4%
Quénia		72	12,7%	98	11,3%	231	29,2%	401	18,0%

A opção por outro modelo menos exaustivo, mantendo um máximo de 6 níveis da árvore de decisão mas, impondo um número mínimo de 100 casos em cada nó raiz e um mínimo de 50 casos por nó folha, proporciona um risco um pouco mais elevado (19.7874), mas que explica, ainda de forma satisfatória, a variância do modelo (cerca de 67,46%) e o resultado apresenta-se como uma solução mais simples (Figura 1 no anexo 2). A consistência do modelo também se verifica no facto de as variáveis que definem a melhor agregação ou mais homogénea não se alterarem.

NEW TRENDS IN MARKETING MANAGEMENT

Quadro 2: Gain Summary

					Target variable: Ward Met 3									
					Statistics									
Node	Node: n	Node: %	Gain	Index (%)	Node	Node: n	Node: %	Gain	Index (%)	Node	Node: n	Node: %	Gain	Index (%)
3	416	18,71	3,00	142,74	28	26	1,17	2,12	100,65	17	38	1,71	1,58	75,13
14	30	1,35	3,00	142,74	5	23	1,03	2,00	95,16	31	39	1,75	1,54	73,20
9	44	1,98	2,98	141,66	42	117	5,26	2,00	95,16	33	20	0,90	1,45	68,99
7	24	1,08	2,96	140,76	32	55	2,47	1,98	94,30	56	32	1,44	1,44	68,40
29	23	1,03	2,87	136,54	24	85	3,82	1,96	93,48	65	40	1,80	1,33	63,05
18	64	2,88	2,83	134,57	37	34	1,53	1,94	92,36	40	36	1,62	1,31	62,12
15	44	1,98	2,73	129,77	43	45	2,02	1,91	90,93	58	40	1,80	1,25	59,48
8	39	1,75	2,69	128,10	57	55	2,47	1,85	88,24	46	48	2,16	1,21	57,49
27	55	2,47	2,62	124,58	47	48	2,16	1,81	86,24	61	20	0,90	1,15	54,72
51	38	1,71	2,42	115,20	71	53	2,38	1,79	85,29	68	53	2,38	1,11	52,97
70	41	1,84	2,37	112,57	16	32	1,44	1,78	84,75	67	73	3,28	1,01	48,23
54	32	1,44	2,19	104,08	39	53	2,38	1,74	82,59	62	61	2,74	1,00	47,58
12	36	1,62	2,17	103,09	53	40	1,80	1,70	80,89	66	27	1,21	1,00	47,58
23	39	1,75	2,15	102,48	45	41	1,84	1,63	77,75					
34	28	1,26	2,14	101,96	64	36	1,62	1,58	75,34					

Resultou destes pressupostos uma árvore de decisão com variáveis simples e claras que definem e agregam os casos da forma mais homogênea. As principais variáveis a proporcionarem a agregação mais homogênea dos dados passam então a ser os escalões de despesas dos jovens *Despesas em vestuário*, *Despesas em livros* e a *Despesas em alimentos de conveniência*, mantendo-se assim as mesmas variáveis resultantes do primeiro

modelo, com uma percentagem de variância explicada, também, satisfatória e com um ganho substantivo em simplicidade para a leitura do *output*.

Para assumir um modelo para uma população em geral, com base nas características principais de uma amostra é necessária a sua replicação noutras amostras e verificar se esse modelo continua a classificar bem essas amostras, pelo que devemos considerar estes pressupostos para validação do último modelo resultante. Procedi ao processo de validação através do um método validação *V-Fold* que particiona a amostra original em *V* sub-amostras de dimensões aproximadas e com uma distribuição semelhante à variável dependente, apesar de menos frequente para o algoritmo CHAID. Mas, segundo McDaniel e Grubb (2002) devemos validar a árvore de decisão através do método *Ten-Fold* quando temos poucos dados, como é o nosso caso.

Solicitámos então a construção da árvore, mas impondo a validação do modelo através do método de validação cruzada (10-Fold). Ao utilizar este método para a validação do nosso modelo verificámos que as regras obtidas nas anteriores simulações sem validação se mantêm e que o risco (média dos riscos) associado à validação cruzada, de 23.0487, é próximo do valor obtido sem validação, pelo que, poderíamos concluir que o modelo resultante se ajustaria de forma conveniente aos dados e também, pelo facto de explicar a variância em cerca de 62,10%.

Resultou assim, uma árvore de regressão ainda menos complexa, com apenas 25 “nós folha” terminais, com uma representatividade mais expressiva em cada “nó folha” e com riscos similares ao modelo antes da validação. Este explica cerca de 67% da variância total, com um nível de risco aceitável. Os valores de risco muito próximos entre as amostras de validação (22.8557) e a amostra de teste (19.8233) permitem concluir pelo bom ajustamento do modelo, com as regras daí resultantes.

Considerando ainda a relevância das regras que nos é proporcionada pela leitura dos índices resultantes do sumário de ganhos verificamos que apenas nos restam 7 “nó folha” para interpretar (Quadro 3).

Quadro 2: Gain Summary

(Node	Node: n	Node: %	Gain	Index (%)
3	416	18,71	3,00	142,74
6	50	2,25	2,90	137,99
10	64	2,88	2,83	134,57
5	80	3,60	2,60	123,71
15	104	4,68	2,55	121,24
8	116	5,22	2,53	120,59
33	52	2,34	2,29	108,89

6. CONSIDERAÇÕES FINAIS

Aa pré-classificação obtida através dos padrões de consumo estabelecidos com base em despesas de consumo e hábitos ou preferências de locais de compra obtiveram-se algumas regras quanto ao tipo de despesas desta juventude que se diferenciam sobretudo não pela sua origem ou país, não pelo seu sexo, mas através de uma característica que passa por gastar quantias mínimas em livros (cross-validação, com baixo risco).

O modelo obtido permite estabelecer relações simples que resultam, sobretudo, dos comportamentos dos jovens relativamente às despesas, ou seja, são sobretudo os montante de despesas envolvidas na compra de vestuário, livros e alimentos de conveniência que explicam o perfil dos jovens que efectuam compras nos diversos tipos de estabelecimentos comerciais. Relativamente às variáveis demográficas, estas não reflectem ou justificam os diferentes comportamentos, demonstrando uma clara tendência de uniformização de comportamentos do segmento da juventude.

Neste sentido o mais importante deste estudo não são as regras que resultam do modelo, mas sim o facto de as diferenças de comportamento se explicarem não pelas características demográficas e culturais, mas assentarem nas diferenças de níveis de consumo ou padrões de consumo da juventude, embora possam conter alguma explicação pela diferença de preço.

BIBLIOGRAFIA:

- BERRY, MICHAEL; LINOFF, GORDON (1997); *Data Mining Techniques – for Marketing, Sales and Customer Support*; John Wiley & Sons, Inc.; New York.
- BOCK, TIMOTHY; UNCLES, MARK (2002); “A Taxonomy of Differences Between Consumers for Market Segmentation”; *International Journal of Research in Marketing*; vol.19; pp.215-224.
- CARDOS, MARGARIDA (2003); “Modelos Discriminantes Lógicos na Caracterização de uma Estrutura de Segmentos” in *Temas em Métodos Quantitativos 3*; eds. Elizabeth Reis e Manuela Hill; Sílabo; Lisboa.
- CARDOSO, MARGARIDA (2001); “Modelos de Segmentos Latentes: Aplicações em Marketing” in *Métodos Quantitativos 2* ed. Manuel Alberto Ferreira, Rui Menezes e Margarida Cardoso; Ed. Sílabo, Lisboa.
- CARDOSO, MARGARIDA; MOUTINHO, LUIZ (2003); “A Logical Type Discriminant Model for Profiling a Segment Structure”; *Journal of Targeting, Measurement and Analysis for Marketing*; vol.12; Iss.1; Sep.; 27-41.
- CURRY, BRUCE; MOUTINHO, LUIZ (1994); “Intelligent Computer Models for Marketing Decisions”; *Management Decision*; vol.32; Iss.4; pp.30-35.
- GRAHAM, LAWRENCE; HAMDAN, LAWRENCE (1987); *Youthrends – Capturing the \$200 Billion Youth Market*; St Martin’s Press; New York.
- HELPER, J.P.; ORSONI, J. (1996); *Marketing*; Edições Sílabo; Lisboa.
- HINI, DEAN; GENDALL, PHILIP; KEARNS, ZANE (1995); “The Link Between Environmental Attitudes and Behaviour”; *Marketing Bulletin*; vol.6; pp.22-31.
- JOHNSON, ROBERT E. (2001); “The Role of Cluster Analysis in Assessing Comparability Under the U.S. Transfer Pricing Regulations”; *Business Economics*; vol.36; Iss.2; Apr.; pp.30-38.
- KOTLER, PHILIP; ARMSTRONG, GARY (1987); *Marketing – An Introduction*; 3ª edição; Prentice-Hall International; New Jersey.
- LENDREVIE, J. et al. (1997); *Mercator: teoria e prática do Marketing*; 6ª ed.; Pub. Dom Quixote; Lisboa.
- LEVITT, THEODORE (1983); *Le marketing Imagination*; Free Press, New York.
- LIMA, MARIA L. (1993); “Atitudes”; in *Psicologia Social*; eds. Jorge Vala e Maria B. Monteiro; Fundação Calouste Gulbenkian; pp.167-199.
- MCDANIEL, MICHAEL; GRUB III, LEE (2002); “CART 4.0”; *Organizational Research Methods*; vol.5; Iss.1; Jan.; 129-132.
- MOSES, ELISSA (2000); *The \$100 Billion Allowance: Accessing the Global Teen Market*; John Wiley & Sons; New York.
- ROBERTS, SHIRLEY (1998); “Harness the Future: the 9 Keys to Emerging Consumer Behaviour”; *The Canadian Manager*; vol.23; Iss.3; Fall; pp.11-13.
- STEENKAMP, JAN-BENEDICT; HOFSTEDE, FRENKEL (2002); “International Market Segmentation: Issues and Perspectives”; *International Journal of Research in Marketing*; vol.19; pp.185-213.
- YAVAS, UGUR; VERHAGE, BRONISLAW; GREEN, ROBERT (1992); “Global Consumer Segmentation versus Local Market Orientation: Empirical Findings”; *Management International Review*; vol.32; Iss.3; Third Quarter; Feb.; pp.265-272.

Anexo 1

Project File

Number of Cases

	Weighted	Unweighted
Cases	2223,00	2223,00

Partition Information

Partition	Off
-----------	-----

Cross Validation Information

Cross-Validation	Off
------------------	-----

Tree Growing Criteria

Growing Method	CHAID
----------------	-------

Algorithm Specifications:

Alpha for splitting:	0,05
----------------------	------

Alpha for merging:	0,05
--------------------	------

Chi-squared statistic:	Likelihood Ratio
------------------------	------------------

Allow splitting of merged criteria:	Off
-------------------------------------	-----

Use of Bonferroni adjustment:	On
-------------------------------	----

Stopping Rules

Maximum tree depth:	6
---------------------	---

Minimum no. of cases for parent node:	50
---------------------------------------	----

Minimum no. of cases for child nodes:	20
---------------------------------------	----

Model

Target Variable

Name	CLU3_1
------	--------

Label	Ward Met 3
-------	------------

Type	Numeric
------	---------

Measurement Level	Continuous
-------------------	------------

Predictors

Name	Type	Level	Label
------	------	-------	-------

COUNTRY	NumericOrdinal		COUNTRY
---------	----------------	--	---------

Q1	NumericOrdinal		Sexo
----	----------------	--	------

Q13A	NumericOrdinal		Compra nos mesmos locais?
------	----------------	--	---------------------------

Q13B_B	NumericOrdinal		Compra em Centros Comerciais?
--------	----------------	--	-------------------------------

Q13B_C	NumericOrdinal		Compra em Super/Hipermercados?
--------	----------------	--	--------------------------------

Q13B_D	NumericOrdinal		Compra no Comércio Local?
--------	----------------	--	---------------------------

Q13B_E	NumericOrdinal		Outro local?
--------	----------------	--	--------------

Q2A	NumericOrdinal		grupo etário (3 escalões)
-----	----------------	--	---------------------------

Q10A3	NumericOrdinal		Despesas Vestuário (3 escalões)
-------	----------------	--	---------------------------------

Q10B3	NumericOrdinal		Despesas Produtos de Higiene (3 escalões)
-------	----------------	--	---

Q10C3	NumericOrdinal		Despesas Alimentação (3 escalões)
-------	----------------	--	-----------------------------------

Q10D3	NumericOrdinal		Despesas Calçado (3 escalões)
-------	----------------	--	-------------------------------

Q10E3	NumericOrdinal		Despesas Artigos de Luxo (3 escalões)
-------	----------------	--	---------------------------------------

Q10F3	NumericOrdinal		Despesas Diversão (3 escalões)
-------	----------------	--	--------------------------------

Q10G3	NumericOrdinal		Despesas Livros (3 escalões)
-------	----------------	--	------------------------------

Q10H3	NumericOrdinal		Despesas Alimentos de Conveniência (3 escalões)
-------	----------------	--	---

Q10I3	NumericOrdinal		Despesas Roupas criança (3 escalões)
-------	----------------	--	--------------------------------------

Q10J3	NumericOrdinal	Despesas Artigos Lar (3 escalões)
Q10K3	NumericOrdinal	Despesas Combustíveis (3 escalões)
Q10L3	NumericOrdinal	Despesas Acessórios de Automóveis (3 escalões)
Resulting Tree		
Size		
Total number of nodes	72	
Total number of levels	6	
Total number of terminal nodes	43	
Resubstitution		
Risk Estimate	0,155938	
SE of Risk Estimate	0,0069706	

Anexo 3

/* Node 3*/

SELECT * FROM <TABLE>

Quando as Despesas de vestuário são do valor do 1º escalão, as despesas em livros são =1 e as despesas em Alimentos de conveniência =1;

/* Node 6*/

SELECT * FROM <TABLE>

Quando as Despesas de vestuário é =1, as despesas em livros são =1, as despesas em Alimentos de conveniência >1 e não respondem a despesas em artigos para o lar;

/* Node 10*/

Quando as Despesas de vestuário é =1 e não respondem às despesas em livros;

/* Node 5*/

Quando as Despesas de vestuário é =1, as despesas em livros são =1, as despesas em Alimentos de conveniência >1 e registam despesas em artigos para o lar;

/* Node 15*/

Quando as Despesas de vestuário são do 2º escalão, as despesas em livros são do 1º escalão, as despesas em Alimentos de conveniência >1 e não compra ou não responde à compra em centros comerciais;

/* Node 8*/

Quando as Despesas de vestuário são do valor do 1º escalão, as despesas em livros são >1 e as despesas em Alimentos de conveniência =1;

/* Node 33*/

Quando as Despesas de vestuário são elevadas (3º escalão), as despesas em Alimentos de conveniência do 1º escalão, as despesas em livros são =1 então não faz as compras no centro comercial;

CITIES IN COMPETITION

Anexo 2

Figura 1: A árvore de Decisão

