



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

ERIVALDO LOPES DE SOUZA

**MODELOS ADITIVOS GENERALIZADOS PARA A AVALIAÇÃO DA INTENÇÃO
DE COMPRA DE CONSUMIDORES**

JOÃO PESSOA - PB

2012

ERIVALDO LOPES DE SOUZA

**MODELOS ADITIVOS GENERALIZADOS PARA A AVALIAÇÃO DA INTENÇÃO
DE COMPRA DE CONSUMIDORES**

Dissertação apresentado ao Programa de Pós-Graduação em Engenharia de Produção do Centro de Tecnologia da Universidade Federal da Paraíba (UFPB), em cumprimento às exigências para obtenção do grau de Mestre em Engenharia de Produção.

Área de Concentração: Gerência da produção de bens e serviços

Orientador: Prof. Dr. Luiz Bueno da Silva

JOÃO PESSOA

2012

ERIVALDO LOPES DE SOUZA

**MODELOS ADITIVOS GENERALIZADOS PARA AVALIAÇÃO DA INTENÇÃO DE
COMPRA DE CONSUMIDORES**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal da Paraíba (UFPB), em cumprimento às exigências para obtenção do grau de Mestre em Engenharia de Produção. Área de concentração: Gerência da produção de bens e serviços.

Aprovada em: ____/____/____

BANCA EXAMINADORA

Prof. Dr. Luiz Bueno da Silva
Universidade Federal da Paraíba - UFPB
Orientador

Prof. Dr. Márcio Botelho da Fonseca Lima
Universidade Federal da Paraíba - UFPB
Examinador interno

Prof. Dr. Eufrásio de Andrade Lima Neto
Universidade Federal da Paraíba - UFPB
Examinador externo

Prof. Dr. Jonas Alves de Paiva
Universidade Federal da Paraíba - UFPB
Examinador externo

Ao meu pai, Aprício Alves (in memoriam).

AGRADECIMENTOS

A Deus, pela oportunidade concedida de conclusão deste curso.

À minha mãe, Rita, pelo incentivo aos meus estudos, por todo o esforço e dedicação.

Aos meus irmãos, João Paulo e Ana Paula, pela paciência e companheirismo.

À minha namorada, Sawana, pela paciência e compreensão nos momentos que precisei me ausentar para elaboração e conclusão deste trabalho, bem como pelas palavras de apoio.

Às correções e à paciência do meu orientador, Prof. Luiz Bueno, que foram fundamentais para a conclusão deste trabalho.

A todos os professores do Programa de Pós-Graduação em Engenharia de Produção da UFPB, pelos ensinamentos, e em especial, ao Prof. Eufrásio Lima Neto, pelos esclarecimentos e relevantes contribuições a este trabalho.

Ao Prof. Márcio Botelho, pelo apoio no momento oportuno.

A todos os amigos (as) conquistados ao longo do curso e, em especial, a Anderson, Almeida, Clarissa, Fábio, Juliana, Manoel, Matheus, Nádia, Priscila, Roberta Guedes, Roberta de Lourdes, Geraldo e Tarso, pelos conhecimentos compartilhados.

Àos funcionários do PPGEF, pelos serviços prestados.

A todos que, ao responderem prontamente os questionários, contribuíram com esta Dissertação.

RESUMO

Nos últimos anos, vários estudos foram publicados versando sobre fatores que influenciam a intenção de compra do consumidor em diversos setores econômicos. Nesta linha de trabalho, procurou-se, especificamente para o setor de compra coletivas, obter modelos de regressão que pudessem contribuir para o estudo da relação entre a intenção de compra e as características de segmentos de mercado. Visa-se com isso auxiliar na inclusão da variável intenção de compras no processo de escolha de um público-alvo, orientando decisões para satisfazer com maior eficiência consumidores do serviço. Para alcançar o objetivo, entrevistaram-se inicialmente 384 usuários de Internet da cidade de João Pessoa, Paraíba, Brasil. Em seguida os dados obtidos a partir de entrevistas, foram usados para estimar aqueles modelos. Esses modelos foram baseados em pressupostos de teorias da abordagem cognitiva do comportamento do consumidor, especialmente da Teoria da Ação Racional. O instrumento usado para a coleta de dados foi um questionário de pesquisa de mercado contendo questões ligadas a fatores psicológicos, sócio-culturais e situacionais do consumidor. O modelo mais bem sucedido foi um modelo aditivo generalizado com nove variáveis e com um termo não paramétrico, obtido a partir do método de suavização splines. Esse modelo apresentou um *pseudo-R²* igual a 0,89 e possibilitou alcançar um percentual de acertos nos julgamentos das observações da amostra igual a 94%. Com o auxílio de simulações, verificou-se de que modo o tipo de modelo proposto é capaz de auxiliar na escolha de um público-alvo com maior interesse no uso do serviço. Apresentou-se ainda a maneira pela qual o modelo pode ser usado para avaliar sistemas produtivos, em relação ao atendimento mais eficiente de clientes que têm a intenção de utilizar o serviço. Os modelos aditivos generalizados mostraram-se eficientes para identificar a presença de relações não lineares e foram capazes de gerar um poder explicativo alto da propensão de indivíduos para utilizar um serviço específico.

Palavras-chave: Comportamento do consumidor. Intenção de compra. Modelos Aditivos Generalizados. Segmentação de mercado.

ABSTRACT

In recent years, several studies have been published dealing on factors that influence consumer purchase intent in various economic sectors. In this line of work, we are specifically for the sector of collective buying, obtain regression models that could contribute to the study of the relationship between the purchase intention and the characteristics of market segments. The aim is to assist in the inclusion of this variable purchase intention in the process of choosing a target group, guiding decisions to meet more effectively service consumers. To achieve this goal, initially interviewed 384 Internet users in the city of João Pessoa, Paraíba, Brazil. Then the data obtained from interviews, were used to estimate those models. These models were based on assumptions of theories of the cognitive approach of consumer behavior, especially the Theory of Reasoned Action. The instrument used for data collection was a questionnaire containing market research questions related to psychological factors, socio-cultural and situational consumer. The most successful model was a generalized additive model with nine variables and nonparametric one end, obtained from smoothing splines. This model had a *pseudo-R*² of 0,89 and allowed to reach a percentage of correct trials of the observations of the sample equal to 94%. With the aid of simulations, it was observed how the proposed model type is capable of assisting in selection of a target with a higher interest in the use of the service. It was also shown how the model can be used to evaluate production systems, in relation to more efficient service to customers intend to use the service. The generalized additive models were effective for identifying the presence of nonlinear relationships and were able to generate a high explanatory power of the propensity of individuals to use specific service.

Keywords: Consumer behavior. Purchase intent. Generalized Additive Models. Market segmentation.

LISTA DE FIGURAS

Figura 1 - TAR aplicada ao comportamento de compra	29
Figura 2 - Fatores determinantes das normas subjetivas e atitudes relativas ao comportamento de compra	29
Figura 3 - Fatores determinantes da intenção e decisão de compra	30
Figura 4 - Curva do efeito da renda por dependente no logaritmo da chance do modelo pelo método de suavização polinomial local (loess)	92
Figura 5 - Curva do efeito da renda por dependente no logaritmo da chance do modelo pelo método de suavização splines.....	92
Figura 6 - Curva ROC para MLG com variável resposta binária.....	96
Figura 7 - Curva ROC para MAG com variável resposta binária	97
Figura 8 - Julgamentos do modelo com pontos de corte 0,65 e 0,35 para o público de pessoas indecisas quanto ao uso do serviço.....	101
Figura 9 - Julgamentos do modelo com pontos de corte 0,65 e 0,35 para o público de não compradores quanto ao uso do serviço.....	102
Figura 10 - Julgamentos do modelo com pontos de corte 0,65 e 0,35 para o público de pessoas indecisas quanto ao uso do serviço.....	102
Figura 11- Gráfico das probabilidades dos perfis de um segmento pelo MLG multinomial para a especificação (1) do serviço.....	123
Figura 12 - Gráfico das probabilidades dos perfis de um segmento pelo MAG para especificação (1).....	124
Figura 13 - Gráfico das probabilidades dos perfis de um segmento pelo MLG multinomial para a especificação (2) do serviço.....	129
Figura 14 - Gráfico das probabilidades dos perfis de um segmento pelo MAG para a especificação (2) do serviço	130
Figura 15 - Intenção dos consumidores de comprar em sites de compra coletiva	152
Figura 16 - Gráfico de caixas das variáveis contínuas	153
Figura 17 - Histogramas das variáveis contínuas	155
Figura 18 - Gráficos de dispersão dos preditores	156
Figura 19 - Gráficos de setores para variáveis categóricas da pesquisa.....	157
Figura 20 - Gráfico de setores do nível de importância atribuído a fatores relacionados ao serviço de compras coletivas	159
Figura 21 - Gráfico de setores do nível de importância atribuído a fatores relacionados ao serviço de compras coletivas	160
Figura 22 - Gráficos de setores para as formas pelas quais o entrevistado já tomou conhecimento de ofertas de sites de compra coletiva.....	162
Figura 23 - Gráficos de setores para os tipos de produtos e serviços que os entrevistados consomem com regularidade	163
Figura 24 - Fatores que influenciam na intenção de compra, variáveis da pesquisa e os itens do questionário	168

LISTA DE QUADROS

Quadro 1 - Objetivo das seções na fundamentação teórica	20
Quadro 2 – Matriz de confusão para classificações em duas categorias	54
Quadro 3 - Variáveis selecionadas para o estudo	77
Quadro 4 - Variáveis obtidas por meio de mudanças e combinações das variáveis do questionário	81
Quadro 5 - Ligação entre as variáveis de decisão do problema e as variáveis do modelo.....	111
Quadro 6 - Caracterização dos segmentos de mercado	117
Quadro 7 - Caracterização das alternativas de especificações para um serviço de compras coletivas.....	118
Quadro 8 - Matriz de julgamentos das combinações entre especificações do serviço e segmentos de mercado usando o MAG	131
Quadro 9 - Matriz de julgamentos das combinações entre especificações do serviço e segmentos de mercado usando os dois modelos.....	132
Quadro 10 - Matriz de correlações das variáveis contínuas	147
Quadro 11 - Distribuição de Frequência amostral em relação aos bairros dos entrevistados	148
Quadro 12 - Tabela de contingência das variáveis Sexo, Renda e Idade	150
Quadro 13 - Tabela de contingência para as variáveis Idade e Sexo	150
Quadro 14 – Tabela de contingência entre as variáveis Idade e Renda	151
Quadro 15 - Tabela de contingência entre as variáveis Renda e Sexo	151

LISTA DE TABELAS

Tabela 1 - Valores da <i>odds ratio</i> para o modelo multinomial ordinal.....	83
Tabela 2 - Valores da odds ratio para o MLG binomial.....	87
Tabela 3 - Valores dos coeficientes e da odds ratio para cada uma das variáveis dos MAG's	94
Tabela 4 - Indicadores da capacidade de previsão do MLG binomial para diferentes pontos de corte	98
Tabela 5 - Matriz de confusão para o MLG binomial com ponto de corte igual a 0,9339 para classificar um indivíduo como comprador	98
Tabela 6 - Indicadores da capacidade de previsão do MAG binomial para diferentes pontos de corte	98
Tabela 7 - Matriz de confusão para o MAG binomial com ponto de corte igual a 0,8848 para classificar um indivíduo como comprador	99
Tabela 8 - Indicadores da capacidade de previsão do MLG multinomial ordinal para diferentes pontos de corte.....	100
Tabela 9 - Matriz de confusão para o modelo com pontos de corte 0,15 e 0,85 para o julgamento respectivamente de indecisos e compradores	100
Tabela 10 - Matriz de confusão para o modelo com pontos de corte 0,35 e 0,65 para o julgamento respectivamente de indecisos e compradores	101
Tabela 11 - Valores do p-valor para cada variável do MLG multinomial	103
Tabela 12 - Valores do p-valor de cada uma das variáveis do modelo	104
Tabela 13 - Valores do p-valor para cada variável dos MAG's com termos não paramétricos obtidos por dois métodos de suavização distintos	105
Tabela 14 - Medidas diagnósticas dos modelos	106
Tabela 15 - Valores mais favoráveis e mais desfavoráveis ao julgamento de um perfil como comprador.....	115
Tabela 16 - Probabilidades e julgamentos dos segmentos de mercado de acordo com os modelos para indivíduos menos interessados no serviço	120
Tabela 17 - Probabilidades e julgamentos dos segmentos de mercado de acordo com os modelos para indivíduos menos interessados no serviço	121
Tabela 18 - Probabilidades e julgamentos das especificações do serviço de acordo com os modelos para os perfis de indivíduos menos interessados em usar o serviço	126
Tabela 19 - Probabilidades e julgamentos das especificações do serviço de acordo com os modelos, para os perfis de indivíduos mais interessados em usar o serviço	127
Tabela 20 - Medidas descritivas das variáveis contínuas da amostra.....	146
Tabela 21 - Atitudes dos entrevistados em relação a características do serviço de compras coletivas.....	149

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO	11
1.1 Definição do tema e problema de pesquisa	11
1.2 Justificativa	15
1.3 Objetivos.....	17
CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA	19
2.1. Segmentação de mercado e escolha do público-alvo	20
2.2. Comportamento do consumidor e a decisão de compra	22
2.2.1. Abordagens teóricas do comportamento do consumidor	24
2.2.2. Fatores que influenciam na decisão de compra	26
2.2.3. Intenção e decisão de compra	28
2.4. Modelos paramétricos de regressão	39
2.4.1. Modelo linear clássico	40
2.4.2. Modelos lineares generalizados	44
2.5. Modelos não paramétricos e semiparamétricos de regressão	56
2.5.1. Modelos não paramétricos <i>scatterplot</i>	57
2.5.2. Modelos aditivos generalizados	66
CAPÍTULO 3 - METODOLOGIA	72
3.1. Classificação da pesquisa	72
3.2. Estudo de caso.....	72
3.2.1. O modelo de avaliação da intenção de compras e a fonte de informação	73
3.2.2. O sistema de produção do serviço de compras coletivas pela internet.....	73
3.3. Definição da amostra e das variáveis de pesquisa	74
CAPÍTULO 4 - RESULTADOS.....	80
4.1. Especificação e conclusões dos modelos	80
4.1.1. Modelo com variável resposta ordinal.....	82
4.1.2. Modelos com variável resposta binária	86
4.2. Diagnóstico dos modelos	95
4.3. Simulação da aplicação dos modelos	106
4.4.1. Classificações de segmentos e especificações do serviço pelos modelos	107
4.4.2. Descrição do exemplo	114
4.4.3. Avaliação as escolhas de públicos-alvo.....	119
CAPÍTULO 5 - CONSIDERAÇÕES FINAIS	134
REFERÊNCIAS	137
APÊNDICE A - Questionário.....	144
APÊNDICE B - Descrição da amostra	146
APÊNDICE C - Rotinas usadas para obter e analisar os modelos	165
APÊNDICE D - Fatores determinantes da intenção no questionário	168

CAPÍTULO 1 - INTRODUÇÃO

1.1 Definição do tema e problema de pesquisa

É importante salientar que nas atividades de projeto e de gestão do produto, do processo de produção e da distribuição dos produtos sejam levadas em consideração as opiniões, os anseios e as necessidades dos consumidores, ou seja, é necessário que haja mais conhecimentos acerca do consumidor que se deseja atender e que esses conhecimentos sejam levados em consideração nas atividades de projeto e gestão do sistema de produção. Saber qual grupo de consumidores serão atendidos pela organização é uma primeira etapa para buscar conhecê-los. Logo, faz-se necessário saber com clareza, em um primeiro estágio do planejamento de um sistema de produção, quais clientes são o “alvo” da organização.

A segmentação do mercado consumidor é uma primeira etapa da definição dos clientes que mais interessam à organização. Kotler e Armstrong (2003) afirmam que algumas empresas estão em melhores condições de atender a determinados grupos de consumidores. Nesse caso, é necessário dividir o mercado, escolher os melhores segmentos e procurar estratégias para atender aos segmentos específicos de maneira mais eficiente que os concorrentes.

Souza e Cardozo (2009) complementam tal argumentação ao especificarem que a dificuldade de atender bem todo o mercado reside na existência de grande heterogeneidade do mesmo com relação às características dos clientes que o compõe. Esses autores reforçam essa opinião ao afirmarem que as organizações devem identificar grupos de consumidores (segmentos de mercado) com características semelhantes para que possam concentrar seus esforços de marketing e atendê-los de maneira mais eficaz.

Assim, de acordo com as declarações dos autores supramencionados, a segmentação de mercado pode ser considerada como um procedimento utilizado para dividir um mercado heterogêneo (em relação a determinadas características dos consumidores) em grupos homogêneos. Esses grupos são chamados de segmentos de mercado.

Ao escolher o segmento que se deseja atender, é esperado que se evite escolher de consumidores que apresentem uma menor pretensão de comprar um produto, concentrando uma maior atenção sobre aqueles indivíduos que possuam um perfil mais favorável à compra, ou seja, que apresentem uma maior intenção de realizar a compra. Isso é esperado porque é reconhecido que a escolha de clientes com um perfil bastante averso à compra de um tipo de

produto específico pode requerer um grande esforço para efetuarla, ou seja, sabe-se que a intenção de compra exerce alguma influência sobre as vendas dos produtos.

Bendixen, Bukasa e Abratt (2004) reconhecem a importância de conhecer a intenção de compra de clientes ao destacar que a mesma está associada à probabilidade do indivíduo ficar mais tempo com a empresa e com a sensibilidade em relação às mudanças no preço. Para esses autores, os consumidores que apresentam uma maior intenção de comprar um determinado produto têm uma maior probabilidade de permanecer comprando-o durante mais tempo e são menos sensíveis a variações no preço.

Então, levar em conta a intenção de compra, ao escolher segmentos de mercado a serem atendidos, não é apenas importante porque tal intenção tem alguma influência sobre as vendas, mas também devido ao fato de que essa intenção também influencia na fidelidade do consumidor e na sua sensibilidade em relação ao preço.

Zeithaml, Berry e Parasuraman (1996) definem a intenção de compra como um sinal de escolha real relativo à compra. Kie e Xie (2009) utilizam a definição daqueles autores como referência para tratar da intenção comportamental de consumidores, o que indica que é um conceito ainda usado recentemente.

Segundo Bagozzi, Yi e Baumgartner (1990), uma intenção qualquer desempenha um papel na relação entre um comportamento e uma atitude. Ramayah, Lee e Mohamad (2010) ressaltam essa observação ao tratar da intenção de compras de produtos ambientalmente corretos, acrescentando que a mesma é uma função do nível de esforço requerido para executar uma determinada ação decorrente do comportamento, como, por exemplo, comprar um produto.

Fishbein e Ajzen (1975) definem que uma intenção é a probabilidade de um comportamento específico vir a ser realizado e propõem a Teoria da Ação Racional que trata a intenção como antecedente imediato de um comportamento real. Ramayah, Lee e Mohamad (2010) tomam essa definição como referência em sua abordagem sobre intenção de consumo. Quanto ao comportamento de compra define-se, então, a intenção de compra como sendo a probabilidade de um produto vir a ser adquirido.

Para Xu, Summers e Belleau (2004) consideram que a intenção comportamental dos consumidores, tal como a intenção de compra, é algo importante no desenvolvimento de uma estratégia de marketing para qualquer tipo de produto ou serviço, mesmo aqueles incomuns. Lin (2007) destaca essa importância para o caso específico de websites comerciais, afirmando que a proliferação desses tipos de Websites tem aumentado a relevância de entender os determinantes da intenção de compras em lojas online. Vijayarathy (2004) destaca que nem

sempre é fácil ou prático obter uma medida ideal do grau de engajamento em um comportamento. Para ele, há suporte teórico e empírico de que há uma forte correlação entre a intenção e o comportamento real, de modo que o primeiro pode ser considerado um bom indicador do segundo. Canniere, Pelsmarcker e Geuens (2009) reforçam essa observação ao afirmarem que muitos estudos têm sido feitos para explicar as intenções. Esses estudos geralmente assumem que as intenções são bons preditores de comportamentos relacionados.

A partir dos conceitos supracitados pode-se afirmar que a intenção de compra pode ser definida como uma disposição apresentada para adquirir um produto ou serviço específico. Essa disposição é um sinal de escolha real e pode ser expressa em termos de probabilidade de um produto vir a ser comprado. Tem-se como pressuposto que, quanto maior for o sinal real de escolha (a disposição apresentada), maior será a probabilidade de adquirir o produto e menor será o esforço necessário por parte da empresa/organização para que seja efetuada a compra.

Considerar a intenção de compra na escolha dos clientes-alvo, classificando-os quanto a esse critério, é muitas vezes fácil e pode ser realizada de forma intuitiva, observando o perfil do cliente e o seu interesse para produtos semelhantes. No entanto, há situações onde essa definição não é tão clara. Quando isso ocorre, surge a necessidade de modelos que forneçam uma medida da intenção de compra para auxiliar na classificação dos clientes quanto àquela.

Uma forma de considerar a intenção de compra, na tomada de decisão acerca dos segmentos de mercado a serem escolhidos, é relacioná-la a características desses grupos de clientes. Considerando tais características como variáveis, o problema de obter uma medida de intenção de compra para segmentos passa a ser um estudo da natureza da relação entre uma única variável dependente e um conjunto de outras variáveis independentes. Sempre que se procura estudar a natureza de uma relação desse tipo, os modelos de regressão apresentam-se como uma alternativa a ser considerada.

Os modelos de regressão são muito utilizados no estudo de relações entre uma variável específica, dita dependente, e um conjunto de variáveis associadas, chamadas de independentes.

Johnson e Wichern (1992) compartilham essa ideia ao analisar que a regressão é uma metodologia estatística usada para estudar relações e fazer a predição de valores de uma ou mais variáveis respostas (dependente) a partir de uma coleção de valores de variáveis explicativas (independentes), podendo ser utilizada para avaliar os efeitos das variáveis explicativas nas respostas.

Há várias classes de modelos de regressão disponíveis na literatura, sendo cada uma dessas adequada para modelar um tipo específico de situação. Uma questão fundamental para escolher a classe do modelo de regressão adequada a um determinado estudo é identificar o tipo de variável dependente que está sendo considerado. Essa variável pode ser métrica (contínua ou discreta) ou pode ser categórica.

A variável dependente de interesse pode ser expressa como uma variável qualitativa, cujas categorias indicam níveis de convicção de um indivíduo para comprar um produto ou utilizar um serviço. Optando por expressar a intenção de compra desse modo, os modelos de regressão que passam a ser adequados para este estudo são modelos aditivos generalizados para variáveis categóricas. Essa classe de modelos de regressão inclui modelos lineares generalizados (paramétricos) e modelos de regressão não-paramétrica, ambos para variáveis dependentes categóricas.

Os modelos lineares generalizados são modelos estatísticos compostos apenas por parâmetros e variáveis (sendo assim chamados de modelos paramétricos) que, embora a priori não apresentem uma equação com uma forma linear, podem assumir essa forma através de uma transformação funcional chamada de função de ligação.

Para McCullagh e Nelder (1989), os modelos lineares generalizados são uma extensão dos modelos lineares clássicos e são compostos por três partes específicas: uma variável dependente que é a componente aleatória; uma equação composta por parâmetros e variáveis, dita componente sistemática, e que está relacionada com a componente aleatória; e uma função de ligação que torna a componente sistemática uma equação linear.

Hardle *et. al.* (2008) mostram que modelos não paramétricos de regressão são aqueles que não envolvem uma equação com parâmetros e variáveis, sendo obtidos através da estimação direta do comportamento da função que explica a relação entre uma variável dependente e outras independentes. Para os mesmos autores, modelos semiparamétricos são aqueles que apresentam uma parte do modelo sendo paramétrica e outra parte sendo não paramétrica.

Bock, Coussement e Poel (2010) afirmam que os modelos aditivos generalizados (MAG's) são uma generalização dos modelos lineares generalizados (MLG's). Histie e Tibishirani (1990) explicam essa generalização declarando que os primeiros substituem a forma linear dos segundos por uma soma de funções suaves. Em termos práticos, esses modelos apresentam semelhanças com os MLG's, distinguindo-se destes por não precisarem ser necessariamente paramétricos, ou seja, podem apresentar as características de um MLG, mas serem não paramétricos ou semiparamétricos.

Young *et. al.* (2011) complementam tal argumentação ao afirmarem que os MAG's apresentam uma vantagem adicional em relação aos MLG's, pois permitem fazer inferências sobre a relação entre uma variável dependente e variáveis independentes sem colocar restrições de parâmetros nas associações. Os parâmetros são partes componentes da equação funcional de um modelo estatístico que precisam ter seus valores estimados. Esses parâmetros apresentam pressupostos que podem limitar o modelo quanto à validade e quanto à aplicação.

Modelos aditivos generalizados para variáveis dependentes categóricas são uma alternativa adequada para fornecer uma medida da intenção de compra de grupos de consumidores. Ao optar-se por utilizá-los, o problema da dificuldade de efetuar a classificação de clientes quanto à intenção de compra passa a ser especificamente a obtenção de um modelo desse tipo. Sendo assim, tem-se a seguinte questão de investigação: como a intenção de compra pode ser relacionada às características de segmentos de mercado por meio de modelos aditivos generalizados?

1.2 Justificativa

O presente estudo direcionar-se-á às empresas que dispõem de informações acerca dos consumidores (características, opiniões, preferências, desejos e necessidades) e informações sobre as vendas que foram realizadas ou que podem ser realizadas. Todas essas informações podem ser obtidas por meio de pesquisas de mercado ou por meio de um sistema de informação estruturado para coletá-las e armazená-las. Assim, empresas que realizam pesquisas de mercado ou usam um sistema de informação desse tipo podem ser beneficiadas, independentemente do porte.

O tema escolhido nesta pesquisa contribui cientificamente ao destacar a importância, no projeto e na gestão de sistemas produtivos, de priorizarem-se aqueles clientes que tendem a consumir um produto ou serviço. Além disso, chama a atenção para a necessidade de obtenção de formas para identificar as características comuns dos consumidores que tendem a consumir um item.

No campo da engenharia de produção, a pesquisa contribui cientificamente ao tratar do projeto e gestão de sistemas produtivos sob a perspectiva do cliente. Sugere-se que é possível definir as características do sistema de produção levando em consideração, como informação adicional, a intenção de comprar um produto ou de usar um serviço. Na literatura sobre intenção de compra, considera-se que essa variável é importante quando decisões são tomadas com base no consumidor.

Clientes com maior intenção de comprar um produto tendem a consumi-lo com menor esforço da empresa que o disponibiliza do que aqueles com menor intenção, como sugerem Ramayah, Lee e Mohamad (2010). Esse esforço pode ser refletido em padrões de qualidade mais rigorosos, maior criatividade no desenvolvimento de produtos e maiores investimentos em todo o sistema produtivo. Xu, Summers e Belleau (2004), afirmam que a intenção de compra é importante no desenvolvimento de uma estratégia de marketing. A estratégia de marketing é, por sua vez, algo determinante nas decisões acerca da distribuição e do desenvolvimento de produtos.

Para obter uma medida da intenção de compra de acordo com o segmento de mercado, propõe-se um modelo e um procedimento para utilizá-lo. O modelo representa uma tecnologia de gestão resultante da conjugação de teorias do comportamento do consumidor, do princípio da segmentação de mercado e da teoria de modelos estatísticos causais.

O modelo proposto por esta pesquisa também é capaz de orientar atividades de projeto e de gestão dos sistemas de produção para que os produtos colocados no mercado sejam ainda mais adequados às características e preferências daqueles que realmente irão comprar, aumentando sua qualidade e tornando-os mais competitivos. O modelo proposto pode evitar decisões sejam tomadas de modo que recursos sejam gastos com aquilo que os clientes que tendem a comprar consideram como desnecessário.

Ao fornecer meios de identificar quais clientes tendem a comprar um produto ou usar um serviço, aponta-se às empresas a necessidade de disponibilizar produtos que sejam os mais adequados possíveis para aqueles consumidores. Se investimentos feitos para satisfazer clientes que não usam um produto ou serviço representam um desperdício, então a difusão da prática de considerar a intenção de compra na tomada de decisões pode reduzir o consumo desnecessário de recursos e pode proporcionar produtos mais adequados ao que deseja indivíduos de uma localidade, gerando uma maior satisfação.

Outra justificativa deste trabalho decorre do fato de indicar uma forma de utilização de dados provenientes de pesquisas de mercado para obter uma informação adicional importante acerca do mercado consumidor. Essa informação consiste em conhecer de que maneira as características dos grupos de clientes estão relacionadas à intenção dos mesmos de adquirir um produto.

A pesquisa de mercado é um instrumento importante e tende a ser cada vez mais popular entre as empresas, além de movimentar uma grande quantidade de recursos através de empresas especializadas.

Young e Javalgi (2007) consideram a pesquisa de mercado como uma ligação vital entre a organização e seus consumidores. Eles afirmam que desde 1995 o negócio de pesquisas internacionais de mercado tem crescido bastante. Nesse ano, as cinco maiores organizações do setor tiveram a soma de suas receitas igual a \$ 5,7 bilhões e 45% de suas receitas foram provenientes de pesquisas para companhias fora seus países de origem. Em 2004 a soma das receitas tinha crescido 133% e chegava a \$ 13,3 bilhões, com 67% dessa receita resultante de pesquisas com companhias fora de seus países de origem.

Lin (2007) reconhece a relevância de estudos sobre a intenção de compra em websites comerciais. Para ele, a proliferação desses sites tem tornado cada vez mais importante o entendimento dos determinantes da intenção de compras em lojas online.

O setor econômico estudado consiste em websites de compras coletivas. Embora o setor tenha tido crescimento nos últimos anos, nenhum estudo foi identificado na literatura sobre intenção de compra em sites de compras coletivas. Benazzi e Pedra (2011) afirmam que um dos maiores sites do setor, o GRUPON, entrou para história do mundo corporativo como uma das empresas que mais rapidamente atingiu um faturamento de US\$ 1 bilhão. Uma marca semelhante somente tinha sido alcançada pelo Youtube. Aqueles autores ainda citam que em 2010 o faturamento dos sites de compra coletiva no Brasil foi superior a US\$ 500 milhões, alcançando um número de 5,6 milhões de usuários. Segundo Silva *et. al.* (2011), entre 2010 e 2011, houve um crescimento de 153% no número de sites de compras coletivas no Brasil, chegando ao número de 1000 websites desse tipo.

1.3 Objetivos

- **Objetivo geral**

Obter modelos aditivos generalizados que possam contribuir para o estudo da relação entre a intenção de compra de um segmento de mercado e as suas características.

- **Objetivos específicos**

- a) definir formas de expressar a intenção de compra de clientes para obter modelos aditivos generalizados alternativos;

- b) verificar as variáveis que representam as características de segmentos e que podem estar relacionadas à intenção de compra;
- c) estimar modelos aditivos generalizados alternativos para relacionar a variável dependente (intenção de compra) com variáveis independentes associadas às características de segmentos.

CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA

O modelo sugerido propõe definir as características do sistema de produção de modo que sejam oferecidos produtos ou serviços em conformidade com consumidores, de um dado segmento de mercado, dispostos a utilizá-los. Sendo assim, o entendimento do processo de segmentação de mercado.

A escolha dos fatores utilizados como covariáveis do modelo foi baseada em teorias do comportamento do consumidor. A sua abordagem cognitiva é aquela, dentre as disponíveis na literatura, que considera a decisão de compra como resultante de um número mais amplo de tipos de fatores. Dentre as teorias da abordagem cognitiva, a teoria da ação racional (TAR) sugere uma explicação para a relação entre os fatores que influenciam na intenção de compra e considera essa última como um determinante imediato da decisão de compra. Por essa razão, a relação entre aqueles fatores e a decisão de compra é baseada TAR.

As relações sugeridas no modelo podem ser entendidas com base na abordagem cognitiva do comportamento do consumidor, especialmente na TAR. Além disso, há vários estudos, mostrados na subseção 2.3.3, onde a relação entre a variável intenção de compra com fatores psicológicos, sócio culturais e situacionais é considerada.

O conceito de intenção de compra também deve ser abordado uma vez que essa é a variável dependente utilizada pelo modelo. A compreensão da relação direta entre a intenção e a decisão de compra é necessária para que o modelo possa ser aplicado para explicar o consumo, pois se uma relação direta entre esses fatores não é assumida, então não é possível sugerir que a pretensão de comprar seja concretizada. Nesse caso a intenção de compra não pode ser usada como indicador do consumo.

As relações sugeridas (entre a variável resposta e cada uma das covariáveis) foram verificadas estatisticamente com base em modelos paramétricos e semiparamétricos de regressão, especificamente utilizando modelos lineares generalizados e modelos aditivos generalizados. A compreensão dos modelos propostos requer o conhecimento dessas classes de modelos estatísticos.

O quadro 1 apresenta uma síntese dos objetivos de cada uma das seções expostas na fundamentação teórica.

Quadro 1 - Objetivo das seções na fundamentação teórica

Objetivo da seção	Seção
Fundamentar a formulação conceitual das relações entre a variável resposta do modelo e as covariáveis	2.1. Segmentação de mercado e escolha do público alvo
	2.2. Comportamento do consumidor e decisão de compra
Fundamentar a verificação estatística das relações propostas e os modelos a serem obtidos	2.3. Modelos paramétricos de regressão
	2.4. Modelos não-paramétricos e semiparamétricos de regressão

Fonte: Elaboração própria

2.1. Segmentação de mercado e escolha do público-alvo

Segundo Ming-Chih, Yi-Ting e Ching-Wei (2011), a segmentação de mercado é definida como o ato de dividir um mercado em grupos distintos e significativos de clientes de modo a identificar grupos de consumidores cujos requisitos de compra e resposta aos planos de marketing sejam semelhantes.

Kotler e Armstrong (2003) declaram que as empresas sabem que não podem satisfazer a todos os clientes de um mercado. Há um grande número de consumidores com necessidades diferentes. Algumas empresas estão em melhores condições de atender a determinados segmentos de mercado. Nesse caso, é necessário dividir o mercado, escolher os melhores segmentos e procurar estratégias de atender aos segmentos específicos de maneira mais eficiente que os concorrentes.

Lopes (2002) afirma que o fundamento da segmentação de mercado é muito simples. Tem como base o fato de um único produto não satisfazer às necessidades e anseios de todos os clientes. Para esse autor, o centro de toda a discussão sobre segmentação é o fato de existir diferenças entre os consumidores, que precisam ser consideradas nas tomadas de decisão por parte da empresa.

Chiu *et. al.* (2009) defendem que a segmentação de mercado aumenta a rentabilidade ou eficácia da organização na medida em que seus benefícios econômicos em relação ao mercado consumidor excedem os custos desse processo. Kotler (2003) acredita que a empresa pode criar uma melhor sintonia entre produto ou serviço ofertado e preço adequadamente cobrado para o segmento-alvo.

Hung e Tsai (2008) defendem que a necessidade de segmentação de mercado existe porque, geralmente, uma empresa não consegue concentrar seus esforços em todo o mercado disponível. Em outras palavras, o mercado contém muitos clientes atuais e potenciais com uma rica diversidade.

As decisões acerca do projeto e gestão de sistemas de produção não podem levar em conta todos os clientes disponíveis. Desse modo, se o mercado é heterogêneo segundo diversas características, a segmentação de mercado torna-se um procedimento essencial para que o sistema de produção concebido desempenhe bem sua função de atender aos clientes.

Ritzman e Krajewski (2005) corroboram essa idéia ao considerarem a segmentação de mercado como uma primeira etapa do processo de conhecimento dos clientes de uma organização. Esse conhecimento é fundamental para o sucesso na formulação de uma estratégia de operações para o sistema produtivo, que tem como fim atender aos consumidores de forma superior a da concorrência.

Segundo esses autores, compreender o conjunto de benefícios ao cliente para um segmento de mercado possibilita que gestores identifiquem meios para ganhar vantagem competitiva. Eles complementam sua abordagem afirmando que cada segmento de mercado possui necessidades de mercado que podem ser relacionados aos atributos de produto/serviço de um sistema de produção.

Para Jacobs, Chase e Aquilano (2006) uma primeira etapa para desenvolver uma estratégia de manufatura para um sistema produtivo é segmentar o mercado de acordo com o grupo de produto, definindo o público alvo para cada elemento de seu portfólio.

Os autores acima tratam de um elemento importante da estratégia de produção que está diretamente ligado ao processo de segmentação de mercado, que é o posicionamento estratégico dos sistemas de produção. Para tais autores, definir esse posicionamento significa definir as características do produto e do processo de produção de forma que o sistema produtivo esteja voltado para um público-alvo específico.

Mesmo quando esse posicionamento não é planejado de maneira formal, é algo intrínseco ao projeto e gestão do sistema produtivo. As decisões de projeto e gestão definem com o tempo as características específicas do sistema de produção, realizando, mesmo que informalmente, um posicionamento do sistema de produção e direcionando o sistema para vender mais para parcelas específicas do mercado. Isso quer dizer que a segmentação de mercado é algo que é realizado quando as decisões de projeto, planejamento e controle são tomadas.

Corrêa e Corrêa (2004) argumentam que a competição internacional mais intensa tem levado à formação de mercados cada vez mais fragmentados e exigentes, levando o cliente a tornar-se mais atento a nuances e detalhes nos produtos que lhe são oferecidos. Essa crescente fragmentação pode ser um empecilho para organizações que ainda procuram projetar e gerir seus sistemas produtivos sem focar-se em segmentos específicos.

2.2. Comportamento do consumidor e a decisão de compra

Segundo Guerra (2005), o comportamento de compra do consumidor representa um dos elementos essenciais da gestão estratégica de qualquer empresa seja qual for a atividade de negócio. Para esse autor é, por isso, fundamental que a empresa conheça os seus públicos de consumidores atuais e potenciais, seus comportamentos, atitudes e processos de escolha, para que possa adequar as suas atividades de gestão e fornecer a melhor resposta às solicitações dos mercados onde opera.

Para Lamb, Hair e McDaniel (2004), o comportamento do consumidor consiste nos processos que o consumidor percorre para tomar decisões de compra, bem como para usar e dispor de bens ou serviços. Trata, assim, da forma como os consumidores tomam decisões de compra, utilizam e descartam mercadorias e serviços adquiridos. Tais autores destacam que o estudo do comportamento do consumidor também inclui a análise dos fatores que influenciam as decisões de compra e o uso do produto.

Kotler e Keller (2006) compartilham a mesma ideia ao afirmarem que o campo do comportamento do consumidor estuda como as pessoas, grupos e organizações selecionam, compram, usam e descartam artigos, serviços, ideias ou experiências para satisfazer suas necessidades e desejos.

Solomon (2008) corrobora com Kotler e Keller (2006) acrescentando que o comportamento do consumidor abrange uma ampla área onde se estudam os processos envolvidos no consumo. Esse autor enfatiza que esse é um processo contínuo e que não se restringe ao que acontece no instante em que o consumidor realiza a compra. Embora a compra, para ele, seja um elemento importante do comportamento do consumidor, a visão mais abrangente enfatiza o processo de consumo, o que inclui as questões que influenciam o consumidor antes, durante e depois da compra.

O processo de decisão de compra é um elemento importante do estudo do comportamento do consumidor. Segundo Lamb, Hair e McDaniel (2004), esse processo envolve cinco etapas: reconhecimento da necessidade, busca de informações, avaliação de

alternativas, decisão de compra e comportamento pós-compra. Os autores acima destacam que todas as etapas são afetadas por fatores culturais, sociais, pessoais e psicológicos. Para Kotler e Keller (2006) complementam tal argumentação ao afirmarem que os fatores culturais exercem a maior e mais profunda influência.

Hofmann, Strack e Deutsch (2008) entendem que os consumidores são dirigidos por um conflito entre o impulso para comprar e a tomada de decisões racionais. Segundo esses autores, a melhor forma de entender o comportamento do consumidor é compreender como suas atitudes são formadas diante daquele conflito. Sendo assim, do entendimento desses autores, é possível concluir que as variáveis que determinam seu comportamento são relacionadas a impulsos (ligação afetiva, atratividade das características etc.) e também a decisões racionais (importância e necessidade efetiva, preços de bens concorrentes etc).

Rucker, Galinsky e Dubois (2011) ressaltam que uma questão crítica do comportamento do consumidor é o entendimento de que atributos são desejáveis pelo mesmo e que valor está disposto a pagar por um produto ou atributo. Sendo assim, é fundamental em qualquer pesquisa relativa ao comportamento do consumidor tratar de questões relativas a atributos do produto valorizados pelo cliente, quando houver disponível algum conhecimento prévio sobre o assunto.

Para Casillas e Martínez-López (2009), é comum em marketing que as informações sobre variáveis relativas ao comportamento do consumidor sejam obtidas por meio de questionários. Para tais autores, também é frequente o uso de variáveis diretamente não observadas para modelar o comportamento do consumidor. Esse tipo de modelagem, segundo esses autores, é feito através da utilização de outras variáveis observadas que estão associadas àquelas.

Engel, Blackwell, Miniardi (2000), Giglio (2004), Kasarklian (2000); Kotler e Armstrong (2003), Solomon (2008), Pinheiro (2006) tratam de teorias do consumo e fatores que influenciam na decisão de compra. A análise do comportamento do consumidor feita por esses autores é baseada em diferentes abordagens teóricas que possibilitam listar alguns fatores ligados à decisão de compra.

2.2.1. Abordagens teóricas do comportamento do consumidor

As teorias do comportamento do consumidor procuram entender como o valor é gerado para consumidores, procurando identificar aqueles valores que determinam as suas escolhas.

Karsaklian (2005) cita cinco abordagens na qual se baseiam as teorias da motivação do consumidor: comportamental, cognitivista, psicanalítica e humanista. O autor reconhece as contribuições de abordagens da sociologia e antropologia ao tratar das relações do consumidor na sociedade. Giglio (2005) complementa essas abordagens ao citar mais duas outras abordagens: racionalidade econômica e a existencialista.

Pinheiro (2006) trata acerca de cinco abordagens relativas à teoria do comportamento do consumidor: racionalidade econômica, comportamental, psicanalítica, cognitivista e social/antropológica. Essas cinco dessas abordagens são tratadas a seguir.

Abordagem da racionalidade econômica

Aborda a relação custo-benefício de um comportamento do consumidor. O princípio dessa teoria consiste em considerar que um consumidor procura obter o maior benefício pelo menor custo.

A teoria é baseada na maximização da utilidade de uma decisão de compra. A maximização da utilidade resulta na otimização do prazer e satisfação de um indivíduo.

A limitação dessa teoria está na simplificação excessiva de uma análise profunda da mente de uma pessoa no momento de optar por um comportamento específico. Os processos psicológicos na mente de um indivíduo no momento de uma decisão de consumo envolvem diferenças individuais, culturais e sociais. Essas diferenças não são consideradas na teoria da racionalidade econômica.

Segundo a teoria da racionalidade econômica, as chances de um cliente consumir mais um produto ou serviço ocorrem quando os custos da aquisição são reduzidos e o prazer proporcionado pelo consumo é maior (a utilidade). Uma pesquisa em um setor industrial sobre a decisão de compra baseada nessa teoria deve identificar aqueles fatores que representam prazer para o cliente no consumo nesse setor específico e os custos que ele está disposto a assumir. A dificuldade de fazer uma pesquisa desse tipo é a de entender como o

prazer é determinado para o consumidor. A teoria não orienta de que modo o prazer para o cliente é definido.

Abordagem comportamental ou behaviorista

Trata da influência do ambiente na decisão de consumir, verificando as reações positivas e negativas em relação a um produto ou serviço. A teoria enfatiza o papel da aprendizagem e do ambiente na maximização da intenção de comprar um produto específico.

Uma pesquisa baseada na teoria comportamental deve investigar os estímulos que produzem reações positivas e negativas no ambiente, afetando a decisão de consumir um produto ou serviço. Nesse sentido, as opiniões e preferências de clientes acerca de características relacionadas ao produto/serviço e relacionadas à imagem das empresas são fundamentais para determinação da decisão de comprar. A dificuldade de efetuar uma pesquisa desse tipo consiste no fato que esta mantém o foco da análise sobre os estímulos e reações produzidos sem investigar que processos mentais determinam esses estímulos. Por essa razão, é possível que a relação entre os estímulos e reações possa ser interpretada equivocadamente.

Abordagem psicanalítica

Procura conhecer a dinâmica psicológica do consumo através da psicanálise, desenvolvida por Freud. Segundo Freud, a mente de um indivíduo é dividida em consciente e inconsciente. Os desejos íntimos são realçados no inconsciente, mas esses desejos são distorcidos na consciência e essa distorção determina um comportamento a ser assumido.

No momento em que uma pessoa vai efetuar uma compra, as suas decisões envolvem suas expectativas, angústias e conflitos na tentativa de obter uma satisfação mesmo que seja temporária ou parcial. Os desejos de satisfação são explicados pelo inconsciente. Por essa razão, é necessário utilizar a psicanálise para explicar a relação que deve existir entre a imagem e o conceito de produtos/serviços, visando atender aos desejos inconscientes por status, beleza, felicidade, etc.

Uma pesquisa com base nessa teoria deve considerar que os desejos do cliente não são determinados apenas pelo consciente, sendo necessário utilizar-se de instrumentos que permitam considerar também desejos mais íntimos presentes no inconsciente.

Abordagens sociais

Essas teorias estudam o consumo como um processo social determinado por fatores históricos, culturais e sociais. Segundo essas teorias, o consumo não é um ato apenas racional ou individual, envolvendo também processos sociais. Por essa razão, para essas teorias, é importante compreender de forma mais profunda a dinâmica social e cultural.

A pesquisa que se baseia nessas teorias deve considerar elementos culturais, históricos e sociais. Sendo assim, essas teorias complementam as demais ao incluir elementos que envolvem os grupos e organizações dos indivíduos.

Abordagem cognitivista

Considera que o consumo é determinado por um complexo processo de tomada de decisão que considera informações sócio culturais, individuais (psicológicos) e do ambiente (situacionais). A teoria integra o indivíduo, o produto/serviço e o ambiente.

Essa teoria fornece um entendimento mais detalhado do comportamento do consumidor, unindo vários elementos abordados por outras teorias. Por essa razão, é uma das teorias mais usadas na prática dos estudos e pesquisas do mercado consumidor.

Para Pinheiro (2005), pensar no comportamento de consumo como um processo de tomada de decisão implica ver o consumidor como aquele que opta por diferentes produtos, tendo por pano de fundo a influência de fatores cognitivos tais como percepção, motivação, aprendizagem, memória, atitudes, valores e personalidade, assim como os socioculturais, isto é, influência de grupo, família, cultura e classe social, bem como fatores situacionais.

Uma conhecida teoria da abordagem cognitiva do comportamento do consumidor é aquela proposta por Ajzen e Fishbein (1975). Essa teoria é utilizada na literatura atualmente para obter modelos que descrevam o comportamento do consumidor.

2.2.2. Fatores que influenciam na decisão de compra

A identificação do perfil do consumidor permite o desenvolvimento de estratégias com eficácia para satisfazer os clientes, criando valor para eles, e gerando bons resultados para o empreendimento. Este perfil é traçado através de características individuais, sociais, culturais

e situacionais. Essas características podem ser agrupadas em três níveis: psicológicas, socioculturais e situacionais.

Os fatores psicológicos dizem respeito a pensamentos, sentimentos e comportamento. Nesse nível o consumidor é estudado como indivíduo tomador de decisões com base em suas características psicológicas. São considerados os seguintes fatores psicológicos:

- Percepção – Por meio da percepção, os estímulos são percebidos e organizados de acordo com um sistema pessoal de valores e crenças.
- Motivação – Identificação dos motivos que levam o consumidor a comprar é importante, pois são várias as razões que conduzem o cliente ao ato da compra.
- Aprendizagem – Mudança do comportamento que ocorre por experiências (SOLOMON, 2002).
- Memória – Conjunto de meios que uma pessoa utiliza para recorrer a um conhecimento adquirido no passado a fim de utilizá-lo no presente.
- Atitude – Predisposição, tendência ou sentimento preexistente para avaliar uma característica do produto como positiva ou negativa. Para Pinheiro (2005), a intenção de compra é baseada na avaliação que depende da atitude.
- Personalidade e estilo de vida – A personalidade é a formação psicológica de um indivíduo e o modo como ele reage ao ambiente. Por sua vez, o estilo de vida é a forma como um indivíduo gasta seu tempo e dinheiro (SOLOMON, 2002). Kotler e Keller (2006) consideram como personalidade traços psicológicos distintos que levam a reações coerentes e contínuas a um estímulo do ambiente e define estilo de vida como o padrão de vida de uma pessoa expresso por atividades, interesses e opiniões.

Os fatores sócio-culturais são determinados pela influência do meio social sobre um indivíduo. São considerados os seguintes fatores desse nível:

- Grupos sociais – Solomon (2002) afirma que os indivíduos pertencem e admiram diversos grupos. Esses grupos exercem influência sobre várias decisões de compra.
- Cultura – Conjunto dos valores e atitudes que permitem um indivíduo comunicar-se, interpretar e avaliar como membros de uma sociedade.
- Classe social – Solomon (2002) cita diferentes variáveis que determinam a classe social, dentre estas estão a educação e a renda.

Os fatores situacionais envolvem aqueles que são circunstanciais e momentâneos. Segundo Engel, Blackwell e Miniard (2000), há três tipos principais de situações:

- As situações de comunicação – Locais onde o consumidor é exposto a comunicações pessoais (conversas com terceiros) e não pessoais (comunicação por diferentes meios, como internet, TV ou cartazes publicitários).
- As situações de compra – Cenários onde o consumidor compra produtos ou serviços. Essas situações englobam os fatores ambientais de informação, os fatores de varejo (como as características do layout do local da compra) e o fator tempo de exposição ao processo de compra.
- As situações de uso – Cenários onde o consumidor consome seu produto ou serviço.

Além dos fatores situacionais, psicológicos e sócio-culturais, outros fatores são importantes para o consumidor, tais como a qualidade percebida, a imagem da empresa (ou mesmo do produto/serviço) e o risco percebido. Kotler e Keller (2008) consideram os fatores pessoais, tais como idade e características da família, como importantes influenciadores das decisões do comprador. Para esses autores, o gosto no que diz respeito ao lazer em geral, como roupas e móveis, se relaciona com a idade. Os padrões de consumo, para eles, dependem do número, idade e sexo de membros familiares.

2.2.3. Intenção e decisão de compra

A intenção de compra é um indicador do ato da compra. Embora haja estudos que indiquem que essa correlação nem sempre é perfeita, muitas pesquisas sugerem que a intenção de compra é um bom indicador da decisão de compra. A Teoria da Ação Racional considera essa correlação sempre existente ao definir, de forma genérica, que a intenção de realizar uma ação é um determinante direto da decisão de efetuar-la.

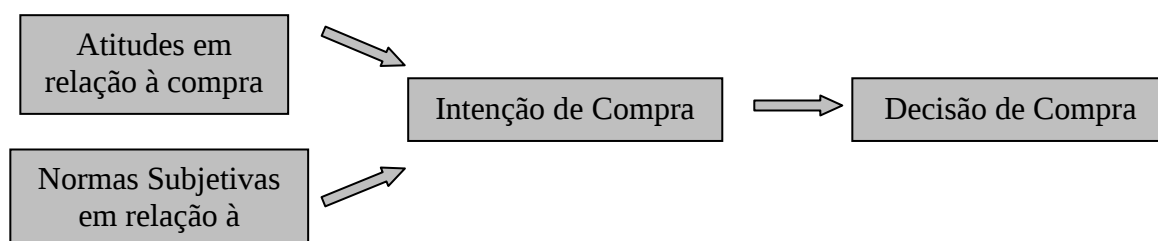
Sob a abordagem cognitiva do comportamento do consumidor a relação direta entre a intenção de compra e a decisão de compra estabelece que os fatores que determinam a o ato da compra, também determinam a intenção de realizá-lo.

A Teoria da Ação Racional (TAR), intenção e decisão de compra

As teorias do comportamento do consumidor podem, então, ser utilizadas para explicar a decisão de compra de clientes a partir de suas características. O consumo de um indivíduo depende de um complexo processo de tomada de decisão que envolve fatores psicológicos, sociais, demográficos e situacionais.

A Teoria da Ação Racional foi proposta por Ajzen e Fishbein (1975) para descrever o comportamento de forma genérica. Trata a intenção comportamental como um antecedente direto da decisão de um comportamento. Considerando a TAR aplicada ao comportamento de compra (ver Figura 1), isso significa dizer que, se um indivíduo tem a intenção de efetuar um comportamento de compra, então é esperado que esta ocorra.

Figura 1 - TAR aplicada ao comportamento de compra

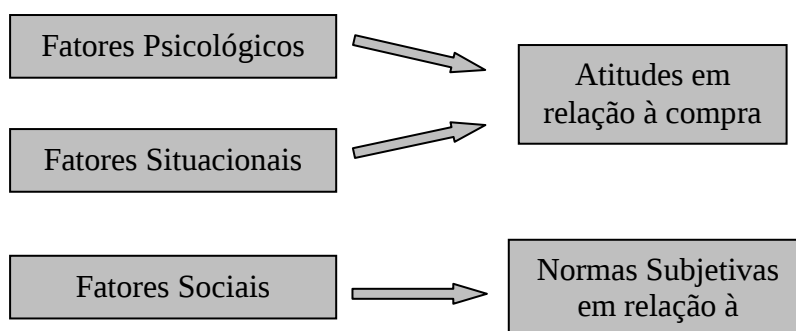


Fonte: Fishbein e Ajzen (1975) com adaptações

A intenção comportamental para TAR é determinada por sua vez por atitudes e normas subjetivas relativas ao comportamento. As normas subjetivas consistem na pressão social percebida para realizar um comportamento e as atitudes são avaliações individuais de um comportamento.

As normas subjetivas são influenciadas principalmente por fatores sócio-culturais, pois dependem da forma como o ambiente social se apresenta. As atitudes são influenciadas principalmente por fatores psicológicos e situacionais, pois dependem das características individuais (ver Figura 2).

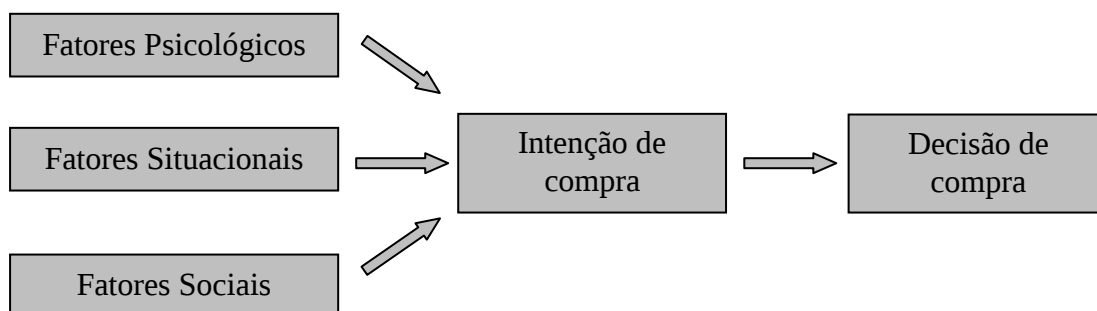
Figura 2 - Fatores determinantes das normas subjetivas e atitudes relativas ao comportamento de compra



Fonte: Elaboração própria

Os fatores característicos ligados a um indivíduo determinam as suas atitudes em relação às compra e as normas subjetivas a que estão submetidos. Essas, por sua vez, moldam a intenção de um indivíduo comprar e, conseqüentemente, sua decisão de compra. Por essa razão, é possível dizer que os fatores característicos ligados a um indivíduo são determinantes indiretos da decisão de comprar. A explicação da intenção de compra pode então ser feita por meio de fatores psicológicos, situacionais e sociais (ver Figura 3).

Figura 3 - Fatores determinantes da intenção e decisão de compra



Fonte: Elaboração própria

Todas as variáveis utilizadas no questionário que fornecerá o banco de dados para o estudo de caso são fatores característicos que podem ser classificados em uma das três classes de fatores. A Figura 24 - Fatores que influenciam na intenção de compra, variáveis da pesquisa e os itens do questionário (Apêndice D) mostra em que grupo se encontra cada uma das variáveis envolvidas no banco de dados usado na pesquisa. A figura também mostra os itens do questionário usado para obter os dados de cada variável.

Relação entre decisão e intenção de compra

Ramayah, Lee e Mohamad (2010) afirmam que qualquer intenção é um determinante de um ato. Para esses autores, a intenção de compra, especificamente, tem sido estudada com o objetivo de utilizá-la como um preditor de um comportamento real. Consideram ainda que esta intenção é uma ligação entre o comportamento e a atitude, e que é influenciada pelo esforço despendido para que ocorra o exercício do comportamento.

Kumar, Lee e Kim (2009) indicam que valores emocionais e percepção de aspectos ligados à percepção da qualidade de um produto ou serviço influenciam no comportamento de compra de clientes ou na intenção de compra, em um modelo que envolva aspectos cognitivos e afetivos. Os autores supracitados tomam como referência o conceito de intenção de compra

estabelecido por Yoo, Donthu e Lee (2000). Para estes últimos, a intenção de compra refere-se a uma tendência do consumidor para comprar rotineiramente uma determinada marca no futuro.

Para Morwitz, Steckel e Gupta (2007), a intenção de compras é utilizada comumente para prever as vendas. Os autores supracitados pesquisam fatores que estão relacionados ao aumento ou decréscimo da correlação entre a intenção de compra e a ocorrência efetiva da compra.

Sun e Morwitz (2010) destacam que vários autores na literatura utilizam a intenção de compras para prever as compras efetivas, partindo do pressuposto de que a mesma é um bom indicador para o comportamento real relativo às compras. No entanto, esses autores ressaltam que existem desvios entre a intenção de compras e a efetivação das compras. Para os mesmos, as características do modelo empregado para modelar a intenção de compras e as características do instrumento usado na coleta de dados são determinantes na redução do viés.

A intenção de compra consiste em qualquer sinal real dado pelo consumidor, indicando que pretende adquirir um produto ou utilizar um serviço. Representa um indicativo do cliente de que, caso o produto ou serviço estivesse disponível no mercado, ele o compraria. Esse indicativo embora sempre contenha uma parcela de incerteza, pode orientar, de forma efetiva, muitas decisões relativas ao sistema produtivo.

Zeithaml, Berry e Parasuraman (1996) reconhecem a importância de informações sobre intenção de compras ao destacarem que as intenções comportamentais de consumidores são sinais reais de uma escolha relacionados à compra e, por isso, é desejável que sejam monitoradas.

É difícil para qualquer organização garantir que a decisão de comprar o produto/serviço irá concretizar-se antes que a mesma ocorra. A intenção de compra geralmente é o melhor sinal disponível sobre a efetivação da compra. São as informações sobre a mesma que muitas vezes orientam as decisões do ambiente interno do sistema produtivo voltadas para o mercado.

Harrison, Mykytyn e Riemenschneider (1997) e Costa (2006) destacam que a intenção pode ser o melhor definidor de um comportamento, caso não ocorra nada no ambiente que possa interferir nos planos. Assim, ao considerar a compra como o comportamento em questão, seria a intenção de compra um bom definidor da decisão de comprar.

Para Xu, Summers e Belleau (2004), determinar a intenção comportamental dos consumidores, tal como a intenção de compra, é algo importante no desenvolvimento de uma estratégia de marketing para qualquer tipo de produto ou serviço, mesmo aqueles incomuns.

Lin (2007) destaca essa importância para o caso específico de websites comerciais, afirmando que a proliferação desses tipos de websites tem aumentado a relevância de entender os determinantes da intenção de compras em lojas online.

As previsões de demanda podem ser vistas como baseadas em intenções de compra. Ao realizá-las as organizações muitas vezes tomam como base dados históricos. Embora esses dados refiram-se à efetivação da compra num momento do passado, não é possível garantir que o ato da compra de cada cliente sempre irá ocorrer em momentos no futuro. No entanto, a sequência de aquisições passadas dos clientes representa um sinal de que o consumidor irá adquirir o produto/serviço no futuro.

Indicações nas pesquisas de mercado feitas junto aos consumidores, através de questionários, também podem ser vistas como um sinal real, embora incerto, de que o consumidor pretende comprar um produto/serviço. Na prática, mesmo não garantindo a efetivação da compra, declarações de clientes em potencial de que pretendem adquirir um produto/serviço são relevantes para tomadas de decisões de projeto e gestão de sistemas produtivos.

O desenvolvimento de sistemas de informações permite que a realização de uma compra seja uma fonte de informação sobre a intenção de clientes de comprarem um produto/serviço. O histórico de compras e de suas características conjugadas com informações presentes em cadastros sobre as características de clientes podem orientar também decisões de projeto e planejamento do sistema de produção, mesmo representando sinais dotados de incerteza sobre vendas futuras de produtos/serviços.

O histórico das compras obtidas a partir de sistemas de informação representa um conjunto de informações sobre intenções de compra. Todas as previsões e inferências feitas para vendas futuras nesse caso são baseadas em intenções de compras.

De acordo com Kotler e Keller (2006), no processo de compra, diversas características são determinantes na tomada de decisão de comprar. Tomando como referência a Teoria do Comportamento do Consumidor, ele declara que as características demográficas, culturais, físicas e psicográficas têm influência na decisão de compra.

Se essas características influenciam na decisão da compra e no processo de compra então também são determinantes da intenção de compra. Assim, é possível fazer inferências e previsões sobre a intenção de compra de clientes a partir das características dos mesmos ou a partir de opiniões e preferências efetivamente associadas àquelas.

Costa (2006) analisa os fatores que podem influenciar organizações na decisão da compra de softwares profissionais. Esse autor propõe um modelo conceitual para explicar a

relação entre a intenção das organizações de adquirir softwares com um conjunto de características da estrutura da organização, das pessoas envolvidas no processo de utilização dos softwares, das relações entre essas pessoas e dos benefícios potenciais percebidos da adoção dos softwares.

Relação entre intenção de compras e fatores determinantes da decisão de compra

Teng e Laroshe (2007) utilizam a análise de equações estruturais para testar um modelo teórico que relaciona as reações de indivíduos a anúncios de publicidade com a intenção de compras. O teste é feito para indivíduos de dois países com culturas bastante distintas (China e Estados Unidos) e mostra que o modo como a intenção de compra é afetada depende de características culturais dos indivíduos.

Chen (2007) utiliza a análise de regressão para relacionar a intenção de compras de alimentos orgânicos com as características sensoriais e nutritivas desses alimentos, bem como com seu preço, com características culturais e hábitos de consumo de consumidores em Taiwan. Ele mostra que duas características ligadas à personalidade exercem efeitos sobre a relação entre a escolha de certos alimentos e as atitudes relativas ao alimento orgânico. Mostra ainda que há seis motivos (humor, conteúdo natural do alimento, proteção ambiental, bem estar animal, valores políticos e religião) que determinam se a atitude em relação a produtos orgânicos é positiva ou não.

O autor supracitado procura determinar as variáveis que influenciam as atitudes em relação a produtos orgânicos porque toma esse fator psicológico como um dos principais influenciadores da intenção de compra dos produtos. Reconhece, porém, que variáveis ligadas a grupos sociais e à cultura (tais como a religião e valores políticos) são importantes para formação de uma atitude em relação aos produtos orgânicos e, conseqüentemente, para determinar a intenção de compra.

Dehua, Yaobin e Deyi (2008) utilizam um modelo de equações estruturais para mostrar que a intenção de compra de produtos eletrônicos em sites, em um site C2C (consumer-to-consumer), está relacionada com as atitudes relativas aos sites onde os produtos são divulgados (atitudes quanto à facilidade de uso dos websites), com fatores situacionais de compra (atitudes dos vendedores ao efetuar a venda) e de comunicação (a recomendação da compra por terceiros). Os autores acima consideram, então, que há fatores psicológicos (atitudes) e situacionais que influenciam a intenção de compra em sites C2C.

Verhagen e Dolen (2009) testam as hipóteses de vários aspectos relacionados à imagem de lojas da internet influenciar à intenção de compra de produtos online. Esses aspectos pesquisados estão relacionados à confiabilidade dos sites, facilidade de navegação, interface, serviço *offline* das empresas etc. Os pesquisadores concluem que a imagem das lojas relacionadas às impressões *online* e *offline*, de fato, tem influência na intenção de compra de produtos online nas lojas.

Wu, Yeh e Hsiao (2011) investigam os efeitos diretos e indiretos da imagem das lojas e da qualidade do serviço das mesmas sobre a intenção de compra. As hipóteses foram investigadas através da modelagem de equações estruturais. Esses autores concluíram que há uma forte relação entre a imagem das lojas e a intenção de compras. Verificaram também que a qualidade do serviço tem um efeito sobre a imagem das lojas e que o risco incorrido ao comprar produtos de uma determinada marca tem um efeito mediano na relação entre a imagem e a intenção de compras. Os autores acima reconhecem então que há fatores psicológicos que influenciam a intenção de compra, uma vez que as avaliações dos consumidores quanto à imagem, risco e qualidade dos serviços da loja são atitudes em relação a essa.

Shukla (2011) verifica que fatores relativos à imagem e à origem de uma marca, bem como os fatores de influência interpessoal, são influenciadores da intenção de compra de produtos de luxo. A pesquisa foi feita com consumidores indianos e britânicos, e as hipóteses foram testadas através da análise fatorial confirmatória. Os fatores de influência interpessoal investigados referem-se a aspectos relacionados ao interesse de indivíduos na imagem que os mesmos passam para terceiros.

As influências interpessoais consideradas por Shukla (2011) são influências de grupos sociais, sendo esses fatores socioculturais. Esse autor ainda considera que há também fatores psicológicos (atitudes em relação à imagem e à origem da marca) que influenciam à intenção de compra.

Poddar, Donthu e Wei (2009) constroem um modelo de equações estruturais para testar a hipótese de influência da orientação do site (estilo do site) sobre a intenção de compras nos mesmos. Os autores acima consideram que a orientação do site tem efeitos indiretos (ao influenciar na qualidade percebida) e diretos sobre a intenção de compra. A orientação do site é determinada a partir das opiniões de consumidores sobre as características dos sites. Essas opiniões representam atitudes (fatores psicológicos) dos clientes em relação às aquelas características

Chen, Hsu e Lin (2010) verificam atributos dos websites que podem aumentar a intenção de compras dos consumidores através de um modelo de análise conjunta. Esses atributos investigados são relacionados a questões de segurança de informações, privacidade, usabilidade dos websites, conveniência, confiança, entrega e valor. Todos esses atributos são atitudes em relação a características de websites, logo representando fatores psicológicos.

Ramayah, Lee e Mohamad (2010) utilizam a modelagem de equações estruturais para mostrar pela definição de um modelo conceitual que os valores, que são fatores culturais, e atitudes dos indivíduos, que são fatores psicológicos, têm influência sobre a intenção de compra de produtos verdes (ambientalmente corretos).

Kumar, Lee e Kim (2009) testam também um modelo conceitual por meio de equações estruturais com o objetivo de mostrar que a intenção de compra de consumidores indianos depende das suas características individuais e de variáveis associadas a marcas, como a qualidade percebida e valores emocionais. A intenção de compras é estudada com relação a produtos americanos e indianos.

Lu e Hsiao (2010) investigaram como os valores de usuários e perspectiva de satisfação, juntamente com a qualidade percebida e o preço percebido, influenciam à intenção de indivíduos pagarem por serviços de redes sociais na internet. O objetivo principal é avaliar como valores relacionados à extroversão/introversão afetam a intenção de usar o serviço. Esses autores estimam um modelo de equações estruturais usando como variáveis de controle a idade, o gênero e a renda. O modelo indica que os valores emocionais e sociais, bem como o preço e a qualidade do serviço, influenciam diretamente o valor percebido e indiretamente na satisfação dos clientes e na intenção de pagar pelos serviços de redes sociais.

Os autores supracitados mostraram então que há fatores psicológicos (valores sociais e emocionais; atitude em relação à qualidade e o preço) que afetam a intenção de clientes pagarem pelo serviço.

Vijayasathya (2004) explica a intenção do consumidor de utilizar lojas *online* acrescentando novos fatores explicativos ao modelo de aceitação de tecnologias (*Technology Acceptance Model* - TAM). Esse modelo considera que a utilidade e a facilidade de uso de uma tecnologia determinam a atitude em relação a esta que, por sua vez, afeta a intenção de uso. As novas variáveis propostas para o modelo são crenças em normas sociais, autoeficácia (uso dos produtos comprados nas lojas), privacidade, compatibilidade com o estilo de vida e segurança das informações pessoais. O autor utiliza a análise de componentes principais para agrupar as variáveis nos fatores investigados e a regressão linear múltipla para avaliar a influência dos fatores.

O autor acima mostra que todos os fatores complementam o TAM, exceto a privacidade, pois afetam a intenção de clientes de usar lojas online. Esse autor também enfatiza que há fatores sociais (crenças normativas) e psicológicos (atitudes em relação às características das lojas online e em relação aos produtos oferecidos).

Shim *et. al.* (2001) verificaram se a pretensão de um indivíduo procurar informações sobre os produtos pode ser usada como um elemento-chave para explicar a intenção de compra na internet. Eles mostraram que a pretensão de usar a Internet para buscar informações não apenas é o mais forte preditor da intenção de compras online, mas também é um importante mediador da relação entre a intenção de compras e outros preditores, tais como atitudes, normas subjetivas, controle percebido e experiência de compra. A determinação desses quatro fatores é feita por meio de variáveis como atitudes em relação às características dos sites (segurança, privacidade, conveniência, facilidade de navegar, variedade de produtos, dentre outros) e memória de experiências de compras. Os autores acima utilizaram um modelo de equações estruturais de dois estágios para testar as hipóteses de influência, e a pesquisa foi realizada em zonas metropolitanas dos Estados Unidos.

Os resultados obtidos por Shim *et. al.* (2001) também indicaram a presença de fatores psicológicos (atitudes em relação a características dos sites e memória em relação a alguns aspectos de experiências passadas) e sociais (normas subjetivas) como importantes preditores da intenção de comprar na internet.

Xu, Summers e Belleau (2004) mostraram que a intenção de compras de um produto controverso (incomum) pode ser explicada parcialmente por meio de normas subjetivas, percepções do produto como controverso, percepções acerca do preço, características demográficas (idade, estado civil, renda, escolaridade e ocupação), aspectos da personalidade e atitudes em relação às características do produto. A pesquisa foi realizada também em zonas metropolitanas dos Estados Unidos e as hipóteses foram testadas usando modelos lineares generalizados (MLG's). Embora os parâmetros do modelo tenham sido significativos, este apresentou um ajuste regular ($R^2=0,49$).

A investigação da intenção de compras de produtos incomuns realizada por Xu, Summers e Belleau (2004) reforçou que fatores psicológicos (atitudes em relação ao produto, personalidade e percepções) e sociais (normas subjetivas) são fatores determinantes da intenção de compras. A pesquisa desenvolvida pelos autores também destacou a importância das variáveis demográficas pessoais para explicar a intenção de comprar o produto.

Gonzalez, Comesaña e Brea (2007) utilizaram um modelo de regressão para mostrar como a qualidade percebida e a satisfação do consumidor estão relacionadas com a intenção

dos consumidores de hospedar-se em hotéis. A pesquisa foi realizada pelos autores acima em cidades da Espanha. Embora os parâmetros dos preditores tenham sido significativos, o ajuste do modelo foi razoável ($R^2=0,51$).

A qualidade percebida é determinada pela percepção de clientes em relação a diferentes características de hotéis relacionados à qualidade. A satisfação do consumidor é determinada pela memória de experiências anteriores em hospedagens. Sendo assim, o estudo contribui também por mostrar que há fatores psicológicos (percepção e memória) capazes de explicar parcialmente a intenção comportamental.

Belanche, Casalo e Guinalíu (2012) tratam da influência de atitudes em relação a características de usabilidade de um Website na intenção de utilizá-lo. Esses autores agrupam constructos para determinar dois fatores: usabilidade dos websites e satisfação de consumidores. Em seguida, utilizam a análise de equações estruturais para mostrar que as atitudes em relação à usabilidade dos sites influenciam indiretamente na intenção de utilizá-lo por meio da satisfação do usuário.

Mesmo que a hipótese de influência direta das atitudes em relação à usabilidade de sites não tenha sido comprovada, a influência foi demonstrada por meio do efeito indireto através da satisfação proporcionada. Belanche, Casalo e Guinalíu (2012) mostram, então, também a influência de fatores psicológicos (atitude em relação a características de usabilidade e memória representada pela satisfação em experiências anteriores) na intenção de usar um serviço online.

Kim e Kim (2004) demonstraram a influência de variáveis pessoais demográficas na intenção de comprar roupas e acessórios de roupas na internet. Os autores utilizaram um modelo de equações estruturais para mostrar que o gênero, a renda e o número de filhos influenciam diretamente na intenção de compras online enquanto o nível de escolaridade influencia indiretamente por meio de programas de incentivo à compra. O modelo estimado por eles não foi capaz de mostrar a influência da idade e variáveis relacionadas ao projeto dos Websites na intenção de comprar.

Diallo (2012) verifica a influência de aspectos relacionados à imagem das lojas, à imagem dos preços e ao risco percebido na intenção de compras em lojas de marca em um mercado emergente. O fator “imagem de lojas de marca” é determinado a partir de percepções das características das lojas (instalações, layout, facilidade de encontrar promoções, capacitação dos empregados, dentre outros). Esse autor mostra que os aspectos relacionados à imagem influenciam diretamente e indiretamente através do risco percebido.

Essa pesquisa foi realizada no Brasil e demonstrou também a capacidade de explicar parcialmente a intenção de compras por meio de fatores psicológicos como percepções (percepções relativas à imagem das lojas e dos preços) e motivações (aspectos relacionados ao risco percebido).

Lin (2007) compara a capacidade de três diferentes modelos de equações estruturais para prever a intenção de compras de produtos na internet com indivíduos de Taiwan. Dois dos modelos alternativos são baseados na Teoria do Comportamento Planejado (*Theory Planned Behavior* - TAB), que determina que a intenção comportamental de compra e a decisão real de compra são explicadas por meio de três fatores: atitudes, normas subjetivas e controle percebido. O outro modelo é baseado no Modelo de Aceitação de Tecnologia (*Technology Acceptance Model* - TAM) e considera que a intenção e a decisão de compra são determinadas pela facilidade de usar o serviço para compra e pela utilidade percebida deste mesmo serviço.

A diferença entre os dois modelos baseados na TAB está no número de estágios adotados no modelo de equações estruturais, sendo um desses modelos de dois estágios e o outro com apenas um estágio. O modelo de dois estágios baseado na TAB mostrou-se superior na capacidade de explicar a intenção e decisão de compra. Esse modelo foi baseado em sete fatores: utilidade percebida, facilidade de uso, influência de terceiros, autoeficácia, compatibilidade com as necessidades pessoais, influência interpessoal e influência externa.

Lin (2007) mostrou que o melhor dos modelos comparados explica a intenção de compras por meio de fatores psicológicos (atitudes em relação à compra e motivações representadas pelo controle comportamental percebido) e fatores sociais (normas subjetivas).

Moon, Chadee e Tikoo (2008) realizaram uma pesquisa com indivíduos da Nova Zelândia para verificar se há indícios da influência de aspectos culturais (individualismo, aversão à incerteza, aspectos relativos à masculinidade e aspectos relativos ao poder), do preço e do tipo de produto na intenção de comprarem produtos personalizados. A pesquisa utilizou a análise de variância (ANOVA) e testes de hipótese, mostrando que há diferenças na intenção de compra para diferentes níveis de individualismo, para diferentes preços e também para diferentes tipos de produto.

Há, então, indícios de que para o grupo estudado, há fatores situacionais (tipo de produto e preço) e fatores psicológicos (personalidade individualista) que podem explicar em parte a intenção de compra de produtos personalizados.

Bian e Forsythe (2012) demonstram, por meio de um modelo de equações estruturais, a existência de efeitos das características individuais e fatores associados a marcas na intenção

de compras de produtos de marcas de luxo, entre estudantes americanos e chineses. As características individuais consideradas na pesquisa consistem em dois fatores psicológicos relacionados à atitude em relação ao meio social e à personalidade no que diz respeito à capacidade de automonitoramento. As variáveis associadas à marca são opiniões dos indivíduos acerca da imagem das marcas.

Hartmann e Apaloaza-Ibanez (2012) mostram por meio da modelagem de equações estruturais que atitudes em relação à energia ambientalmente correta estão relacionadas à intenção de consumi-la. A pesquisa realizada com entrevistados na Espanha também mostra que fatores psicológicos são determinantes na intenção de compra.

Hausmann e Siekpe (2009) pesquisaram os efeitos das características de interface dos sites na intenção de compras na internet nos Estados Unidos. Os autores testam uma ampla escala de elementos de projeto de sites para determinar aqueles elementos humanos e computacionais importantes. Os elementos são então ligados usando a Teoria das Gratificações, o Modelo de Aceitação de Tecnologia (TAM) e o conceito de fluxo, para explicar a intenção de compras. O método usado para testar as hipóteses de influência foi a modelagem de equações estruturais.

As características dos Websites podem ser consideradas fatores situacionais (situações de compra ou situações de uso). Essa conclusão possibilita identificar que Hausmann e Siekpe (2009) mostraram que a intenção de compras na internet pode ser explicada em parte utilizando fatores situacionais.

2.4. Modelos paramétricos de regressão

Os modelos paramétricos de regressão englobam o grupo de modelos que relaciona uma única variável dependente a um conjunto de variáveis explicativas ou independentes, através de uma função que apresenta uma forma conhecida e definida através de parâmetros.

De acordo com Florencio (2010), na regressão paramétrica, as distribuições condicionais da relação entre as variáveis independentes e a variável dependente são modeladas por meio de famílias de distribuições paramétricas, cuja forma da função distribucional é conhecida, sendo desconhecidos apenas os valores dos seus parâmetros. Esse autor ressalta que a forma funcional (matemática) desconhecida da relação entre a variável resposta e as covariáveis pode ser determinada mediante o ajuste de curvas aos dados e estimação dos coeficientes da equação de regressão.

Aquele autor ainda afirma que um modelo é linear paramétrico quando a forma funcional que representa a relação entre regressando e regressores, apresenta uma forma linear, ou seja, $\sum_{i=1}^n \beta_i g_i(X)$ para um conjunto de funções $g_1, g_2, \dots, g_n$. Nesse caso, são determinados os coeficientes de uma forma funcional definida previamente.

2.4.1. Modelo linear clássico

Os modelos lineares de regressão são modelos paramétricos que apresentam uma função linear relacionando a variável dependente às variáveis explicativas. Hardle, Muller e Sperlich (2008) afirmam que em um modelo de regressão linear a média de uma variável dependente, Y , está relacionada a um conjunto de variáveis exploratórias $X_1, X_2, \dots, X_n$, da seguinte forma:

$$E(Y / X_1, X_2, \dots, X_n) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_d X_d \quad (2.1)$$

Estimação

Um modelo de regressão linear clássico simples apresenta a seguinte forma:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon \quad (2.2)$$

O modelo é determinado ao estimar os parâmetros. No modelo linear clássico, esses parâmetros podem ser determinados pelo método dos mínimos quadrados. Esse método consiste em determinar os valores dos parâmetros que minimizam a soma dos quadrados dos resíduos (SQR). O resíduo de uma observação i (ε_i) é a diferença entre o valor observado da variável resposta e o seu valor segundo o modelo.

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_{1i})$$

$$\varepsilon_i^2 = (y_i - (\beta_0 + \beta_1 x_{1i}))^2$$

$$SQR = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{1i}))^2$$

$$\frac{\partial(SQR)}{\partial\beta_0} = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{1i}))^2 = 0$$

$$\frac{\partial(SQR)}{\partial\beta_1} = \sum_{i=1}^n x_i (y_i - (\beta_0 + \beta_1 x_{1i}))^2 = 0$$

$$\begin{cases} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{1i}))^2 = 0 \\ \sum_{i=1}^n x_i (y_i - (\beta_0 + \beta_1 x_{1i}))^2 = 0 \end{cases}$$

O sistema composto pelas duas equações pode ser expresso na forma matricial:

$$X'(Y - X\hat{\beta}) = 0$$

Assim, a matriz de parâmetros que minimiza a equação matricial dos mínimos quadrados é:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2.3)$$

Os valores ajustados do modelo são dados por:

$$\hat{Y} = X\hat{\beta}$$

$$\hat{Y} = X(X'X)^{-1}X'Y$$

$$\hat{Y} = HY$$

Onde:

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}, H = \begin{pmatrix} 1 & h_1 \\ 1 & h_2 \\ \vdots & \vdots \\ 1 & h_n \end{pmatrix}$$

A matriz H é chamada de matriz chapéu (hat matrix). A multiplicação dessa matriz pela matriz de valores observados da variável resposta fornece a matriz de valores ajustados para as observações.

$$H = X(X'X)^{-1}X'$$

O procedimento para estimar os coeficientes de um modelo de regressão linear múltipla é o mesmo descrito. A diferença ocorrerá na matriz de dados dos preditores, que terá mais dimensões. Assim, a solução para uma forma geral do modelo de regressão linear clássico é:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2.4)$$

$$H = X(X'X)^{-1}X' \quad (2.5)$$

, onde:

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} \text{ e } H = \begin{pmatrix} 1 & h_{11} & \cdots & h_{k1} \\ 1 & h_{12} & \cdots & h_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & h_{1n} & \cdots & h_{kn} \end{pmatrix}$$

Os elementos da matriz chapéu são chamados de alavancagens e indicam a contribuição de cada observação para a estimativa dos parâmetros. Em modelos não paramétricos essa matriz assume um papel fundamental para determinação de valores estimados para variável resposta e recebe o nome de matriz de suavização.

Testes de significância

Para investigar modelos de regressão, as hipóteses de relação tomadas como existentes para formulação do modelo devem ser testadas estatisticamente. O objetivo de usar estes testes, é mostrar que a probabilidade das relações sugeridas existirem é suficientemente alta, respaldando o modelo proposto.

Em modelos de regressão é comum utilizarem-se dois testes estatísticos de hipóteses. Em um destes testes é verificado a hipótese de que não é possível formular um modelo (de uma determinada classe de modelos de regressão) que relacione uma variável dependente com um conjunto de variáveis independentes investigadas, ou seja, é verificado se todos os coeficientes β_i do modelo de regressão são iguais a zero. Se os coeficientes forem iguais a zero no modelo, alterações nas variáveis independentes não afetam a variável dependente pois sempre estarão sendo multiplicadas por zero. A estatística de testes usada para testar uma hipótese desse tipo em um modelo de regressão linear clássico é dada por:

$$F_c = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / p}{\sum_{i=1}^n (\hat{y}_i - y_i)^2 / (n - p - 1)} \quad (2.6)$$

, onde p é o número de variáveis independentes no modelo e n é o número de observações. Se a variável dependente tem um comportamento aproximado de uma distribuição normal, então F_c segue uma distribuição F com graus de liberdade p e $n-p-1$.

A hipótese de nulidade conjunta de todos os coeficientes é então testada comparando F_c e o valor na distribuição F para os graus de liberdade indicados. Se $F_c > F_{(p, n-p-1)}$ então rejeita-se para um nível de significância α (probabilidade de rejeitar a hipótese de teste sendo a mesma verdadeira) a hipótese de que todos os coeficientes são nulos, não existindo uma relação entre a variável resposta e as covariáveis.

O outro tipo de hipótese a ser testada verifica separadamente se cada um dos coeficientes é diferente de zero, ou seja, verifica se a hipótese de não haver relação entre a variável dependente Y e a variável independente X_j não pode ser rejeitada. A estatística de teste usada para verificar essa hipótese para j -ésima variável é:

$$t_c = \frac{\beta_j}{\left(\hat{\sigma} / \sum_{j=1}^n x_j^2 \right)} = \frac{\beta_j}{se(\beta_j)} \quad (2.7)$$

, onde $se(\beta_j)$ é o erro padrão do coeficiente β_j .

Se a variável dependente tem uma distribuição normal, t_c segue uma distribuição t-student com $n-2$ graus de liberdade, sendo n o número de observações. Se $t_c > t_{n-2}$ para um nível de significância α , então rejeita-se a hipótese de que não existe uma relação entre as variáveis.

Os testes de hipóteses podem ser avaliados em termos do p-valor. Este é o nível de significância mínimo a partir do qual a hipótese de teste pode ser rejeitada. A hipótese de teste somente é rejeitada se o p-valor do modelo é inferior ao nível de significância adotado. No teste F , o p-valor é a probabilidade de $F_c < F_{(p, n-p-1)}$ e no teste t é a probabilidade de $t_c < t_{n-2}$.

2.4.2. Modelos lineares generalizados

McCullagh e Nelder (1989) resumem o modelo linear clássico, indicando que os componentes de Y são considerados variáveis com distribuição normal independentes entre si e com média $\mu = X\beta$ e variância constante σ^2 .

Hastie e Tibshirani (1990) destacam que os modelos lineares generalizados (MLG's) são extensões matemáticas de modelos lineares que permitem identificar estruturas de variância não constante e não lineares nos dados.

Guisan, Edwards, Hastie (2002) ressaltam que essa classe de modelos é baseada numa relação (denominada de função de ligação) assumida entre a média de uma variável resposta e uma combinação linear de variáveis exploratórias. Além disso, os dados podem ter sua distribuição aproximada para uma variedade de famílias de distribuições de probabilidade tais como: normal, binomial, Poisson, binomial negativa e gamma. Lô e Ronchetti (2009) complementam essa abordagem, afirmando que os MLG's podem ser usados para modelar a relação entre preditores e uma função da média de uma variável resposta discreta ou contínua.

Cordeiro e Andrade (2009) afirmam que os MLG's são baseados em distribuições da família exponencial da variável e que incluem um algoritmo geral para computação de estimativas de máxima verossimilhança. Segundo esses autores, essa generalização dos modelos lineares permite o ajuste de diferentes modelos para diferentes tipos de escala.

Para Berg (2007), a presença de uma função de ligação nos MLG's e a possibilidade de considerar a variável resposta como sendo uma distribuição da família exponencial, são as características do modelo que permitem estendê-lo para aplicações onde não seria possível utilizar o modelo linear padrão.

A generalização dos modelos lineares é mostrada por McCullagh e Nelder (1989). Essa generalização é possível a partir da especificação de três partes em um modelo linear, a saber:

1. A componente aleatória: as componentes de Y apresentam distribuições da família exponencial independentes com $E(Y) = \mu$ e com variância constante σ^2 .
2. A componente sistemática: as variáveis $X_1, X_2, \dots, X_d$, produzem um preditor linear

$$\eta = \sum_{j=1}^d X_j \beta_j.$$

3. Uma função: que faz a ligação entre a componente aleatória e a componente sistemática $g(\cdot)$, ou seja, que $\eta_i = g(\mu_i)$. Essa função é chamada de função de ligação.

Segundo a generalização feita pelos autores abaixo, o modelo linear clássico é uma particularidade de um modelo mais amplo. Para esse modelo clássico, a componente aleatória segue uma distribuição normal ou gaussiana (componente 1) e a função de ligação (componente 3) que relaciona a componente aleatória e a componente sistemática é uma função identidade $\eta_i = \mu_i$ (MCCULLAGH e NELDER, 1989).

A generalização possibilita que modelos lineares de regressão sejam bem ajustados mesmo quando a variável dependente não tem uma distribuição próxima de uma distribuição normal. Possibilita ainda que modelos lineares sejam usados mesmo quando a variável dependente não é contínua.

Agresti (2002) também ressalta que a componente aleatória de um modelo linear generalizado apresenta observações seguindo uma distribuição da família exponencial. Tal autor destaca que diversas distribuições importantes são casos especiais da família exponencial. Sendo assim, se uma distribuição pode ter sua função de densidade ou função de massa expressa como a equação (2.8), então a mesma pode ser adequada para representar a distribuição das observações da componente aleatória de um modelo linear generalizado.

Nelder e Wedderburn (1972) formularam os modelos lineares generalizados, apresentando a família de distribuições exponenciais sob a forma da equação (2.8). Essa forma bastante utilizada nos dias atuais. Berg (2007) utiliza essa formulação como base para propor um modelo aditivo generalizado de previsão.

$$f(y; \theta, \phi) = \exp\left(\frac{[y\theta - b(\theta)]}{a(\phi)} - c(y, \phi)\right), \quad (2.8)$$

sendo que θ é chamado de parâmetro natural, ϕ é chamado de parâmetro de dispersão e que $b(\cdot)$ e $c(\cdot)$, são funções especificadas, sendo $b(\cdot)$ diferenciável duas vezes e $\phi^{-1} = a(\phi) > 0$.

Estimação

Nos modelos lineares generalizados os parâmetros são estimados pelo método da máxima verossimilhança. O método consiste em obter os valores do vetor de parâmetro β que maximize a função de verossimilhança do modelo. A função de verossimilhança fornece a

probabilidade conjunta das observações previstas do modelo assumirem os valores observados na prática.

Segundo Wood (2006, p.66), as estimativas de máxima verossimilhança podem ser encontradas utilizando-se de um método iterativo conhecido como Método dos Mínimos Quadrados Reponderados Iterativamente (*Iteratively Re-weighted Least Squares* – IRLS ou IRWLS). Esse método é baseado numa aproximação quadrática do log da função de verossimilhança.

Os modelos lineares generalizados (MLG's) apresentam a forma da equação (2.9).

$$\begin{aligned}\eta &= \beta_0 + \beta_1 X_1 + \dots + \beta_d X_d \\ \eta &= g(\mu) \\ \mu &= E(Y / X_1, \dots, X_d),\end{aligned}\tag{2.9}$$

sendo η um preditor linear e $E(Y / X_1, \dots, X_d)$ o valor esperado de Y , dados os valores das variáveis independentes X_j para $j = 1, 2, \dots, d$.

A variável resposta dos MLG's apresenta uma distribuição de probabilidade que tem a forma da equação (2.8).

Nelder e McCullagh (1989) mostram que se a variável Y tem uma função de distribuição de probabilidade da forma da equação (2.8), então os valores de sua média e de sua variância obedecem às relações da equação (2.10) e (2.11). Na equação (2.11), V é chamada de função de variância.

$$\mu = \frac{\partial b(\theta)}{\partial \theta}\tag{2.10}$$

$$Var(Y) = \phi^{-1} \frac{\partial(\mu)}{\partial \theta} = \phi^{-1} V\tag{2.11}$$

A estimação do vetor de parâmetros $\beta = (\beta_1, \beta_2, \beta_3, \dots, \beta_d)$ para modelos lineares generalizados é feita a partir do método da máxima verossimilhança. As estimativas são obtidas a partir da resolução da equação (2.12).

$$\sum_{i=1}^n x_{ij} \left(\frac{\partial(\mu_i)}{\partial \eta_i} \right) V_i^{-1} (y_i - \mu_i) = 0 \quad (2.12)$$

Para Nelder e McCullagh (1989), sendo n o tamanho da amostra e d o número de variáveis, $d+1$ parâmetros podem ser estimados. As equações podem ser resolvidas pelo método scoring de Fisher. Gonçalves (2009) afirma que o método *IRWLS* fornece um resultado semelhante àquele obtido pela estimação do método scoring de Fisher. Esse autor descreve a aplicação daquele método na busca das soluções de um MLG.

O método inicia-se a partir de um vetor inicial $\beta(0)$. Calcula-se, em seguida, a resposta modificada $z_i^{(m)}$ e os pesos $w_i^{(m)}$ na m -ésima iteração, usando as equações (2.13) e (2.14).

$$z_i^{(m)} = \eta_i^{(m)} + (y_i - \mu_i^{(m)}) \left(\frac{\partial(\mu_i)}{\partial \eta_i} \right)^{(m)} \quad (2.13)$$

$$w_i^{(m)} = \left(\frac{\partial(\mu_i)}{\partial \eta_i} \right) (V_i^{(m)})^{-1}, \text{ para } i=1, 2, \dots, n. \quad (2.14)$$

Os valores $\eta_i^{(m)}$, $\mu_i^{(m)}$ e $V_i^{(m)}$ são obtidos a partir das equações (2.15), (2.16) e (2.17).

$$\eta_i^{(m)} = \beta_0^{(m-1)} + \sum_{j=1}^d \beta_j^{(m-1)} x_{ij} \quad (2.15)$$

$$\mu_i^{(m)} = g^{-1}(\eta_i^{(m)}) \quad (2.16)$$

$$V_i^{(m)} = \left(\frac{\partial(\mu_i)}{\partial \theta_i} \right)^{(m)} \quad (2.17)$$

Neste método, os valores obtidos para $z_i^{(m)}$ e $w_i^{(m)}$ em cada iteração m são usados juntamente com a matriz de dados X , na equação (2.18), para obter o vetor de coeficientes β . A matriz $z^{(m)}$ representa um vetor $n \times 1$ composto pelos n valores obtidos para $z_i^{(m)}$ na m -ésima iteração, e $W^{(m)}$ representa a matriz diagonal com os n valores obtidos para $w_i^{(m)}$.

$$\beta = (X'W^{(m)}X)^{-1}X'W^{(m)}Z^{(m)} \quad (2.18)$$

O critério de parada das iterações pode ser pré-definido. Define-se previamente um valor para $\delta > 0$ e as iterações são interrompidas quando a desigualdade (2.19) é satisfeita.

$$\frac{\sum_{j=1}^d \|\beta_j^{(m)} - \beta_j^{(m-1)}\|}{\sum_{j=1}^d \|\beta_j^{(m-1)}\|} \leq \delta \quad (2.19)$$

Testes de significância

Há testes de hipóteses que desempenham o mesmo papel do teste F e do teste t nos MLG's, mas estes apresentam uma lógica um pouco distinta. Nos MLG's o teste conjunto da significância dos coeficientes é feito por meio do teste χ^2 . Sendo $l(\beta_{\text{modelo}})$ a estimativa de máxima verossimilhança de um vetor de parâmetros estimados para um modelo corrente; e $l(\beta_{\text{saturado}})$, a estimativa de máxima verossimilhança de um modelo saturado (modelo onde o número de coeficientes é igual ao número de observações, fornecendo o melhor ajuste possível para os dados), define-se o desvio (*deviance*) para o modelo corrente com a equação (2.20).

$$D_{\text{modelo}} = 2[l(\beta_{\text{saturado}}) - l(\beta_{\text{modelo}})] \quad (2.20)$$

O desvio escalonado, por sua vez, é obtido dividindo-se o desvio pelo parâmetro escalar, como mostrado na equação (2.21).

$$D_{\text{modelo}}^* = \frac{D_{\text{modelo}}}{\phi} = 2[l(\beta_{\text{saturado}}) - l(\beta_{\text{modelo}})] \quad (2.21)$$

sendo ϕ o parâmetro escalar do modelo.

O desvio do modelo é usado para calcular a estatística de teste utilizada no teste χ^2 . A lógica utilizada nesse teste consiste em verificar se a diferença entre o desvio de um modelo

de interesse e um modelo nulo (modelo que possui apenas um termo constante) é significativa. A distribuição χ^2 pode ser usada de acordo com (2.22).

$$(D_{\text{modelo}}^* - D_{\text{nulo}}^*) \sim \chi_{p-1}^2 \quad (2.22)$$

Assim a significância conjunta dos coeficientes em um MLG é testada usando a estatística de teste (2.23).

$$\chi_c^2 = (D_{\text{modelo}}^* - D_{\text{nulo}}^*) \quad (2.23)$$

Se $\chi_c^2 > \chi_{p-1}^2$, para um nível de significância especificado, rejeita-se a hipótese de teste (não significância) e os coeficientes podem conjuntamente ser considerados significantes.

O teste estatístico que desempenha para um MLG um papel semelhante ao teste t na regressão linear ordinária é o teste de Wald. Neste teste a estatística usada é calculada pela equação (2.24).

$$z_c = \left(\frac{\beta_j}{se(\beta_j)} \right)^2 \quad (2.24)$$

Sendo $z_{\alpha/2}$ o valor em uma distribuição normal padronizada, para um nível de significância α . Rejeita-se a hipótese de não significância do coeficiente β_j se $z_c > z_{\alpha/2}$.

Os resultados dos testes de significância dos coeficientes de um MLG podem também ser expressos em termos do p-valor. Se $p\text{-valor} < \alpha$, rejeita-se a hipótese de teste.

É importante notar que a estatística *Deviance* também é útil na verificação do ajuste do modelo. Faraway (2006, p.45) trata sobre a avaliação do ajuste geral de um MLG. Para esse autor, a proporção da variância explicada (R^2) é uma importante medida de ajuste de um modelo linear padrão. Segundo ele, é possível usar a aplicação do mesmo conceito para verificar a proporção da estatística *Deviance* (que é uma medida de desvio do modelo) que é explicada por um MLG. Este sugere o uso da equação (2.25) para calcular o valor da medida de ajuste. A equação é baseada no tamanho da amostra n , e nos valores da estatística

Deviance para um modelo nulo (modelo que apresenta apenas o intercepto) e para o modelo encontrado.

$$R^2 = \frac{1 - \exp((D_{residual} - D_{nulo})/n)}{1 - \exp(-D_{nulo}/n)}, \quad (2.25)$$

2.4.3. Modelos lineares generalizados para respostas binárias e ordinais

Para McCullagh e Nelder (1989, p.108), há uma ampla variedade de funções de ligações que podem ser utilizadas para modelos com variáveis respostas binárias. Três destas, segundo eles, são comumente usadas na prática: função logística (*logit*), função normal inversa (*probit*) e a função complementar *log-log*. Para os modelos binários, é comum usar uma função de ligação *logit*.

No caso de modelos com variável resposta categórica ordinal, um modelo muito usado apresenta uma função de ligação *logit*, com a regressão paralela da escala escolhida e com as probabilidades associadas a uma dada categoria sendo acumuladas. De acordo com McCullagh e Nelder (1989, p.152) esse é o tipo de modelo mais simples usado quando uma variável resposta ordinal está envolvida na análise.

A distribuição usada para modelar a distribuição dos erros da variável dependente (intenção de compra) binária foi a distribuição binomial e a usada na modelagem dos erros da variável dependente sob a forma ordinal foi a distribuição multinomial.

Adaptação do método *IRWLS*

Quando a variável a ser explicada em um experimento, apresenta uma forma binária (variável com duas categorias), então a distribuição binomial mostra-se como uma alternativa para descrever uma aproximação para o comportamento dessa variável. Essa distribuição binomial apresenta a forma da equação (2.26).

$$P(x = k; n) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (2.26)$$

Se a variável analisada é uma variável com um número de categorias superior a dois, então a distribuição multinomial pode ser uma boa alternativa para a descrição do comportamento dessa variável. Essa distribuição tem a forma indicada na equação (2.27).

$$P(X_1 = k_1, X_2 = k_2, \dots, X_d = k_d; n) = \frac{n!}{k_1! k_2! \dots k_d!} p^{k_1} p^{k_2} \dots p^{k_d} \quad (2.27)$$

O valor esperado de uma distribuição binomial para uma única observação é $\mu = np$ e o valor esperado de uma distribuição multinomial considerando uma observação é $\mu_j = np_j$. O valor esperado da distribuição multinomial é associado a cada uma das categorias da distribuição.

Tanto a distribuição binomial quanto a distribuição multinomial são partes da família exponencial de distribuições, que apresenta a forma geral da equação (2.8). Por essa razão, é plausível estimar modelos lineares generalizados cujo comportamento da variável resposta é representado por essas distribuições.

As equações (2.28), (2.29) e (2.30), mostram algumas características da distribuição binomial na família exponencial.

$$b(\theta) = \log(1 + \exp(\theta)) \quad (2.28)$$

$$V(\mu) = \mu(1 - \mu) \quad (2.29)$$

$$\theta = \log \text{it}(\mu) = \ln\left(\frac{\mu}{1 - \mu}\right) \quad (2.30)$$

Essas características são importantes para a determinação dos parâmetros de um MLG pelo método *IRWLS*, quando não são utilizadas ligações canônicas. Esse método requer a determinação das duas derivadas de primeira ordem indicadas nas equações (2.14), (2.15) e (2.18).

$$\frac{\partial \mu}{\partial \theta} = b''(\theta) \quad (2.31)$$

$$\frac{\partial \mu}{\partial \eta} = \frac{\partial}{\partial \eta} \left(\ln\left(\frac{\eta}{1 - \eta}\right) \right) \quad (2.32)$$

As derivadas de primeira ordem usadas no método iterativo de estimação são obtidas a partir da função de ligação. As derivadas presentes nas equações utilizadas no método *IRWLS*

são calculadas utilizando-se das equações (2.31) e (2.32), para um modelo de regressão logística. Essas derivadas podem ser expressas como funções de μ (ou de η).

A derivada apresentada na equação (2.31) é obtida a partir da função $b(\theta)$ do parâmetro natural, que é uma característica da distribuição de probabilidade usada para descrever o comportamento da variável resposta do modelo. No caso específico em que se utiliza a distribuição binomial, a equação (2.28) pode ser usada no cálculo da derivada. A derivada da equação (2.32) pode ser calculada utilizando-se da função de ligação.

No caso específico dos modelos com a função de ligação *logit* (equação 2.33) e com a distribuição binomial ou multinomial como aproximações para o comportamento da variável resposta, há a particularidade de utilizar uma ligação canônica. Segundo Faraway (2008, p.126), ao usar uma ligação canônica, tem-se $\eta = g(\mu) = \theta$. Isso conduz ao caso particular onde as derivadas de primeira ordem usadas no *IRWLS* são equivalentes e podem ser calculadas também usando apenas a função de ligação ou apenas as características da distribuição.

$$g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right) \quad (2.33)$$

Função de verossimilhança e estatística *Deviance*

Considere o modelo logístico onde a variável resposta binária assume os valores $y_i = 0$ ou $y_i = 1$. A função de distribuição de probabilidade dessa função é mostrada na equação (2.34).

$$f(y_i, \pi_i) = \pi_i^{y_i} (1 - \pi_i)^{(1-y_i)} \quad (2.34)$$

A função de verossimilhança é obtida a partir do produto dos valores da função de distribuição de probabilidade para cada observação. De forma geral, a função de verossimilhança tem a forma mostrada na equação (2.35).

$$l(y_i, \pi_i) = \ln(L(y_i, \pi_i)) = \ln\left(\prod_{i=1}^n f(y_i, \pi_i)\right) \quad (2.35)$$

Para um modelo cuja distribuição da variável resposta é binomial, a função de verossimilhança é dada pela equação (2.36), sendo β o vetor de parâmetros do modelo.

$$l(\beta) = \ln(L(\beta)) = \ln\left(\prod_{i=1}^n \pi^{y_i} (1 - \pi)^{(1-y_i)}\right) \quad (2.36)$$

No caso da estatística *deviance*, cuja forma geral é mostrada na equação (2.37), ela é obtida por meio do logaritmo da função de verossimilhança. Sendo assim, para um modelo logit binomial essa estatística pode ser expressa pela equação (2.38).

$$D_{\text{modelo}} = 2[l(\beta_{\text{saturado}}) - l(\beta_{\text{modelo}})] \quad (2.37)$$

$$D_{\text{modelo}} = -2 \sum_{i=1}^n \left[y_i \ln\left(\frac{\pi_i}{y_i}\right) - (1 - y_i) \ln\left(\frac{1 - \pi_i}{1 - y_i}\right) \right] \quad (2.38)$$

O modelo multinomial utilizado na pesquisa apresenta três categorias. Segundo Hosmer e Lemeshow (2000, p.262), a função de verossimilhança de um modelo desse tipo é dada pela equação (2.39). A função de verossimilhança pode então ser usada para calcular a estatística Deviance para um modelo logit desse tipo.

$$l(\beta) = \ln(L(\beta)) = \ln\left(\prod_{i=1}^n \left(\prod_{j=0}^2 (\pi_j(x_i))^{y_{ij}}\right)\right) \quad (2.39)$$

Os valores da estatística *Deviance* são utilizados para o cálculo do pseudo- R^2 e das estatísticas de teste, que são, por sua vez, usados para avaliar o modelo.

Matriz de confusão e Curva ROC

De maneira geral, a curva característica de operação do receptor (*Receiver operating characteristic* - ROC) é uma técnica usada para avaliação e seleção de classificadores. A análise dessa técnica é feita por meio de gráficos da curva ROC, que exibem o número de acertos e alarmes falsos dos classificadores.

Os acertos e alarmes falsos são determinados a partir de um método de avaliação de classificadores chamado de matriz de confusão. A matriz de confusão é um quadro que expressa o número de acertos e erros cometidos ao classificar observações em um grupo de categorias.

O quadro 1 mostra a forma de uma matriz de confusão para classificações envolvendo duas categorias.

Quadro 2 – Matriz de confusão para classificações em duas categorias

Classificação	Julgamento	
	Verdadeiro	Falso
Positivo	Número de Verdadeiros Positivos (NVP)	Número de Falsos Positivos (NFP)
Negativo	Número de Verdadeiros Negativos (NVN)	Número de Falsos Negativos (NFN)

Fonte: elaboração própria.

Os números que preenchem a matriz são calculados aplicando-se um determinado classificador a uma série de observações e agrupando-as de acordo com cada uma das quatro possibilidades apresentadas na matriz. A partir dos números que preenchem a matriz, calcula-se a taxa de verdadeiros positivos (TVP) e falsos positivos (TFP). A TVP é também chamada de sensibilidade e a TFP é chamada também de especificidade.

$$TVP = \frac{NVP}{NVP + NVN} \quad (2.40)$$

$$TVP = \frac{NFP}{NFP + NFN} \quad (2.41)$$

A taxa de verdadeiros positivos (calculada pela equação 2.40) representa o percentual de acertos cuja classificação foi “positivo”, enquanto a taxa de falsos positivos (calculada pela equação 2.41) representa o percentual de erros cuja classificação foi positiva. O uso dessas

duas medidas permite que a matriz de confusão também seja usada para avaliar o poder de explicação de um determinado classificador.

A curva ROC é esboçada em um espaço bidimensional, sendo estabelecido um gráfico da taxa de verdadeiros positivos em função da taxa de verdadeiros negativos, sendo baseada na matriz de confusão. O espaço onde é esboçada a curva é chamado de espaço ROC.

Aqueles pontos da curva que estão muito próximos do ponto $(0,0)$ do espaço ROC representam estratégias de classificação que raramente geram uma classificação positiva. Nesses pontos, temos um classificador que nunca comete erros indicando falsos positivos, mas que também nunca os classifica como positivos.

Aqueles pontos do espaço que estão próximos do ponto $(1,1)$ indicam uma estratégia oposta, pois classificam sempre como positivos, mas muitos erros são cometidos ao indicar muitos falsos positivos.

Aqueles pontos do espaço ROC que são próximos de $(0,1)$ são classificadores perfeitos, pois classificam quase todos os pontos como positivos e acerta praticamente todos. De modo geral, quanto mais próximo estiver desse ponto, melhor será o ponto do espaço ROC e melhor será como estratégia.

Aqueles classificadores que estão próximo do eixo TVP são chamados de classificadores conservadores porque classificam um item como positivo somente quando apresentam poucas chances de erro. Isso é refletido em taxas de falsos negativos menores. Os classificadores mais afastados do eixo TVP são denominados classificadores liberais porque fazem classificações positivas com poucas evidências, cometendo muitos erros.

Se um classificador é representado como um ponto sobre a reta que liga os pontos $(0,0)$ e $(1,1)$, este é dito aleatório. Qualquer ponto, representando um classificador, que esteja abaixo dessa reta indica que o classificador é pior do que um classificador aleatório. Isso indica que uma classificação completamente contrária à desse classificador produz resultados melhores.

Quando a comparação entre classificadores está restrita àqueles baseados num mesmo modelo binomial de regressão logística que se distinguem apenas pelo ponto de corte (valor a partir do qual uma observação é classificada como sucesso ou fracasso) adotado, então a seleção dos classificadores é feita por meio da determinação desse ponto de corte. O valor a ser escolhido a ser escolhido deverá ser o ponto sobre a curva que apresente maior proximidade com o ponto $(0,1)$ do espaço ROC.

O uso da curva ROC torna-se complexo quando os classificadores envolvem um modelo logístico cuja variável resposta apresenta n categorias, pois a matriz mostrada no

quadro 1 torna-se uma matriz $n \times n$. Nesse caso, a diagonal principal apresenta um tamanho n e há $n^2 - n$ possíveis erros. A manipulação pode, então, ser feita considerando uma curva ROC para cada categoria, o que conduz a definição de um ponto de corte para cada categoria.

A seleção do ponto de corte para um modelo logístico multinomial pode ser bem mais simples se consistir na busca por bons valores usando um procedimento empírico de tentativas.

2.5. Modelos não paramétricos e semiparamétricos de regressão

Florencio (2010) ainda afirma que, na análise não-paramétrica, em contraste com o método paramétrico, estima-se uma função média sem referência a uma forma funcional estabelecida previamente e o pesquisador precisa apenas definir o espaço de funções apropriado aos dados disponíveis, ao qual espera que a função pertença. Esse autor ressalta que tal característica é importante porque muitas vezes a análise teórica não estabelece a estrutura funcional entre as variáveis ou estabelece apenas formas alternativas.

Para Aldman (1992), a regressão não-paramétrica é um conjunto de técnicas para ajustar uma curva quando há pouco conhecimento a priori sobre a forma da mesma. Ryan (2009) enfatiza essa definição, mas considera que, da mesma forma dos procedimentos não-paramétricos gerais, os métodos de regressão não-paramétrica não serão tão eficientes quanto os métodos paramétricos de regressão, onde um modelo apropriado é assumido previamente.

You e Zhou (2006) ressaltam que, embora modelos paramétricos tenham sido amplamente utilizados, a sua má especificação pode ter um custo alto ao incluir parâmetros ineficientes ou até mesmo inconsistentes. Eles declaram que a evolução em termos de computação tem levado a um aumento do interesse e dos esforços para o desenvolvimento e aplicação de modelos não paramétricos nas três últimas décadas. Segundo esses autores, muitas técnicas úteis têm sido desenvolvidas tais como a suavização *kernel*, a aproximação por séries, a regressão *splines*, a regressão polinomial local, dentre várias outras. Consideram ainda que essas e outras abordagens não paramétricas são úteis para explorar estruturas ocultas dos dados, fornecendo por meio da flexibilidade conclusões bastante concisas.

Yao (2012) considera que modelos de regressão não paramétrica fornecem flexíveis e poderosas ferramentas para estimação de funções de regressão. Ele destaca o fato desses

modelos serem atualmente usados em vários campos e também cita como os métodos mais usados a regressão *kernel*, *splines*, polinomial local e por séries ortogonais.

Duran, Akdeniz e Hu (2011) tratam da eficiência da estimação de modelos semiparamétricos. Eles chamam a atenção para o fato de que, desde que foram propostos, esses modelos passaram a ser amplamente estudados na literatura e vêm recebendo uma grande quantidade de pesquisas nas últimas duas décadas. Uma das razões desse interesse, segundo eles, decorre da vantagem desses modelos serem muito mais flexíveis do que o modelo linear padrão por combinarem componentes paramétricas e não paramétricas. Outra vantagem dos modelos semiparamétricos é a que torna mais fácil interpretar o efeito de cada variável quando comparada aos modelos não paramétricos.

Segundo Florencio (2010), verifica-se que os modelos de regressão semiparamétrica são adequados para situações onde, dentre k variáveis analisadas, há conhecimento a priori da forma funcional da relação entre a variável dependente e g das k variáveis independentes. Para as $k - g$ variáveis restantes, nenhum conhecimento há sobre a forma funcional.

O mesmo autor complementa sua abordagem ao afirmar que a regressão semiparamétrica permite a modelagem de formas funcionais complexas que contemplam simultaneamente componentes paramétricos e não-paramétricos. Para ele, trata-se de uma alternativa mais flexível do que a abordagem clássica e menos restritiva para a estimação de uma curva desconhecida.

Boente e Rodriguez (2010) declaram que modelos semiparamétricos podem ser introduzidos quando um modelo linear é insuficiente para explicar a relação entre a variável dependente e as variáveis independentes associadas. Os autores supracitados ainda afirmam que essa abordagem tem sido usada para estender modelos lineares generalizados e para permitir que a maior parte dos preditores seja modelado linearmente, enquanto um número menor dela entre no modelo não parametricamente.

2.5.1. Modelos não paramétricos *scatterplot*

Histie e Tibishirani (1987) indicam que um modelo de regressão não paramétrica *scatterplot* pode ser usado para detectar desvios da linearidade no comportamento dos dados ou sugerir uma forma paramétrica adequada a ser utilizada na modelagem da relação entre

uma variável resposta e uma covariável. Alternativamente, o próprio modelo não paramétrico pode ser usado para fazer previsões.

Hardle, Muller e Sperlich (2008) afirmam que, num modelo de regressão não paramétrica *sacatterplot*, procura-se relacionar uma variável dependente Y a uma variável X . Essa relação é feita por uma determinada função, como indicado na equação (2.42).

$$Y = m(x) + \varepsilon \quad (2.42)$$

Em muitos casos, a teoria não coloca nenhum tipo de restrição na forma da função $m(\cdot)$, ou seja, se a mesma é linear, quadrática, logarítmica, etc. Por isso, a análise empírica utiliza os dados para encontrar $m(\cdot)$ (HARDLE, MULLER e SPERLICH, 2008). Nos casos em que a curva da função que relaciona o regressando e os regressores são limitados, a regressão não paramétrica apresenta-se como uma alternativa para a análise dos dados.

A forma da curva em um modelo não paramétrico, como o mostrado na equação (2.42), é determinada por meio de métodos de suavização. Formalmente, um suavizador é uma função de Y e X , cujo valor possui o mesmo domínio da variável X . Em termos práticos, a suavização é realizada por uma matriz de suavizadores que representa a matriz chapéu do modelo (*hat matrix*).

Na literatura estão disponíveis vários tipos de suavizadores. Alguns dos suavizadores são descritos sucintamente a seguir:

- **Suavizadores de banda (*Bin smoothers*)**: também conhecido como regressograma de um suavizador categórico por particionamento; calcula a média da variável resposta em cada região. No uso dos suavizadores de banda, são escolhidos pontos de corte $c_0 < c_1 < \dots < c_k$ e define-se a região onde se encontram os valores x_i , ou seja, $c_k < x_i < c_{k+1}$. Nesse método de suavização, definem-se regiões (bandas) de forma que o mesmo número de pontos esteja presente em cada uma das regiões.
- **Suavizadores de média móvel e de linha móvel (*running-mean smoothers e running-line smoothers*)**: No método de suavização *running-mean* escolhe-se um dos valores de x disponíveis entre os dados e calcula-se a média para uma vizinhança próxima. Normalmente escolhe-se uma região simétrica onde o x_i é o centro da região. Esse procedimento é chamado de vizinhança simétrica mais próxima. O índice destes pontos mais próximos é denotado por $N^s(x_i)$ e os valores estimados para a variável dependente y para um dado ponto x_i é calculado como a média da região simétrica

mais próxima ($média(N^s(x_i))$). Nesse método a região não precisa necessariamente ser simétrica, havendo outras formas de definir a região. No método *running-line*, a média é estimada pela utilização do método dos mínimos quadrados usado para cada região. O parâmetro que controla a suavização da linha é o tamanho da região (k). Se for desejável que a região de pontos analisada seja dividida em n partes, e que cada parte contenha uma proporção w dos pontos, então o tamanho da região é dado por $k=(wn-1)/2$. O número de pontos na região $N^s(x_i)$ é denotado por $[N^s(x_i)]$.

- **Suavizadores Kernel (Kernel smoothers):** Nesse método utilizam-se pesos locais, definidos por uma função de ponderação (função *kernel*) para produzir uma estimativa de y para um valor alvo x_0 . São diversas as funções *kernel* que podem ser utilizadas (*gaussian kernel*, *Epanechnikov kernel*, *uniform kernel*, *quartic kernel*, *triangle kernel*, etc.). Sendo x_0 o valor alvo, c_0 uma constante de ponderação, x_j o j -ésimo ponto distante do valor alvo, λ o parâmetro que controla o quanto as ponderações devem ser concentradas próximas do valor alvo e $d(\cdot)$ a função *kernel* escolhida, temos que os pesos atribuídos por $S_{0j}=(c_0/\lambda)*d(|(x_0-x_j)/\lambda|)$. Duas funções *kernel* muito usadas são o *Gaussian kernel* e o *Epanichnikov kernel*.
- **Suavizadores splines:** Nesse método o intervalo onde os dados referentes às variáveis independentes é dividido em partições e procura-se funções polinomiais para cada partição de modo que as mesmas se aproximem dos pontos presentes em sua partição. Na prática, o objetivo é encontrar uma função $f(x)$ com m derivadas contínuas que minimiza a soma do quadrado de resíduos penalizada. A primeira parcela da equação representa um critério de bondade do ajuste da função obtida aos dados. A segunda parcela representa uma penalização para falta de suavidade do ajuste a partir das derivadas da função f . Na equação, λ representa o parâmetro de suavização e $f^m(x)$ representa a derivada de ordem m da função estimada.
- **Suavizadores ponderados localmente (LOESS):** são baseados no ajuste sucessivo de polinômios pelo método dos mínimos quadrados ponderados usando pesos que são determinados por uma função específica. Cada modelo é ajustado considerando os valores de uma variável dependente y cujos valores de x pertencem a uma vizinhança de uma observação (x_i, y_i) especificada, chamada de ponto alvo. Ao considerar n observações de pontos alvo, obtém-se a curva suavizada cujos valores previstos são $\hat{f}(x_i)$.

Visto que os métodos de suavização usados neste trabalho foram os métodos de suavização *splines* e ponderados localmente, serão estes vistos de forma mais detalhada. A escolha por esses métodos foi feita por conveniência, já que eram aqueles disponíveis no comando do software R usado na estimação.

Suavização por funções base *splines*

Um método comum de suavização por funções base consiste em representar uma função como uma combinação linear de k funções base.

$$x(t) = \sum_{k=1}^K \beta_k \phi_k(t), \quad (2.43)$$

onde β_i representa o i -ésimo coeficiente e $\phi_i(t)$ representa a i -ésima função base.

Na forma vetorial, a equação (2.43) pode ser reescrita como (2.44).

$$x(t) = \Phi \beta, \quad (2.44)$$

sendo ambos os vetores de dimensão k .

Nesse método, o grau de suavidade é dado pelo número de funções base utilizado. Quanto mais funções base forem utilizadas, melhor será a representação dos dados e pior será a capacidade de explicar o comportamento destes. Por outro lado, quanto menor for o número de funções base, melhor explicado será o comportamento das observações e pior será a representação dos dados.

Um modelo linear ordinário pode ser explicado como uma soma de funções base. Para isso, reescreve-se a soma do quadrado dos resíduos na forma da equação (2.45).

$$SQR = \sum_{i=1}^n \left[y_i - \sum_{k=1}^K \beta_k \phi_k(t) \right]^2 \quad (2.45)$$

Em termos matriciais, a equação (2.45) pode ser escrita na forma da equação (2.46).

$$SQR = (Y - \Phi\beta)'(Y - \Phi\beta) \quad (2.46)$$

A solução é então dada pela equação (2.47).

$$S = \Phi(\Phi'\Phi)\Phi' \quad (2.47)$$

Para um modelo linear ordinário temos $\Phi = X$ e a matriz da equação (2.47) é a matriz chapéu que fornece os valores ajustados para o modelo.

Modelos baseados no método dos mínimos quadrados ponderados também são comuns. O método propõe uma adaptação no critério da soma dos quadrados dos erros, incluindo um termo W (matriz de ponderação) de ponderação das observações como mostrado na equação (2.48).

$$SQR = (Y - \Phi\beta)'W(Y - \Phi\beta) \quad (2.48)$$

Nesse método, a matriz chapéu é dada pela equação (2.49).

$$S = \Phi(\Phi'\Phi)^{-1}\Phi'W \quad (2.49)$$

É importante salientar que as funções base apresentem características semelhantes às aquelas que possuem a função a ser estimada a partir delas. Isso significa, por exemplo, que funções monótonas devem ser representadas por bases com as mesmas características. A escolha adequada pode permitir obter funções que representem bem os dados usando menos funções base.

Existem vários tipos de funções base. Um tipo de função comumente utilizada em suavização como soma de funções base é aquela chamada de função *splines*. Esse tipo de função apresenta a forma da equação (2.50).

$$s(t) = \sum_{i=0}^{m-1} \theta_i t^i + \sum_{i=1}^k \delta_i (t - x_i)_+^{m-1} \quad (2.50)$$

Qualquer função splines é uma combinação linear de $m-k$ funções base. É fácil notar que, considerando um conjunto de nós $(x_1, x_2, \dots, x_k)$, as funções base são:

$$\{1, t, t^2, \dots, t^{m-1}, (t - x_1)_+^{m-1}, \dots, (t - x_k)_+^{m-1}\}$$

O método de suavização *splines* pode ser usado para estimar uma função não paramétrica *scatterplot* do tipo da equação (2.51).

$$y_j = x(t_j) + \varepsilon_j \quad (2.51)$$

O objetivo é equilibrar o *tradeoff* entre o viés e a variância. Deseja-se obter uma curva que apresente um ajuste bom. Porém, quanto melhor for o ajuste menor o viés e maior a variância. Quanto maior a variância, menos suave será a curva e mais difícil será interpretar o seu comportamento. Desse modo, procura-se mas que esse ajuste não seja tão bom ao ponto de prejudicar a interpretação do comportamento dos dados.

Uma maneira comum de quantificar a noção de não suavidade de uma função é utilizando o quadrado da segunda derivada desta função, chamado de curvatura de uma função. Assim, uma medida natural da não suavidade é obtida dividindo-se o domínio da função em intervalos de comprimento próximo de zero, calculando-se o quadrado da segunda derivada para cada intervalo e somando os resultados. Isso é equivalente a resolver a integral da equação (2.57)

$$p = \int_a^b f''(x)^2 dx \quad (2.52)$$

Os mínimos quadrados podem ser modificados para incluir a penalização.

$$SQRP = \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 + \lambda \int_a^b [f''(x)]^2 dx \quad (2.53)$$

A ordem da derivada usada no termo de controle da suavização na equação (p) está relacionada àquela ordem do polinômio que se deseja usar. Para polinômios cúbicos, que são

usados geralmente, a ordem 2 é adequada. Se for necessário usar polinômios de ordem mais alta, é preciso usadas derivadas também de ordem mais altas.

Um teorema proposto por De Boor (2001) diz que SQR é minimizado por um *splines* cúbico com nós nas observações. Por essa razão, são amplamente usados na estimação não paramétrica.

No que diz respeito ao cálculo, a técnica computacional mais comum para suavização *splines* é utilizar uma expansão por bases B-splines de ordem 4, com nós nos pontos de observação para minimizar SQR.

Como foi indicado anteriormente:

$$x(t) = \Phi\beta$$

Sem a penalização, a matriz chapéu que minimiza a SQR é dada pela equação (2.54).

$$S = \Phi(\Phi'W\Phi)^{-1}\Phi'W, \quad (2.54)$$

onde W é a matriz de ponderação usada para corrigir alguma relação de dependência entre as observações.

É possível reescrever o termo de penalização como se segue na equação (2.55).

$$\begin{aligned} p &= \int (x''(t))^2 dt = \int (c'\phi''(t))^2 dt = c' \int (\phi''(t))^2 dt \\ p &= c'R, \end{aligned} \quad (2.55)$$

de modo que:

$$R = \int (\phi''(t))^2 dt$$

A matriz chapéu que minimiza SQR é dada pela equação (2.56).

$$S = \Phi(\Phi'W\Phi + \lambda R)^{-1}\Phi'W \quad (2.56)$$

Suavização por ponderação polinomial local

No método de ponderação local, é definido um número de pontos $r = \lambda n$ a ser considerado, onde λ é o parâmetro de suavização que define a proporção dos n pontos da amostra que serão usados numa estimação local de uma vizinhança.

Não existe um critério rígido para a escolha de λ , sendo comum a escolha empírica do valor do parâmetro. O parâmetro determina o viés de um modelo não paramétrico e sua variância, controlando o *tradeoff* entre essas duas propriedades. Em termos práticos, modelos com valor de λ mais altos apresentam-se como curvas menos onduladas e adequadamente pouco adaptadas aos valores mais altos das observações da variável resposta, enquanto modelos com valores mais baixos de λ geram curvas muito onduladas e com menos capacidade de captar tendências.

O grau do polinômio usado na ponderação local depende do comportamento dos pontos. Se os pontos presentes em um gráfico de dispersão mostram um comportamento onde há tendências com pontos de máximos e mínimos locais, então se adotam polinômios com grau maior ou igual a 2. No entanto, se um gráfico de dispersão mostra ausência de pontos críticos (máximos ou mínimos locais) então se adota um polinômio de grau 1.

Sendo (x_i, y_i) o ponto alvo, (x_j, y_j) o j -ésimo ponto distante do ponto alvo e h_i é o valor que ocupa a r -ésima posição da sequência crescente ordenada da diferença $|x_i - x_j|$, a função usada para atribuir os pesos usados no ajuste de cada polinômio local é dada pela equação (2.57).

$$u_{ij} = U(h_i^{-1}(x_j - x_i)) \quad (2.57)$$

A função $U(g)$ usada pode ser a função tricúbica, que é muitas vezes adequada. Essa função é mostrada na equação (2.58)

$$\begin{cases} U(g) = (1 - |g|^3)^3, & \text{para } |g| < 1 \\ U(g) = 0 & , \text{para } |g| \geq 1 \end{cases} \quad (2.58)$$

Essa função fornece os valores para uma matriz diagonal de pesos referentes ao ponto alvo, denotada por:

$$U = \begin{bmatrix} u_{i1} & 0 & \cdots & 0 \\ 0 & u_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & u_{in} \end{bmatrix}$$

Sendo assim, a observação no ponto alvo recebe um peso igual a 1 e a medida que os pontos são mais afastados atribui-se um peso cada vez menor. Utilizando-se os valores da matriz U e os valores da matriz de dados da variável independente, são obtidos os valores suavizados $\hat{f}(x_i)$. A matriz de suavizadores é calculada pela equação (2.59).

$$S = X(X'UX)X'U \quad (2.59)$$

Sendo Y a matriz de valores observados, a matriz dos valores preditos pelo modelo é dada pela equação (2.60).

$$\hat{Y} = S'Y \quad (2.60)$$

Estimação do parâmetro de suavização

Para Souza (2008), há duas abordagens para a escolha do parâmetro de suavização λ . Em uma dessas abordagens, considera-se que a escolha do parâmetro de suavização de forma livre é uma parte importante do procedimento. Nessa abordagem, diferentes parâmetros são usados e opta-se por aquele que melhor ajustou a função estimada aos dados.

A outra abordagem, de acordo com a mesma autora, considera importante que a escolha do parâmetro λ seja feita automaticamente com base nos dados. Um procedimento automático muito utilizado para a seleção dos parâmetros de suavização em um MAG com p termos não paramétricos é o *cross validation* (validação cruzada).

Nesse método, o objetivo é encontrar os valores dos parâmetros de suavização que minimizem o valor da função VC, expressa na equação (2.61). Gonçalves (2009) ressalta, porém, que esse método requer um custo computacional muito alto já que são necessárias as repetições de um procedimento de estimação iterativo (como o Local Scoring) para cada valor fixado dos parâmetros de suavização.

$$VC = \frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_\lambda^{-1}) \quad (2.61)$$

Hastie e Tibshirani (1990) citam como alternativa para seleção dos parâmetros de suavização o uso do Critério de Informação de Akaike (AIC), adaptado para a situação de interesse. Gonçalves (2009) ressalta que a seleção por meio do AIC representa um ganho computacional já que apenas uma interação para cada parâmetro é necessária.

$$AIC = \frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_\lambda^{-1}) - \frac{2}{n} \text{tr}(R)\Phi \quad (2.62)$$

2.5.2. Modelos aditivos generalizados

Para Stein (2010), os modelos aditivos consistem em somas de funções não paramétricas que não têm sua forma especificada e que são estimadas a partir de curvas de suavização.

Florencio (2010) afirma que a regressão semiparamétrica tem uma relação intrínseca com os modelos aditivos. Estes, por sua vez, são uma forma de generalização de modelos lineares, pois nos modelos aditivos o efeito incremental de uma variável independente é aditivo em relação ao das outras variáveis. Isso indica que em modelos aditivos é possível analisar a contribuição de cada variável na explicação da variável dependente de interesse. Nessa classe de modelos, três abordagens podem ser seguidas: estritamente paramétrica, estritamente não paramétrica e semiparamétrica.

Como é feito em qualquer modelo de regressão, os modelos aditivos podem ser representados, para o caso onde existe um conjunto de variáveis $(X_1, X_2, \dots, X_d)$, pela equação (2.63).

$$Y_i = f(X_1, X_2, \dots, X_d) + \varepsilon_i \quad (2.63)$$

Para Hastie e Tibshirani (1990), a relação entre a variável dependente Y e cada uma das variáveis independentes pode ser representada por uma função que é independente das demais (equação 2.64).

$$E(Y / X_1, X_2, \dots, X_d) = f(X_1, X_2, \dots, X_d) = f(X_1) + f(X_2) + \dots + f_d(X_d) \quad (2.64)$$

Se no modelo não são estimados parâmetros, para nenhuma das d funções, o modelo da equação anterior é dito não paramétrico. Cada uma das funções, que são elementos aditivos do modelo, é estimada separadamente por um procedimento de suavização.

Quando os modelos aditivos combinam formas paramétricas para um número g de variáveis e não-paramétricas para um número $k - g$ de variáveis, como na equação (2.65), o mesmo modelo é dito aditivo semiparamétrico.

$$E(Y / X_1, X_2, \dots, X_d) = \beta_0 + \beta_1 X_1 + \dots + \beta_g X_g + f_1(X_{g+1}) + \dots + f_{d-g}(X_d) \quad (2.65)$$

Hardle *et. al.* (2001) apresentam uma forma compacta da generalização dos modelos aditivos (equação 2.66), sendo $G(\cdot)$ a função de ligação usada no modelo e α o intercepto.

$$E(Y / X; T) = G(X^T \beta + \alpha + m_1(T_1) + \dots + m(T_d)) \quad (2.66)$$

A equação pode ser escrita também na forma mostrada na equação (2.67). Sendo $g(\cdot)$ a função de ligação e μ o valor esperado da variável resposta. O intercepto nessa equação é representado por β_0 .

$$\mu = g(\beta_0 + \beta_1 X_1 + \dots + \beta_r X_r + f_1(X_{r+1}) + \dots + f_{d-r}(X_d)) \quad (2.67)$$

Underwood (2009) afirma que a diferença essencial entre os modelos aditivos generalizados e os modelos lineares generalizados está nos termos da componente sistemática. Segundo esse autor, o efeito provocado sobre uma variável dependente a partir de uma variação em uma variável independente é definido com base no parâmetro estimado nos MLG's, enquanto nos MAG's esse efeito também é definido por funções suavizadas.

De acordo com Bock, Coussement e Poel (2010), os MAG's são generalizações dos MLG's e constituem uma técnica estatística flexível e poderosa, que tem seu sucesso comprovado em capturar relações não lineares entre variáveis exploratórias e a variável resposta em diversos domínios. Hastie e Tibishirani (1990) também destacam que os MAG's

generalizam os MLG's por substituírem formas funcionais lineares por funções suavizadas, possibilitando a descoberta de ajustes não lineares entre regressando e regressores.

Young *et. al.* (2011) complementam a abordagem anterior ao ressaltar que os MAG's têm mostrado vantagens distintas quando comparados aos MLG's, por permitir que investigadores façam inferências sobre associações entre variáveis sem colocar restrições paramétricas sobre as mesmas.

Hastie e Tibishirani (1990) declaram que os MAG's são extensões semiparamétricas dos MLG's e que, como os MLG's, usam uma função de ligação para estabelecer uma relação entre a média da variável resposta e funções suavizadas de variáveis exploratórias. Para tais autores, a força dessa técnica está na habilidade de tratar com relações muito distintas daquelas lineares e monotônicas entre a média de uma variável dependente e funções suavizadas de variáveis independentes.

De acordo com Yee e Wild (1996), MAG's em vez de restringir a relação entre a variável resposta e as covariáveis a forma linear como ocorre nos MLG's, restringem a relação a uma relação suave. Isso, segundo os autores, permite que as características não lineares dos dados se revelem para o analista.

Guisan, Edwards e Hastie (2002) afirmam que os MAG's são extensões semiparamétricas dos MLG's, possuindo como únicos pressupostos a aditividade dos termos e a suavidade das funções não paramétricas das covariáveis. Acrescentam que a força dos MAG's está na sua habilidade de lidar com relações não lineares e não monotônicas entre a variável resposta e as covariáveis.

Estimação

Um modelo aditivo generalizado apresenta a forma da equação (2.68).

$$\mu = g(\beta_0 + \beta_1 X_1 + \dots + \beta_r X_r + f_1(X_{r+1}) + \dots + f_{d-r}(X_d)) \quad (2.68)$$

Os procedimentos usados na estimação dos MAG's são análogos àqueles usados para estimação dos MLG's. Porém, algumas modificações são feitas de forma que as funções f sejam estimadas a partir da utilização de suavizadores. Um procedimento de estimação usado para ajustar um modelo aditivo generalizado é uma combinação de dois procedimentos

iterativos. Esses procedimentos são o procedimento de ponderação local (*local scoring* ou PPL) e o retroajuste (*backfitting*).

O uso da combinação desses dois procedimentos gera os mesmos resultados de um procedimento semelhante ao *IRWLS* usado na estimação dos MLG's (GONÇALVES, 2009). Esse procedimento é chamado de *IRWLS* com penalização (*P-IRWLS*). A autora descreve como é feita a estimação pelo *P-IRWLS*.

O método é iniciado a partir do vetor inicial $\beta^{(0)}$ e valores iniciais para as funções não paramétricas $f_1^{(0)} = f_2^{(0)} = \dots = f_d^{(0)}$. Calcula-se em seguida a resposta modificada $z_i^{(m)}$ e os pesos $w_i^{(m)}$ na m-ésima iteração, usando as equações (2.74) e (2.75).

$$z_i^{(m)} = \eta_i^{(m)} + (y_i - \mu_i^{(m)}) \left(\frac{\partial(\mu_i)}{\partial \eta_i} \right)^{(m)} \quad (2.69)$$

$$w_i^{(m)} = \left(\frac{\partial(\mu_i)}{\partial \eta_i} \right) (V_i^{(m)})^{-1}, \text{ para } i=1,2,\dots,n. \quad (2.70)$$

Os valores $\eta_i^{(m)}$, $\mu_i^{(m)}$ e $V_i^{(m)}$ são obtidos a partir das equações (2.71), (2.72) e (2.73). É importante notar que $\eta_i^{(m)}$ para os MAG's não é um preditor linear. Rigby e Stasinopoulos (2009) utilizam η para representar um preditor, seja o mesmo linear ou não. Lima, André e Singer (2006) denominam preditor aditivo o termo η quando utilizado em MAG's. As derivadas de primeira ordem usadas no *P-IRWLS*, quando utilizada uma função de ligação logit (ligação canônica), podem ser determinadas utilizando a função de ligação. Essas derivadas poderão ser escritas, nesse caso específico, em termos do preditor $\eta_i^{(m)}$.

$$\eta_i^{(m)} = \beta_0^{(m-1)} + \sum_{j=1}^d \beta_j^{(m-1)} x_{ij} + \sum_{j=r+1}^d f_j^{(m-1)}(x_{ij}) \quad (2.71)$$

$$\mu_i^{(m)} = g^{-1}(\eta_i^{(m)}) \quad (2.72)$$

$$V_i^{(m)} = \left(\frac{\partial(\mu_i)}{\partial \theta_i} \right)^{(m)} \quad (2.73)$$

No método, os valores obtidos para $z_i^{(m)}$ e $w_i^{(m)}$ em cada iteração são usados juntamente com a matriz de dados X , nas equações (2.74) e (2.75), para obter o vetor de coeficientes $\beta^{(m)}$ e as funções não paramétricas $f_j^{(m)}$.

$$\beta^{(m)} = (X'W^{(m)}X)^{-1}X'W^{(m)}\left(z^{(m)} - \sum_{i=r+1}^d f_i^{(m)}\right) \quad (2.74)$$

$$f_j^{(m)} = S_j^{(m)}\left(z^{(m)} - X\beta - \sum_{\substack{i=r+1 \\ i \neq j}}^d f_i^{(m)}\right), \text{ sendo } j = r+1, \dots, d \quad (2.75)$$

A equação usada para determinação da matriz $S_j^{(m)}$ de suavizadores em cada iteração é especificada quando se escolhe um método de suavização, tal como o de ponderação local ou como o por funções base *splines*.

O critério de parada das iterações pode ser pré-definido. Define-se previamente um valor para $\delta > 0$, e as iterações são interrompidas quando a desigualdade das equações (2.76) e (2.77) é satisfeita.

$$\frac{\sum_{j=r+1}^d \|f_j^{(m)} - f_j^{(m-1)}\|}{\sum_{j=r+1}^d \|f_j^{(m-1)}\|} \leq \delta \quad (2.76)$$

$$\frac{\sum_{k=1}^r \|\beta_k^{(m)} - \beta_k^{(m-1)}\|}{\sum_{k=1}^r \|\beta_k^{(m-1)}\|} \leq \delta \quad (2.77)$$

Testes de significância e *pseudo-R*²

Wood (2006, p.163) mostra que um MAG pode ser visto como um MLG e afirma que isso torna possível construir uma função de verossimilhança para aquele tipo de modelo. Isso permite estimar o modelo utilizando o método da máxima verossimilhança, acrescentando um termo de penalização na função de verossimilhança. As penalidades usadas na função devem

ser projetadas para que a falta de suavidade das estimativas f_j sejam reduzidas. Para ele, sendo $l(\beta)$ a função de verossimilhança do vetor de parâmetros do modelo, a função de verossimilhança penalizada pode ser escrita na forma da equação (2.78).

$$l_p(\beta) = l(\beta) - \frac{1}{2} \sum_j \lambda_j \beta^T S_j \beta \quad (2.78)$$

, onde S_j é a matriz de suavização usada para estimar os valores da j -ésima função não paramétrica do modelo e λ_j é o parâmetro de suavização usado para essa função.

Sendo esta uma função de verossimilhança do MAG, é possível utilizá-la para determinar uma estatística *Deviance* relacionada. Por meio dessa estatística, utilizam-se alguns métodos similares àqueles usados para testar a significância dos coeficientes de MLG's, como o teste χ^2 . Além disso, é possível calcular também o valor do pseudo- R^2 conforme sugere Faraway (2006).

Wood (2006) ainda acrescenta que os termos não paramétricos do preditor de um MAG podem ser expressos na forma de termos lineares. Esse autor sugere que uma função não paramétrica pode ser expressa usando um conjunto de funções base, b_{ji} , de modo que esta se apresentem de uma forma semelhante àquela da equação (2.84), sendo $\tilde{\beta}_{ij}$ os coeficientes das bases de suavização. Desse modo, para ele, é fácil criar uma matriz de dados $\tilde{X}_j$ para cada suavizador de maneira que a equação (k) possa ser reescrita como na equação (2.79).

$$f_j(x_j) = \sum_{i=1}^q \tilde{\beta}_{ij} b_{ji}(x_j) \quad (2.79)$$

$$f_j = \tilde{X}_j \tilde{\beta}_j \quad (2.80)$$

, onde $\tilde{X}_{j,ik} = b_{jk}(x_{ji})$

Sob a forma da equação (2.80), é possível testar a significância de uma função não paramétrica, verificando a significância do coeficiente $\tilde{\beta}_{ij}$. Se não é possível rejeitar a hipótese de que o coeficiente é diferente de zero, então a relação entre a variável resposta e a função base da covariável não é significativa.

CAPÍTULO 3 - METODOLOGIA

3.1. Classificação da pesquisa

No que se refere à natureza da pesquisa, este trabalho se classifica como uma pesquisa básica, uma vez que não tem finalidades imediatas. Quanto à forma de abordagem, a pesquisa é classificada como quantitativa, pois procura traduzir opiniões e características em termos quantitativos com o objetivo de obter um modelo de regressão que forneça uma medida de probabilidade (GIL, 1994).

Uma pesquisa que procura identificar os fatores que estão relacionados com uma variável de interesse pode ser classificada como explicativa. Se uma pesquisa procura manipular os fatores relacionados a uma variável de interesse para obter informações sobre esta última, pode ser caracterizada como uma pesquisa explicativa e experimental (FORTIN, 2009). Assim, do ponto de vista dos objetivos, o estudo é classificado como explicativo e experimental, já que busca identificar fatores que estão relacionados com a intenção de compra e manipular esses fatores de controle para obter informações sobre consumidores ou mesmo sobre segmentos de mercado.

3.2. Estudo de caso

Como foi discutido na seção 2.4, a intenção de compras é um sinal de que um determinado indivíduo pode vir a consumir um produto ou serviço. Esse sinal pode ser uma declaração (em uma pesquisa de mercado) de que consumiria determinado produto, ou mesmo uma tendência identificada a partir do histórico de vendas presente em um banco de dados de um sistema de informação. É expressa comumente em termos de probabilidade, mas uma medida de previsão do consumo de um bem também pode expressar a intenção de compra. Desse modo, há distintas formas de obter informações sobre a intenção de compra de indivíduos e a mesma pode também ser expressa de diferentes maneiras.

Nesta seção serão indicadas as fontes de informação usadas nessa pesquisa para obter um modelo de avaliação da intenção de compra de consumidores do serviço de compras coletivas na cidade de João Pessoa. Será abordada também a aplicação do modelo e serão definidas algumas características do sistema de produção de interesse da pesquisa.

3.2.1. O modelo de avaliação da intenção de compras e a fonte de informação

O estudo de caso adotou como fonte de informações uma pesquisa de mercado realizada junto a pessoas da cidade de João Pessoa. A intenção de compras é expressa em termos de probabilidade de um determinado indivíduo comprar um serviço específico, que é o serviço de compras coletivas pela internet. Embora o modelo de regressão obtido seja orientado para esse caso específico, não há nenhuma razão a priori para afirmar que um modelo semelhante não possa ser obtido para outras situações, com a intenção de compra expressa de maneira distinta e utilizando fontes de informações sobre o perfil de clientes que sejam diferentes daquela usada como base para esse trabalho.

Este estudo de caso consistiu em utilizar informações, obtidas a partir de uma pesquisa de mercado, para obter um modelo de regressão que possa identificar qual o perfil das pessoas que tendem a consumir produtos ou serviços através dos sites de compras coletivas. A pesquisa de mercado foi baseada no questionário do Apêndice C.

Procurou-se identificar aquelas características que devem ter os serviços de compras coletivas que atuam na cidade de João Pessoa, para que seja disponibilizado um serviço em conformidade com o que desejam as pessoas que tendem a ser clientes da empresa.

O modelo de regressão resultante poderá também orientar as empresas a darem maior atenção para aquelas variáveis que são mais determinantes na diferenciação do perfil das pessoas que tendem a comprá-las daquele das pessoas que, embora opinem sobre suas preferências, dificilmente iriam consumir o serviço. Essas variáveis são relevantes porque podem direcionar o projeto e a gestão do serviço de compra coletiva atuante em João Pessoa para o melhor aproveitamento dos seus recursos disponíveis, evitando investimentos desnecessários para melhorar o serviço para pessoas que dificilmente o consumiriam e direcionando os mesmos para indivíduos que tendem a consumir os produtos.

3.2.2. O sistema de produção do serviço de compras coletivas pela internet

O serviço de compras coletivas tem como produto final o fato de disponibilizar ofertas a clientes de vários serviços. Para obter esse produto final, os sites utilizam as pessoas da empresa para fazerem contatos com empresas e negociarem os descontos. Uma vez obtidos os descontos desejáveis junto às empresas, as ofertas são disponibilizadas nos sites e divulgadas via internet, através de vários outros sites, correio eletrônico e sites de relacionamentos. Após

a compra da oferta (produto ou serviço oferecido com o desconto), o serviço disponibiliza cupons que podem ser usados para o consumo nas empresas que possuem a oferta.

Então, o serviço utiliza os seus recursos humanos para entrar em contato com empresas, para fazer divulgações, negociar descontos, definir as características dos sites, dentre outras funções.

Um subsistema de controle informal acaba existindo com o objetivo de conciliar a demanda por descontos em produtos ou serviços específicos. Pressupõe-se que, se o número de ofertas no nível que o cliente deseja é insuficiente para suprir a demanda, então haverá um risco da perda de clientes para empresas concorrentes e, se muitas ofertas são divulgadas e a demanda é baixa, então recursos são usados desnecessariamente na divulgação e negociação com os fornecedores, podendo os fornecedores disponibilizar as ofertas em outros sites que sejam mais eficazes na obtenção de consumidores para os produtos ou serviços ofertados.

O processo de produção do serviço de compra coletiva consiste em toda a dinâmica realizada para disponibilizar as ofertas para os clientes. Considera-se que esse processo de produção pode ser dividido em três etapas básicas: identificação dos fornecedores, negociação de descontos e divulgação das ofertas. A receita gerada pelo sistema de produção vem da cobrança de um percentual das vendas de ofertas.

Pressupõe-se, ainda, que a qualidade do serviço prestado é um ponto fundamental para que uma empresa de compras coletivas mantenha-se no mercado. Em termos de processo, essa qualidade pode ser medida em termos da conformidade das informações fornecidas, sobre as ofertas, com o que de fato é disponibilizado por empresas. Em termos de cliente, a qualidade representa principalmente a disponibilidade de níveis de descontos mínimos requeridos para linhas de produtos ou serviços desejados.

3.3. Definição da amostra e das variáveis de pesquisa

O estudo foi baseado num questionário de pesquisa de mercado objetivando coletar informações acerca de suas características, preferências e opiniões, que possam estar relacionadas com a utilização do serviço em estudo.

A população estudada envolve os usuários de internet da cidade de João Pessoa-Paraíba-Brasil, com idades entre 18 e 59 anos, com renda superior a R\$ 620,00. A população da cidade na faixa etária considerada é de 343.654 pessoas, segundo o censo de 2010 do IBGE.

Foi escolhida a parcela da população de João Pessoa com essas características porque se tomou como pressuposto que as pessoas que menos utilizam a internet e que possuem menos liberdade e poder de compra para decidir comprar estão nas faixas de idade e renda que não foram consideradas.

Devido às restrições de tempo, a dificuldade de encontrar participantes com determinados perfis (tais como usuários de internet com mais de 40 anos que apresentassem rendas mais baixas) e a dificuldade de encontrar informações sobre a população relativas a variáveis que deveriam ser usadas para estratificação (como a renda familiar por indivíduo), optou-se por utilizar um método de amostragem por conveniência.

Deve-se ressaltar que o principal objetivo dessa pesquisa é sugerir um procedimento que possibilite modelar a relação entre a intenção de compra e variáveis ligadas ao comportamento do consumidor, de modo que seja possível utilizar o modelo resultante para orientar o projeto e gestão de sistemas produtivos levando em conta o nível e interesse de clientes por um produto ou serviço disponibilizado.

O tipo de modelo resultante tem como base uma abordagem teórica que está consolidada no meio acadêmico e industrial para análise do comportamento dos consumidores (abordagem cognitiva do comportamento do consumidor) e está respaldado em vários artigos, apresentados na fundamentação teórica, que sugerem a existência de relações entre a intenção de compra e variáveis ligadas ao comportamento de clientes.

A principal função da amostra coletada foi mostrar como a modelagem sugerida poderia ser aplicada com dados reais. Outra função importante foi apresentar a utilização de um ferramental estatístico disponível para buscar ajustes melhores a dados coletados, de forma que fosse possível viabilizar a predição de perfis de clientes que tendiam a consumir um produto ou serviço do sistema produtivo. Sendo estas as principais funções do uso da amostra na pesquisa, a amostragem por conveniência se justifica.

Mesmo sendo uma amostra obtida por conveniência, procurou-se determinar um tamanho amostral adequado para representar a população em termos numéricos. Martins (2002) apresenta a equação (3.1) disponibilizada para o cálculo de uma amostra para população finita.

$$n = \frac{z^2 p(1-p)N}{e^2(N-1) + z^2 p(1-p)} \quad (3.1)$$

Onde z é a abscissa da distribuição normal padrão, p é a proporção de indivíduos que tem uma dada característica, N é o tamanho da população e e é o erro máximo tolerável. No caso da presente pesquisa a proporção p de interesse é a proporção de pessoas que marcarão uma alternativa no questionário. Há várias alternativas nesse questionário e nenhuma informação a priori é conhecida acerca das mesmas. Por essa razão, foi usada a proporção que fornece o máximo tamanho da amostra, que é 0,5.

A equação de determinação da amostra fornece um tamanho de 383 observações para um nível de significância de 0,95. Com base nas variáveis renda, sexo e idade foram definidos o tamanho de oito estratos, mostrados na tabela 3.

Outra questão analisada, no que diz respeito à amostra, foi o número de observações mínimas que precisam ser coletadas para que possa ser usado o método de análise de regressão. Hill e Hill (2010) sugerem a utilização de uma regra prática muito usada por pesquisadores experientes. Para a análise de regressão, o número de observações precisa ser no mínimo igual a cinco vezes o número de variáveis incluídas nos modelos. Se 25 variáveis são consideradas na análise, o tamanho mínimo tolerável da amostra é 125. Esse número é inferior ao definido anteriormente na estimação das proporções das respostas.

As variáveis que poderiam ser consideradas na pesquisa foram listadas a partir de consultas a periódicos que tratavam das variáveis determinantes da intenção de compra de consumidores em Websites (Dehua, Yaobin & Deyi, 2008; Verhagen & Dolen, 2009; Chen, Hsu & Lin, 2010; Lu & Hsiao, 2010; Vijayasarathy, 2004; Shim et al, 2001; Xu, Summers & Belleau, 2004; Belanche, Casalo & Guinalíu, 2012; Kim & Kim, 2004; Lin, 2007; Hausmann & Siekpe, 2009).

Em seguida, foram feitas adaptações nas variáveis de maneira a adequá-las aos objetivos da pesquisa. Posteriormente, foram selecionadas aquelas que, a priori, poderiam estar mais relacionadas à decisão de compra e que apresentavam maior importância para o projeto e planejamento do sistema de produção do serviço de compras coletivas pela internet.

As variáveis selecionadas para este estudo são listadas no quadro 3, que apresenta a escala considerada, o item do questionário usado na coleta das observações, a classificação da variável e a simbologia usada para representar as mesmas.

Ao final, foram escolhidas 25 variáveis, sendo duas variáveis métricas discretas, 15 variáveis categóricas ordinais, três categóricas nominais e cinco variáveis categóricas dicotômicas.

Quadro 3 - Variáveis selecionadas para o estudo

Item do Questionário	Variáveis	Símbolo	Tipo de variável	Escala
4.9	Intenção de compra	IC	Ordinal	(3) Certamente compraria; (2) Provavelmente Compraria; (1) Provavelmente não compraria; (0) Certamente não compraria
2.1	Sexo	Sx	Dicotômico	(0) Feminino; (1) Masculino
2.3	Idade	Id	Discreta	-
2.4	Escolaridade	Aesc	Discreta	-
2.5	Renda	Rd	Contínua	-
2.6	Número de dependentes da renda domiciliar	Dep	Discreta	-
2.5/2.6	Renda por Dependente	Rddep	Contínua	-
2.2	Estado civil	Ec	Nominal	(1) Solteiro; (2) Casado; (3) Divorciado; (4) Viúvo
2.7	Bairro	Bair	-	-
3.1	Importância da facilidade de obter ajuda no site	Faj	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância
3.2	Importância da facilidade de navegação	Fan	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância
3.3	Importância da quantidade de informação sobre o produto	QInfo	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância
3.4	Importância da aparência do site	Ap	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância
3.5	Importância das empresas parceiras serem conhecidas	Econ	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância
3.6	Importância do site anunciante da oferta ser conhecido	Acon	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância
3.7	Importância da qualidade do serviço de atendimento ao consumidor	Qsac	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância
3.8	Importância da variedade de produtos ou serviços ofertados	Vprod	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância

Fonte: elaboração própria.

Continuação

Item do Questionário	Variáveis	Símbolo	Tipo de variável	Escala
3.9	Importância do tempo que as ofertas ficam disponíveis	Tdisp	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância
3.10	Importância da variedade de formas de pagamento	Vpag	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância
3.11	Importância da velocidade de navegação do site	Vnav	Ordinal	(0) Sem importância; (1) De pouca importância; (2) Importante; (3) De grande importância
4.1	Frequência com que visita um site de compra coletiva	Visit	Nominal	(0) Nunca; (1) Poucas vezes no ano; (2) Todo mês; (3) Toda semana; (4) Todo dia
4.2	Tempo que fica na internet por dia	Tint	Contínua	-
4.3	Compra em lojas virtuais	Cmint	Dicotômica	(0) Nunca; (1) Pouco; (2) Bastante
4.4	Acesso a ofertas	Of	Dicotômica	(0) Não; (1) Sim
4.5.1	Acesso a ofertas diretamente no site	Dir	Dicotômica	(0) Não; (1) Sim
4.5.2	Acesso a ofertas por outros sites	Posit	Dicotômica	(0) Não; (1) Sim
4.5.3	Acesso a ofertas por redes sociais	Rsoc	Dicotômica	(0) Não; (1) Sim
4.5.4	Acesso a ofertas por email	Email	Dicotômica	(0) Não; (1) Sim
4.5.5	Acesso a ofertas pela comunicação com terceiros	Terc	Dicotômica	(0) Não; (1) Sim
4.5.6	Acesso a ofertas por sites de busca	Sbus	Dicotômica	(0) Não; (1) Sim
4.6.1	Consumo regular de serviços de lazer	Slaz	Dicotômica	(0) Não; (1) Sim
4.6.2	Consumo regular de produtos e serviços de estética	Estet	Dicotômica	(0) Não; (1) Sim
4.6.3	Consumo regular de eletro-eletrônicos	Eletro	Dicotômica	(0) Não; (1) Sim
4.6.4	Consumo regular de serviços de turismo	Tur	Dicotômica	(0) Não; (1) Sim
4.6.5	Consumo regular de vestuários	Vest	Dicotômica	(0) Não; (1) Sim

Conclusão

Item do Questionário	Variáveis	Símbolo	Tipo de variável	Escala
4.6.6	Consumo regular de livros	Liv	Dicotômica	(0) Não; (1)Sim
4.6.7	Consumo regular de revistas	Ver	Dicotômica	(0) Não; (1)Sim
4.8	Desconto mínimo que levaria o cliente a comprar	Desc	Contínua	-

Fonte: elaboração própria

O questionário (ver no Apêndice A) foi aplicado em pontos de grande circulação de pessoas em distintos locais da cidade de João Pessoa. Esses locais foram lojas, supermercados, shoppings, mercados públicos e praças. Os itens do questionário foram escolhidos de forma que fosse possível coletar dados sobre os fatores característicos dos indivíduos que influenciam na intenção de compra (fatores psicológicos, socioculturais e situacionais). A figura 21 apresentada no Apêndice E, mostra de que forma cada item contribuiu para obter esses dados.

O método usado na análise de dados foi a análise de regressão com modelos aditivos generalizados. Todos os modelos foram obtidos através do programa computacional R, versão 2.13.2 (CRAN,2012), e posteriormente foram selecionados os melhores, considerando critérios de ajuste aos dados.

CAPÍTULO 4 - RESULTADOS

4.1. Especificação e conclusões dos modelos

Utilizando todas as variáveis na forma em que aparecem no questionário, o melhor modelo obtido possui quatro variáveis e apresenta um *pseudo-R*² igual a 0,25. Para alcançar o objetivo da pesquisa era necessário obter um modelo com bom ajuste aos dados, de modo que fosse possível usá-lo para fazer previsões. Por isso, foi necessário utilizar alguns artifícios para obter um ajuste melhor. Alguns desses artifícios foram:

1. Agrupar categorias, transformando variáveis nominais ou ordinais em variáveis dicotômicas. Isso foi feito, por exemplo, para as variáveis *Faj*, *Cmint*, *Ap*, *Vprod* e *Visit*.
2. Definir um valor para algumas variáveis contínuas ou discretas de modo que divida o conjunto de valores dessa variável em duas categorias (dois intervalos), tornando-as dicotômicas. Isso foi feito para as variáveis *Rd*, *Dep* (discreta) e *Id*, pra obter *FRd*, *Dep* (dicotômica) e *FId*.
3. Combinar duas variáveis dicotômicas agrupando suas categorias de modo a obter uma terceira variável dicotômica. Esse artifício foi usado para obter *FDep*, *FId*, *FSx*, *FFan* e *FLiv*.
4. Utilizar como uma única variável contínua o produto ou a razão de duas variáveis quantitativas. Foram utilizados nos modelos como única variável “*Aesc/Id*”, “*Aesc*Rd*”, “*Rd/Tint*” e “*Aesc/Tint*”.

Utilizando, então, as mudanças nas categorias e as combinações de variáveis modelos mais bem ajustados puderam ser obtidos.

O quadro 3 mostra as variáveis obtidas a partir das variáveis originais presentes no questionário.

Quadro 4 - Variáveis obtidas por meio de mudanças e combinações das variáveis do questionário

Variáveis	Símbolo	Tipo de variável	Escala
Intenção de compra (2)	IC2	Dicotômica	(0) Certamente não compraria; (1) Provavelmente não compraria, provavelmente compraria ou certamente compraria
Intenção de compra (3)	IC3	Ordinal	(0) Certamente não compraria; (1) Provavelmente não compraria, provavelmente compraria; (2) Certamente compraria
Importância da facilidade de obter ajuda no site	Faj	Dicotômica	(0) Sem importância ou de pouca importância; (1) Importante ou de grande importância
Compra em lojas virtuais	Cmint	Dicotômica	(0) Nunca comprou em lojas virtuais; (1) Comprou em lojas virtuais
Importância da variedade de formas de pagamento	Vpag	Dicotômica	(0) Sem importância ou de pouca importância; (1) Importante ou de grande importância
Importância da aparência do site	Ap	Dicotômica	(0) Sem importância ou de pouca importância; (1) Importante ou de grande importância
Importância da variedade de produtos ou serviços ofertados	Vprod	Dicotômica	(0) Sem importância ou de pouca importância; (1) Importante ou de grande importância
Acessos anteriores a sites de compras coletivas	Visit	Dicotômica	(0) Nunca visitou sites de compra coletiva; (1) Visitou em sites de compra coletiva
Dependentes da renda domiciliar	Dep	Dicotômica	(0) É o único dependente de a renda domiciliar; (1) Há outros dependentes da renda
Fator Dependente	FDep	Dicotômica	(1) Tem renda domiciliar superior a R\$2.000,00 e é o único dependente dessa renda; (0) Não tem renda domiciliar superior a R\$2.000,00 e/ou não é o único dependente dessa renda
Fator Renda	FRd	Dicotômica	(0) Não possui renda superior a R\$2.000,00; (1) Possui renda superior a R\$2.000,00
Fator Idade	FId	Dicotômica	(0) Não possui mais de 35 anos e/ou não possui renda domiciliar superior a R\$2.000,00; (1) Possui mais de 35 anos e possui renda domiciliar superior a R\$2.000,00
Fator Sexo	FSx	Dicotômico	(0) Não é do sexo masculino e/ou não visita sites de compra coletiva com frequência semanal; (1) É do sexo masculino e visita sites de compra coletiva com frequência semanal
Fator Importância da Facilidade de navegação	FFan	Dicotômico	(0) Não possui renda domiciliar superior a R\$2.000,00 e/ou não considera a facilidade de navegação muito importante; (1) Possui renda domiciliar superior a R\$2.000,00 e considera a facilidade de navegação muito importante.
Fator Consumo de Livros	Fliv	Dicotômico	(0) Não possui renda domiciliar superior a R\$2.000,00 e/ou não compra livros regularmente; (1) Possui renda domiciliar superior a R\$2.000,00 e compra livros regularmente

Fonte: Elaboração própria

4.1.1. Modelo com variável resposta ordinal

Nessa subseção são apresentados, inicialmente, os modelos de regressão obtidos a partir dos dados e são expostas as conclusões, fornecidas a partir dos modelos, relativas à influência das covariáveis na variável resposta. Posteriormente indicam-se razões possíveis que sejam capazes de explicar a natureza das relações entre cada uma das covariáveis e a variável resposta.

O modelo que apresentou o melhor ajuste dentre aqueles que fornecem informações mais detalhadas sobre a intenção de compra foi um modelo de *odds* proporcional. O modelo apresentou um *pseudo-R*² igual a 0,525. O modelo indica as probabilidades de que a variável intenção de compras assuma uma das três categorias. Percebe-se que a única diferença entre as equações das probabilidades está nos interceptos.

$$P(IC3 \geq 1) = \frac{e^{-1,292+\gamma}}{1 + e^{-1,292+\gamma}}$$

$$P(IC3 \geq 2) = \frac{e^{-3,593+\gamma}}{1 + e^{-3,593+\gamma}}, \quad (4.1)$$

onde:

$$\gamma = 3,517 \left(\frac{Aesc}{Id} \right) - 0,0001(Aesc * Rd) + 0,93Terc + 2,151C \min t - 1,452FRd + 1,207Ap +$$

$$+ 1,089Vpag + 1,359Vprod - 0,951Dep + 1,024Visit - 0,848FFan + 0,947FLiv$$

Os valores da *odds ratio* (razão da chance) permitem obter conclusões sobre a natureza das relações entre cada covariável do modelo e a variável resposta. Os valores da *odds* são mostrados na tabela 1.

A influência da escolaridade na intenção de compra foi mais bem explicada em combinação com outras variáveis. Embora a razão da chance tenha sido maior para um dos termos que inclui essa variável, é preciso analisar o seu efeito de forma mais detalhada. É importante notar primeiramente que a idade sempre é superior aos anos de escolaridade. Sendo assim, a *odds* relacionada à escolaridade é sempre inferior a 33,68.

A *odds* do fator escolaridade-idade indica que quanto menor for a diferença entre a idade e os anos de escolaridade maior será a chance de um indivíduo ter um interesse maior em utilizar o serviço. Quanto mais tempo um indivíduo passa sem elevar seu nível de escolaridade, menos chances esse indivíduo terá de ter um nível de interesse maior no uso do serviço. Como

para a maior parte da população o nível de escolaridade é interrompido, o efeito que é mais comum é as chances de alguém usar o serviço diminuir com a idade.

Tabela 1 - Valores da *odds ratio* para o modelo multinomial ordinal

Variáveis	Odds
Aesc/Id	33,68
Aesc*Rd	0,99
Terc	2,53
Cmint	8,59
FxRd	0,234
Ap	3,34
Vpag	2,97
Vprod	3,11
Dep	0,39
Visit	2,78
Ffan	0,42
Fliv	2,58

Fonte: elaboração própria.

A escolaridade geralmente acompanha o aumento da idade para pessoas mais jovens. Isso também indica que pessoas mais jovens tem maior propensão a estarem mais interessadas pelo serviço de compras coletivas. Dentre as pessoas mais jovens, aquelas que possuem um nível de escolaridade mais alto têm maior intenção de utilizar o serviço.

Outro termo do modelo indica um efeito conjunto de outra variável com a escolaridade. O termo escolaridade-renda (*Aesc*Rd*) indica que dentre as pessoas com maior nível de escolaridade aquelas que possuem renda mais alta tendem a ter menor interesse nas compras em sites de compra coletiva. É importante notar que embora a razão da chance desse termo indique uma variação pequena (redução de 1%) na chance de um indivíduo interessar-se pelo serviço, para rendas mais altas (acima de R\$ 5.000,00) a redução na chance pode ser expressiva.

O efeito da renda não apenas existe, reduzindo o efeito da escolaridade. O termo da variável faixa de renda indica que pessoas com renda superior a R\$ 2.000,00 possuem uma chance menor de interessar-se pelo serviço de compras coletivas do que aqueles que possuem

renda inferior. Porém, esse efeito é pouco expressivo uma vez que não aumenta juntamente com a elevação da renda. Desse modo, o impacto maior da renda ocorre reduzindo o efeito do nível de escolaridade.

É possível que indivíduos com níveis de renda mais altos tenham menor preocupação com os preços dos produtos, o que é o principal atrativo dos serviços de compra coletiva online. Sendo assim, estariam menos dispostos a despendar tempo para buscar ofertas na internet e acabariam tendo menor interesse no serviço.

O modelo também indica que pessoas que tomam conhecimento das ofertas por terceiros ($Terc=1$) têm 2,53 vezes mais chances de usar o serviço do que aquelas que não costumam tomar conhecimento desse modo. Isso se torna esperado se for tomado como pressuposto que muitas daquelas pessoas que tomam conhecimento de ofertas dessa forma, geralmente conversam com terceiros sobre o serviço de compra coletiva. Se as ofertas de compras coletivas fazem parte de diálogos de um indivíduo, é porque provavelmente há um interesse por trás dessa ação.

A *odds ratio* da variável $Cmint$ indica que pessoas que comprem na internet têm 8,59 vezes mais chances de interessarem-se pelos sites de compra coletiva. O hábito de comprar na internet tende a reduzir a resistência do indivíduo ao consumo nessa nova modalidade de serviço online.

Outro hábito que também indicou ser positivo para reduzir a resistência de uma pessoa ao consumo em sites de compras coletivas foi a visita a sites desse tipo, representado pela variável $Visit$. Aqueles que já visitaram sites de compra coletiva têm 2,78 vezes mais chances de usar o serviço do que as pessoas que nunca visitaram.

O modelo também revela que aquelas pessoas que possuem dependentes ($Dep=1$) têm uma menor chance de utilizar o serviço. O efeito dessa variável é justificável se for tomado como pressuposto que o fato de haver dependentes leva normalmente um indivíduo a dividir os seus gastos com os mesmos e que as decisões de compra geralmente levam em conta opiniões desses dependentes. Esses dois motivos tendem a proporcionar menor liberdade para realizar uma compra momentânea pelo impulso ao ver um preço mais baixo sendo oferecido.

A importância atribuída à variedade de produtos oferecidos nos sites de compras coletivas ($Vprod=1$) é um fator que eleva a chance de um indivíduo interessar-se pelo serviço em 3,11 vezes. Embora novos sites de compra coletiva tenham surgido voltados para nichos de mercado bem específicos, a maioria dos usuários que utilizam o serviço de compras coletivas ainda considera importante a sua forma tradicional dos serviços, onde uma grande variedade de

produtos e serviços é oferecida. De fato, os grandes sites de compra coletiva atualmente no Brasil mantêm a variedade grande das ofertas.

Os indivíduos que consideram a variedade de formas de pagamento como importante possuem 2,97 vezes mais chances de interessarem-se mais pelo serviço do que aquelas que não consideram esse um fator relevante.

Se for assumido que essa variedade de formas de pagamento é um fator que tende a ser considerado importante por pessoas que consomem produtos na internet e que essas pessoas têm menor resistência a usar sites de compras coletivas, então é esperado que pessoas que consideram a variedade de formas de pagamento importante sejam aquelas mais propensas a interessar-se pelo serviço.

O site de compras coletivas tem como principal característica a conquista de cliente por meio da compra por impulso, haja vista que as ofertas, geralmente, são disponibilizadas por um preço abaixo do praticado no mercado e por um tempo que é relativamente curto. Quanto mais racional for um cliente no momento da compra, menos propenso será a comprar por impulso. Considerar a aparência do site de compra coletiva um aspecto importante na hora da compra é uma característica menos racional. Isso poderia ser uma explicação para o fato de pessoas que consideram a aparência do site como algo importante ($Ap=1$) terem 3,34 vezes mais chances de apresentarem um interesse maior pelo serviço do que aqueles que não julgam essa como uma característica irrelevante.

Ffan é um fator que indica conjuntamente se um indivíduo ligado a uma faixa de renda superior a R\$ 2.000,00, considera a facilidade de navegar no site como algo muito importante para o consumo em um site de compra coletiva. Pessoas com nível de renda superior a R\$ 2.000,00 e que consideram a facilidade de navegação de um site fundamental, possuem uma chance relativamente menor de utilizar o serviço.

Partindo-se do pressuposto de que pessoas que possuem renda superior a R\$ 2.000,00 estão dentro de um público que em sua maioria possuem uma renda que os proporciona ter um computador e acesso a internet em sua residência. Devido a isso, para esse público, há mais facilidade de navegar na internet. Se houver maior dificuldade de navegar em sites para pessoas nessa faixa de renda, isso pode ocorrer porque há o costume de utilizarem menos do que o habitual a rede mundial de computadores, o que é algo que acaba contribuindo para terem maior resistência ao uso do serviço de compras coletivas.

Além disso, o uso de sites de compras coletivas tende a ser maior para pessoas que apresentam interesse pelo serviço. Considera-se que o fato de um indivíduo considerar a facilidade de navegação como algo bastante importante para uso de um site, indica, por si só,

uma tendência de haver interesse menor pelo serviço online do que aqueles que não consideram isso um fator determinante quando comparada às vantagens do uso de um site. Isso é especialmente válido para pessoas cuja renda é maior.

É possível que uma pessoa com renda inferior a R\$ 2.000,00 indique que a facilidade de navegação é algo muito importante porque sua renda não proporciona o uso habitual da internet ou mesmo do computador em sua residência. Pressupõe-se que essa atitude com relação a facilidade de navegar no site tende a ser mais relacionado à chance desse público usar a internet com menos frequência.

Fliv também é um fator relacionado a mesma faixa de renda discutida anteriormente. O fator indica conjuntamente se um indivíduo com nível de renda superior a R\$ 2.000,00 possui o hábito de comprar livros. As pessoas com o hábito de comprar livros são em sua maioria as que possuem maior escolaridade e que possuem renda capaz de proporcionar esse hábito. O modelo indica que a escolaridade é um fator importante na explicação da intenção de um indivíduo ter interesse de usar o serviço de compras coletivas. Isso pode explicar o fato de pessoas com renda superior a R\$ 2.000,00, e que compram livros regularmente, ter maiores chances de usar o serviço.

Há, no entanto, outras razões para o hábito de comprar livros explicar a relação desse costume com o uso de sites de compra coletiva. O hábito de adquirir livros pela internet tem se tornado cada vez mais comum, principalmente para pessoas que possuem poder de compra para possuir um computador e usar a internet com regularidade na sua residência. Se as pessoas que compram livros o adquirem na internet, então há uma maior disposição desse público para comprar pela internet. Isso tende a reduzir a resistência ao uso de sites de compra coletiva.

4.1.2. Modelos com variável resposta binária

Dentre os modelos investigados, o melhor modelo logit binomial paramétrico é o mostrado na equação (4.2). A *odds ratio* de cada uma das variáveis desse modelo é mostrada na tabela 2.

$$P(IC2 = 1) = \frac{e^{\eta}}{1 + e^{\eta}} \quad (4.2)$$

$$\eta = -0,001Rddep + 1,385Terc + 1,46Faj + 2,833Cmin t + 1,514Vpag + 2,017Visit - \\ - 2,75FDep - 1,66FSx - 4,383FId - 0,007\left(\frac{Rd}{Tint}\right)$$

Tabela 2 - Valores da odds ratio para o MLG binomial

Variáveis	Odds
Rddep	0,99
Terc	3,99
Faj	4,305
Cmint	16,996
Vpag	4,545
Visit	7,515
FDep	0,063
FSx	0,19
FId	0,012
Rd/Tint	0,993

Fonte: elaboração própria

O modelo apresentou um *pseudo R²* igual a 0,87. Todos os coeficientes são significativos para um nível de significância igual a 0,05.

Melhores ajustes puderam ser obtidos para os modelos quando o nível de detalhe das informações proporcionadas pelo modelo, ao explicar a variável dependente, foi reduzido. Isso ocorreu porque a intenção de compra passou a ser observada como uma variável dicotômica, assumindo o valor 0 (zero) se um entrevistado afirmasse que certamente não utilizaria o serviço e o valor 1 (um) se a resposta fosse distinta desta. Considerou-se que mesmo que um indivíduo tenha respondido que provavelmente não utilizaria o serviço, o mesmo não descartou a possibilidade de consumir em sites de compras coletivas.

Considerando a variável resposta como sendo uma variável binária, os melhores modelos obtidos foram um modelo onde todas as variáveis foram modeladas parametricamente e outro onde apenas a variável renda por dependente entrou no modelo numa função suave. O uso de uma função suave para uma das variáveis do modelo permitiu uma melhoria no ajuste do modelo e indicou significância do efeito de variáveis que na forma paramétrica não podiam ser explicados de forma significativa.

Para o modelo binomial com componentes paramétricas, pessoas com rendas por dependente mais altas, possuem menos chances de serem usuários do serviço de compras

coletivas do que aquelas que apresentam uma renda por dependente mais baixa. A *odds* dessa variável indica que para cada R\$ 1,00 de aumento na renda por dependente, a chance do indivíduo ser um usuário reduz-se em 1% ($1-0,01=0,99$). Isso indica que um aumento expressivo de R\$ 500,00 na renda por dependente de um indivíduo, faz com que a chance dele continuar a ser um usuário seja cinco vezes menor.

Uma possível explicação para essa informação é o fato de pessoas com rendas mais altas por dependente tem menor preocupação com os preços dos produtos e estão menos dispostos a despender tempo para buscar ofertas na internet. Além disso, o fato de haver dependentes da renda conduz o indivíduo a ter menos liberdade para tomar decisões de compra por impulso como ocorre em sites de compras coletivas.

Pessoas que tomam conhecimento de ofertas de sites de compras coletivas por meio de terceiros possuem 3,99 vezes mais chances de ser um usuário de sites de compras coletivas do que aqueles que não tomam conhecimento dessa forma. Isso pode ser explicado porque as pessoas que tomam conhecimento por meio de terceiros, mantém algum diálogo sobre o serviço e isso é algo que tende a ser mais presente para pessoas que estão dispostos a usar os sites.

Os indivíduos que atribuem uma importância à facilidade de obter ajuda nos sites possuem 4,305 vezes mais chance de serem usuários de acordo com a intenção de compra do que aqueles que acham esse fator sem importância ou mesmo pouco importante. Esses indivíduos provavelmente utilizam ajudas em sites e isso indica que há uma disposição de continuar usando um site mesmo quando dúvidas sobre o serviço existem. Essa característica é comum a pessoas que utilizam o serviço de compras coletivas online que necessitam comumente de mais informações sobre o site, a compra e os produtos antes de concluir uma transação.

Aqueles que compram na internet têm 16,996 vezes mais chances de usar o serviço do que os que não fazem compras online. Esse é um resultado esperado já que as pessoas que fazem transações pela internet têm menor resistência a usar um serviço online.

O grupo de pessoas que consideram importante que um serviço disponibilize uma variedade de formas de pagamento tem 4,545 vezes mais chances de adquirir um produto do que aqueles que acham esse fator irrelevante ou mesmo pouco relevante. Essa também é uma característica comum de pessoas que compram na internet. Assim, esse fator geralmente está ligado a pessoas que têm alguma disposição para usar o serviço de comprar coletiva.

O fato de um indivíduo já ter visitado sites de compras coletivas online eleva sua chance em 7,515 vezes. Isso tende a ser algo favorável para reduzir a resistência de uma pessoa ao serviço de compras coletivas.

FDep é um fator que avalia duas características conjuntamente: o fato de uma pessoa ter uma renda superior a R\$ 2000,00 e possuir dependentes. As pessoas que possuem essas duas características têm uma chance 94% (1-0,94) menor de serem usuários de compras coletivas do que aqueles que não apresentam.

A combinação da variável renda com o número de dependentes dessa renda possui um efeito na definição da intenção de compra de um indivíduo. No entanto, esse efeito pode ser maior em casos específicos. Por exemplo, aquelas pessoas que possuem dependentes e têm renda superior a R\$ 2.000,00 têm ainda menos chance de serem futuros usuários do serviço.

FSx é um outro fator que indica a combinação de duas características: o fato de um indivíduo ser do sexo masculino e o fato deste visitar sites de compras coletivas com frequência superior a uma vez por semana. Esse fator não indica que o grupo de pessoas que mais visitam sites de compra coletiva têm menos chances de usar o serviço. Indica que dentro do grupo de pessoas que visitam os sites de compra coletiva com mais frequência, aqueles do sexo masculino possuem menos chances de ser usuário do que uma pessoa do sexo feminino embora o fato do indivíduo visitar sites com frequência, mesmo que do sexo masculino, sempre indica que o mesmo tem mais chances de usar o serviço.

Se um indivíduo visita sites de compras coletivas semanalmente, possui 7,515 vezes mais chances de ser usuário do que aquele que não visita com frequência alta (semanal ou diária). Porém, esse impacto da característica na chance depende do sexo. Se o indivíduo é do sexo masculino sua chance é 1,427 vezes maior do que aqueles que não visitam com tanta frequência.

O fator *FSx* ainda indica um grupo que embora visite sites de compras coletivas com uma regularidade mensal, possuem até menos chances de ser usuário do que aqueles que não visitam com essa frequência. Esse grupo envolve pessoas que acessam os sites uma vez por mês e são do sexo masculino. De fato, durante as entrevistas houve pessoas que comentaram que não comprariam nos sites de compras coletivas, embora acessassem por curiosidade para ver os preços. O modelo mostrou que isso é mais frequente entre os homens.

O fator *FId* indica a presença de duas características si indica que dependendo das características do indivíduo relativas à idade, o fato de possuir uma renda inferior a R\$ 2.000,00 ou mesmo possuir uma renda superior sendo o único dependente da renda faz com que tenha bem mais chance de ser usuário do que aqueles que não apresentam essa

característica. Segundo o modelo, pessoas com idade superior a 35 anos e que têm renda superior a R\$ 2.000,00 possuem 99% menos chances de usar o serviço do que aquelas que não apresentam essas duas características.

Esse fator mostra que pessoas com idade superior a 35 anos têm bem menos chances de usar o serviço do que aqueles mais jovens, mas isso é válido apenas para pessoas com renda superior a R\$ 2.000,00.

Os fatores *FSx* e *FId* foram as variáveis menos significativas do modelo. Isso provavelmente ocorreu porque ao definir esses fatores adotaram-se limites para renda, para frequência de visitas a sites de compra coletiva e para idade. Esses limites foram escolhidos de forma aleatória. Os fatores podem ser mais bem definidos testando outros limites.

O último termo apresentado no modelo envolve duas variáveis contínuas: renda e tempo que regularmente o entrevistado passa na internet. Indica que um aumento de uma unidade na razão “*Rd/Tint*” reduz a chance de um indivíduo ser usuário do serviço em 0,7% ($1 - 0,993 = 0,007$). Em termos práticos, a *odds* indica que quanto menor o acesso à internet menor será a chance de um indivíduo ser usuário do serviço. Da mesma forma, quanto maior a renda também é menor a chance do indivíduo ser usuário.

É importante notar que a influência do tempo diário que se passa na internet sobre a chance está condicionada à renda. Segundo o modelo, é fato que passar mais tempo na internet eleva a chance de um indivíduo ser um usuário do serviço de compras coletivas, mas o impacto dessa variável sobre a chance depende da renda do indivíduo. Se um indivíduo possui uma renda muito alta, passar mais tempo regularmente na internet representa um aumento expressivo na chance do indivíduo ser um usuário do serviço. No entanto, se o indivíduo apresenta uma renda baixa então passar mais tempo na internet aumentará pouco a chance de ser um usuário.

Nota-se que a influência da renda na chance do indivíduo ser um usuário ocorre de forma complexa e está condicionada ao comportamento de outras variáveis. O modelo foi capaz de identificar a ligação com três variáveis: número de dependentes da renda, idade e tempo diário na internet.

Um modelo aditivo *logit* binomial usando as mesmas variáveis do modelo anterior apresentou uma variável com p-valor (*Vpag*) igual a 0,0437 e com o p-valor para função paramétrica igual a 0,029. Modelos aditivos com termos mais significativos foram obtidos fazendo uma mudança no termo “*Rd/Tint*” do modelo por “*Aesc/Tint*”. A correlação entre esses termos é igual a 0,62 e a correlação entre as variáveis *Aesc* e *Rd* é igual a 0,43. O fato das variáveis e dos termos apresentarem uma correlação moderada possibilita que ambos os

termos forneçam modelos com resultados semelhantes, ficando a escolha a critério das melhorias do modelo.

O MAG com melhor ajuste obtido indicou resultados semelhantes para todos os termos do modelo, exceto para o termo não paramétrico (cuja variável é *Rddep*) e o termo que foi substituído. Desse modo, conclusões semelhantes podem ser obtidas para os MAG's obtidos, fazendo-se necessárias explicações adicionais apenas sobre os dois termos citados.

Ao aplicar o método de suavização spline para modelar a contribuição da variável renda por dependente (*Rddep*) para o logaritmo da chance do modelo, obtém-se o modelo com a equação (4.3).

$$P(IC2 = 1) = \frac{e^{\gamma_1}}{1 + e^{\gamma_1}} \quad (4.3)$$

sendo:

$$\begin{aligned} \gamma_1 = & f_1(Rddep) + 1,442Terc + 1,304Faj + 2,820C \min t + 1,796Vpag + 1,947Visit - \\ & - 2,043FDep - 2,2FSx - 1,365FId - 4,456 \left(\frac{Aesc}{T \text{ int}} \right) \end{aligned}$$

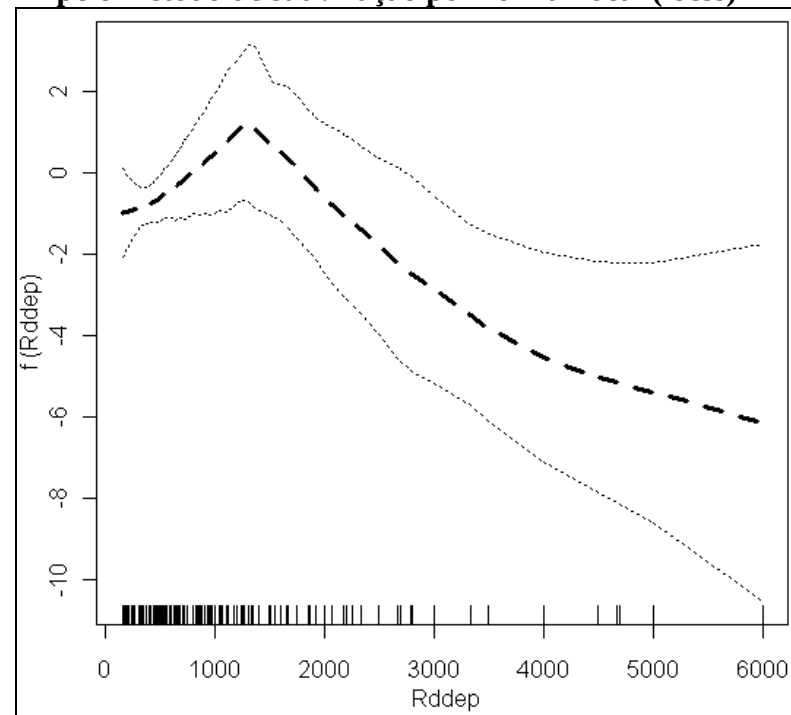
Se o método de suavização polinomial local é utilizado, obtém-se a equação (4.4).

$$P(IC2 = 1) = \frac{e^{\gamma_2}}{1 + e^{\gamma_2}} \quad (4.4)$$

$$\begin{aligned} \gamma_2 = & f_2(Rddep) + 1,498Terc + 1,311Faj + 2,636C \min t + 1,672Vpag + 1,936Visit - \\ & - 1,909FDep - 1,974FSx - 1,357FId - 4,603 \left(\frac{Aesc}{T \text{ int}} \right) \end{aligned}$$

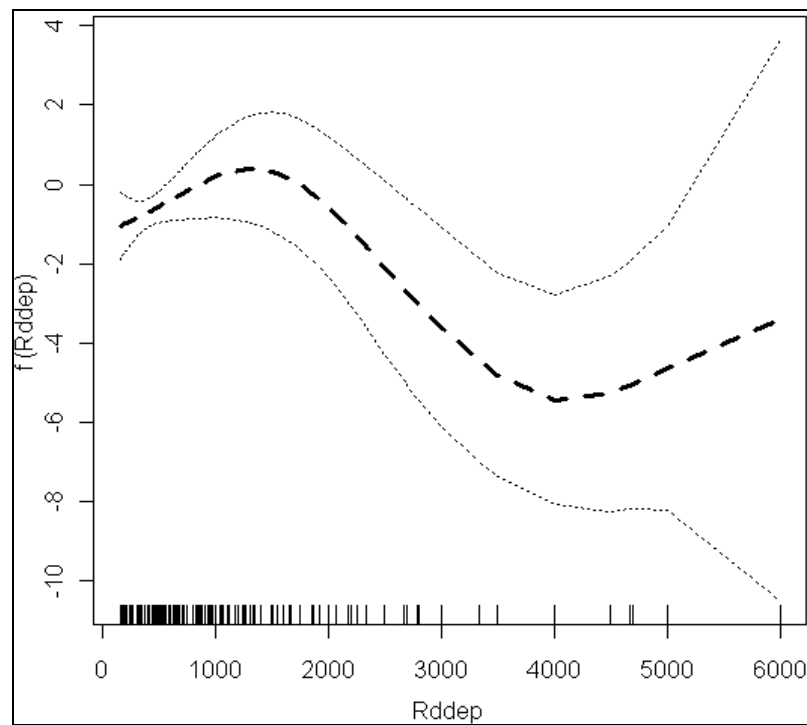
As figuras 4 e 5 mostram a curva que descreve o comportamento da variável *Rddep*, obtidas, respectivamente, por meio da suavização polinomial local e por suavização splines. Com base nessas curvas é possível identificar o incremento na chance do modelo ao aumentar ou diminuir a variável *Rddep*.

Figura 4 - Curva do efeito da renda por dependente no logaritmo da chance do modelo pelo método de suavização polinomial local (loess)



Fonte: elaboração própria.

Figura 5 - Curva do efeito da renda por dependente no logaritmo da chance do modelo pelo método de suavização splines



Fonte: elaboração própria.

Deve-se notar que a razão da chance indica que há um aumento constante “ e^β ” para cada incremento de uma unidade na variável ligada ao coeficiente “ β ”. Esse aumento constante indica que a variação no logaritmo da chance é linear. Por essa razão, o coeficiente angular é constante (taxa de variação da curva constante).

No MLG binomial obtido inicialmente, a curva do logaritmo da chance apresentará uma taxa de variação constante e negativa (curva linear decrescente). Nos MAG’s estimados, as curvas não paramétricas, presentes nos modelos, indicaram que essa variação no logaritmo não é sempre constante e negativa. Ao usar a suavização polinomial local, obteve-se a descrição a seguir do comportamento do logaritmo da chance. A seguir também são descritas as conclusões acerca da chance de um indivíduo, com renda por dependente específica, ser usuário do serviço.

- Para rendas por dependente inferiores até um valor próximo de R\$ 1.300,00, o logaritmo da chance apresenta-se como positivo e praticamente constante, pois a curva apresenta-se próxima de uma reta e é crescente. Nessa faixa, duas pessoas com rendas por dependente cuja diferença é de R\$ 100,00, tem uma diferença na chance de serem usuários que é praticamente a mesma, independente desses grupos apresentarem rendas mais altas ou mais baixas.
- A partir de um valor próximo de R\$ 1.300,00, o logaritmo da chance apresenta-se negativo e varia quase linearmente até um valor da renda por dependente próximo de R\$ 3.500,00. Isso indica que nessa faixa, quanto maior for a renda por dependente de um indivíduo, menor será a chance de ele ser um usuário do serviço. A diferença da chance de duas pessoas com uma determinada diferença na renda por dependente é praticamente a mesma independente de apresentarem rendas mais altas ou não.
- Para rendas por dependente superiores a um valor próximo de R\$ 3.500,00, valores mais altos ainda representam menos chance de usar o serviço. Porém, a diferença da chance de duas pessoas com uma determinada diferença na renda por dependente está condicionada ao fato de serem essas rendas mais altas ou mais baixas. Se as rendas por dependente são altas há pouca diferença na chance, mas se são mais baixas (próximas de R\$ 3.500,00) então a diferença é maior.

Tabela 3 - Valores dos coeficientes e da odds ratio para cada uma das variáveis dos MAG's

Variáveis	Métodos de Suavização Utilizados			
	Splines		Polinomial Local	
	Coeficientes	Odds	Coeficientes	Odds
f(Rddep)	-	-	-	-
Terc	1,442	4,223	1,498	4,473
Faj	1,304	3,684	1,311	3,71
Cmint	2,82	16,776	2,636	13,957
Vpag	1,796	6,025	1,672	5,323
Visit	1,947	6,958	1,936	6,931
FDep	-2,043	0,13	-1,909	0,148
FSx	-2,2	0,111	-1,974	0,139
Aesc/Tint	-4,456	0,012	-4,603	0,01
FId	-1,365	0,255	-1,357	0,257

Fonte: elaboração própria.

É importante atentar-se para os limites do intervalo de confiança da curva não paramétrica. A partir de um valor próximo de R\$ 4.000,00, há grande incerteza segundo o modelo ao tratar da diferença das chances de ser usuário dos sites de compras coletivas.

A tabela 3 mostra o valor da *odds ratio* de cada uma das variáveis dos MAG's com diferentes métodos de suavização. Os valores apresentam-se aproximados daquelas encontradas para o MLG com variável resposta binária.

Quando é utilizado o método de suavização spline para o termo não paramétrico, algumas diferenças são identificadas na descrição do comportamento desse termo. Porém, conclusões semelhantes são obtidas para boa parte do domínio da função não-paramétrica. As diferenças entre os métodos no que diz respeito às conclusões, são descritas a seguir, indicando-se os intervalos de valores onde essas diferenças existem.

- Ao contrário da função não paramétrica obtida pelo método de suavização polinomial local, não há um valor específico a partir do qual a curva do logaritmo da chance muda abruptamente seu comportamento passando a ser decrescente. Essa mudança ocorre de maneira gradual num determinado intervalo, onde a curva do log da chance é reduzido,

assume o valor zero e passa a ser decrescente. Em termos práticos, há um intervalo, cujos limites aproximam-se dos valores R\$ 1.300,00 e R\$ 1.800,00, onde praticamente não há diferença na chance do indivíduo ser um usuário com relação à renda por dependente. Nesse intervalo, pessoas com renda por dependente maiores têm praticamente a mesma chance de serem usuários do que pessoas com rendas menores.

- Numa faixa de renda cujos limites estão próximos de R\$ 3.900,00 e R\$ 4.300,00, também há pouca variação da curva de logaritmo da chance. Isso indica que nesse intervalo praticamente também não há diferença na chance de indivíduos serem usuários de acordo com a renda por dependente.
- A partir de um valor próximo de R\$ 4.300,00, pessoas com rendas por dependente maiores têm chances maiores de serem usuários do que aquelas com rendas mais baixas (próximas de R\$ 4.300,00).

Para o método de suavização splines também é fundamental notar que a partir de um valor próximo de R\$ 4.000,00 o intervalo de confiança para curva do logaritmo da chance é grande. Na prática, isso indica que há maior incerteza nas conclusões obtidas pelo método de suavização splines a partir desse valor próximo de R\$ 4.000,00.

O intervalo de confiança é ainda maior para rendas por dependente superiores a R\$ 5.000,00. Nesse intervalo, as conclusões apresentam-se bastante incertas.

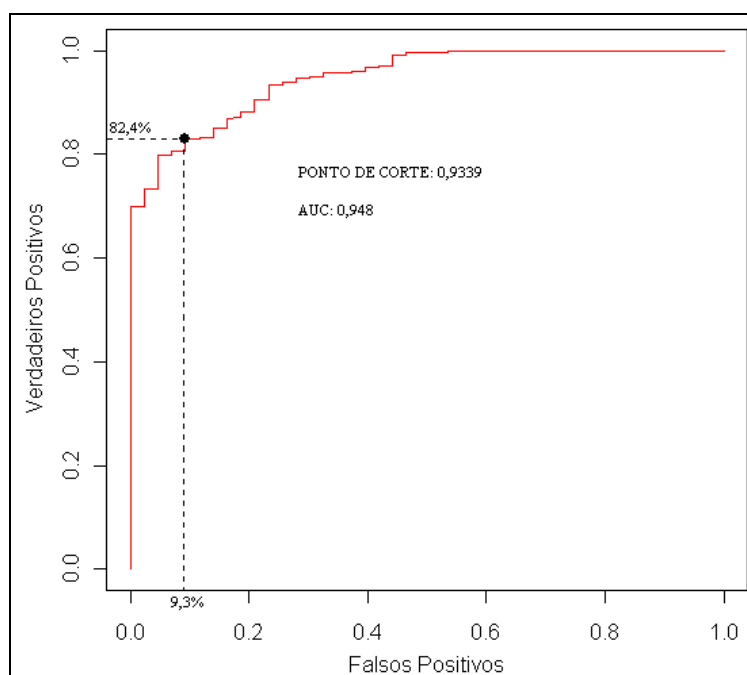
4.2. Diagnóstico dos modelos

O diagnóstico dos modelos aditivos generalizados com variável resposta consistem na verificação da qualidade do ajuste aos dados e nos testes de significância dos termos do modelos (parâmetros e funções não paramétricas). Nessa seção, a qualidade do ajuste dos dados será verificada com base nas medidas *pseudo-R²* e acurácia (percentual de acertos no total de julgamentos realizados). Os testes de significância utilizados serão o teste Wald e o teste χ^2 .

Curva ROC e Matriz de Confusão

A definição do ponto de corte é importante quando se utilizam modelos com função de ligação logit e há uma disparidade entre o número de observações das duas categorias da variável resposta.

Figura 6 - Curva ROC para MLG com variável resposta binária



Fonte: elaboração própria

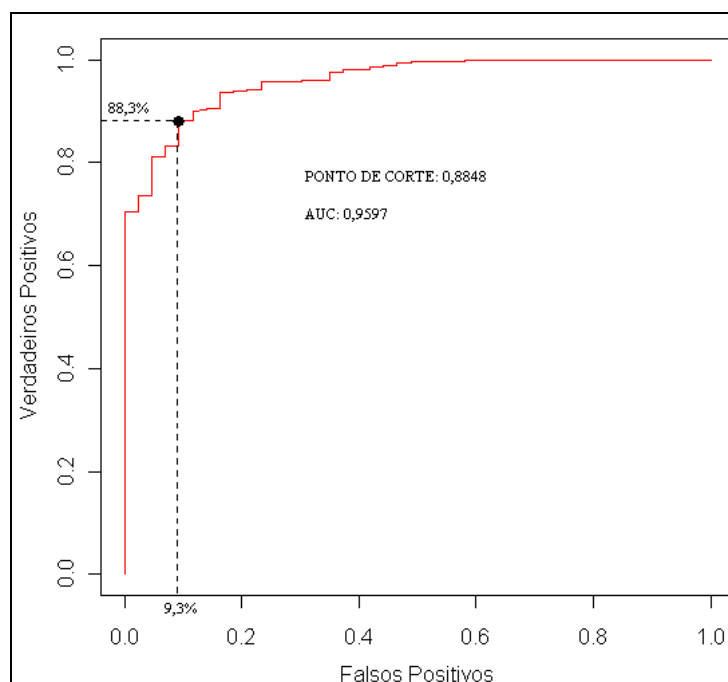
Uma escolha do ponto de corte pode ser feita por meio de uma análise gráfica da curva ROC. Da figura 6, por exemplo, nota-se que o ponto de corte igual 0,9339 para o MLG binomial mostrou bons resultados, levando a uma taxa de erros baixa tanto no julgamento de indivíduos com perfil de “compradores” como para indivíduos com perfil de “não compradores”.

Da figura 7 é possível perceber que para o MAG um bom ponto de corte em termos da capacidade de previsão para os dois grupos foi 0,8848. A tabela 5 mostra a matriz de confusão para o MAG com esse ponto de corte.

Valores de pontos de corte foram comparados em termos de indicadores da capacidade de previsão (tabelas 4 e 6). A acurácia determina a taxa de acertos geral das previsões. Para os modelos binomiais, há pontos de corte que proporcionam uma capacidade de previsão geral melhor do que aqueles citados anteriormente. As tabelas 8 e 9 apresentam informações sobre

as taxas de acertos para diferentes pontos de corte, para os modelos de variável resposta binária. Adotando-se um ponto de corte igual a 0,56 para o MAG, é possível obter uma taxa de acertos total dos julgamentos para amostra (acurácia) igual a aproximadamente 0,94.

Figura 7 - Curva ROC para MAG com variável resposta binária



Fonte: elaboração própria.

A matriz de confusão elaborada para cada um dos modelos (tabelas 5 e 7) mostra como os julgamentos foram feitos para o público entrevistado. Para o MLG, quatro dos indivíduos que eram compradores, segundo o critério estabelecido, foram julgados como não compradores e 60 que não compradores foram classificados como compradores. Isso significa que ao fazer previsões para o grupo de compradores 82,41% foram corretas e para o grupo de não compradores 90,7% das previsões foram corretas. Considerando as previsões para todos os entrevistados, 83,33% foram corretas.

Tabela 4 - Indicadores da capacidade de previsão do MLG binomial para diferentes pontos de corte

Ponto de Corte	Acurácia	Julgamentos certos	
		Compradores	Não compradores
0,9614	0,763	73,90%	95,35%
0,9569	0,7969	77,71%	95,35%
0,9466	0,8203	80,65%	93,02%
0,9339	0,8333	82,41%	90,70%
0,9249	0,8359	83,28%	86,05%
0,88	0,862	86,51%	83,72%
0,851	0,8854	89,74%	79,07%
0,8052	0,9115	92,96%	76,74%
0,7068	0,9193	94,43%	72,09%
0,5539	0,9219	96,19%	60,47%

Fonte: elaboração própria

Tabela 5 - Matriz de confusão para o MLG binomial com ponto de corte igual a 0,9339 para classificar um indivíduo como comprador

Predito	Real		Total
	Comprador	Não Comprador	
Comprador	281	4	285
Não Comprador	60	39	99
Total	341	43	384

Fonte: elaboração própria

Tabela 6 - Indicadores da capacidade de previsão do MAG binomial para diferentes pontos de corte

Ponto de Corte	Acurácia	Julgamentos certos	
		Compradores	Não compradores
0,9777	0,7448	71,55%	97,67%
0,9653	0,7734	75,07%	95,35%
0,9561	0,8021	78,30%	95,35%
0,9482	0,8203	80,35%	95,35%
0,9301	0,8411	82,99%	93,02%
0,9147	0,8542	84,75%	90,70%
0,8848	0,8854	88,27%	90,70%
0,8185	0,9036	91,20%	83,72%
0,7107	0,9297	95,01%	76,74%
0,5609	0,9375	97,65%	62,79%

Fonte: elaboração própria

A matriz de confusão para o MAG indica resultados um pouco melhores. O número de erros nas predições para o grupo de compradores foi o mesmo, mas o número de erros nas predições para o grupo de não compradores foi igual a 40. Sendo assim, o MAG estimado classificou corretamente 88,48% dos casos onde os indivíduos eram compradores e 99,7% dos casos daqueles indivíduos que eram não compradores. Isso significa que 88,54% das predições realizadas para os entrevistados foram corretas.

Tabela 7 - Matriz de confusão para o MAG binomial com ponto de corte igual a 0,8848 para classificar um indivíduo como comprador

Predito	Real		Total
	Comprador	Não Comprador	
Comprador	301	4	305
Não Comprador	40	39	79
Total	341	43	384

Fonte: elaboração própria

A escolha do ponto de corte e a análise da capacidade de classificar corretamente do modelo, usando a matriz de confusão, também foram realizadas para o modelo multinomial ordinal.

Mesmo utilizando diferentes pontos de corte, o modelo multinomial estimado apresenta apenas um desempenho regular. Os pontos de corte que geraram o melhor resultado de classificação foram 0,15 (para julgar uma observação como do grupo de indecisos) e 0,8 (para julgar uma observação do grupo de compradores). Ao adotar esses valores, o modelo proporciona uma taxa de acertos de 52,6% do total de julgamentos.

Nos julgamentos feitos para o grupo de compradores, 86,92% das classificações foram corretas quando utilizados esses pontos de corte. No entanto, apenas 38,39% das classificações no grupo de indecisos foram corretas e somente 18,6% foram corretas para o grupo de não compradores. Os resultados das classificações são mostrados na matriz de confusão do modelo (tabela 9).

Tabela 8 - Indicadores da capacidade de previsão do MLG multinomial ordinal para diferentes pontos de corte

Pontos de corte		Acurácia	Julgamentos certos		
Indecisos	Compradores		Compradores	Indecisos	Não compradores
0,3	0,6	0,4557	51,16%	14,69%	93,85%
0,3	0,7	0,4583	60,47%	14,69%	91,54%
0,3	0,8	0,4427	60,47%	14,69%	86,92%
0,2	0,8	0,4896	23,26%	30,81%	86,92%
0,2	0,9	0,4297	27,91%	30,81%	67,69%
0,15	0,8	0,526	18,60%	38,39%	86,92%
0,15	0,9	0,4714	23,26%	39,34%	67,69%
0,35	0,65	0,4479	72,09%	9,48%	93,08%
0,25	0,65	0,474	39,53%	20,85%	93,08%

Fonte: elaboração própria

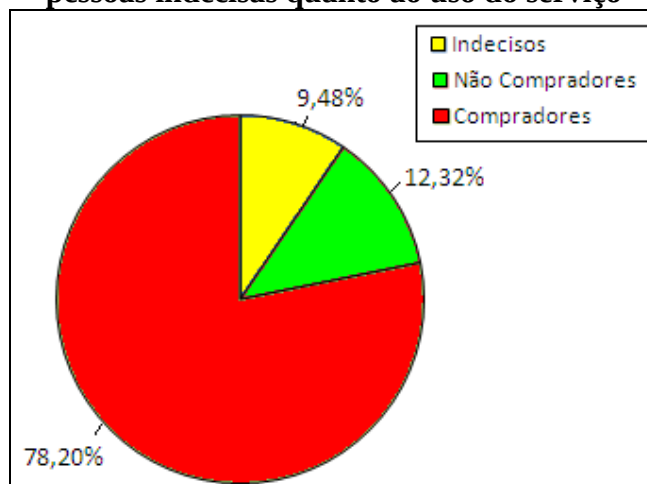
Tabela 9 - Matriz de confusão para o modelo com pontos de corte 0,15 e 0,85 para o julgamento respectivamente de indecisos e compradores

Predito	Real			Total
	Não Compradores	Indecisos	Compradores	
Não Compradores	8	0	0	8
Indecisos	33	81	17	131
Compradores	2	130	113	245
Total	43	211	130	384

Fonte: elaboração própria

Para os pontos de corte 0,35 e 0,65, a taxa geral de acertos não foi alta, pois o modelo apresente baixa capacidade de previsão para julgar pessoas que estão indecisas quanto ao uso do serviço. Porém, se não houver problemas em errar nas previsões do grupo de pessoas indecisas, então o uso desses pontos de corte conduz o modelo a boas taxas de acertos para o restante do publico. 72,093% das previsões feitas para o grupo de compradores são corretas e 93,077% do grupo de não compradores são corretas.

Figura 8 - Julgamentos do modelo com pontos de corte 0,65 e 0,35 para o público de pessoas indecisas quanto ao uso do serviço



Fonte: elaboração própria.

Utilizando os pontos de corte 0,65 para classificar indivíduos como compradores e 0,35 para julgar como indecisos, o modelo acertará apenas 9,48% quando usado para classificar indivíduos indecisos quanto ao uso dos sites de compras coletivas. Para esse público específico, o modelo apresenta-se tendencioso e julga erradamente maior parte como compradores (78,2%). A tabela 10 mostra a matriz de confusão para esses pontos de corte.

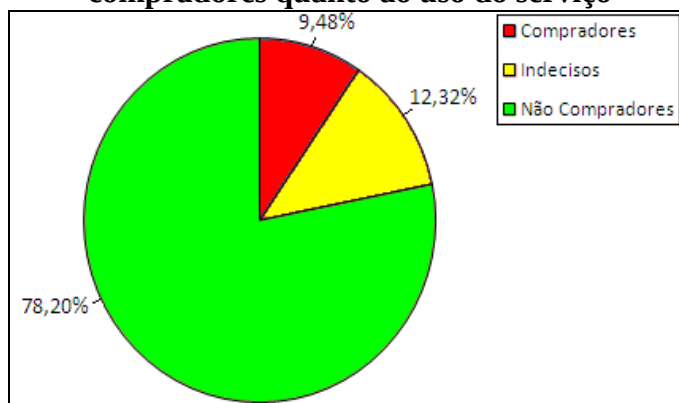
Tabela 10 - Matriz de confusão para o modelo com pontos de corte 0,35 e 0,65 para o julgamento respectivamente de indecisos e compradores

Predito	Real			Total
	Não Compradores	Indecisos	Compradores	
Não Compradores	31	26	4	61
Indecisos	8	20	5	33
Compradores	4	165	121	290
Total	43	211	130	384

Fonte: elaboração própria

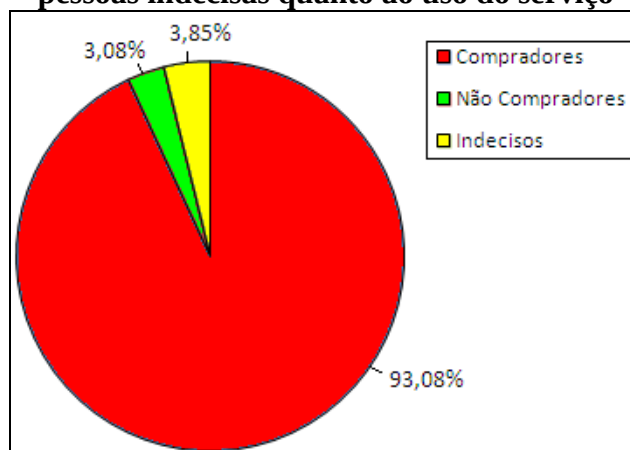
De fato, classificar um indivíduo que está indeciso quanto a usar um site de compra coletiva como um potencial comprador não é algo problemático, porque é desejável que o modelo descarte (julgar como não compradores) o mínimo de indivíduos para orientar o projeto e gestão do sistema de produção. A maior parte dos erros de previsão do modelo são julgamentos de indivíduos indecisos como sendo potenciais compradores.

Figura 9 - Julgamentos do modelo com pontos de corte 0,65 e 0,35 para o publico de não compradores quanto ao uso do serviço



Fonte: elaboração própria.

Figura 10 - Julgamentos do modelo com pontos de corte 0,65 e 0,35 para o publico de pessoas indecisas quanto ao uso do serviço



Fonte: elaboração própria.

Considerando as categorias de indecisos e de compradores como uma única categoria, é possível perceber que o modelo multinomial erra mais nos julgamentos do que os modelos binomiais (figura 9). No entanto, nas classificações do grupo de compradores, o MLG multinomial mostrou-se superior e acerta 96,93% ($93,08\% + 3,85\%$) das classificações desse público (figura 10). Em termos práticos, isso significa que aqueles indivíduos julgados pelo modelo multinomial como sendo compradores têm uma chance muito baixa de ser na realidade um não comprador. Esse é um tipo de erro indesejável e o risco de cometê-lo usando o MLG multinomial é inferior aquele que se expõe ao usar o MAG binomial.

O modelo multinomial com os pontos de corte 0,35 e 0,65 é o mais recomendado numa situação onde um proprietário de um site de compra coletiva tem como principal preocupação classificar corretamente indivíduos que são compradores em potencial, ou mesmo indecisos, quanto ao uso do site. Por exemplo, considere a situação onde o

proprietário escolhe um segmento para manter o foco de seu serviço e deseja essencialmente verificar se esse segmento está mesmo interessado no site (sendo um comprador em potencial). Se o modelo multinomial estimado indica que as características do segmento apontam o mesmo como sendo de um público de compradores, haverá pouca chance dessa predição estar errada.

O fato da medida de acurácia mostrar que o modelo multinomial, com os pontos de corte referidos, possui uma capacidade de predição apenas regular, ocorre por aquele apresentar uma capacidade de acertos muito baixa ao classificar indivíduos como indecisos. Uma classificação do modelo que julgue a indecisão deve ser vista com certa desconfiança, pois as chances do indivíduo julgado como indeciso ser na verdade um comprador em potencial é alta. No entanto, aquelas classificações que se referem a um dado perfil como sendo de um comprador ou como sendo de um não comprador apresenta credibilidade, pois as chances de haverem erros nessas classificações são baixas.

Testes de significância e medidas de qualidade do ajuste

O p-valor do teste de Wald para cada variável do modelo multinomial ordinal é mostrado na tabela 11. É possível perceber que as variáveis do modelo são significativas. A variável que se mostrou menos significativa foi *FFan*, mas mesmo essa variável apresenta um p-valor inferior a 0,04. O p-valor do teste de Wald indica a probabilidade de errar ao rejeitar a hipótese de que o coeficiente da variável é igual a zero. Se o coeficiente da variável for igual a zero, então essa variável não influencia na variável resposta do modelo.

Tabela 11 - Valores do p-valor para cada variável do MLG multinomial

Variáveis	p-valor
Aesc/Id	0,0023
Aesc*Rd	0,0385
Terc	0,0045
Cmint	0,0000
FxRd	0,0019

Cont.

Variáveis	p-valor
Ap	0,0004
Vpag	0,0182
Vprod	0,0004
Dep	0,0229
Visit	0,0079
Ffan	0,0356
Fliv	0,0294

Fonte: elaboração própria.

A aplicação do teste de Wald aos coeficientes do MLG cuja variável resposta é modelada de acordo com uma variável binomial gera os resultados apresentados na tabela 12. A variável menos significativa do modelo é o fator *FId*. Essa variável apresenta um p-valor inferior a 0,03.

Tabela 12 - Valores do p-valor de cada uma das variáveis do modelo

Variáveis	p-valor
Rddep	0,008551
Terc	0,012693
Faj	0,004847
Cmint	0,000024
Vpag	0,014118
Visit	0,000251
FDep	0,000160
FSx	0,027019
FId	0,029790
Rd/Tint	0,025085

Fonte: elaboração própria

A significância dos termos do modelo também foi verificada para o MAG. A tabela 13 apresenta os resultados do teste para utilização de dois métodos diferentes de suavização do termo não paramétrico. A variável que apresentou o coeficiente menos significativo do modelo foi o fator *FDep*. Esse coeficiente, mesmo sendo o menos significativo, possui p-valor inferior a 0,03. Isso indica que a probabilidade de erra ao afirmar que cada uma das variáveis influencia no modelo é inferior a 0,03.

O teste aplicado ao termo não paramétrico testa a significância da variação da curva estimada. Se não for possível rejeitar a hipótese de essa variação é igual à zero, então é possível afirmar que não há relação entre a variável resposta e a covariável em questão. No caso do MAG estimado essa variação é significativa.

Tabela 13 - Valores do p-valor para cada variável dos MAG's com termos não paramétricos obtidos por dois métodos de suavização distintos

Variáveis	Método de suavização	
	Splines	Polinomial local
	p-valor	p-valor
f(Rddep)	0,00116	0,0028
Terc	0,01281	0,0115
Faj	0,01564	0,01593
Cmint	0,00002	0,00005
Vpag	0,00953	0,01779
Visit	0,00059	0,00062
FDep	0,00258	0,00546
FSx	0,00793	0,01728
Aesc/Tint	0,00205	0,00128
Fid	0,02257	0,02185

Fonte: elaboração própria.

A tabela 14 apresenta três medidas diagnósticas gerais do modelo. Utilizando diferentes pontos de corte para classificar as observações, obteve-se diferentes valores da acurácia do modelo. Os maiores valores obtidos são mostrados na tabela. A acurácia indica o percentual acerto do total de classificações efetuadas considerando os elementos da amostra (384 observações).

O $\text{pseudo-}R^2$ utilizado foi o *Nagelkerke- R^2* . Essa medida indica a proporção da estatística *Deviance* que o modelo é capaz de explicar. A medida varia entre 0 e 1. Quanto mais próximo do valor 1, melhor é considerado o ajuste do modelo em relação aos dados.

O modelo que apresentou os valores mais altos da acurácia e do $\text{pseudo-}R^2$ foi o MAG. O modelo multinomial apresentou os valores mais baixos.

Uma terceira medida apresentada na tabela 14 é o p-valor do teste χ^2 . Esse teste verifica a significância do conjunto dos coeficientes. Testa a hipótese de que todos os coeficientes do modelo são iguais à zero, ou seja, testa a hipótese de que não existe uma relação entre as covariáveis do modelo e a variável resposta. A probabilidade de errar ao rejeitar essa hipótese é inferior a 0,0001 para os três modelos.

Tabela 14 - Medidas diagnósticas dos modelos

Modelo	Maior acurácia encontrada	pseudo- R^2	p-valor (teste χ^2)
MLG binomial	0,92	0,87	< 0,0001
MAG binomial	0,94	0,89	< 0,0001
MLG multinomial ordinal	0,53	0,53	< 0,0001

Fonte: elaboração própria.

4.3. Simulação da aplicação dos modelos

Nessa seção serão propostas duas formas de aplicar o modelo para orientar decisões acerca do sistema de produção estudado de modo que considerem a variável intenção de compra quando forem tomadas. A primeira dessas aplicações consiste na orientação da escolha de um público-alvo considerando que as decisões acerca da especificação do sistema de produção já tenham sido tomadas. A segunda aplicação consiste na orientação das decisões relativas à especificação do sistema produtivo, considerando que o público-alvo já tenha sido escolhido previamente.

4.4.1. Classificações de segmentos e especificações do serviço pelos modelos

Nessa subseção serão escolhidos os modelos usados nos exemplos. Três dos modelos estimados foram tratados anteriormente. Para tratar de forma mais sucinta a aplicação desses modelos apenas dois serão considerados. Os dois modelos serão usados porque algumas variáveis estão presentes em apenas um deles e isso permite que sejam usados de maneira que se complementam. Para que possam ser tratados dessa maneira, três métodos. Por meio de um desses métodos é possível aplicar os modelos conjuntamente.

Posteriormente, nessa subseção, serão tratadas das variáveis de decisão consideradas no exemplo. A aplicação dos modelos para orientação de decisões acerca das especificações do sistema de produção requer que as variáveis do modelo sejam ligadas as variáveis de decisão do serviço de compras coletivas. Isso é necessário porque as variáveis do modelo indicam características do consumidor relativas ao serviço de compras coletivas. Embora muitas decisões sejam baseadas nessas variáveis, elas não representam diretamente variáveis de decisão.

Modelos usados

Maior ênfase foi dada a três tipos de modelos distintos obtidos a partir do banco de dados da pesquisa. Dois desses modelos apresentam a variável resposta modelada segundo uma distribuição binomial e apresentam poucas diferenças com relação às covariáveis, tendo o MAG se ajustado melhor aos dados do que o MLG. O outro modelo enfatizado foi um MLG cuja variável resposta foi modelada segundo uma distribuição multinomial. O modelo apresenta o pior ajuste dentre os três modelos destacados quando se deseja classificar as observações em três categorias. Porém, o ajuste é razoavelmente bom quando são definidos os pontos de corte 0,65 e 0,35, e na classificação não é feita distinção entre as categorias 2 (compradores) e 1 (indecisos). A principal vantagem de usar esse modelo está na possibilidade de analisar um número maior de variáveis.

O modelo multinomial inclui três variáveis que não são consideradas no MAG binomial (variedade de produtos oferecidos no site, facilidade de navegar no website e a sua aparência). Embora o MAG tenha apresentado o melhor ajuste, se um analista desejasse considerar em sua análise um número maior de variáveis, poderia também usar, em conjunto com o MAG, o MLG multinomial. Para utilizá-los conjuntamente é necessário definir um método para classificar as observações realizadas. A seguir são descritos três métodos:

- **Método 1**: Aquele perfil que ambos os modelos julgassem como de um “comprador” seriam classificados desse modo. Caso apenas um dos modelos classificasse assim, a classificação seria como “não comprador”. Ao julgar dessa forma, as classificações dos “compradores” seria mais rigorosa e a classificação dos “não compradores” seria pouco rigorosa. O critério seria adequado para o analista que estivesse preocupado apenas com a identificação de “compradores”. Por exemplo, isso ocorreria numa situação onde a preocupação principal fosse excluir o mínimo de pessoas possível de um projeto de um produto baseado principalmente no cliente. Nessa situação, o analista não teria tanta preocupação em errar ao afirmar que um indivíduo verdadeiramente “comprador” seria um “não comprador”, pois esse erro somente faria com que mais clientes pudessem ser usados para orientar o projeto. Por outro lado, haveria grande preocupação em errar julgando como “comprador” um indivíduo que na realidade é “não comprador”, porque ao fazer isso estaria o analista excluindo clientes equivocadamente.
- **Método 2**: Aquele perfil que ambos os modelos julgassem como sendo de um “não comprador” seria classificado desse modo. Caso apenas um dos modelos classificasse assim, o julgamento seria “comprador”. O critério utilizado é o inverso do anterior. A preocupação primordial do analista é não errar nos julgamentos como “não comprador”. Um exemplo em que seria mais adequado esse critério consiste naquela situação onde um analista procura orientar a instalação de um empreendimento de alto custo em uma região específica e há grande preocupação em não instalá-lo em locais onde maior parte da população apresenta um perfil de pessoas que tendem a não ser consumidores. Se o perfil dominante do local que seria na verdade de compradores é julgado como “não comprador”, o empreendimento tenderia a não ser instalado e a empresa responsável deixaria de fazer a instalação num local onde as pessoas apresentariam interesse. Talvez a empresa poderia deixar de ter resultados financeiros positivos. Por outro lado, se julgasse o perfil dominante como de “comprador” quando na verdade seria de “não compradores”, o pouco interesse das pessoas poderia levar a resultados financeiros desfavoráveis e o risco de grandes prejuízos poderia existir.
- **Método 3**: Julgar como “comprador” quando os dois modelos julgassem como “comprador” e julgar como “não comprador” quando ambos julgassem como “não comprador”. Se os modelos não apresentassem a mesma classificação, o julgamento

poderia ser “inconclusivo” e as classificações do modelo não poderiam ser consideradas pelo analista. Nesse critério, haveria mais rigor para ambas as classificações, mas o uso conjunto dos modelos não seria aplicado para analisar todas as escolhas de perfis. O critério poderia ser adequado, por exemplo, para uma situação onde uma empresa coloca no mercado novos produtos com grande regularidade e sua linha de produtos possui itens para diferentes perfis. Se o analista considera muito importante tanto evitar excluir clientes equivocadamente do projeto dos produtos como também considerar nos projetos aqueles que apresentam pouco interesse, então os acertos para ambas as classificações são relevantes. Caso ambos os modelos diferissem na classificação, então a decisão poderia ser baseada, por exemplo, apenas em opiniões de analistas experientes no setor dos produtos lançados.

Definição das variáveis

Na situação a ser exemplificada nessa seção, algumas variáveis de decisão acerca do serviço de compras coletivas online foram escolhidas a fim de mostrar de que modo podem ser ligadas àquelas do modelo. As variáveis envolvidas no processo de segmentação são aquelas variáveis demográficas incluídas diretamente no modelo. Um resumo das variáveis de decisão relacionadas a especificações de aspectos do sistema de produção (serviço de compras coletivas) e daquelas usadas no processo de segmentação de mercado é mostrado no quadro 5, presente na página posterior. Esse quadro também mostra as variáveis do modelo que foram ligadas a cada uma das variáveis de decisão listadas.

As dez variáveis de decisão utilizadas para exemplificar o uso do modelo são as seguintes:

- **Simplicidade do site**: Utilizar maior número de *links* (janelas) e maior número de níveis por *link* no site. A variável *Fan* é a variável da pesquisa ligada a essa característica do serviço. Sites que apresentam menor número de janelas e de níveis por janelas apresentam maior simplicidade e são mais fáceis de navegar. Isso será um ponto positivo para aquelas pessoas que consideram a facilidade de navegar como um item importante ou muito importante do serviço. A única variável (obtida por meio da variável *Fan*) ligada à estrutura do site nos modelos é o fator *FFan*, presente apenas no modelo multinomial.

- **Abertura a novatos em compras coletivas:** Incluir elementos explicativos que facilitem a navegação por parte de pessoas que nunca utilizaram um site de compra coletiva. Essa característica está ligada a variável *Visit* dos modelos. Se um site apresenta poucos elementos explicativos desse tipo, então aqueles indivíduos que nunca utilizaram o serviço tenderão a apresentar maior dificuldade para usá-lo. Esse passa a ser um ponto negativo para aqueles indivíduos.
- **Período das ofertas:** Tempo de disponibilidade das ofertas. Dentre as variáveis do modelo, aquela que apresenta ligação com essa variável é o *Tint*, pois pessoas que passam pouco tempo na internet tendem a se expor menos a divulgação das ofertas na internet e precisam que as ofertas passem mais tempo disponíveis. Se um indivíduo passa pouco tempo online e geralmente executa outras atividades na internet, tenderá a acessar pouco os sites e perderá muitas das boas ofertas. O pouco tempo disponível dessas ofertas seria então um ponto negativo para esse tipo de cliente.
- **Detalhamento do guia de ajuda do site:** Elaboração de um guia de ajuda detalhado. Essa variável de decisão está ligada a variável *Faj* presente no MAG. Uma empresa proprietária do site de compra coletiva pode optar por usar um guia de ajuda ao uso do serviço, mais simples por julgar que os clientes para o qual está voltado apresentar facilidade para usar o serviço. Para aqueles clientes que costumam usar regularmente os guias de ajuda por apresentarem maior dificuldade de usar o serviço, verão essa característica do site como negativa já que tendem a considerar a facilidade de obter ajuda como um elemento importante.
- **Abertura a novatos em compra online:** Incluir elementos que facilitem as compras para aqueles que nunca compraram online. Essa característica do serviço de compras coletivas está ligada a variável *Cmint*. Considerando que uma prestadora do serviço de compra coletiva online não inclua elementos desse tipo no seu website, aqueles clientes inexperientes em compras online que acessarem o site apresentarão maior dificuldade de adquirir as ofertas. Para esses clientes a falta de elementos que orientem melhor os novatos em compras online, será um ponto negativo do serviço.

Quadro 5 - Ligação entre as variáveis de decisão do problema e as variáveis do modelo

Área de Decisão	Variáveis de decisão	Variável ligada à variável de decisão	Variável do modelo ligada à variável de decisão	
			MAG binomial	MLG multinomial
Características do sistema de produção	Estrutura do site	Fan	-	FFan
	Abertura a novatos em compras coletivas	Visit	Visit,FSx	Visit
	Período das ofertas	Tint	Tint	-
	Detalhamento da ajuda	Faj	Faj	-
	Abertura a novatos em compras online	Cmint	Cmint	Cmint
	Elementos ligados à aparência do site	Ap	-	Ap
	Diversidade de produtos	Vprod	-	Vprod
	Descontos por indicação	Terc	Terc	Terc
	Formas de pagamento	Vpag	Vpag	Vpag
	Inclusão de livros na linha de produtos	Liv	-	Fliv
Características do público-alvo a ser escolhido	Sexo	Sx	FSx	-
	Idade	Id	FId	Id
	Renda	Rd	FSx; Fid	Rd; FRd; Fliv; FFan
	Anos de estudo	Aesc	Aesc	Aesc
	Renda por membro familiar	Rddep	Rddep;Dep	FDep

Fonte: Elaboração própria

- **Aparência do site:** Incluir elementos que melhorem a aparência do site, tais como maior número de imagens mais bem elaboradas. Somente no modelo multinomial foi incluída a variável *Ap*, que apresenta ligação direta com essa característica dos sites. A inclusão de elementos, tais como imagens mais bem elaboradas, que melhoram a aparência de um site acaba o tornando mais lento durante a navegação. Porém, se um cliente considera este como sendo um elemento importante, então a falta dos elementos que melhoram a aparência será um ponto negativo do serviço para aqueles.
- **Diversidade de produtos e serviços oferecidos:** Oferecer uma variedade maior de produtos e serviços. Essa característica do serviço está ligada a variável *Vprod* presente apenas no modelo multinomial proposto. Para aqueles clientes que consideram essa variedade como algo importante ou muito importante, quanto maior a variedade de itens que ofereça mais positiva tende a ser a avaliação do cliente no que diz respeito a esse quesito.
- **Descontos por indicação de novos clientes:** Conceder descontos maiores para pessoas que trouxerem novos clientes. A variável *Terc* do modelo é determinante para inclusão dessa estratégia no serviço. Se para o público que usa o serviço de compras coletivas online, a comunicação das ofertas por terceiros é uma forma pelo qual os indivíduos costumam tomar conhecimento dessas ofertas, então o uso da estratégia tende a gerar melhores resultados.
- **Diversidade de formas de pagamento:** Quantidade de formas de pagamento disponibilizadas. Essa variável está diretamente ligada a variável *Vpag* presente no modelo. Se um serviço de compras coletivas online disponibiliza muitas formas de pagamento então neste sentido será avaliado positivamente por aqueles clientes que consideram relevante este item.
- **Inclusão de linha de produto ou serviço no portfólio:** Consiste na decisão de incluir algum tipo de produtos dentre aqueles com que trabalha o site. A única variável do modelo ligada a decisão de incluir uma linha de produtos foi a variável *Liv*, presente no modelo multinomial como o fator *FLiv*. A variável *Liv* indica se um indivíduo

compra livros com regularidade ou não. Se um serviço de compras coletivas inclui os livros dentre as ofertas de produtos disponibilizados, então para aqueles clientes que comprem com regularidade este tem a ser um ponto positivo.

Cinco variáveis dos modelos foram consideradas para o processo de segmentação de mercado: sexo, renda, escolaridade, renda por membro familiar e idade. Um resumo das ligações entre as variáveis envolvidas no exemplo e aquelas presentes no modelo é mostrado na quadro 5, apresentado na página posterior.

Duas dificuldades precisam ser superadas ao usar o modelo. A segmentação baseada em variáveis quantitativas requer a definição de uma faixa de valores em vez de um valor específico, pois a escolha baseada num único valor seria muito restrita. Outra dificuldade seria lidar com variáveis que não estão envolvidas no processo de segmentação, pois se não forem especificadas supõe-se que possam assumir quaisquer dos valores. As dificuldades existem nas duas situações citadas porque o modelo só fornecerá uma medida de intenção de compra se um valor específico for fornecido para cada variável, de modo que surge o problema de escolher esses valores para que conclusões plausíveis possam ser obtidas para o segmento ou para uma especificação do serviço. Há diferentes formas de lidar com esse problema.

Uma opção seria escolher os valores dessas variáveis da maneira mais desfavorável possível para classificar como “comprador”, a fim de incluir o mínimo de indivíduos no grupo de compradores. Uma segunda opção seria utilizar os valores mais favoráveis possíveis, de modo que fosse incluído o máximo de indivíduos no grupo de compradores. Uma terceira opção seria considerar tanto valores favoráveis quanto desfavoráveis, de modo que seja possível verificar um intervalo de probabilidades. Esse critério será considerado na situação aqui exemplificada.

Com base no modelo, identificaram-se, para cada variável, os valores dentro da amostra que forneceriam a menor chance de um indivíduo ser comprador e aqueles que forneceriam a maior chance. Os valores são listados na tabela 15. São úteis para a situação aqui descrita aqueles valores referentes às variáveis “escolaridade” e “sexo”, que não fazem parte do processo de segmentação, e aqueles referentes às variáveis idade, renda e renda por dependente, que são quantitativas.

4.4.2. Descrição do exemplo

O site de compra coletiva 1 é um website bastante simples de navegar. Apresenta poucas janelas além daquelas relacionadas aos produtos e cada janela apresenta poucos níveis hierárquicos ($Fan = 1$). O site é autoexplicativo de modo que foi projetado para mesmo os indivíduos que nunca o acessaram, sejam capazes de navegar no mesmo com facilidade ($Visit=1$). As ofertas ficam geralmente disponíveis por tempo mais longo, pois se procura proporcionar ofertas com um grande número de cupons. Sendo assim, mesmo pessoas que costumam passar pouco tempo na internet acabam tendo a possibilidade de tomar conhecimento das ofertas antes mesmo que estas se encerrem (Valor de $Tint$ baixo).

O Website apresenta um guia de ajuda detalhado ($Faj=1$) e voltado especialmente para aqueles clientes novatos que apresentam pouca experiência ou nenhuma com compras na internet ($Cmint=0$). Coerente com as demais características o Website apresenta poucos elementos que visam melhorar a aparência do site ($Ap=0$). Por exemplo, há poucas imagens no site além daquelas necessárias e pouco se investe para conhecer preferências de clientes com relação à aparência do site.

A empresa procura manter um menor número de produtos ($Vprod=0$), apresentando um interesse maior em conseguir disponibilizar um número maior de cupons para o número mais limitado de itens que trabalha. Nenhum tipo de incentivo é dado para que clientes divulguem para outras pessoas os serviços do site, como descontos ainda maiores para aqueles que trouxerem novos clientes ($Terc=0$).

Outra decisão em relação a produtos e serviços oferecidos diz respeito a inclusão de um novo produto no portfólio. A empresa que durante um tempo apresentou resistência a incluir os livros no seu portfólio de produtos, optou por disponibilizar ofertas desses itens ($Liv=1$).

A empresa responsável pelo site é de pequeno porte e o website ainda é novo no mercado. Por essa razão disponibiliza poucas formas de pagamento. Os pagamentos são feitos apenas por meio de cartões de débito ($Vpag=0$).

Tabela 15 - Valores mais favoráveis e mais desfavoráveis ao julgamento de um perfil como comprador

Variável	Modelo Multinomial		MAG	
	Valor mais favorável	Valor mais desfavorável	Valor mais favorável	Valor mais desfavorável
Fan	0	1	-	-
Visit	1	0	1	0
Tint	-	-	Menor do intervalo	Maior do intervalo
Faj	-	-	1	0
Cmint	1	0	1	0
Ap	1	0	-	-
Vprod	1	0	-	-
Terc	1	0	1	0
Vpag	1	0	1	0
Sx	-	-	0	1
Id	<35	>35	Menor do intervalo	Maior do intervalo
Rd	Menor do intervalo	Maior do intervalo	Menor do intervalo	Maior do intervalo
Rddep	-	-	Próximo de 1240	Distante de 1240
Aesc	Menor do intervalo	Maior do intervalo	Menor do intervalo	Maior do intervalo
Liv	1	0	-	-

Fonte: elaboração própria.

Considerando que o proprietário do novo site de compras coletivas, cujas características foram citadas anteriormente, decida priorizar pessoas com idade não superior a 30 anos, com uma renda na faixa de R\$ 620,00 até R\$ 6.000,00 e com uma renda por membro familiar não superior a R\$ 3.000,00. Coerente com a escolha, o proprietário direciona investimentos em publicidade para esse público, define critérios para escolha de produtos e fornecedores de modo a melhor atender ao segmento.

Deseja-se utilizar os modelos propostos para verificar algumas questões acerca da relação entre a especificação do serviço de compras coletivas e a escolha de um público alvo a ser atendido.

- (1) Verificar se o público-alvo escolhido (segmento 1) apresenta, segundo os modelos, apenas indivíduos que provavelmente não utilizarão o serviço.
- (2) O quadro 4 apresenta alternativas de segmentos de mercado que poderiam ser escolhidos. Supondo que, no local da pesquisa, uma parcela considerável do público apresentasse as características do segmento 9, verificar se esse segmento seria uma boa escolha considerando os dois modelos tratados no exemplo.
- (3) Supondo que o segmento 1 seja considerado um público-alvo importante para a empresa responsável pelo site de compra coletiva e que a empresa deseja projetar seu serviço de modo que esteja em conformidade com as características, desejos e necessidades de uma parcela desse segmento que irá de fato usar o serviço. Verificar quais grupos de especificações, dentre aqueles listados no quadro 5, são mais recomendadas para atingir esse objetivo.
- (4) Supondo que a empresa pretenda atuar para atender subgrupos de pessoas interessadas no serviço, pertencentes aos 15 segmentos listados no quadro 4. Verificar que grupos de especificações seriam recomendados e quais deveriam ser evitados.

Quadro 6 - Caracterização dos segmentos de mercado

Segmento	Variáveis do processo de segmentação				
	Anos de estudo	Sexo	Renda	Idade	Renda por dependente
1	-	-	Entre 1200 e 6000	Até 30	Entre 1240 e 3000
2	-	-	Até 2480	Até 25	Até 1240
3	Acima de 15	-	Entre 2480 e 8000	Entre 25 e 35	Entre 2480 e 4000
4	Até 11	-	Até 980	Até 25	Até 980
5	Até 13	Feminino	Entre 1860 e 8000	Até 30	Entre 1860 e 4000
6	Acima de 11	Feminino	Até 1860	Até 25	Até 1860
7	Acima de 13	-	Entre 1200 e 6000	Até 35	Até 3000
8	Até 11	-	Entre 3000 e 6000	Acima de 35	Entre 1240 e 3000
9	-	-	Entre 2480 e 5000	Acima de 35	Entre 2480 e 4000
10	-	Feminino	Até 1860	Acima de 35	Até 1860
11	Até 11	-	Até 1860	Entre 25 e 35	Até 1860
12	Acima de 15	-	Entre 3000 e 5000	Acima de 35	Acima de 1860
13	Até 11 anos	-	Até 1860	Até 25	Até 1860
14	-	-	Entre 3000 e 5000	Até 25	Entre 1200 e 3000
15	Até 13 anos	Masculino	Até 3000	Até 25	Até 3000

Fonte: elaboração própria

Quadro 7 - Caracterização das alternativas de especificações para um serviço de compras coletivas

Serviço de compra coletiva	Estrutura do site	Abertura a novatos em compras coletivas	Período das ofertas	Detalhamento da ajuda	Abertura a novatos em compras online	Elementos ligados à aparência do site	Variedade de produtos e serviços	Descontos por indicação	Inclusão de livros na linha de produtos	Variedade de formas de pagamento
1	Simplex	Muita	Longo	Detalhado	Muita	Pouca	Baixa	Não Há	Sim	Pouca
2	Complexa	Muita	Curto	Detalhado	Pouca	Muita	Baixa	Há	Sim	Muita
3	Simplex	Pouca	Curto	Detalhado	Pouca	Muita	Alta	Há	Não	Muita
4	Complexa	Muita	Curto	Não Detalhado	Pouca	Muita	Alta	Há	Sim	Muita
5	Simplex	Muita	Longo	Detalhado	Muita	Muita	Alta	Não Há	Sim	Muita
6	Complexa	Pouca	Curto	Não Detalhado	Pouca	Pouca	Baixa	Há	Não	Muita
7	Simplex	Pouca	Longo	Não Detalhado	Muita	Pouca	Baixa	Não Há	Sim	Pouca
8	Simplex	Pouca	Curto	Detalhado	Pouca	Pouca	Alta	Não Há	Não	Pouca
9	Complexa	Muita	Longo	Não Detalhado	Pouca	Muita	Baixa	Não Há	Sim	Muita
10	Complexa	Pouca	Curto	Não Detalhado	Pouca	Muita	Baixa	Há	Não	Pouca
11	Complexa	Muita	Longo	Detalhado	Muita	Pouca	Alta	Não Há	Não	Muita
12	Simplex	Pouca	Curto	Não Detalhado	Pouca	Pouca	Baixa	Há	Sim	Pouca
13	Complexa	Pouca	Longo	Detalhado	Pouca	Muita	Baixa	Não Há	Não	Muita
14	Simplex	Muita	Curto	Detalhado	Pouca	Muita	Alta	Há	Sim	Muita
15	Complexa	Pouca	Longo	Não Detalhado	Muita	Pouca	Baixa	Não Há	Não	Pouca

Fonte: elaboração própria

4.4.3. Avaliação as escolhas de públicos-alvo

Há no problema duas variáveis determinantes da intenção de compra (sexo e escolaridade) que não fazem parte do processo de segmentação para definir o segmento 1. Essas variáveis podem assumir qualquer valor. Quatro das variáveis envolvidas no problema são contínuas (*Rd*, *Rddep*, *Id* e *Tint*).

Para a variável *Tint*, apenas dois valores são atribuídos: 120 min por dia se o período de ofertas é curto e 15 min por dia se o tempo é longo. Esses valores foram atribuídos porque, considerou-se que um indivíduo passa muito tempo na internet se acessa a rede por um período superior a duas horas e passa pouco tempo se acessa por menos de 120 min por dia. Como segundo um dos modelos, quanto menor o tempo que um indivíduo passa na internet menor a chance de ser “comprador” então se escolheu os valores da amostra que fornecessem a menor chance possível nos dois intervalos em questão.

Para as demais variáveis contínuas e para as variáveis categóricas para o qual não foram estabelecidos valores identificou-se os valores que forneciam a maior chance (valor mais favorável) e a menor chance (valor mais desfavorável) de um dado perfil ser julgado como de um “comprador”.

As outras variáveis do modelo têm seus valores definidos a partir da definição das características do sistema de produção, estabelecida por meio das variáveis de decisão do serviço.

Atribuídos os valores é possível obter uma medida de probabilidade de um grupo de indivíduos do segmento 1, para o qual está voltado o sistema produtivo, utilizarem o serviço. A probabilidade desse grupo de indivíduos pode ser inferior a 0,01 considerando o perfil de indivíduos com menores chances de serem “compradores” e pode chegar a 0,92 para o perfil de indivíduos com maiores chances dentro do segmento.

Como a probabilidade mínima para julgar um perfil como sendo de um comprador é 0,81, então há, no segmento, grupos com perfil favorável e desfavorável a compra cujo sistema de produção apresenta conformidade com as suas características. Isso indica que, sobre a questão (1), no segmento 1 não há apenas grupos de indivíduos “não compradores” com cujas características o serviço apresenta conformidade. A mesma conclusão é feita tanto para o MAG quanto para o MLG ao considerar a categoria 2 (compradores) e 1 (indecisos) como sendo uma única categoria.

Tabela 16 - Probabilidades e julgamentos dos segmentos de mercado de acordo com os modelos para indivíduos menos interessados no serviço

Segmento	MAG	Julgamento	Modelo multinomial		Julgamento	Julgamento Conjunto
	P=1		P(Cm=1)	P(Cm=2)		
1	<0,01	Não Comprador	0,25	0,04	Não Comprador	Não Comprador
2	<0,01	Não Comprador	0,51	0,19	Comprador	Inconclusivo
3	<0,01	Não Comprador	0,1	0,01	Não Comprador	Não Comprador
4	0,26	Não Comprador	0,44	0,11	Comprador	Inconclusivo
5	<0,01	Não Comprador	0,09	0,01	Não Comprador	Não Comprador
6	<0,01	Não Comprador	0,38	0,53	Comprador	Inconclusivo
7	<0,01	Não Comprador	0,18	0,02	Não Comprador	Não Comprador
8	<0,01	Não Comprador	0,09	0,01	Não Comprador	Não Comprador
9	<0,01	Não Comprador	0,22	0,03	Não Comprador	Não Comprador
10	<0,01	Não Comprador	0,51	0,25	Comprador	Inconclusivo
11	0,26	Não Comprador	0,35	0,06	Comprador	Inconclusivo
12	<0,01	Não Comprador	0,08	0,01	Não Comprador	Não Comprador
13	0,27	Não Comprador	0,42	0,1	Comprador	Inconclusivo
14	<0,01	Não Comprador	0,43	0,1	Comprador	Inconclusivo
15	<0,01	Não Comprador	0,37	0,07	Comprador	Inconclusivo

Fonte: elaboração própria

Tabela 17 - Probabilidades e julgamentos dos segmentos de mercado de acordo com os modelos para indivíduos menos interessados no serviço

Segmento	MAG	Julgamento	MLG multinomial		Julgamento	Julgamento Conjunto
	P(IC=1)		P(IC=1)	P(IC=2)		
1	0,92	Comprador	0,48	0,14	Comprador	Comprador
2	0,92	Comprador	0,48	0,14	Comprador	Comprador
3	0,04	Não Comprador	0,47	0,42	Comprador	Inconclusivo
4	0,88	Comprador	0,47	0,14	Comprador	Comprador
5	0,87	Comprador	0,47	0,13	Comprador	Comprador
6	0,61	Não Comprador	0,48	0,36	Comprador	Inconclusivo
7	0,47	Não Comprador	0,43	0,45	Comprador	Inconclusivo
8	0,76	Não Comprador	0,08	0,01	Não Comprador	Não Comprador
9	0,21	Não Comprador	0,18	0,03	Não Comprador	Não Comprador
10	0,92	Comprador	0,42	0,1	Comprador	Comprador
11	0,92	Comprador	0,44	0,12	Comprador	Comprador
12	0,06	Não Comprador	0,15	0,02	Não Comprador	Não Comprador
13	0,93	Comprador	0,47	0,14	Comprador	Comprador
14	0,93	Comprador	0,11	0,01	Não Comprador	Inconclusivo
15	0,58	Não Comprador	0,11	0,01	Não Comprador	Não Comprador

Fonte: elaboração própria.

A tabela 16 mostra as probabilidades do perfil mais favorável de cada segmento ser julgado como “comprador”, dentre aqueles perfis de indivíduos com cujas características a especificação (1) apresenta conformidade. A tabela 17 mostra as probabilidades para os perfis mais desfavoráveis. As probabilidades usadas para classificar os grupos de pessoas, foram calculadas para os dois modelos aqui considerados. As tabelas 16 e 17 ainda mostram a classificação segundo os critérios estabelecidos para os dois modelos.

A partir dos valores da tabela foram feitos gráficos (figuras 11 e 12) indicando a maior probabilidade e a menor probabilidade dentre aqueles perfis do segmento para o qual há conformidade entre suas características e o grupo 1 de especificações. Considerando apenas todos os perfis cujas especificações estão em conformidade, o gráfico mostra, para cada segmento, o intervalo onde se encontram todas as probabilidades dos grupos de indivíduos utilizarem o serviço.

Por exemplo, considerando o gráfico da figura 12, no segmento 12 o grupo de especificações 1 definidas para o serviço está em conformidade com as características de vários indivíduos dentro do segmento e a probabilidade desses indivíduos serem “compradores” varia de um valor inferior a 0,01 até 0,04. Se o serviço é especificado desse modo e as demais decisões acerca do sistema de produção são voltadas para o segmento 3, então a empresa responsável pelo site de compras coletivas está tomando decisões equivocadas porque o serviço somente está em conformidade com indivíduos pouco interessados em compras coletivas. É necessário, então, ou especificar melhor o serviço ou voltar as decisões para outro segmento de mercado. A mesma conclusão pode ser tomada a partir do gráfico da figura 1, que indica que aqueles indivíduos não podem ser classificados como “indecisos” e nem como “compradores”.

Com base no MAG, de acordo com a figura 12, se o serviço com as mesmas especificações é planejado para atender a pessoas do segmento 4, todos os indivíduos com cujas características o serviço está em conformidade possuem probabilidade de serem “compradores” entre 0,26 e 0,88. Isso indica que dependendo das decisões tomadas com base no segmento, o serviço pode estar voltado para um público que não tem interesse no serviço ou pode estar voltado para um público de “compradores”. Numa situação desse tipo, considerar uma análise mais aprofundada do modelo (por exemplo, com a avaliação da *odds* de variáveis) pode ser importante.

Figura 11- Gráfico das probabilidades dos perfis de um segmento pelo MLG multinomial para a especificação (1) do serviço

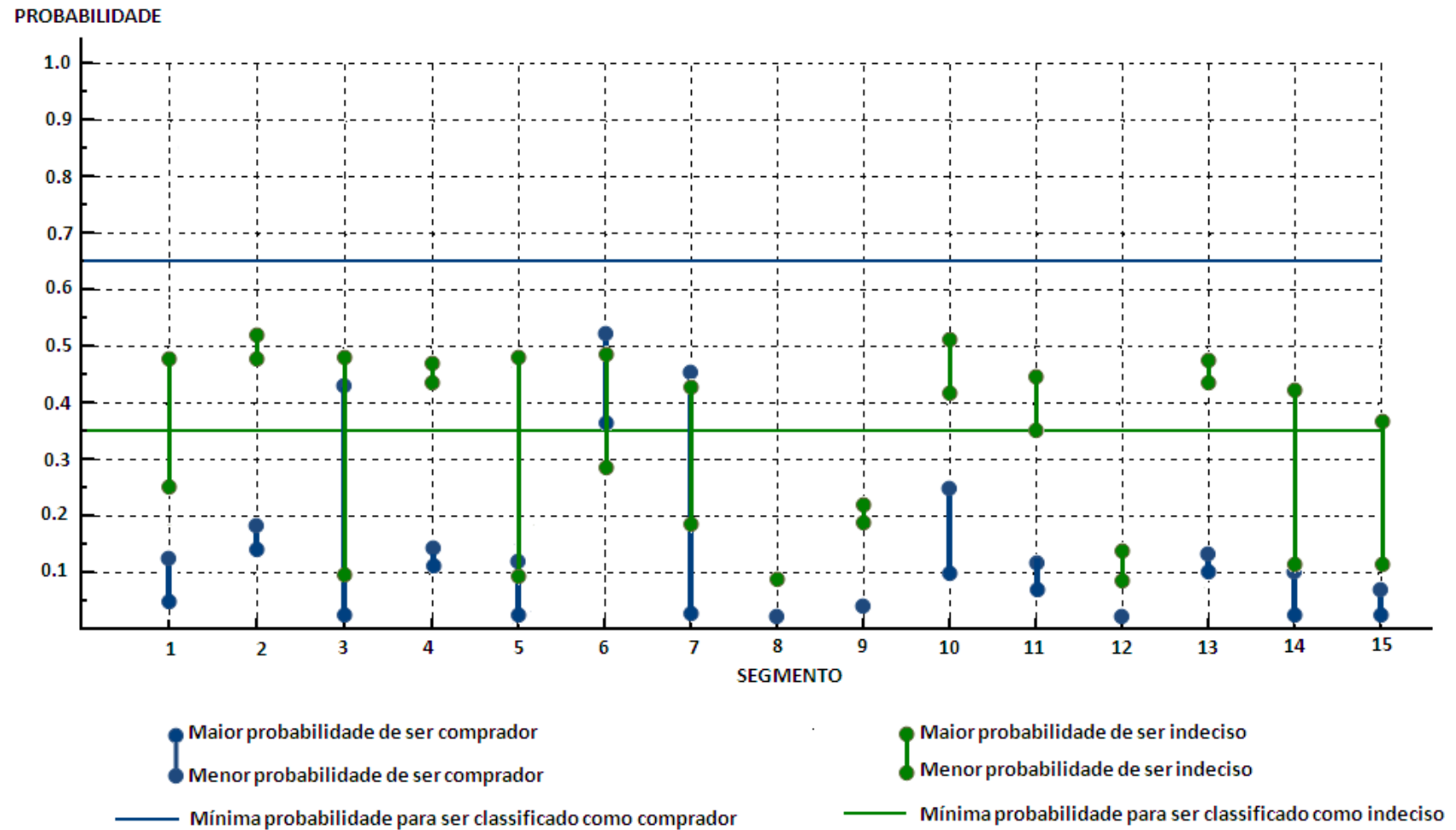
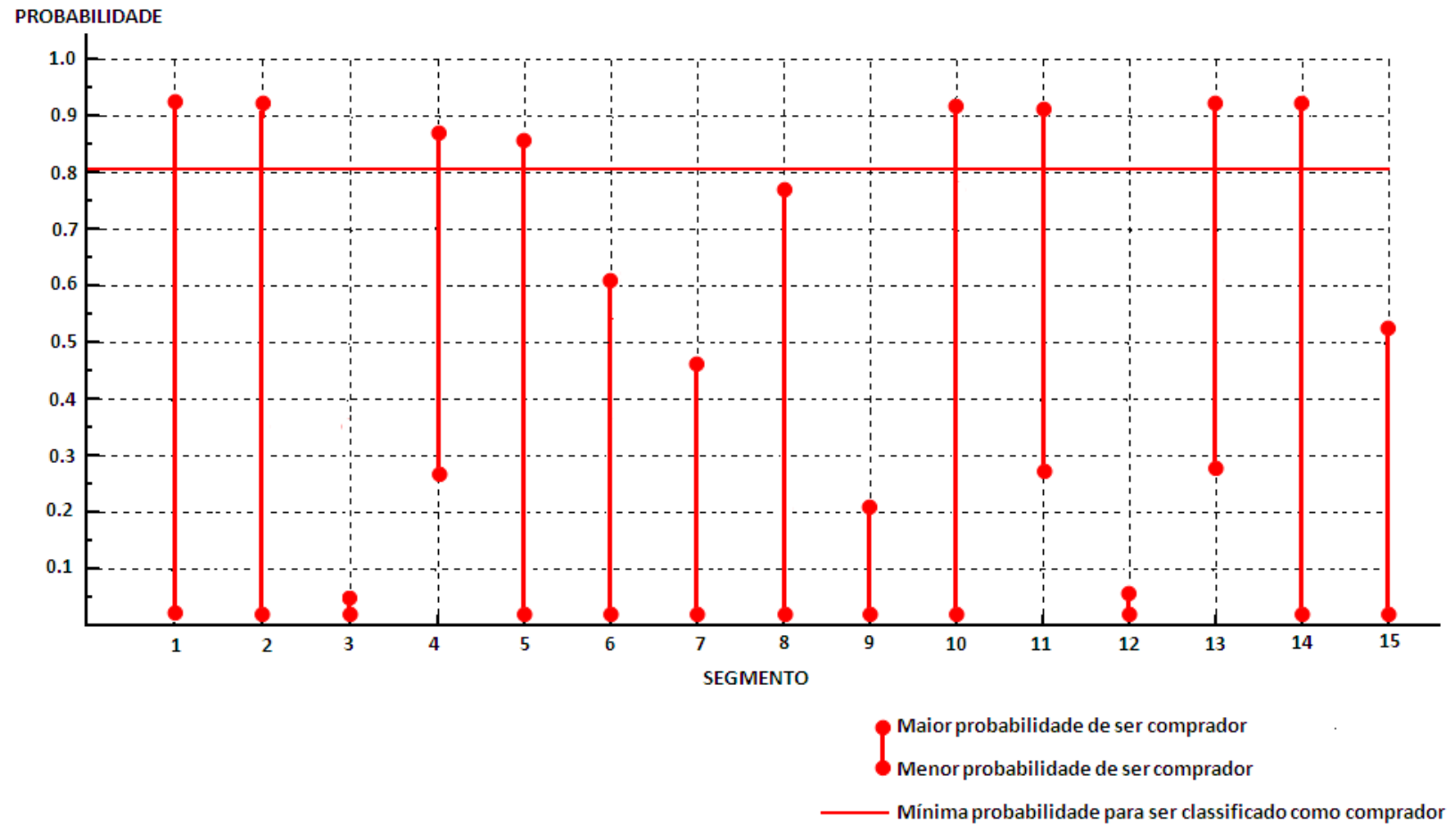


Figura 12 - Gráfico das probabilidades dos perfis de um segmento pelo MAG para especificação (1)



O MLG julga os mesmos grupos de indivíduos do segmento como possuindo apenas “indecisos”. Se não houver distinção entre pessoas classificadas como indecisas e aquelas julgadas como compradoras, então esse modelo indicará que o grupo possui apenas compradores. Ao considerar um julgamento conjunto não seria possível ter uma conclusão clara para o segmento 4, de modo que um outro meio fosse necessário para analisar a intenção de compra no segmento. Porém, os dois modelos compartilham a ideia de que a especificação e a escolha do segmento não estão totalmente equivocadas já que é possível com base nessas decisões direcionar o serviço para um público de “compradores”. Para o segmento 3 ou mesmo para o segmento 9 – respondendo a questão (2) – isso não seria possível.

A partir dos gráficos da figura 11 e 12, dada a especificação do serviço (especificação 1), é possível avaliar quais dos 15 segmentos listados no quadro 6 deveriam ser evitados, considerando a probabilidade do sistema produtivo estar em conformidade com um público-alvo de “compradores”. Pelo MLG, os segmentos 8, 9 e 12 são considerados segmentos que deveriam ser evitados. De acordo com o MAG, além dos segmentos 8, 9 e 12, são os segmentos 3, 6, 7 e 15 também não deveriam ser escolhidos se o serviço apresenta a especificação 1.

4.4.4. Avaliando as alternativas de especificações do sistema de produção

Uma aplicação dos modelos diferente da anterior é proposta nesta subseção. Procura-se agora avaliar que especificações do serviço seriam adequadas para atender a indivíduos de cada segmento. Para isso, avaliam-se mudanças nas variáveis de decisão ligadas as características do sistema produtivo. A tabela 7 mostra quinze alternativas de grupos de especificações que poderiam ser utilizados para definir as características do serviço de compras coletivas pela internet.

Considerando que o segmento 1 foi escolhido, é possível avaliar quais especificações estariam em conformidade com características dos indivíduo desse segmento que tendem a, de fato, usar o serviço. As probabilidades novamente foram calculadas para aquele grupo com perfil mais desfavorável possível a ser classificado como “comprador” e para aquele com perfil mais favorável. Os resultados são apresentados nas quadros 18 e 19.

Tabela 18 - Probabilidades e julgamentos das especificações do serviço de acordo com os modelos para os perfis de indivíduos menos interessados em usar o serviço

Especificação	MAG	Julgamento	Modelo multinomial		Julgamento	Julgamento Conjunto
	P=1		P(Cm=1)	P(Cm=2)		
1	<0,01	Não Comprador	0,25	0,04	Não Comprador	Não Comprador
2	0,9	Comprador	0,04	0,95	Comprador	Comprador
3	0,58	Não Comprador	0,15	0,83	Comprador	Inconclusivo
4	0,72	Não Comprador	0,01	0,98	Comprador	Inconclusivo
5	<0,01	Não Comprador	0,32	0,62	Não Comprador	Não Comprador
6	0,27	Não Comprador	0,42	0,47	Comprador	Inconclusivo
7	<0,01	Não Comprador	0,11	0,01	Não Comprador	Não Comprador
8	0,05	Não Comprador	0,5	0,16	Comprador	Inconclusivo
9	<0,01	Não Comprador	0,1	0,89	Comprador	Inconclusivo
10	0,06	Não Comprador	0,41	0,5	Comprador	Inconclusivo
11	<0,01	Não Comprador	0,51	0,3	Comprador	Inconclusivo
12	0,05	Não Comprador	0,52	0,25	Comprador	Inconclusivo
13	<0,01	Não Comprador	0,38	0,54	Comprador	Inconclusivo
14	0,91	Comprador	0,02	0,97	Comprador	Comprador
15	<0,01	Não Comprador	0,1	0,01	Não Comprador	Não Comprador

Fonte: elaboração própria.

Tabela 19 - Probabilidades e julgamentos das especificações do serviço de acordo com os modelos, para os perfis de indivíduos mais interessados em usar o serviço

Especificação	MAG	Julgamento	MLG multinomial		Julgamento	Julgamento Conjunto
	P(IC=1)		P(IC=1)	P(IC=2)		
1	0,92	Comprador	0,48	0,14	Comprador	Comprador
2	>0,99	Comprador	0,06	0,93	Comprador	Comprador
3	>0,99	Comprador	0,04	0,94	Comprador	Comprador
4	>0,99	Comprador	0,01	0,98	Comprador	Comprador
5	0,98	Comprador	0,25	0,7	Comprador	Comprador
6	>0,99	Comprador	0,34	0,58	Não Comprador	Inconclusivo
7	0,33	Não Comprador	0,16	0,02	Não Comprador	Não Comprador
8	0,98	Comprador	0,45	0,45	Comprador	Comprador
9	>0,99	Comprador	0,14	0,84	Comprador	Comprador
10	>0,99	Comprador	0,32	0,61	Não Comprador	Inconclusivo
11	0,98	Comprador	0,46	0,41	Comprador	Comprador
12	>0,99	Comprador	0,5	0,32	Comprador	Comprador
13	>0,99	Comprador	0,29	0,65	Comprador	Comprador
14	>0,99	Comprador	0,02	0,98	Comprador	Comprador
15	0,34	Não Comprador	0,31	0,05	Não Comprador	Não Comprador

Fonte: elaboração própria

Gráficos semelhantes àqueles construídos para avaliar segmentos foram esboçados (figura 13 e 14). Os gráficos são interpretados da mesma forma daqueles presentes na figura 11 e 12.

Dado que o segmento 1 foi escolhido, a especificação 1 está em conformidade com as características de indivíduos que possuem probabilidade de serem “compradores” que varia de um valor inferior a 0,01 até 0,92. Isso indica que a especificação 1 possibilita que o serviço seja direcionado para pessoas que tendem a usar o serviço, porém é possível, dependendo de outras decisões tomadas, que o serviço acabe se voltando para um público que dificilmente usará os sites de compras coletivas.

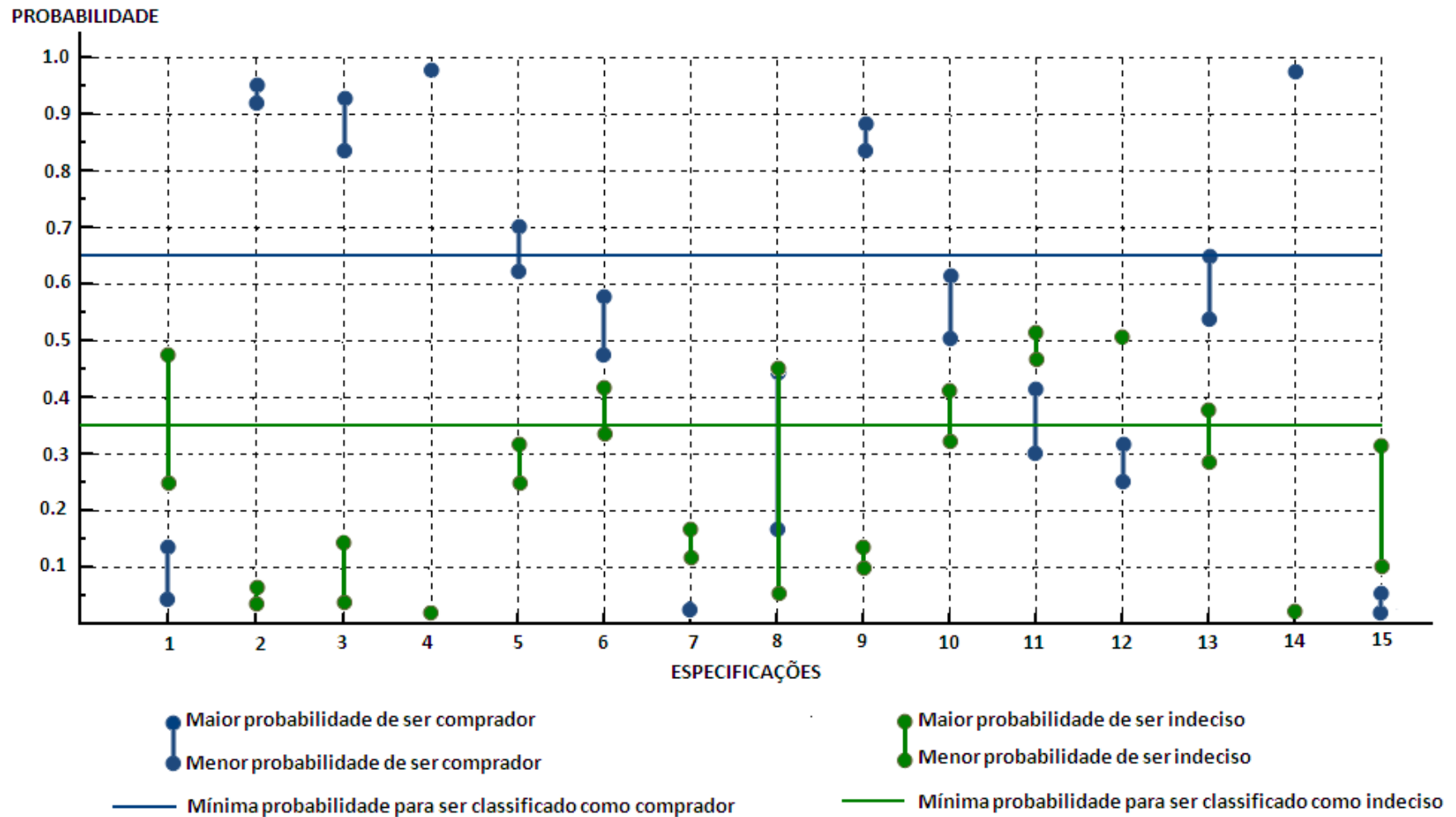
Com base no MAG, observando a figura 4, percebe-se que se o serviço for especificado com o grupo de especificações 2 ou 14, estará então em conformidade com pessoas do segmento 1 que possuem probabilidade de ser comprador entre 0,90 e um valor superior a 0,99. Isso indica que levando em conta a intenção de comprar, as especificações 2 e 14 são adequadas para atender o segmento 1. De acordo com o MLG, a mesma conclusão pode ser estendida também para os segmentos 3, 4, 9, 11 e 12. Como os dois modelos apresentam as mesmas conclusões ao julgar os grupos de especificações 2 e 14, usando conjuntamente ambos, estes poderiam ser considerados os mais recomendados para atendimento do segmento 1. Essa seria a conclusão para questão (3) indicada na subseção 4.2.2.

Os dois modelos indicam que as especificações 7 e 14 não são boas porque adequam o serviço para um público que tende a não usar o serviço. Se há alguma razão para utilizar uma das especificações, de acordo com os modelos, para manter a conformidade com pessoas que usarão o site de compra coletiva seria recomendado direcionar o serviço para um segmento de mercado diferente do segmento 1.

Supondo que o proprietário de um site de compra coletiva deseje direcionar o serviço para grupos de indivíduos com pretensão de comprar nesse tipo de websites, sem optar especificamente por um dos 15 segmentos. A questão (4) considera a verificação dos grupos de especificações recomendadas para que haja conformidade com características de pessoas que utilizarão o serviço em qualquer um dos 15 segmentos.

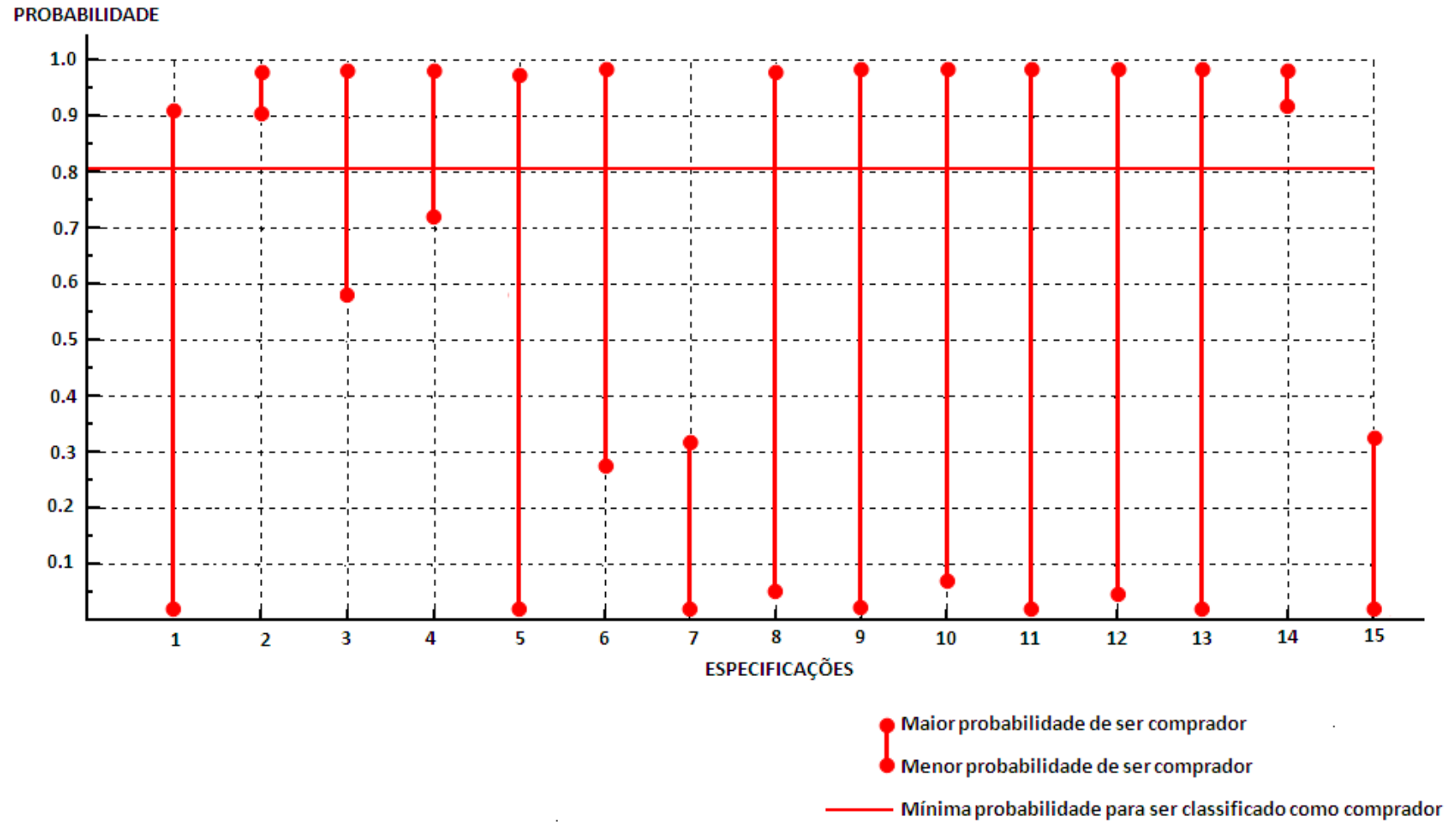
O quadro 8, baseado nos julgamentos do MAG, pode ser usado para decidir sobre o grupo de especificações que seria adequado para alcançar indivíduos “compradores” nos 15 segmentos. Para cada escolha de segmento e escolha de especificação verifica-se se os indivíduos com perfil mais favorável (com maior chance de ser “comprador”) e aqueles com

Figura 13 - Gráfico das probabilidades dos perfis de um segmento pelo MLG multinomial para a especificação (2) do serviço



Fonte: Elaboração própria

Figura 14 - Gráfico das probabilidades dos perfis de um segmento pelo MAG para a especificação (2) do serviço



Fonte: Elaboração própria

Quadro 8 - Matriz de julgamentos das combinações entre especificações do serviço e segmentos de mercado usando o MAG

Especificação	Segmentos de mercado																														Número de Julgamentos	
	Perfis mais favoráveis do segmento															Perfis mais desfavoráveis do segmento																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	C	NC
1	C	C	NC	C	C	NC	NC	NC	NC	C	C	NC	C	C	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	8	22	
2	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	NC	C	C	C	NC	C	C	C	C	C	C	C	28	2	
3	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	NC	C	NC	C	NC	C	NC	NC	NC	C	C	NC	C	C	NC	22	8
4	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	NC	C	NC	C	C	C	NC	NC	NC	C	C	NC	C	C	C	24	6
5	C	C	NC	C	C	C	C	C	C	C	C	NC	C	C	C	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	13	17
6	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	NC	C	NC	NC	NC	C	C	NC	C	NC	NC	23	7
7	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	0	30
8	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	NC	NC	NC	C	NC	C	C	NC	NC	C	C	NC	C	NC	NC	21	9
9	C	NC	C	C	C	C	C	C	C	C	C	C	C	C	C	NC	NC	NC	C	NC	NC	NC	NC	NC	NC	C	NC	C	NC	NC	17	13
10	NC	C	C	C	C	C	C	C	C	C	C	C	C	C	C	NC	NC	NC	C	NC	C	NC	NC	NC	C	C	NC	C	NC	NC	19	11
11	C	C	NC	C	C	C	C	C	C	C	C	NC	C	C	C	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	13	17
12	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	NC	NC	NC	C	NC	C	NC	NC	NC	C	C	NC	C	NC	NC	20	10
13	C	C	NC	C	C	C	C	C	C	C	C	C	C	C	C	NC	NC	NC	C	NC	NC	NC	NC	NC	C	NC	NC	C	NC	NC	17	13
14	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	NC	C	C	NC	C	C	C	C	C	C	C	C	C	28	2
15	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	0	30

Legenda: C - Comprador
NC - Não Comprador

Fonte: Elaboração Própria

Quadro 9 - Matriz de julgamentos das combinações entre especificações do serviço e segmentos de mercado usando os dois modelos

Especificação	Segmentos de mercado																														Número de Julgamentos		
	Perfis mais favoráveis do segmento															Perfis mais desfavoráveis do segmento																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	I	C	NC
1	C	C	I	C	C	I	I	NC	NC	C	C	NC	C	I	C	NC	I	NC	I	NC	I	NC	NC	NC	I	I	NC	I	I	I	12	8	10
2	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	I	C	C	C	I	C	C	C	C	C	C	C	C	2	28	0
3	C	C	C	C	C	C	C	I	C	C	C	C	C	I	C	I	C	NC	C	NC	C	I	NC	I	C	C	I	C	C	I	7	20	3
4	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	I	C	I	C	C	C	I	I	I	C	C	C	C	C	C	5	25	0
5	C	C	I	C	C	C	C	C	I	C	I	C	C	C	C	NC	I	I	I	I	I	I	I	I	I	I	I	I	I	I	17	12	1
6	I	C	C	C	C	C	C	C	C	C	C	I	C	C	C	I	C	I	C	I	C	I	I	I	C	I	I	C	I	NC	11	18	1
7	NC	NC	NC	NC	NC	I	I	NC	NC	NC	NC	NC	NC	NC	NC	NC	I	NC	NC	NC	I	NC	NC	NC	I	NC	NC	NC	NC	NC	5	0	25
8	C	C	C	C	C	C	C	C	C	I	C	C	C	I	C	I	I	NC	I	NC	C	I	C	I	C	C	NC	C	I	I	9	18	3
9	C	C	I	C	C	C	C	C	C	C	C	C	C	C	C	I	I	I	C	I	I	I	I	I	I	C	I	C	I	I	13	17	0
10	I	C	C	C	C	C	C	C	C	C	C	C	I	C	C	I	I	I	C	I	C	I	I	I	C	C	I	C	I	I	12	18	0
11	C	I	I	C	C	C	C	C	C	I	I	I	C	C	C	I	I	I	NC	I	I	I	I	I	I	I	I	I	I	I	19	10	1
12	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	I	I	I	C	I	C	I	I	I	C	C	I	C	I	I	10	20	0
13	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	I	I	I	C	I	I	I	I	I	I	C	I	C	I	I	12	18	0
14	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	I	C	C	C	I	C	C	C	C	C	C	C	C	2	28	0
15	NC	NC	NC	NC	NC	I	I	NC	NC	NC	NC	NC	NC	NC	NC	NC	I	NC	NC	NC	I	NC	NC	NC	I	NC	NC	NC	NC	NC	5	0	25

Legenda: C - Comprador
 NC - Não Comprador
 I - Inconclusivo

Fonte: Elaboração Própria

perfil mais desfavorável (com menor chance de ser “comprador”) são classificados como comprador (C) ou não comprador (NC).

Aquelas escolhas dos grupos de especificações que proporcionam o maior número de classificações como “compradores” são as melhores escolhas. Considerando os 15 segmentos, as melhores especificações para atendê-los seriam aquelas dos grupos 2 e 14. Ambas as escolhas conduzem a classificação de 28 perfis como “compradores”, do total dos 30 perfis analisados.

O quadro 9 é semelhante ao quadro 8, porém o julgamento é feito considerando conjuntamente as classificações feitas pelo MAG e pelo MLG. O perfil classificado de forma diferente pelos dois modelos é considerado inconclusivo (I). Mesmo considerando o julgamento conjunto dos modelos, pelo quadro 9, o grupo de especificações 2 e 14 são os mais indicados para manter a conformidade entre o serviço oferecido e as características daqueles indivíduos que irão usar os sites de compra coletiva, em qualquer dos 15 segmentos.

CAPÍTULO 5 - CONSIDERAÇÕES FINAIS

O modelo proposto tomou como base teórica a abordagem cognitiva do comportamento do consumidor. Dentro da abordagem cognitiva, a Teoria da Ação Racional (TAR) foi utilizada para justificar a relação entre intenção de compra e decisão de compra, bem como para orientar na busca pelos fatores que influenciam na intenção de compra. Percebeu-se que o fato de a intenção de compra ser determinada por atitudes em relação à compra e por normas subjetivas, justifica a existência de uma relação, mesmo que indireta, entre aquela variável e fatores característicos (fatores psicológicos socioculturais e situacionais) associados a um indivíduo. Como as variáveis identificadas são representações desses fatores, então a TAR pode ser usada como base do modelo proposto.

Havendo uma base teórica para justificar o modelo, em uma segunda etapa foram mostradas várias pesquisas recentes, na literatura, com modelos semelhantes ao proposto. Embora a maior parte desses modelos estejam restritos a equações estruturais lineares, todos relacionam, como o modelo proposto, a intenção de compra a fatores psicológicos, sociais e situacionais.

O modelo segue uma abordagem definida e consolidada no campo de estudos do comportamento de consumidores. Pode ser visto sob a ótica de uma teoria proposta há anos e que vem sendo usada até os dias atuais (a TAR). Além disso, está respaldado por muitas publicações que relacionam a intenção e a decisão de compra à fatores característicos associados a indivíduos.

A pesquisa teve como objetivo fundamental obter modelos aditivos generalizados que pudessem contribuir para o estudo da relação entre intenção de compras de um segmento de mercado e as suas características. A motivação para alcançar esse objetivo consiste em encontrar modelos estatísticos capazes de diferenciar o perfil de um público alvo disposto a consumir um produto ou serviço daquele que não apresenta essa intenção.

Para alcançar o objetivo da pesquisa, primeiramente foi escolhido um setor de negócios no qual seria aplicado o estudo. O setor de serviço de compras coletivas online foi escolhido devido ao seu crescimento e à importância que tem assumido. Posteriormente, por meio de uma revisão da literatura, identificaram-se formas de expressar a intenção de compras. Em seguida, recorreu-se novamente à revisão da

literatura para identificar variáveis associadas à intenção de compras em websites. Após a escolha dessas variáveis, um questionário foi elaborado para coletar os dados referentes a cada uma delas. Esses dados foram coletados por meio de entrevistas com uma amostra de 384 usuários de internet em locais de grande circulação da cidade de João Pessoa utilizando os questionários.

De posse dos dados, foram estimados modelos aditivos generalizados alternativos que relacionassem a intenção de compras às características dos indivíduos ligadas ao consumo em websites. Foram então escolhidos os melhores modelos, levando em conta alguns critérios de ajuste aos dados da amostra.

O foco dos estudos disponíveis atualmente na literatura restringe-se à identificação de variáveis ligadas ao consumo. A pesquisa apresenta a contribuição de propor uma aplicação diferente de modelos que relacionam a intenção de compra a características dos indivíduos ligadas ao consumo. Nessa aplicação, o modelo é usado para diferenciar o perfil de pessoas que não estão dispostas a consumir um produto ou serviço, de modo a auxiliar na escolha de um público alvo levando em conta a intenção de compra. Na aplicação proposta, o poder explicativo do modelo passa a ser fundamental. Por essa razão foi necessário utilizar um método estatístico distinto dos modelos de equações estruturais lineares geralmente usados na literatura. Os modelos aditivos generalizados possibilitam melhores ajustes porque permitem identificar também relações não lineares existentes entre as variáveis.

Na revisão da literatura não foi encontrada nenhuma publicação sobre o assunto envolvendo modelos aditivos generalizados. Embora seja um método flexível e poderoso, os MAG's parecem ser pouco usados ainda em pesquisas que buscam explicar a intenção de compra.

Na literatura sobre o assunto também não foram encontradas aplicações dos modelos que explicam a intenção de compra no processo de definição de um público-alvo. Utilizar os modelos nesse tipo de aplicação é importante porque possibilita levar em conta na escolha de um público alvo o conhecimento sobre indivíduos mais propensos a comprar um produto ou serviço disponibilizado a um mercado consumidor. Embora seja reconhecida a importância de considerar a intenção de compras na escolha de um público alvo e a variável seja considerada em geral de maneira intuitiva, o modelo proposto permite considerá-la de maneira formal e com base em dados estatísticos.

O melhor dos modelos obtidos apresentou um *pseudo-R²* igual a 0,89 e possibilitou alcançar um percentual de acertos nos julgamentos feitos para amostra de aproximadamente 94%. O procedimento utilizado para obter modelos com melhores ajustes é importante porque na aplicação proposta os modelos devem ter uma boa capacidade de prever, com base em dados estatísticos, se um grupo de consumidores visado por uma empresa tem a intenção de consumir seu produto ou serviço. É importante ressaltar que em uma revisão da literatura, feita ao longo da pesquisa, não foram identificados modelos com ajuste igual ou superior ao apresentado pelo modelo aditivo generalizado.

Os melhores modelos estimados foram aplicados por meio de uma simulação com os dados da amostra a fim de apresentar um procedimento para utilização na prática. Na simulação foram definidas características de um serviço de compra coletiva ligadas a uma parcela das variáveis do modelo. Uma outra parcela das variáveis foram tomadas como variáveis de decisão no processo de escolha de um público alvo. Os modelos foram então usados para avaliar 15 segmentos de mercado e 15 alternativas de especificação do serviço.

A pesquisa apresentou, no entanto, algumas limitações. A amostra usada foi obtida por um método de amostragem por conveniência e o tipo de modelo proposto não teve sua eficácia comprovada em empresas. Além disso, pode ser difícil aplicar o modelo se apenas uma parcela muito pequena de suas variáveis for tomada na avaliação dos segmentos. Essa dificuldade existe porque o modelo fornecerá nesse caso um intervalo de probabilidades que pode envolver valores muito baixos e também valores muito altos.

Uma sugestão para futuras pesquisas seguindo a mesma linha consiste em testar a eficácia do modelo em um ambiente de negócios real. Isso pode ser feito utilizando o modelo para julgar o público alvo de diferentes empresas. Espera-se que empresas voltadas para atender aqueles grupos de consumidores com alta probabilidade de consumir um produto ou serviço apresentem geralmente bons resultados econômicos, enquanto aquelas voltadas para um público com baixa probabilidade de serem “compradores” tenham resultados econômicos piores.

Ainda pode ser colocada como sugestão para futuras pesquisas a aplicação do método proposto em outros setores econômicos distintos do serviço de websites de compras coletivas.

REFERÊNCIAS

- AGRESTI, Alan. **Categorical data analysis**. Jonh Wiley: New Jersey, 2002.
- AJZEN, I; FISHBEIN, M. **Understanding attitudes and predicting social behavior**. Prentice-Hall: New Jersey, 1975.
- BAGOZZI, Richard. P; YI, Yoajae; BAUMGARTNER, Johann. The level of effort required for behaviour as a moderator of the attitude-behaviour relation. In: **European Journal Social Psychology**, V.20, p.45–59, 1990.
- BELANCHE, Daniel; CASALÓ, Luis V; GUINALÍU, Miguel. Website usability, consumer satisfaction and the intention to use a website: The moderating effect of perceived risk. In: **Journal of Retailing and Consumer Services**, V.19, p.124–132, 2012.
- BENDIXEN, M; BUKASA, K. A; ABRATT, R. Brand Equity in the Business-to-Business Market. In: **Industrial Marketing Management**, V.33, p.371-80, 2004.
- BENAZZI, João R. S. C; PEDRA, Bruno Yagelovic. Compras coletivas: uma análise exploratória de sua utilidade para as empresas anunciantes. In: SIMPÓSIO EM TENOLOGIAS DIGITAIS E SOCIABILIDADE: Salvador, 2011.
- BERG, Daniel. Bankruptcy Prediction by Generalized Additive Models. In: **Applied Stochastic Models in Business and Industry**, V.23, No.2, p.129–143, 2007.
- BIAN, Qin; FORSYTHE, Sandra. Purchase intention for luxury brands: A cross cultural comparison. In: **Journal of Business Research**, V.65, p.1443–1451, 2012.
- BOCK, Koen W. De; COUSSEMENT, Kristof; POEL, Dick Van den. Ensemble classification based on generalized additive models. In: **Computational Statistics and Data Analysis**, V.54, p.1535-1546, 2010.
- BOENTE, Graciela; RODRIGUEZ, Daniela. Robust inference in generalized partially linear models. In: **Computational Statistics and Data Analysis**, V. 54, p. 2942-2966, 2010.
- BOOR, C. de. **A Practical Guide to Splines**. New York: Springer-Verlag, 2001.
- BRUSTOLIN, J. A pesquisa de mercado inteligente. 2007. Disponível em: http://www.manalais.com.br/downloads/insights-manalais_02_mar2007.pdf. Acesso em: 27 dez.2009.
- CANNIÈRE, Marie Hélène De; PELSMACKER, Patrick De; GEUENS, Maggie. Relationship Quality and the Theory of Planned Behavior models of behavioral intentions and purchase behavior. In: **Journal of Business Research**, V.62, p.82–92, 2009.
- CASILLAS, Jorge; MARTÍNEZ-LÓPEZ, Francisco J. Mining uncertain data with multiobjective genetic fuzzy systems to be applied in consumer behaviour modelling. In: **Expert Systems with Applications**, V.36, p.1645–1659, 2009.
- CHASE, Richard B.; JACOBS, F. Robert; AQUILANO, Nicholas J. **Administração da Produção e Operações para vantagens competitivas**. São Paulo: McGraw – Hill, 2006.

CHEN, Mei-Fang. Consumer attitudes and purchase intentions in relation to organic foods in Taiwan: Moderating effects of food-related personality traits. In: **Food Quality and Preference**, V.18, pg. 1008–1021, 2007.

_____, Ying-Hueih; HSU, I-Chieh; LIN, Chia-Chen. Website attributes that increase consumer purchase intention: A conjoint analysis. In: **Journal of Business Research**, V.63, p. 1007–1014, 2010.

CHIU, Chui-Yu; CHEN, Yi-Feng; KUO, I-Ting; CHUN KU, He. An intelligent market segmentation system using k-means and particle swarm optimization. In: **Expert Systems with Applications**, V.36, p. 4558–4565, 2009.

CORDEIRO, Gauss M.; ANDRADE, Marinho G. de. Transformed generalized linear models. In: **Journal of Statistical Planning and Inference**, V. 139, p. 2970 – 2987, 2009.

CORRÊA, H.L.; CORRÊA, C.A. **Administração de Produção e Operações**. São Paulo: Atlas, 2004.

COSTA, Ricardo Simm. **Identificação de fatores que influenciam a decisão de aquisição de software em um contexto profissional**. Dissertação (Mestrado em Administração) – Universidade Federal do Rio Grande do Sul. 2006.

CRAN. R-2.13.2 for Windows (32/64 bit). Disponível em: <http://cran.r-project.org/bin/windows/base/old/2.13.2/>. Acesso em: set/2012.

DEHUA, HE; YAOBIN, LU; DEYI, ZHOU. Empirical Study of Consumers' Purchase Intentions in C2C Electronic Commerce. In: **Tsinghua Science and Technology**, V.13, p. 287-292, 2008.

DIALLO, Mbaye Fall. Effects of store image and store brand price-image on store brand purchase intention: Application to an emerging market. In: **Journal of Retailing and Consumer Services**, V.19, p.360–367, 2012.

DURAN, Esra Akdeniz; AKDENIZ, Fikri; HU, Hongchang. Efficiency of a Liu-type estimator in semiparametric regression models. In: **Journal of Computational and Applied Mathematics**, V.235, p.1418–1428, 2011.

ENGEL, James F.; BLACKWELL, Roger D.; MINIARDI, Paul W. **O Comportamento do Consumidor**. ed.8. Rio de Janeiro: LTC, 2000.

FARAWAY, Julian J. **Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models**. London: Chapman&Hall/CRC, 2006.

FISHBEIN, M; AJZEN, I. **Belief, attitude, intention, and behavior: An introduction to theory and research**. Reading, MA: Addison-Wesley, 1975.

FLORENCIO, Lutemberg de Araújo. **Engenharia de avaliações com base em modelos GAMLSS**. Dissertação (Mestrado em Estatística) – Universidade Federal de Pernambuco. 2010.

FORTIN, Marie-Fabienne. **O processo de investigação**. Montreal: Décaire Éditeur, 2009.

GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 1994.

- GONÇALVES, Nayara F. de Moura. **Bootstrap em modelos auto-regressivos aditivos generalizados**. Dissertação (Mestrado em Estatística) – Universidade Federal de Minas Gerais. 2009.
- GONZÁLEZ, María Elisa Alén; COMESAÑA, Lorenzo Rodríguez; BREA, José Antonio Fraiz. Assessing tourist behavioral intentions through perceived service quality and customer satisfaction. In: **Journal of Business Research**, V.60, p.153–160, 2007.
- GUERRA, Nuno José Martins. Análise do processo de decisão de compra do consumidor de vinho. Dissertação (Mestrado em Gestão de Empresas) – Universidade de Évora. 2005.
- GIGLIO, Ernesto M. **O comportamento do consumidor**. ed.3. São Paulo: Pioneira, 2004.
- GUISAN, Antoine; EDWARDS, Thomas C.; HASTIE, Trevor. Generalized linear and generalized additive models in studies of species distributions: setting the scene. In: **Ecological Modelling**. V.157, p.89-100, 2002.
- JANG, SooCheong; NAMKUNG, Young. Perceived quality, emotions, and behavioral intentions: Application of an extended Mehrabian–Russell model to restaurants. In: **Journal of Business Research**, V.62, p.451–460, 2009.
- HARDLE, Wolfgang; MULLER, Marlene; SPERLICH, Stefan; WERWATZ, Axel. **Nonparametric and semiparametric Models**. Springer: Berlin, 2008.
- HARRISON, David A; MYKYTYN Jr, Peter P; RIEMENSCHNEIDER, Cynthia K. Executive Decisions About Adoption of Information Technology in Small Business: Theory and Empirical Tests. In: **Information Systems Research**, v.8, n.2, p.171-195, 1997.
- HARTMANN, Patrick; APAOLAZA-IBÁÑEZ, Vanessa. Consumer attitude and purchase intention toward green energy brands: The roles of psychological benefits and environmental concern. In: **Journal of Business Research**, V.65, p.1254–1263, 2012.
- HAUSMAN, Angela V; SIEKPE, Jeffrey Sam. The effect of web interface features on consumer online purchase intentions. In: **Journal of Business Research**, V.62, p.5–13, 2009.
- HILL, Manuela Magalhães; HILL, Andrew. **Investigação por questionário**. Lisboa: Sílabo, 2009.
- HISTIE, T. J; TIBISHIRANI, R. J. **Generalized Additive Models**. CHAPMAN&HALL: New York, 1990.
- HOFMANN, Wilhelm; STRACK, Fritz; DEUTSCH, Roland. Free to buy? Explaining self-control and impulse in consumer behavior. In: **Journal of Consumer Psychology**, V.18. p.22–26, 2008.
- HOSMER, D.W.; LEMESHOW, S. **Applied Logistic Regression**. ed.2. Editora John Wiley and Sons, 2000.
- HU, Bo; Shao, Jun. Generalized linear model selection using R^2 . In: **Journal of Statistical Planning and Inference**, V. 138, p.3705 – 3712, 2008.
- HUNG, Chihli; TSAI, Chih-Fong. Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand. In: **Expert Systems with Applications**, V.34, p. 780–787, 2008.

HRUSCHKA, Harald; NATTER, Martin. Comparing performance of feed forward neural nets and K-means for cluster-based market segmentation. In: **European Journal of Operational Research**, V.114, pg. 346-353, 1999.

IBGE. Censo 2010. 2010. Disponível em: <<http://www.censo2010.ibge.gov.br/>>. Acesso em: 10 fev 2012.

JOHNSON, R.A; WICHERN, D.W. **Applied Multivariate Statistical Analysis**.ed.6.New Jersey:Prentice-Hall, 2007.

KARSAKLIAN, Eliane. **Comportamento do Consumidor**. São Paulo: Atlas, 2005.

KEH, Hean Tat; XIE, Yi. Corporate reputation and customer behavioral intentions: The roles of trust, identification and commitment. In: **Industrial Marketing Management**, V.38, p. 732–742. 2009.

KIM, Eun Young; KIM, Youn-Kyung. Predicting online purchase intentions for clothing products. In: **European Journal of Marketing**, V.38, No.7, p.883-897, 2004.

KOTLER, P.; ARMISTRONG, G. **Princípios de Marketing**. 9ed. São Paulo: Pearson, 2003.

_____, Philip; KELLER, Kevin Lane. **Administração de Marketing**. 12ed. São Paulo: Pearson Prentice Hall, 2006.

KUMAR, Archana; LEE, Hyun-Joo; KIM, Youn-Kyung . Indian consumers purchase intention toward a United States versus local brand. In: **Journal of Business Research**, V.62, p. 521–527, 2009.

KUO, R.J.; HO, L.M.; HU, C.M. Integration of self-organizing feature map and K-means algorithm for market segmentation. In: **Computers & Operations Research**, V.29, p. 1475 -1493, 2002.

LAMB JR., Charle W.; HAIR JR., Joseph F.; MCDANIEL, Carl. **Princípios de Marketing**. São Paulo: Pioneira Thomson Learning, 2004.

LIN, Hsiu-Fen. Predicting consumer intentions to shop online: An empirical test of competing theories. In: **Electronic Commerce Research and Applications**, V.6, p.433–442, 2007.

LIMA,Lilian Pereira; ANDRÉ, Carmen Diva; SINGER, Julio da Motta. Modelos Aditivos Generalizados: metodologia e prática. In: **Revista Brasileira de Estatística**, V.62, No.217, p.37-71, 2001.

LÔ, Serigne N.; RONCHETTI, Elvezio. Robust and accurate inference for generalized linear models. In: **Journal of Multivariate Analysis**, V. 100, p. 2126-2136, 2009.

LOPES, Isac José. **A segmentação de mercado como estratégia de marketing**. São Paulo: Makron, 2002.

LU,Hsi-Peng; HSIAO,Kuo-Lun.The influence of extro/introversion on the intention to pay for social networking sites. In: **Information & Management**, V.47, 150–157, 2010.

MALHOTRA, Naresh K. **Pesquisa de Marketing: Uma orientação aplicada**. Porto Alegre: Bookman, 2006.

MCCULLAGH, P; NELDER, J. A. **Generalized Linear Model**. Chapman&Hall/CRC: New York, 1989.

- MING-CHIH, Tsai; YI-TING, Tsai; CHING-WEI, Lien. Generalized linear interactive model for market segmentation: The air freight market. In: **Industrial Marketing Management**, V.40, p. 439–446, 2011.
- MOON, Junyeon; CHADEE, Doren; TIKOO, Surinder. Culture, product type, and price influences on consumer purchase intention to buy personalized products online. In: **Journal of Business Research**. V.61, p.31–39, 2008.
- MORWITZ, Vicki G.; STECKEL, Joel H.; GUPTA, Alok. When do purchase intentions predict sales? In: **International Journal of Forecasting**, V. 23, p. 347–364, 2007.
- NELDER, J.A. ; WEDDERBURN, R.W.M. Generalized Linear Models. In: **Journal Royal Statistics Society**, V. 135, p.370-384, 1972.
- PINHEIRO, Roberto Meireles; CASTRO, Guilherme Caldas e outros. **Comportamento do consumidor e pesquisa de mercado**. Rio de Janeiro: Editora FGV, 2006.
- PODDAR, Amit; DONTU, Naveen; WEI, Yujie. Web site customer orientations, Web site quality, and purchase intentions: The role of Web site personality. In: **Journal of Business Research**, V. 62, p. 441–450, 2009.
- RAMAYAH, T.; LEE, Jason Wai Chow; MOHAMAD, Osman. Green product purchase intention: Some insights from a developing country. In: **Resources, Conservation and Recycling**, V.54, p.1419–1427, 2010.
- RIGBY, Bob; STASINOPOULOS, Mikis. **A flexible regression approach using GAMLSS in R**. Disponível em: <http://gamlss.org/images/stories/papers/Lancaster-booklet.pdf>. Acesso em: set/2012.
- RITZMAN, L.; KRAJEWSKI L. **Administração da Produção e Operações**. S. Paulo: Prentice Hall, 2004.
- RUCKER, Derek D.; GALINSKY, Adam D.; DUBOIS, David. Power and consumer behavior: How power shapes who and what consumers value. In: **Journal of Consumer Psychology** (2011), doi:10.1016/j.jcps.2011.06.001.
- RYAN, Thomas P. **Modern Regression Methods**. Jonh Wiley: New Jersey, 2009.
- SHIM, Soyeon; EASTLICK, Mary Ann; LOTZ, Sherry L; WARRINGTON, Patricia. An online prepurchase intentions model: The role of intention to search. In: **Journal of Retailing**, V.77, p.397–416, 2001.
- SHUKLA, Paurav. Impact of interpersonal influences, brand origin and brand image on luxury purchase intentions: Measuring interfunctional interactions and a cross-national comparison. In: **Journal of World Business**, V. 46, p. 242–252, 2011.
- SILVA, Aline Lisboa da; VIEIRA, Eloy dos Santos; SILVA, Andreza Lisboa da; FELIZOLA, Matheus Pereira Mattos. Compras coletivas: estudo de caso sobre o site alguns trocados em Aracaju. XXXIV Congresso Brasileiro de Ciências da Comunicação: Recife, 2011.
- SOLOMON, Michael R. **O comportamento do consumidor: comprando, possuindo e sendo**. Porto Alegre: Bookman, 2008.
- SOUZA, C. K.; CARDOSO, O. R. **Análise comparativa entre as interfaces do Marketing e Engenharia de Produção**. Disponível em: http://www.peteps.ufsc.br/novo/attachments/069_artigoseprosulmeu.pdf. Acesso em: 26 dez.2009.

_____, Camila Pedroso Estevam. **Testes de hipóteses para dados funcionais baseados em distâncias: um estudo usando splines**. Dissertação (Mestrado em Estatística) – Universidade Estadual de Campinas. 2008.

STEIN, Fabiano Rocha. **Modelagem de produção industrial de celulose KRAFT com modelos aditivos generalizados e redes neurais**. Dissertação (Mestrado em Tecnologia de Celulose e Papel) – Universidade Federal de Viçosa. 2010.

SUN, Baohong; MORWITZ, Vicki G. Stated intentions and purchase behavior: A In; unified model. **International Journal of Research in Marketing**. V. 27, p. 356–366, 2010.

TENG, Lefa; LAROCHE, Michel. Building and testing models of consumer purchase intention in competitive and multicultural environments. In: **Journal of Business Research**, V.60, p. 260–268, 2007.

UNDERWOOD, Fiona M. Describing long-term trends in precipitation using generalized additive models. In: **Journal of Hydrology**. V. 364, p.285–297, 2009.

VERHAGEN, Tibert; DOLEN, Willemijn van. Online purchase intentions: A multi-channel store image perspective. In: **Information & Management**, V. 46 p. 77–82, 2009.

VIJAYASARATHY, Leo R. Predicting consumer intentions to use online shopping: the case for an augmented technology acceptance model. In: **Information & Management**, V.41, p.747–762, 2004.

YAO, Weixin. A bias corrected nonparametric regression estimator. In: **Statistics and Probability Letters**, V.82, p.274–282, 2012.

YEE, T. W.; WILD, C. J. Vector Generalized Additive Models. In: **Journal of the Royal Statistical Society**. V.58, No.3, p.481–493, 1996.

YOO, Boonghee; DONTHU, Naveen; LEE, Sungho. An Examination of Selected Marketing Mix Elements and Brand Equity. In: **Journal of the Academy of Marketing Science**. V.28, No. 2, p.195–211.

YOU, Jinhong; ZHOUB, Xian. Statistical inference in a panel data semiparametric regression model with serially correlated errors. In: **Journal of Multivariate Analysis**, V.97, p.844–873, 2006.

YOUNG, Robert B.; JAVALGI, Rajshekhar G. International marketing research: A global project management perspective. In: **Business Horizons**, V.50, p.113–122, 2007.

YOUNG, Robin L; WEINBERG, Janice; VIEIRA, Veronica; OZONOFF, Al; WEBSTER, Thomas F. Generalized additive models and inflated type I error rates of smoother significance tests. In: **Computational Statistics and Data Analysis**, V.55, p. 366–374, 2011.

WOOD, Simon. **Generalized Additive Models: An introduction with R**. Chapman e Hall/CRC: New York, 2006.

WU, Paul C.S.; YEH, Gary Yeong-Yuh; HSIAO, Chieh-Ru. The effect of store image and service quality on brand image and purchase intention for private label brands. In: **Australasian Marketing Journal**, V. 19, p.30–39, 2011.

XU, Yingjiao; SUMMERS, Teresa A; BELLEAU, Bonnie D. Who buys American alligator? Predicting purchase intention of a controversial product. In: **Journal of Business Research**, V.57, p.1189–1198, 2004.

ZEITHAML, V, BERRY, L;PARASURAMAN, A. The behavioral consequences of service quality. In: **Journal of Marketing**, V.60, p.31-46, 1996.

APÊNDICE A - Questionário



1 - Identificação da pesquisa

Bom dia / boa tarde / boa noite! Essa pesquisa tem como objetivo conhecer melhor os hábitos de consumo em sites de compra coletiva (Sites que oferecem cupons com descontos para produtos e serviços de empreendimentos. Os descontos geralmente são válidos apenas se um número mínimo de pessoas comprarem a oferta). A pesquisa é realizada pelo Programa de Pós-Graduação em Engenharia de Produção (PPGEP) da UFPB. Tal pesquisa será usada para elaboração de um trabalho de dissertação de mestrado. Contamos com o seu apoio para responder ao questionário. A sua resposta para todas as perguntas é muito importante para o êxito do trabalho. Cabe ressaltar que o sigilo das informações será plenamente preservado. Pela sua colaboração especial, antecipamos nossos mais sinceros agradecimentos!

2 - Caracterização do entrevistado

2.1) Sexo: 1.(☐) Masculino 2.(☐) Feminino

2.2) Estado civil: 1.(☐) Solteiro 2.(☐) Casado / União estável 3.(☐) Divorciado / separado 4.(☐) Viúvo

2.3) Idade: _____

2.5) Qual a sua renda domiciliar mensal? _____

2.6) Quantas pessoas dependem dessa renda? _____

2.7) Bairro onde reside: _____

3 – Indique a importância que você atribui as características de sites de compra coletiva.

Quando escolho um site de compra coletiva para comprar considero...	Sem importância	De pouca importância	Importante	De grande importância
3.1) A facilidade de obter ajuda no site	0	1	2	3
3.2) A facilidade de navegar pelo site	0	1	2	3
3.3) Quantidade informação sobre os produtos ou serviços oferecidos no site	0	1	2	3
3.4) A aparência do site	0	1	2	3
3.5) O fato das empresas que oferecem o produto ou serviço no site serem conhecidas	0	1	2	3
3.6) O fato do site que anuncia a oferta ser conhecido	0	1	2	3
3.7) A qualidade do serviço de atendimento ao consumidor	0	1	2	3

3 – Indique a importância que você atribui as características de sites de compra coletiva.

Quando escolho um site de compra coletiva para comprar considero...	Sem importância	De pouca importância	Importante	De grande importância
3.8) A variedade de produtos ou serviços ofertados	0	1	2	3
3.9) A ausência do limite de tempo para compra das ofertas disponíveis	0	1	2	3
3.10) A variedade de formas de pagamento	0	1	2	3
3.11) A velocidade de navegação do site	0	1	2	3

4 – Questões relativas a compras de serviços, compras na internet e sites de compras coletivas

4.1) Com que frequência você visita sites de compras coletivas?

- 1.(☐) Nunca visito 4.(☐) Toda semana
 2.(☐) Poucas vezes no ano 5.(☐) Todo dia
 3.(☐) Todo mês

4.2) Quantas horas ou minutos você costuma passar na internet diariamente? _____**4.3) Você costuma comprar em lojas virtuais?**

- 1.(☐) Bastante 2.(☐) Pouco 3.(☐) Nunca

4.4) Já tomou conhecimento de alguma oferta de sites de compras coletivas?

- 1.(☐) Sim 2.(☐) Não

Se você marcou "Sim" na pergunta anterior (4.4) responda o item 4.5.

4.5) De que forma você já tomou conhecimento das promoções disponíveis em sites de compra coletiva? (pode marcar mais de uma opção)

- 1.(☐) Direto no site 5.(☐) Através de terceiros
 2.(☐) Publicidade em outros sites 6.(☐) Sites de busca
 3.(☐) Em redes sociais 7.(☐) Outro _____
 4.(☐) Por email

4.6) Que tipo(s) de produto(s) ou serviço(s) você costuma consumir com regularidade fora da internet? (pode marcar mais de uma opção)

- 1.(☐) Serviços de lazer (bares, restaurantes, alimentação) 5.(☐) Vestuários
 2.(☐) Produtos e serviços de estética 6.(☐) Livros
 3.(☐) Eletro-eletrônicos 7.(☐) Revistas
 4.(☐) Turismo / Diárias em hotéis ou pousadas / Passagens 8.(☐) Outros _____

4.7) Você compraria produtos e serviços em sites de compras coletivas?

- 1.(☐) Certamente compraria
 2.(☐) Provavelmente Compraria
 3.(☐) Provavelmente não compraria
 4.(☐) Certamente não compraria

Se NÃO marcou o item 4 na pergunta anterior (4.7), responda as perguntas 4.8 e 4.9.

4 – Questões relativas a compras de serviços, compras na internet e sites de compras coletivas**4.8) Qual o mínimo desconto que levaria você a comprar em um site de compra coletiva qualquer produto ou serviço que você costuma consumir com regularidade ? (Citar um valor entre 0 e 100%)**

4.9) Que tipos de produtos ou serviços você consumiria em sites de compras coletivas? (pode marcar mais de uma opção)

- 1.(☐) Serviços de lazer (bares, restaurantes, alimentação) 5.(☐) Vestuários
 2.(☐) Produtos e serviços de estética 6.(☐) Livros
 3.(☐) Eletro-eletrônicos 7.(☐) Revistas
 4.(☐) Turismo / Diárias em hotéis ou pousadas / Passagens 8.(☐) Outros _____

APÊNDICE B - Descrição da amostra

A amostra consiste em 420 observações relativas a 28 variáveis. Dessas observações, 384 foram usadas na estimação dos modelos e 40 foram usadas na validação.

Tabela 20 - Medidas descritivas das variáveis contínuas da amostra

Medidas Descritivas	Variáveis				
	Idade	Renda	Renda por dependente	Anos de escolaridade	Tempo na internet (min)
Média	30	2539	1079	13	187
Mediana	27	1860	750	13	120
Desvio Padrão	10	2172.8	940	3	156
Menor	18	620	155	4	15
Maior	59	15000	6000	30	960
Quartil Inferior	23	1200	533	11	60
Quartil Superior	35	3000	1275	15	240
Número de Outliers	19	39	34	10	13

Fonte: Elaboração própria

A tabela 25 mostra algumas medidas descritivas para as variáveis contínuas da pesquisa. Em nenhuma das variáveis contínuas o valor da mediana apresentou grande diferença em relação ao valor da média, sendo a maior para a variável renda por dependente cuja diferença foi de 30% do valor da média. Sendo assim, as duas medidas de posição têm semelhante representatividade para sumarizar as observações da amostra para quaisquer variáveis.

Três das cinco variáveis contínuas apresentaram uma variabilidade relativamente alta, embora em nenhum dos casos o desvio padrão é superior a 90% do valor da média. Essas variáveis são: Renda, renda por dependente e tempo na internet. Para as variáveis idade e anos de escolaridade, o desvio padrão não foi superior a 35% do valor da média.

As duas variáveis relativas à renda apresentaram um número maior de outliers. Apesar desse número de outliers, 90% das observações dessas variáveis não apresentaram grandes diferenças. Para as demais variáveis o número de outliers não foi superior a 4% das observações. Em qualquer tratamento dos dados é importante

considerar essas observações porque são importantes para representar indivíduos com características semelhantes na população.

O quartil inferior e o menor valor indicam o intervalo da variável que apresenta 25% das observações, enquanto o quartil superior e o maior valor indicam o intervalo que apresenta 75% das observações. Por exemplo, 25% dos entrevistados apresentaram uma idade entre 18 e 23 anos. Da amostra, 75% apresentaram uma idade entre 23 e 35 anos.

Quadro 10 - Matriz de correlações das variáveis contínuas

	Idade	Renda	Renda por dependente	Anos de escolaridade	Tempo na internet (min)
Idade	1,000	0,239	0,186	0,100	-0,139
Renda	0,240	1,000	0,698	0,434	0,029
Renda por dependente	0,186	0,698	1,000	0,459	0,120
Anos de escolaridade	0,100	0,434	0,459	1,000	0,184
Tempo na internet (min)	-0,139	0,029	0,120	0,184	1,000

Dentre os pares de variáveis contínuas a menor correlação ocorreu entre a renda e o tempo que um indivíduo passa na internet (quadro 8). A maior correlação ocorreu entre as variáveis relativas à renda, o que era esperado uma vez que a renda por dependente foi calculada usando os valores da variável renda. As variáveis relativas à renda também apresentaram uma correlação um pouco mais alta com a variável “anos de escolaridade”.

Foram entrevistadas pessoas de 48 bairros da cidade de João Pessoa. Maior parte deles residia em Mangabeira (ver quadro 9). Além desses bairros, apresentaram um maior percentual de entrevistados os bairros: Bancários, Bessa, Centro, Jaguaribe, Geisel, Funcionários e Valentina.

Quadro 11 - Distribuição de Frequência amostral em relação aos bairros dos entrevistados

Bairro	Frequência Absoluta	Frequência Relativa	Bairro	Frequência Absoluta	Frequência Relativa
Água Fria	9	2,34%	Jardim Oceania	1	0,26%
Altiplano	3	0,78%	Jardim Planalto	10	2,60%
Alto do Mateus	5	1,30%	Jardim Veneza	2	0,52%
Bairro das Industrias	12	3,13%	João Paulo II	2	0,52%
Bairro dos Estados	5	1,30%	José Américo	10	2,60%
Bairro dos Ipês	2	0,52%	Manaíra	9	2,34%
Bairro dos Novaes	4	1,04%	Mandacaru	6	1,56%
Bancários	26	6,77%	Mangabeira	39	10,16%
Bessa	17	4,43%	Miramar	1	0,26%
Brisamar	1	0,26%	Muçumagro	1	0,26%
Cabo Branco	1	0,26%	Oitizeiro	5	1,30%
Castelo Branco	12	3,13%	Padre Zé	1	0,26%
Centro	23	5,99%	Penha	2	0,52%
Colibris	2	0,52%	Portal do Sol	2	0,52%
Colinas do Sul	5	1,30%	Rangel	7	1,82%
Costa e Silva	9	2,34%	Roger	6	1,56%
Cristo	14	3,65%	São José	1	0,26%
Cruz das Armas	9	2,34%	Tambaú	10	2,60%
Ernani Sátiro	5	1,30%	Tambauzinho	2	0,52%
Expedicionários	3	0,78%	Tambiá	3	0,78%
Funcionários	19	4,95%	Torre	14	3,65%
Geisel	20	5,21%	Treze de Maio	1	0,26%
Jaguaribe	17	4,43%	Valentina	18	4,69%
Jardim Cid. Universitária	6	1,56%	Varadouro	2	0,52%

Sobre a importância dos fatores relativos às características do serviço de compras coletivas, o único desses que apresentou um percentual de entrevistados considerando sem importância alguma foi a aparência dos websites. Mesmo assim, em todos os casos maior parte dos entrevistados julgaram os fatores tratados na parte 3 do questionário como importante ou muito importantes.

Tabela 21 - Atitudes dos entrevistados em relação a características do serviço de compras coletivas

Fator	Nível de Importância Atribuído			
	Não Importante	Pouco Importante	Importante	Muito Importante
Facilidade de Obter Ajuda no Site	4,69%	20,27%	54,95%	19,79%
Facilidade de Navegar pelo Site	3,84%	14,65%	48,70%	32,81%
Aparência do Site	12,50%	26,30%	37,24%	23,96%
Empresas parceiras serem conhecidas	1,30%	6,51%	30,99%	61,20%
Site de compras coletivas ser conhecido	1,56%	7,03%	36,72%	54,69%
Variedade de produtos/serviços	3,91%	14,32%	50,78%	30,99%
Variedade de formas de pagamento	2,86%	5,73%	37,24%	54,17%
Quantidade de informação disponibilizada sobre os produtos	1,82%	6,51%	35,94%	55,73%
Qualidade do Serviço de Atendimento ao Consumidor	2,34%	10,68%	25,52%	61,46%
Velocidade de navegação do site	3,12%	10,16%	35,42%	51,30%

Fonte: Elaboração própria

As variáveis consideradas como sendo as mais importantes para definição das características desejadas da amostra foram as variáveis: renda, idade e sexo. O quadro 10 mostra uma tabela de contingência com os pesos relativos da combinação de categorias dessas três variáveis. Pelo quadro, é possível perceber que oito dos quarenta grupos indicados apresentaram peso relativo superior a 4% da amostra. Dos quarenta grupos, apenas três deles não apresentou nenhum entrevistado.

Quadro 12 - Tabela de contingência das variáveis Sexo, Renda e Idade

Sexo	Renda (R\$)	Idade			
		18-25	25-35	35-45	Mais de 45
Masculino	620-1240	0,115	0,086	0,01	0,018
	1240-2480	0,07	0,055	0,008	0,01
	2480-3720	0,031	0,013	0,013	0,005
	3720-4960	0,005	0,008	0	0
	Mais de 4960	0,018	0,029	0,013	0,018
Feminino	620-1240	0,07	0,031	0,016	0,003
	1240-2480	0,047	0,049	0,026	0,01
	2480-3720	0,031	0,044	0,029	0,018
	3720-4960	0,005	0,01	0	0,003
	Mais de 4960	0,016	0,031	0,023	0,01

Fonte: Elaboração própria

A tabela de contingência do quadro 11 resume o peso relativo das categorias das variáveis Idade e Sexo. O grupo que apresentou o maior número de entrevistados foi o de indivíduos do sexo masculino com idade entre 18 e 25 anos.

Quadro 13 - Tabela de contingência para as variáveis Idade e Sexo

Sexo	Idade			
	18-25	25-35	35-45	Mais de 45
Masculino	0,240	0,190	0,044	0,052
Feminino	0,169	0,167	0,094	0,044

Fonte: Elaboração própria

O quadro 12 é também uma síntese do quadro 10. Apresenta as frequências relativas de uma tabela de contingência entre as variáveis “Renda” e “Idade”. Percebe-se que maior parte dos entrevistados possui idade entre 18 e 25 anos, tendo renda entre R\$ 620,00 e R\$ 1240,00. Nenhum dos entrevistados apresentou uma idade entre 35 e 45 anos, com uma renda no intervalo entre R\$ 3720,00 e R\$ 4960,00.

Quadro 14 – Tabela de contingência entre as variáveis Idade e Renda

Renda	Idade			
	18-25	25-35	35-45	Mais de 45
620-1240	0,185	0,117	0,026	0,021
1240-2480	0,117	0,104	0,034	0,021
2480-3720	0,063	0,057	0,042	0,023
3720-4960	0,010	0,018	0,000	0,003
Mais de 4960	0,034	0,060	0,036	0,029

Fonte: Elaboração própria

O quadro 13 resume as informações do quadro 10, considerando apenas as variáveis “Renda” e “Sexo”. O maior número de entrevistados está no grupo de indivíduos do sexo masculino e que possui renda entre R\$ 620,00 e R\$ 1240,00. O grupo que apresentou o menor número de entrevistados foi aquele de indivíduos do sexo masculino com renda entre R\$ 3720,00 e R\$ 4960,00.

Quadro 15 - Tabela de contingência entre as variáveis Renda e Sexo

Sexo	Renda				
	620-1240	1240-2480	2480-3720	3720-4960	Mais de 4960
Masculino	0,229	0,143	0,062	0,013	0,078
Feminino	0,120	0,133	0,122	0,018	0,081

Fonte: Elaboração própria

Com relação a variável intenção de compras o gráfico da figura 15 indica o percentual de entrevistados em cada uma das categorias. O gráfico indica que 76,04% mostraram disposição para utilizar o serviço, com 42,19% afirmando que certamente utilizariam os sites de compras coletivas.

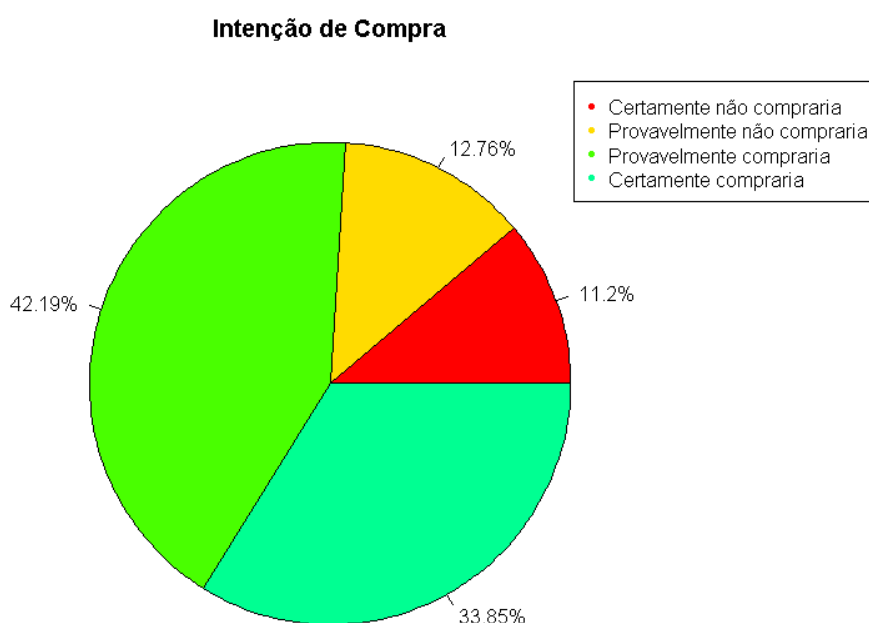
Apenas um grupo de 11,2% descartou a possibilidade de comprar utilizando o serviço de compras coletivas online, respondendo que certamente não utilizariam os sites. É provável que muito esforço seria necessário para tornar esses indivíduos consumidores de um site de compra coletiva.

Das pessoas que participaram da pesquisa 12,76% mostraram pouca disposição para usar o serviço no momento, mas não descartaram a possibilidade de usar o serviço

uma vez que mostraram indecisão ao afirmar que provavelmente não usariam os sites para compra.

As pessoas que declararam que provavelmente comprariam ou que falaram que provavelmente não comprariam, compõem um grupo de pessoas com indecisão ao opinar se usariam ou não serviço. Esse grupo representa 54,95% da amostra. Inclui indivíduos que embora indecisos, não declararam com convicção que não usariam o serviço. Essas pessoas podem tornar-se consumidoras mais facilmente se ações efetivas forem tomadas de modo que apresentem maior interesse pelo serviço.

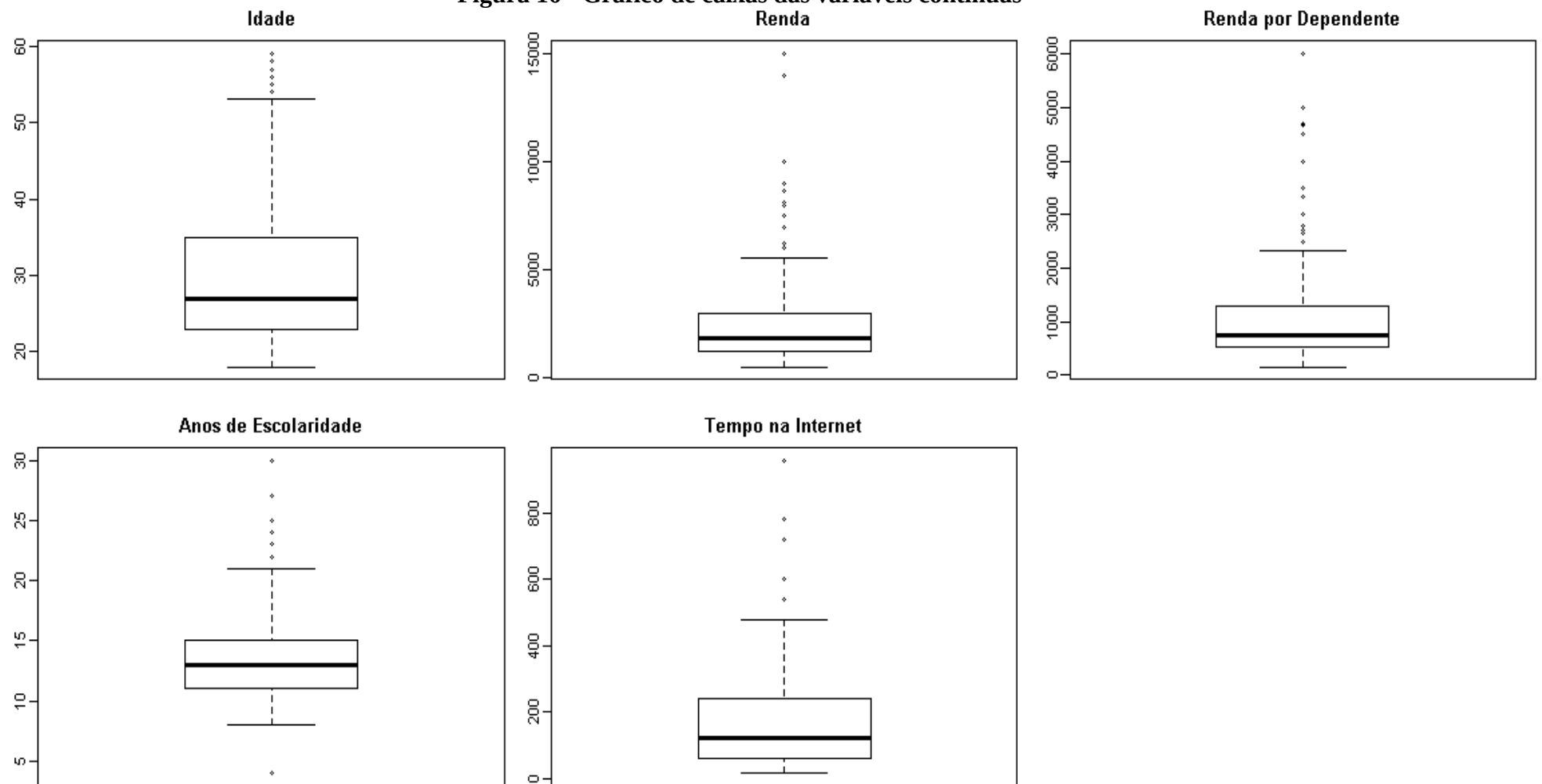
Figura 15 - Intenção dos consumidores de comprar em sites de compra coletiva



Fonte: Elaboração própria

Se uma empresa de serviços de compras coletivas online busca definir grupos de clientes para focar seus esforços, então deveria evitar direcioná-los para as pessoas que tendem a afirmar que certamente não comprariam em sites de compras coletivas, pois fazer com que esse grupo de indivíduos use os sites provavelmente iria requerer grande esforço por parte da empresa. É preferível que a empresa procure incluir, no negócio, elementos para conquistar um público de pessoas indecisas sobre o uso ou de indivíduos que manifestaram desejo de utilizar os sites de compras coletivas. Essas pessoas poderiam representar um grupo de potenciais consumidores do serviço.

Figura 16 - Gráfico de caixas das variáveis contínuas



Os gráficos de caixa apresentados na figura 16 para as cinco variáveis contínuas resumem as informações sobre a mediana, os quartis e a quantidade de outliers. Nos gráficos nota-se que o maior número de outliers encontra-se nas variáveis referentes à renda. A figura 17 mostra os histogramas para as mesmas variáveis.

Com relação a idade, nota-se que maior parte dos participantes da pesquisa possuem idade entre 20 e 30 anos. A pesquisa contou com uma menor participação de pessoas com mais de 40 anos. Quanto à renda residencial, a pesquisa contou com poucos indivíduos com renda superior a R\$ 5000,00. Com relação à renda por indivíduo dependente, maior parte dos entrevistados apresenta renda até 2000 reais. Essa variável considera não apenas a renda numa residência de um indivíduo como também o número de pessoas que utilizam essa renda.

No que diz respeito aos anos de estudo, a maioria dos entrevistados estudou entre 10 e 15 anos. Um número menor de participantes da pesquisa estudou menos de 10 anos e mais de 20 anos.

Quanto ao tempo diário que as pessoas passam na internet, maior parte dos participantes da pesquisa declararam despendar até 3 horas (180 minutos). Poucos entrevistados declararam passar mais do que seis horas diárias (360 minutos).

A figura 18 mostra gráficos de dispersão entre as variáveis contínuas. Nota-se que o único par de variáveis que graficamente parece apresentar uma relação é aquele das variáveis referentes à renda.

A figura 19 mostra os gráficos de setores de variáveis categóricas da pesquisa. É possível notar que 51,56% dos entrevistados afirmaram já ter comprado em websites, mas poucas vezes. 30,47% declararam que nunca haviam comprado na internet e somente 17,97% da amostra afirmaram comprar bastantes nos websites.

Das pessoas que participaram da pesquisa, 52,6% são do sexo feminino e 47,4% são do sexo masculino. Com relação ao estado civil, 60,42% são solteiros, divorciados ou viúvos e 39,58% são casados ou apresentam outro tipo de união estável.

A maioria dos entrevistados é o único dependente da renda de sua residência, além de si mesmo. 76,56% não possuíam dependentes e 23,44% possuíam.

Maior parte das pessoas entrevistadas (57,81%) visitou pouco ou nenhuma vez sites de compras coletivas. 42,97% acessam poucas vezes no ano e 14,84% nunca acessaram um site desse tipo. Sendo assim, conclui-se 42,19% declararam acessar com

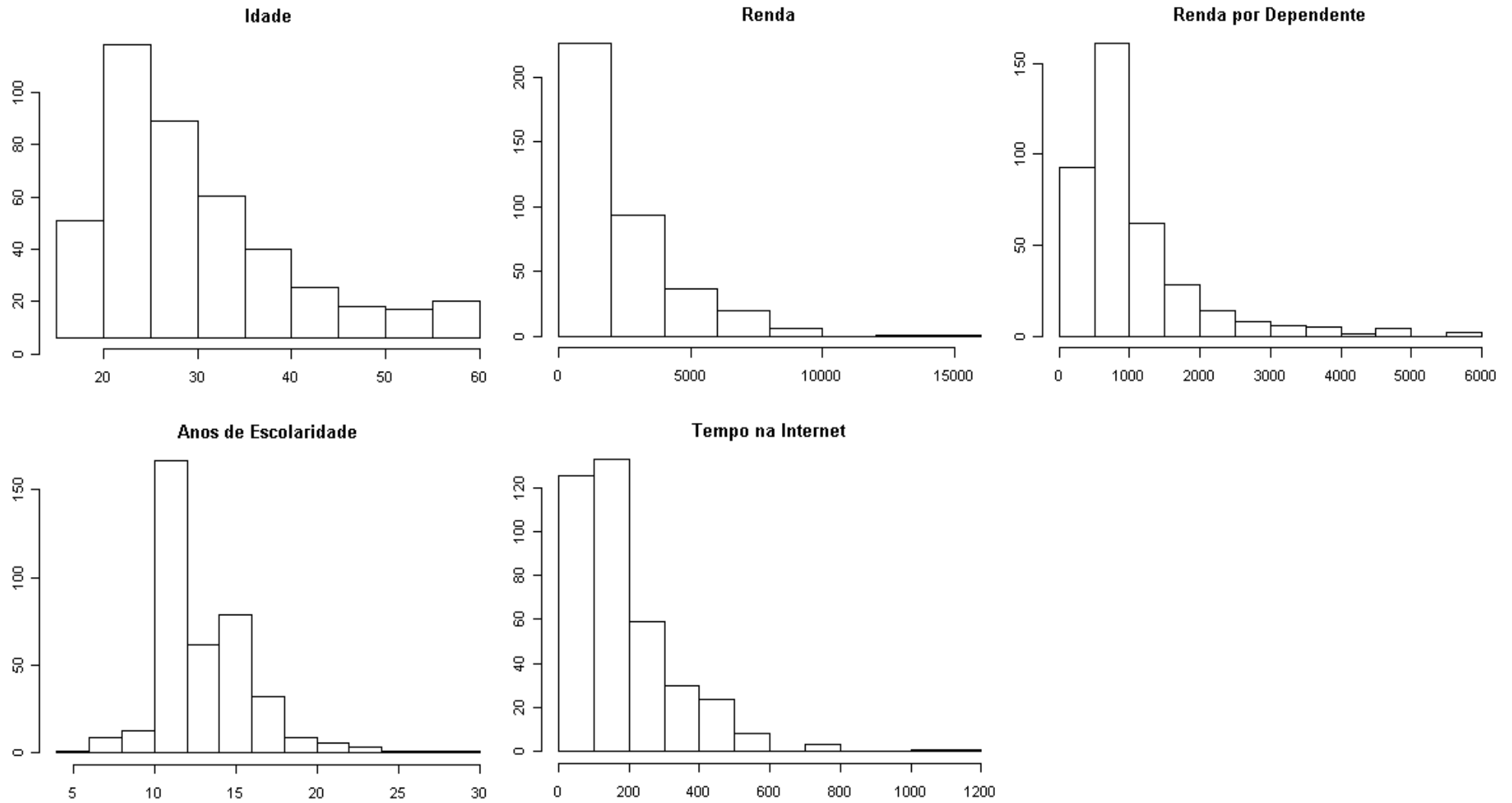
Figura 17 - Histogramas das variáveis contínuas

Figura 18 - Gráficos de dispersão dos preditores

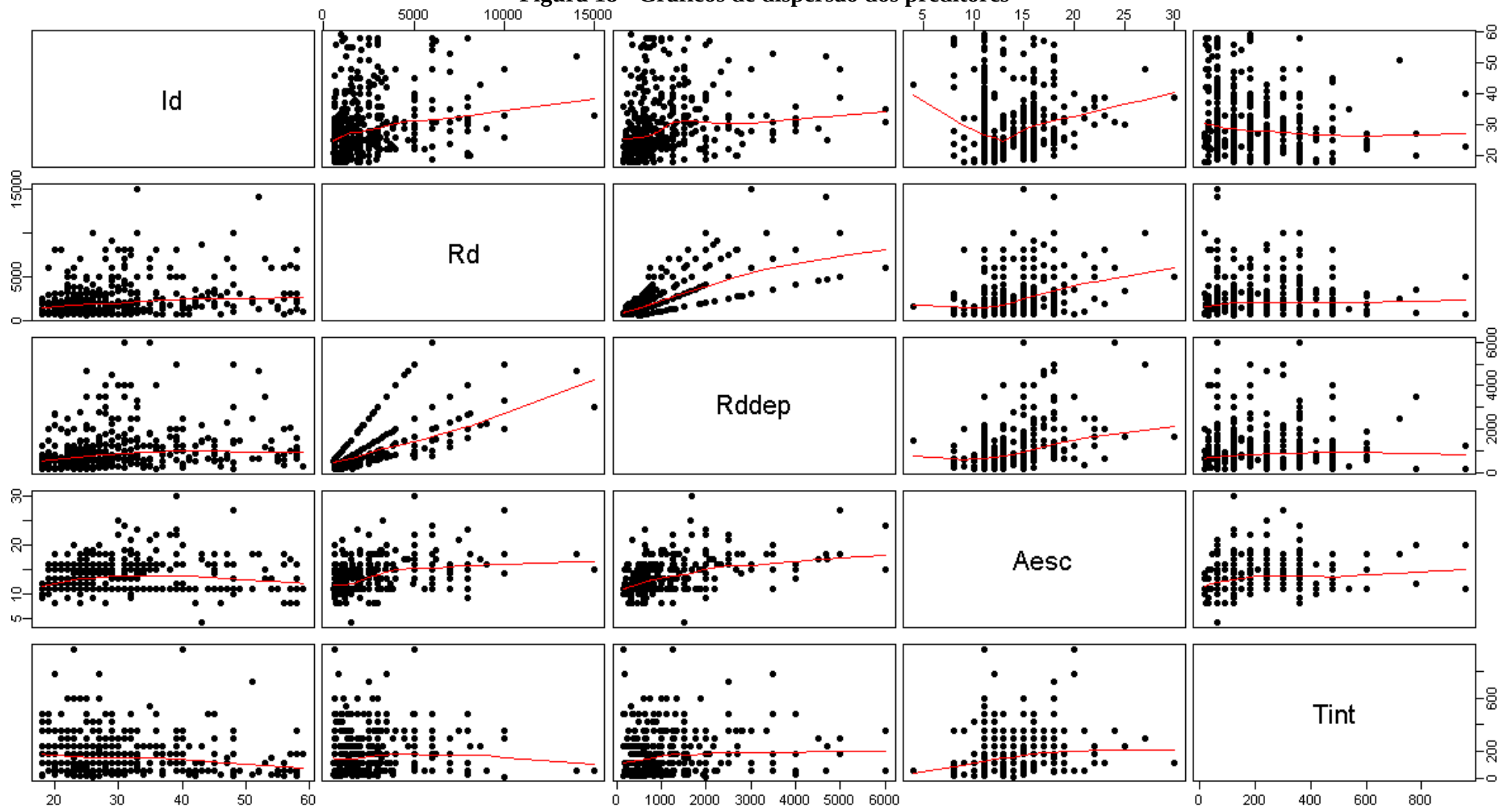
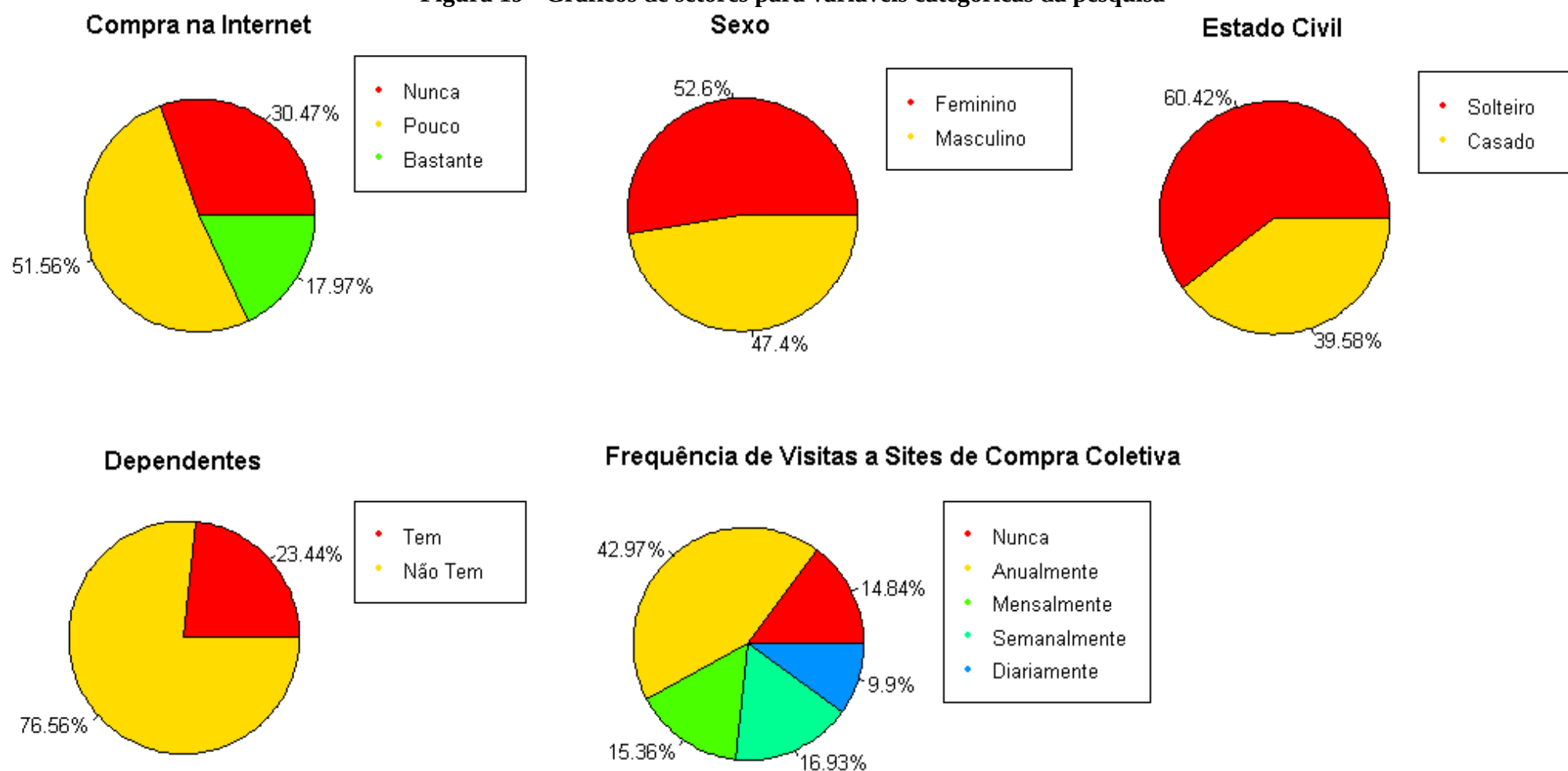


Figura 19 - Gráficos de setores para variáveis categóricas da pesquisa



maior regularidade esse tipo de websites (mensalmente, semanalmente ou diariamente). Apenas 9,9% declararam acessar praticamente todos os dias e 16,93% afirmaram que acessam toda semana.

Os gráficos de setores das figuras 20 e 21 indicam o perfil dos entrevistados relativos a atitudes em relação a fatores característicos do serviço de compras coletivas.

Da figura 6, percebe-se que 54,95% consideram importante que haja facilidade para obter ajuda no site sempre que for necessário. 19,79% atribuem grande importância a essa característica dos sites. Das pessoas entrevistadas, 25,26% atribuem pouca ou nenhuma importância a esse item.

Dos participantes da pesquisa, 48,7% acham que elementos que facilitem a navegação nos sites são importantes e 32,81% consideram a facilidade de navegação muito importante. Somente 18,49% disseram considerar ser essa característica pouco importante.

Com relação às características de aparência dos websites, 37,24% das pessoas entrevistadas consideram um aspecto importante e 23,96% acham esse item muito importante. Essa foi umas das características que os entrevistados atribuíram menor importância, pois 39,1% acham esse aspecto pouco importante ou sem importância.

Na amostra, foi opinião dominante considerar importante ou muito importante que os sites de compras coletivas e também as empresas parceiras sejam conhecidas do público. 92,19% têm essa opinião acerca da popularidade dos sites e 91,41% sustentam essa opinião acerca da popularidade das empresas parceiras.

Com relação à diversificação de produtos ou serviços oferecidos pelos sites de compras coletivas, maior parte dos entrevistados também considera um aspecto importante. Apenas 18,23% afirmaram que essa característica tem pouca ou nenhuma importância. 30,99% acham que é muito importante que os sites ofereçam maior variedade de produtos e serviços e 50,78% declarou achar esse um item importante dos websites.

Quanto à variedade das formas de pagamento (figura 13), houve um consenso para maior parte do público entrevistado. 91,41% acham que a disponibilidade de uma variedade maior de formas de pagamento é um aspecto importante ou muito importante do serviço.

Houve consenso também entre a maioria dos entrevistados no que diz respeito à quantidade de informação divulgada nos sites sobre os produtos e serviços que estão

Figura 20 - Gráfico de setores do nível de importância atribuído a fatores relacionados ao serviço de compras coletivas

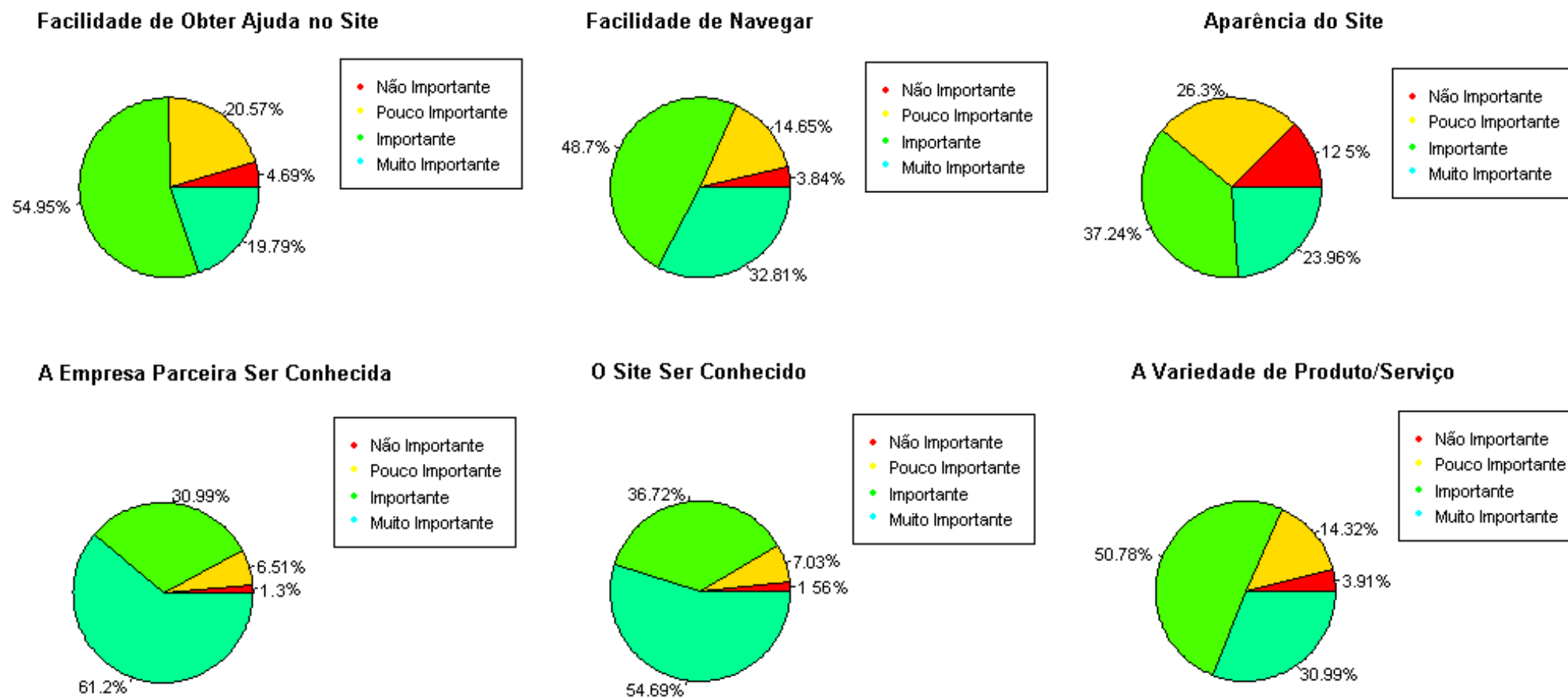
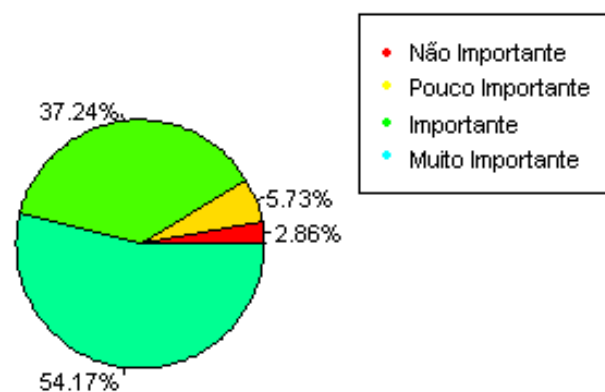
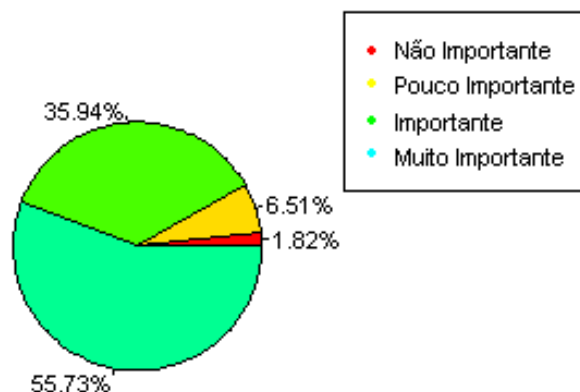


Figura 21 - Gráfico de setores do nível de importância atribuído a fatores relacionados ao serviço de compras coletivas

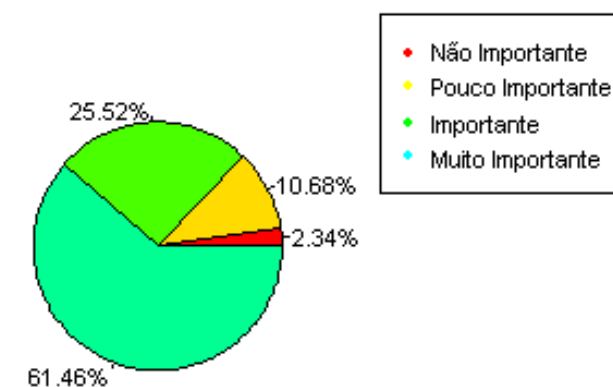
A Variedade de Formas de Pagamento



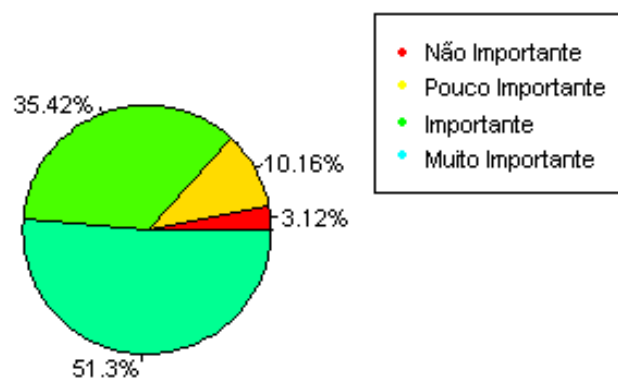
Quantidade de Informação Fornecida



Qualidade do SAC



Velocidade de Navegação do Site



sendo oferecidos. 55,73% considera essa característica como tendo grande importância e 35,94% considera importante. Apenas 1,82% acham esse um aspecto sem nenhuma importância.

Sobre a importância da qualidade do serviço de atendimento do consumidor (SAC), 61,46% julgou como algo muito importante e 25,52% julgou como algo importante. Dos participantes da pesquisa, 13,5% julgaram que a qualidade do SAC é um item de pouca ou nenhuma importância. Alguns entrevistados declararam que raramente usam o SAC quando compram e por isso achavam esse um aspecto menos relevante.

Quanto à velocidade de navegação do site, 51,3% dos participantes da pesquisa atribuíram muita importância a características que tornem a navegação menos lenta. 35,42% acharam que esse era um aspecto importante. 13,28% dos entrevistados acharam que esse item é pouco importante ou que não tem nenhuma importância.

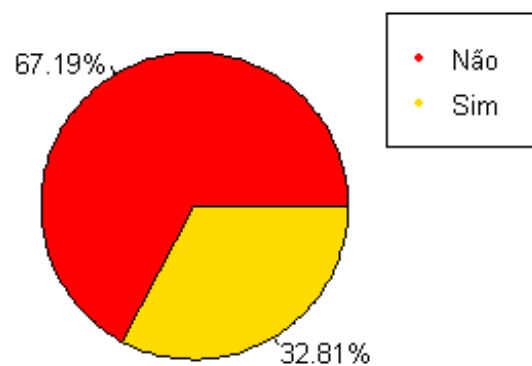
Os gráficos da figura 8 são relativos aos meios de divulgação pelos quais se toma conhecimentos das ofertas de sites de compras coletivas. 32,81% afirmaram que já tomaram conhecimento por meio de publicidade em outros sites, 37,5% declararam que já tomaram conhecimento de ofertas por meio de redes sociais e 37,4% indicaram que já tomaram conhecimento por meio do correio eletrônico. O meio de divulgação que conseguiu atingir o maior número das pessoas que participaram da pesquisa foi a comunicação com terceiros. 48,7% dos entrevistados afirmaram que já haviam tomado conhecimento de ofertas dos sites de compra coletiva por meio de terceiros.

A figura 22 apresenta uma descrição da forma como os entrevistados já tomaram conhecimento de ofertas de sites de compra coletiva. 67,32% já tomaram conhecimento através de publicidade em outros sites, 62,5% já souberam de ofertas através de redes sociais e 62,76% já foram informados de ofertas por meio do endereço eletrônico. Do total de entrevistados 51,3% afirmaram já ter tomado conhecimento de ofertas por meio da comunicação com terceiros.

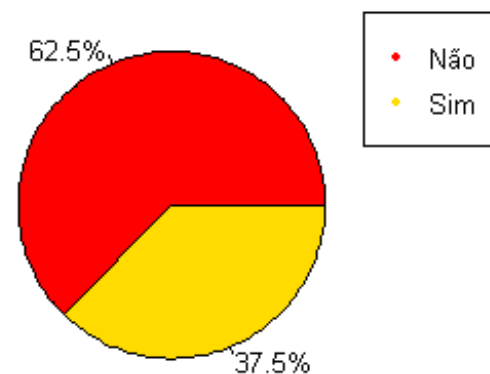
A figura 23 apresenta gráficos relativos aos produtos que as pessoas entrevistadas costumam comprar com regularidade. 70,05% afirmaram que consomem com regularidades serviços de lazer. 63,54% declararam comprar com regularidade itens de vestuário. Esses são os tipos de produtos e serviços mais consumidos pelos indivíduos participantes da pesquisa.

Figura 22 - Gráficos de setores para as formas pelas quais o entrevistado já tomou conhecimento de ofertas de sites de compra coletiva

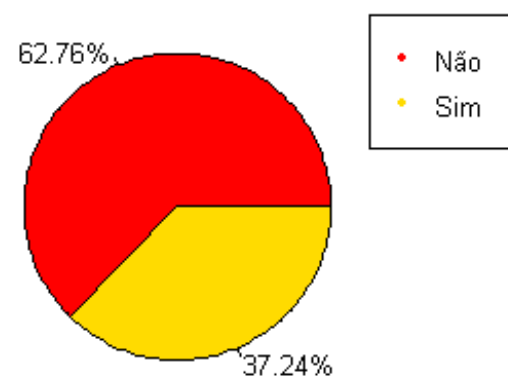
Publicidade em Outros Sites



Redes Sociais



Endereço Eletrônico



Através de Terceiros

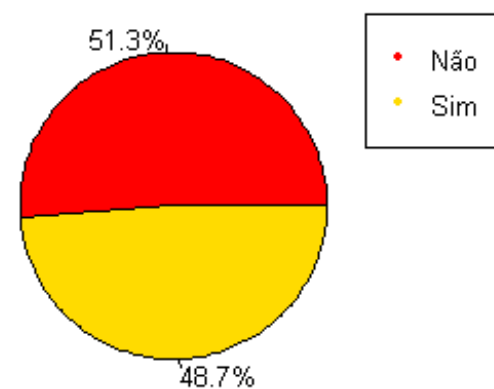
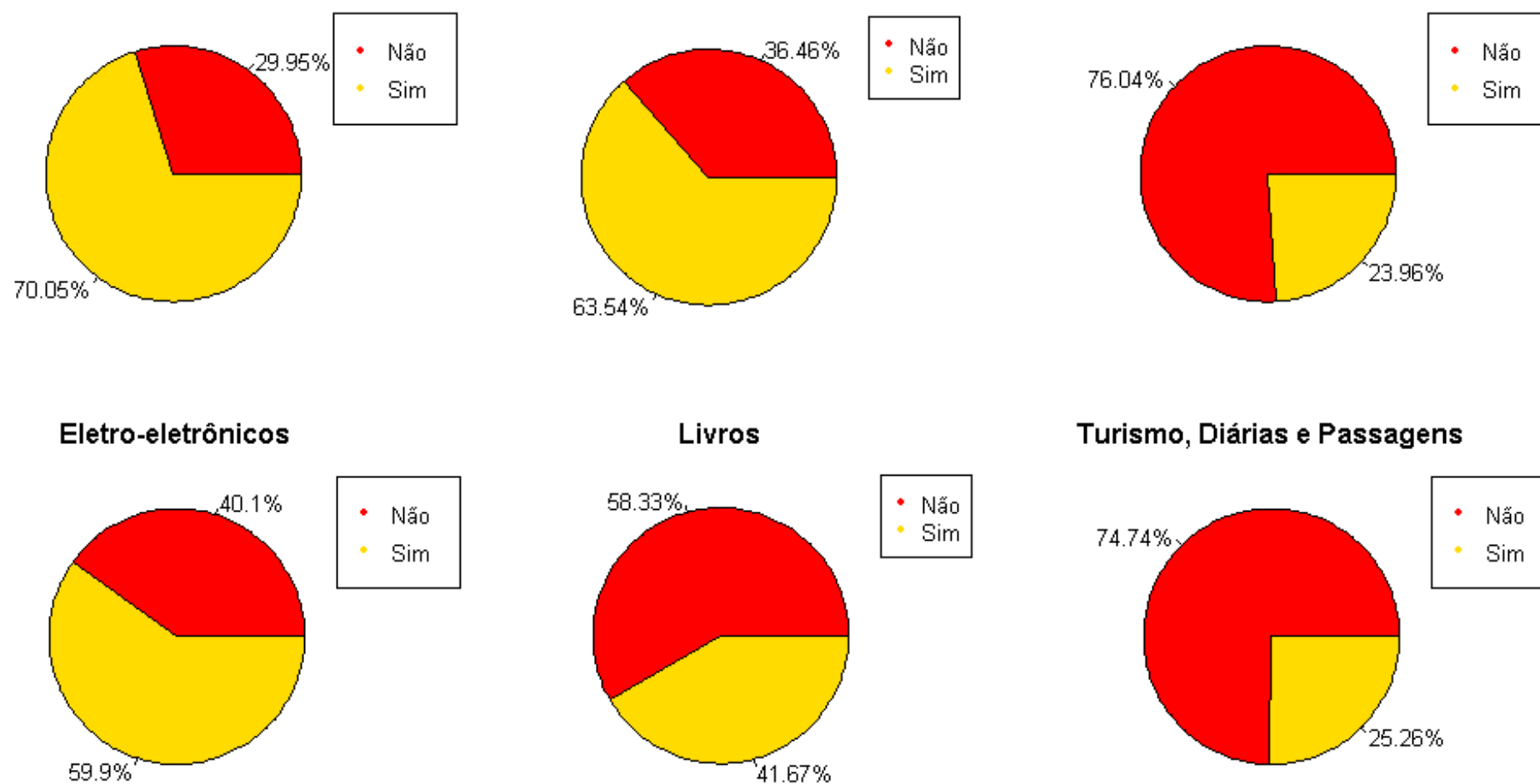


Figura 23 - Gráficos de setores para os tipos de produtos e serviços que os entrevistados consomem com regularidade



Os serviços e produtos de estética, bem como os itens de viagens, foram os tipos de produtos e serviços menos consumidos pelos entrevistados. 76,04% afirmaram que não costumam consumir com regularidade produtos e serviços de estética. Quanto aos itens relacionados a viagens, 74,74% declaram que não consumiam com regularidade.

A figura 23 ainda mostra que, das pessoas que participaram da pesquisa, 59,9% afirmaram comprar com regularidade produtos elétricos/eletrônicos. 41,67% dos entrevistados afirmaram que também compravam livros regularmente.

APÊNDICE C - Rotinas usadas para obter e analisar os modelos

```
#Importando e anexando banco de dados salvo com o nome "dissertação.txt"
d=read.table(file.choose(),head=T)#Escolher o arquivo com os dados
attach(d)#Anexar os dados
```

```
#Instalando e carregando os pacotes necessários
install.packages(Design)#Instalar pacote para usar o comando "lrm"
install.packages(gam)#Instalar pacote para usar o comando "gam"
require(Design)#Carregar pacote "Design"
require(gam)#Carregar o pacote "gam"
```

```
#Obtendo novas variáveis a partir dos dados
v1=Rd*Aesc
FLiv=Fx_Rd3*Liv#Fx_Rd3 assume o valor 0 se o indivíduo possui renda inferior u igual a
R$ 2.000,00 e "1" se for superior.
FId=Fx_Rd3*Fx_Et4#Fx_Et4 assume o valor "0" se o indivíduo possui menos do que 35 anos
e assume o valor 1 se possuir menos de 35 anos.
v19=Aesc/Id
v7=Rd/Tint
v26=Tint/Id
FDep=Fx_Rd3*Dep1#Dep1 assume o valor "0" se um indivíduo é o único dependente de sua
renda e assume o valor "1" se isso não ocorrer
v22=Aesc/Tint
FSx=Sx*Visit3#Visit3 assume o valor "1" se o individuo acessa sites pelo menos uma vez por
semana e "0" caso contrário.
```

```
#MLG com variável resposta apresentando uma distribuição binomial
m1=glm(IC2~0+Rddep+Terc+Faj_I+Cmint_dic+Vpag_I+Visit4+v+vs26+FId+v7,family=binomial(link=logit))
```

```
#Obtendo os valores do p-valor do teste de Wald, do "Null Deviance" e do "Residual Deviance" para o MLG com respostas binárias
summary(m1)
```

```
#MAG com variável resposta apresentando uma distribuição binomial usando a suavização splines para obter o termo não paramétrico
m2.1=gam(IC2~0+s(Rddep)+Terc+Faj_I+Cmint_dic+Vpag_I+Visit4+v+vs24+v22+FId,family=binomial(link=logit))
```

```
#MAG com variável resposta apresentando uma distribuição binomial usando a suavização por ponderação local para obter o termo não paramétrico
m2.2=gam(IC2~0+lo(Rddep)+Terc+Faj_I+Cmint_dic+Vpag_I+Visit4+v+vs24+v22+FId,family=binomial(link=logit))
```

```
#Obtendo os valores do p-valor do teste de Wald, do "Null Deviance" e do "Residual Deviance" para o MAG com respostas binárias
summary(m2.1)
```

```
gam.exact(m2.1)
summary(m2.2)
gam.exact(m2.2)
```

```
#MLG multinomial ordinal
```

```
m3=lrn(as.factor(IC3)~v19+Terc+Cmint_dic+Fx_Rd3+Ap_I+Vpag_I+Vprod_I+Dep1+Visit
4+vfr7+vfr13+v1)
```

```
#Obtendo os valores do p-valor do teste de Wald, do "Null Deviance" e do "Residual
Deviance" para o MLG com respostas categóricas ordinais m3
```

```
#Calculando o pseudo-R2 para os três modelos.Considerar Dnull o valor do "Null Deviance"
e Dmod sendo o valor do "Residual Deviance" para a amostra de tamanho 384
```

```
R2_m1=1-exp((Dmod1-Dnull1)/384)
R2max_m1=1-exp(-Dnull1/384)
NpR2_m1=R2_m1/R2max_m1#R2 do MLG para respostas binárias
```

```
R2_m2.1=1-exp((Dmod2.1-Dnull2.1)/384)
R2max_m2.1=1-exp(-Dnull2.1/384)
NpR2_m2.1=R2_m2.1/R2max_m2.1#R2 do MAG para respostas binárias com suavização
dos dados ligados ao não paramétrico pelo método splines
```

```
R2_m2.2=1-exp((Dmod2.2-Dnull2.2)/384)
R2max_m2.2=1-exp(-Dnull2.2/384)
NpR2_m2.2=R2_m2.2/R2max_m2.2#R2 do MAG para respostas binárias com suavização
dos dados ligados ao não paramétrico pelo método de ponderação local
```

```
m3#O modelo obtido pelo comando "lrn" fornece diretamente o valor do R2
```

```
#Obtendo o p-valor do teste Qui-quadrado para os modelos
```

```
anova(m1,test="Chi")
anova(m2.1,test="Chi")
anova(m2.2,test="Chi")
anova(m3,test="Chi")
```

```
#Obtendo a odds ratio das variáveis do modelo. Ignorar o valor da odds para os termos não
paramétricos
```

```
exp(coefficients(m1))
exp(coefficients(m2.1))
exp(coefficients(m2.2))
exp(coefficients(m3))
```

```
#Obtendo os valores preditos dos modelos para os dados amostrais. Esses valores são usados
para construir a matriz de confusão.
```

```
fittedglm=predict.glm(m1,"response")
fittedgam=predict.gam(m2.1,"response")
predict.gam(m2.2,"response")
predict.lrm(m1,"fitted.ind")
```

```
#Modelos binomiais
```

```
predglm = prediction(fittedglm,IC2)
perfglmp = performance(predglm,"tpr","fpr")
perfglmn = performance(predglm,"tnr","fnr")
perfglmauc = performance(predglm,"auc")
plot(perfglmp,col=rainbow(10),xlab="Falsos Positivos",ylab="Verdadeiros Positivos")
plot(perfglmn,col=rainbow(10),xlab="Falsos Negativos",ylab="Verdadeiros Negativos")
predgam = prediction(fittedgam,IC2)
perfgamp = performance(predgam,"tpr","fpr")
perfgamn = performance(predgam,"tnr","fnr")
perfgamauc = performance(predgam,"auc")
plot(perfgamp,col=rainbow(10),xlab="Falsos Positivos",ylab="Verdadeiros Positivos")
plot(perfgamn,col=rainbow(10),xlab="Falsos Negativos",ylab="Verdadeiros Negativos")
```

```
#Verificando pontos de corte para o GLM
```

```
pc1=unlist(slot(perfglmp,"alpha.values"))
vp1=unlist(slot(perfglmp,"y.values"))
fp1=unlist(slot(perfglmp,"x.values"))
vn1=unlist(slot(perfglmn,"y.values"))
fn1=unlist(slot(perfglmn,"x.values"))
selpc1=cbind(pc1,vp1,fp1,vn1,fn1)
```

```
#Verificando pontos de corte para o GAM
```

```
pc2=unlist(slot(perfgamp,"alpha.values"))
vp2=unlist(slot(perfgamp,"y.values"))
fp2=unlist(slot(perfgamp,"x.values"))
vn2=unlist(slot(perfgamn,"y.values"))
fn2=unlist(slot(perfgamn,"x.values"))
selpc2=cbind(pc2,vp2,fp2,vn2,fn2)
```

APÊNDICE D - Fatores determinantes da intenção no questionário

Figura 24 - Fatores que influenciam na intenção de compra, variáveis da pesquisa e os itens do questionário

