

**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO MECÂNICA**

ERIVALDO LOPES DE SOUZA

**UMA APLICAÇÃO DA REGRESSÃO LOGÍSTICA NA TOMADA DE DECISÃO:
ESTUDO DE CASO**

JOÃO PESSOA

2010

ERIVALDO LOPES DE SOUZA

**UMA APLICAÇÃO DA REGRESSÃO LOGÍSTICA NA TOMADA DE DECISÃO:
ESTUDO DE CASO**

Trabalho de Conclusão de Curso desenvolvido e apresentado no âmbito do Curso de Graduação em Engenharia de Produção Mecânica da Universidade Federal da Paraíba (UFPB), em cumprimento às exigências para a obtenção do título de Graduado em Engenharia de Produção Mecânica.

Orientador: Prof. Dr. Luiz Bueno da Silva

**JOÃO PESSOA - PB
2010**

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO MECÂNICA

AVALIAÇÃO DO TRABALHO DE CONCLUSÃO DE CURSO

Comunicamos à Coordenação do curso de Graduação em Engenharia de Produção Mecânica (Bacharelado) que a monografia do aluno ERIVALDO LOPES DE SOUZA, matrícula 10311336, intitulada **“UMA APLICAÇÃO DA REGRESSÃO LOGÍSTICA NA TOMADA DE DECISÃO: ESTUDO DE CASO”**, foi submetida à apresentação da Comissão Examinadora composta pelos seguintes professores: Luiz Bueno da Silva (Orientador); Aurelia Acunã Idrogo (examinadora); Maria de Lourdes Barreto (Examinadora), no dia ____/____/____, às ____ horas no período letivo 2009.1.

A Monografia foi _____ pela Comissão Examinadora e obteve nota _____.

Reformulações sugeridas: sim () não ()

BANCA EXAMINADORA

Prof. Dr. Luiz Bueno da Silva
Orientador

Profa. Dra. Aurelia Acunã Idrogo
Examinadora

Profa. Dra. Maria de Lourdes Barreto
Examinadora

CIENTE: _____
Erivaldo Lopes de Souza

**Ao meu pai,
Aprígio Alves (*in memoriam*)**

AGRADECIMENTOS

À minha mãe, Rita, pelo incentivo aos meus estudos.

Aos meus irmãos Ana Paula e João Paulo, pela amizade e compreensão.

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

A Santhiago Montenegro por todas as informações compartilhadas.

Aos professores do Departamento de Engenharia de Produção Mecânica.

A todos os amigos (as) conquistados ao longo do curso.

RESUMO

Este trabalho consiste em um estudo de caso onde é abordada uma aplicação da regressão logística como ferramenta auxiliar na tomada de decisão sobre escolha de segmentos de mercado. O seu objetivo central é avaliar uma aplicação específica da regressão logística na segmentação de mercado, tomando como base uma pesquisa de mercado sobre sacola ecológica na cidade de João Pessoa, Paraíba, Brasil. Realizou-se um levantamento da literatura referente a assuntos relacionados à análise de regressão logística e à pesquisa de mercado. Em seguida, em pesquisa de campo, foram aplicados questionários a uma amostra de moradores de diferentes bairros do município de João Pessoa, abordando questões referentes às preferências, opiniões e características dos entrevistados. Os resultados mostraram que é possível obter modelos de regressão logística que indiquem a probabilidade de um determinado grupo de consumidores estar ou não disposto a pagar o valor das sacolas. Verificou-se ainda que esses modelos podem auxiliar na segmentação do mercado. Porém a pesquisa também mostrou que nem sempre esses modelos de regressão logística podem ser obtidos e que os resultados da aplicação dessa técnica estatística variam para diferentes preços e modelos de sacolas, o que indica que as opiniões e características que distinguem as pessoas que pagariam o valor das sacolas daquelas que não o pagariam mudam em função do preço e do tipo de produto.

Palavras-chave: Pesquisa de mercado. Análise de regressão logística. Segmentação de mercado.

ABSTRACT

This paper presents a case study which looks an application of logistic regression as a tool to assist in decision making on choice of market segments. Its main objective is to evaluate a specific application of logistic regression in market segmentation, based on market research on eco-bag in João Pessoa, Paraíba, Brazil. We conducted a survey of the literature pertaining to issues related to logistic regression analysis and market research. Then, in field research, questionnaires were applied to a sample of residents from different neighborhoods of the city of Joao Pessoa, addressing issues related to preferences, opinions and characteristics of respondents. The results showed that it is possible to logistic regression models showing the probability of a particular group of consumers or may not be willing to pay for your bags. It was also found that these models can assist in market segmentation. But the survey also showed that even when these logistic regression models can be obtained and the results of applying this statistical technique vary for different prices and models of bags, which indicates that the opinions and characteristics that distinguish people who would pay the value the bags of those who do not pay the change as a function of cost and type of product.

Keywords: Market research. Logistic regression analysis. Market segmentation.

SUMÁRIO

LISTA DE FIGURAS	9
LISTA DE TABELA.....	10
LISTA DE QUADROS	11
1 INTRODUÇÃO.....	12
1.1 Definição do tema e do problema proposto.....	12
1.2 Justificativa.....	13
1.3 Objetivos.....	16
1.3.1 Objetivo geral	16
1.3.2 Objetivos específicos.....	16
2 ASPECTOS TEÓRICOS.....	17
2.1 Estudo do mercado	17
2.1.1 Mercado	17
2.1.2 Estudo do mercado consumidor	17
2.1.3 Segmentação do mercado	19
2.1.4 Pesquisa de mercado.....	20
2.2 Regressão Logística.....	20
2.2.1 Classificação das variáveis	21
2.2.2 Análise de Regressão.....	23
2.2.3 Modelos Lineares Generalizados (MLG's)	23
2.2.4 Regressão logística	24
2.2.5 Métodos de seleção de modelos	29
2.2.6 Parâmetros de avaliação do modelo	29
3 ASPECTOS METODOLÓGICOS.....	34
4 RESULTADOS	38
4.1 Análise descritiva dos dados.....	38
4.2 Método utilizado e o resultado da estimação.....	39
4.3 Informações sobre o mercado consumidor de sacolas ecológicas.....	56
4.4 Um exemplo da utilização do modelo como auxílio à segmentação.....	58
5 CONSIDERAÇÕES FINAIS	62
6 REFERÊNCIAS	64

LISTA DE FIGURAS

Figura 1 - Classificação das variáveis	22
Figura 2 – Curva em S gerada do modelo de regressão logística.....	26
Figura 3 - Método backward	39
Figura 4 - Comportamento do AIC com a retirada das variáveis menos significativas nos modelos para sacolas de tamanho médio a um preço de R\$ 7,00	40
Figura 5 - Comportamento do pseudo- R^2 com a retirada das variáveis menos significativas dos modelos para sacolas de tamanho médio a R\$ 7,00.....	41
Figura 6 - Comportamento do número de variáveis com $p\text{-valor} > 0,2$ com a retirada das variáveis nos modelos de sacolas de tamanho médio a um preço de R\$ 7,00.....	41
Figura 7 - Comportamento do maior p-valor dos coeficientes com a retirada das variáveis nos modelos para sacolas de tamanho médio a um preço de R\$ 7,00	42
Figura 8 - Comportamento do p-valor com a retirada das variáveis nos modelos de sacolas de tamanho médio a um preço de R\$ 7,00	42
Figura 9 - p-valor dos coeficientes do modelo para sacolas de tamanho médio a um preço de R\$ 7,00	44
Figura 10 - Comportamento do AIC com a retirada de variáveis menos significativas nos modelos para sacolas de tamanho grande a um preço de R\$ 10,00	50
Figura 11 - Comportamento do pseudo- R^2 com a retirada das variáveis menos significativas dos modelos para sacolas de tamanho grande a um preço de R\$10,00	50
Figura 12 - Comportamento do número de variáveis com $p\text{-valor} > 0,2$ com a retirada de variáveis dos modelos para sacolas de tamanho grande a um preço de R\$ 10,00..	51
Figura 13 - Comportamento do maior p-valor dos coeficientes com a retirada de variáveis dos modelos para sacolas de tamanho médio a um preço de R\$ 10,00	51
Figura 14 - Comportamento do p-valor do modelo com a retirada de variáveis nos modelos para sacolas de tamanho grande a um preço de R\$ 10,00	52
Figura 15 - p-valor dos coeficientes do modelo para sacolas de tamanho grande a um preço de R\$ 10,00	54
Figura 16 – Gráfico com os valores das probabilidades dos segmentos de 1 a 10 estarem dispostos a pagar R\$ 7,00 pelas sacolas de tamanho médio.....	60

LISTA DE TABELA

Tabela 1 – Informações sobre os modelos gerados para sacolas ecológicas de tamanho médio ao preço de R\$ 7,00	40
Tabela 2 – Outras informações sobre os modelos gerados para sacolas ecológicas de tamanho médio ao preço de R\$ 7,00	43
Tabela 3 – Valores da <i>odds ratio</i> para cada variável do modelo para sacolas ecológicas de tamanho médio ao preço de R\$ 7,00	46
Tabela 4 – Informações sobre os modelos gerados para sacolas ecológicas de tamanho pequeno ao preço de R\$ 4,00.....	47
Tabela 5 – Informações sobre os modelos gerados para sacolas ecológicas de tamanho médio a R\$ 5,00.....	48
Tabela 6 – Informações sobre os modelos gerados para sacolas ecológicas de tamanho grande ao preço de R\$ 10,00	49
Tabela 7 – Outras informações sobre os modelos gerados para sacolas ecológicas de tamanho grande ao preço de R\$ 10,00	53
Tabela 8 – Valores da <i>odds ratio</i> para cada variável do modelo para as sacolas de tamanho grande ao preço de R\$ 10,00	55

LISTA DE QUADROS

Quadro 1 – Informações a cerca dos coeficientes das variáveis independentes e do intercepto do modelo obtido para sacolas de tamanho médio a R\$ 7,00.....	44
Quadro 2 – Informações a cerca dos coeficientes das variáveis independentes e do intercepto do modelo obtido para sacolas de tamanho médio a R\$ 5,00.....	48
Quadro 3 – Informações a cerca dos coeficientes das variáveis independentes e do intercepto do modelo obtido para sacolas de tamanho grande a R\$ 10,00.....	53
Quadro 4 - Informações acerca dos fatores e valores da probabilidade dos segmentos A e B estarem dispostos a comprar as sacolas de tamanho médio a R\$ 7,00.....	59
Quadro 5 – Informações acerca dos fatores para os segmentos enumerados de 1 a 10	60

1 INTRODUÇÃO

1.1 Definição do tema e do problema proposto

O presente trabalho propõe utilizar a análise de regressão logística e uma pesquisa de mercado para construir um modelo que possa auxiliar na escolha de segmentos de mercado onde uma empresa poderia atuar. Para isso tomou como base um estudo de caso sobre o mercado de sacolas ecológicas na cidade de João Pessoa.

A regressão logística é uma técnica estatística que objetiva discriminar dois grupos de observações dentro de uma amostra, distinguindo-se da regressão linear por usar uma variável dependente binária (PACAGNELLA JÚNIOR *et al*, 2009). Ela é utilizada quando se pretende analisar um conjunto de dados sobre determinados fatores que podem ser divididos em uma variável dependente e um conjunto de variáveis independentes, quando as variáveis dependentes incluem dados categóricos.

Segundo a visão de Johnson e Wichern (1992) *apud* Montenegro (2009), a análise de regressão é uma metodologia estatística usada para 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.

A modelagem regressiva pode ser aplicada a diversas áreas, como na área de saúde, ergonomia, produção, gestão, marketing, dentre outras.

Sobre a modelagem com regressão logística, Penha (2002) afirma que a regressão logística vem sendo usada não apenas em Engenharia de Produção, mas também em várias outras áreas. Hosmer e Lemeshow (2000) compartilham a mesma idéia ao afirmarem que a regressão logística tem sido utilizada em diversas áreas do conhecimento, como, por exemplo, na medicina e em marketing.

Nas áreas de gestão e marketing, a busca de meios para conhecer o consumidor é um problema que tem sido constantemente estudado. Um dos instrumentos disponíveis para obter esse conhecimento é a pesquisa de mercado que consiste num importante instrumento de coleta de informações junto ao consumidor, concorrente ou fornecedor, que tem por finalidade orientar a tomada de decisões de empresários e empreendedores (AZEVEDO, 2009).

Por meio da pesquisa de mercado é possível coletar um conjunto de dados acerca do mercado consumidor que se deseja atuar e posteriormente transformá-los em um conjunto de

informações úteis para decidir a maneira como a empresa pretende tomar suas decisões e que tornem possível conhecer melhor as características do produto que se pretende inserir no mercado.

Tomando como base esses conceitos, procurou-se disponibilizar informações sobre um mercado consumidor específico a partir da aplicação da regressão logística a dados obtidos por meio de uma pesquisa de mercado sobre sacolas ecológicas, realizada na cidade de João Pessoa.

Pretende-se, assim, responder ao seguinte questionamento:

Como os modelos de regressão logística podem ser utilizados para auxiliar a segmentação do mercado?

1.2 Justificativa

O fechamento prematuro das empresas no país tem sido uma das principais preocupações da sociedade, particularmente para as entidades que desenvolvem programas de apoio ao segmento das empresas de pequeno porte (SEBRAE/MG, 2005).

O SEBRAE (2007) realizou uma pesquisa de campo com micro e pequenas empresas extintas, nas 26 unidades federativas do país e no Distrito Federal, no período de 2003 a 2005. Constatou que para 68% dos empresários de empresas extintas as razões para o fechamento da empresa estão em um conjunto de falhas gerenciais na qual destaca-se: o local inadequado, a falta de conhecimentos gerenciais e o desconhecimento do mercado.

A pesquisa de mercado é uma ferramenta comumente utilizada para suprir a carência de informações sobre o mercado, que, por sua vez, podem subsidiar as decisões dos empresários. Segundo Pinheiro (2006 apud TUSSI, 2008) a pesquisa de mercado, como ferramenta de auxílio à administração mercadológica, pode ser um instrumento poderoso de análise de mercado e de interpretação da realidade.

A pesquisa de mercado pode indicar os melhores caminhos para conquistar um público-alvo, mostrando onde está o cliente potencial, quais suas expectativas e necessidades, e ainda o que esse cliente potencial mais valoriza (TUSSI, 2008).

Apesar da importância das pesquisas de mercado, no Brasil, segundo Brustolin (2007) ainda se investe pouco nesse instrumento (250 milhões anualmente), especialmente nas pequenas e médias empresas. Para o referido autor é muito provável que isso ocorra devido à

cultura arraigada dos brasileiros muito acostumados a tomar decisões com base na intuição e na experiência profissional.

O autor destaca ainda que a mortalidade de empresas nos primeiros três anos de vida devido a decisões errôneas, muitas vezes, baseadas em planejamentos empíricos é bastante alta.

Pode-se então inferir que a pesquisa de mercado é um instrumento fundamental para superar uma das principais causas da extinção de muitas empresas no país (o desconhecimento do mercado) e para melhorar o gerenciamento dessas empresas.

Por essa razão, é importante que seu uso seja incentivado e que se investigue formas de melhor explorá-la. Além disso, qualquer instrumento de análise de dados que possa disponibilizar mais informações a partir da pesquisa de mercado é importante para quem costuma e precisa aplicá-las.

As técnicas estatísticas podem contribuir para explorar melhor os dados obtidos por uma pesquisa de mercado disponibilizando informações acerca das preferências dos consumidores com relação a diversos fatores. A regressão logística é uma dessas técnicas que podem trazer contribuições para melhorar o aproveitamento da pesquisa de mercado.

A importância da regressão logística para o tratamento de dados tem sido percebida em diferentes áreas do conhecimento, como, por exemplo, na medicina e em marketing. Essa técnica estatística tem sido utilizada em diversas situações onde se busca fazer previsões de uma variável que resulta em dados categóricos.

Conjugar a regressão logística a pesquisas de mercado pode resultar em contribuições interessantes especialmente para áreas da engenharia de produção que demandam grande conhecimento das preferências dos consumidores, como as áreas que envolvem projetos de produtos e de unidades produtivas.(Fazer a ponte)

As sacolas ecológicas, abordadas no estudo de caso, podem ser consideradas ainda como um produto novo no mercado de João Pessoa. Por essa razão, o mercado de sacolas ecológicas deste município pode mostrar algumas contribuições da conjugação da regressão logística à pesquisa de mercado para as áreas de engenharia de produção citadas anteriormente.

Essas sacolas ecológicas desempenham um papel importante no combate a um grave problema ambiental, que é descarte inadequado e excessivo de sacos plásticos.

O descarte inadequado das sacolas plásticas tem sido um grave problema para sociedade. Os sacos descartados poluem rios, córregos e entopem as bocas-de-lobo das galerias pluviais. Para um consumidor comum, embalar as compras ao sair do supermercado

em dezenas de sacolas plásticas é uma prática habitual. Já para um consumidor consciente, tal ato é uma agressão ao meio ambiente (ROCHA, 2009).

Para se ter noção da dimensão do problema, estima-se que, a cada minuto, um milhão de sacolas plásticas são descartadas no mundo. No Brasil, cerca de 10% do lixo é composto por sacos plásticos. O descarte é tão grande que, no Rio de Janeiro, são gastos R\$ 15 milhões anualmente com dragagem dos rios para a remoção desses agentes potencializadores da poluição (LEITÃO, 2009).

Problemas como esses, citados anteriormente, podem ser amenizados utilizando alternativas aos sacos plásticos. As sacolas ecológicas reutilizáveis, que é uma dessas alternativas, são feitas de um material biodegradável e não liberam nenhum tipo de substância prejudicial ao meio ambiente no seu processo de degradação. A sua disseminação no mercado paraibano pode dar ao cliente a oportunidade de contribuir com a preservação do meio ambiente e conscientizá-lo que existem alternativas para o plástico.

A justificativa deste trabalho, que consiste na aplicação de uma técnica estatística para obter informações adicionais dos dados de uma pesquisa de mercado, decorre das seguintes razões:

- Divulgar a utilização da regressão logística em pesquisas de engenharia de produção onde se estuda a relação de dependência entre fatores que envolvem dados categóricos;
- Propor uma forma de obter informações adicionais de uma pesquisa de mercado, ressaltando a importância da regressão logística no estudo de dados categóricos e indicando meios de conhecer melhor o mercado de sacolas ecológicas na cidade de João Pessoa.
- Contribuir para difusão das sacolas ecológicas reutilizáveis no mercado paraibano ao disponibilizar informações sobre o mercado desse produto na cidade de João Pessoa, para empresas que desejem produzi-las. Sua disseminação no mercado paraibano pode dar ao cliente e a empresa a oportunidade de contribuir com a preservação do meio ambiente ao proporcionar uma alternativa aos sacos plásticos.

1.3 Objetivos

1.3.1 Objetivo geral

Obter modelos de regressão logística para uma pesquisa de mercado, que possam auxiliar na segmentação do mercado de sacolas ecológicas no município de João Pessoa.

1.3.2 Objetivos específicos

- a) Verificar as variáveis que exercem maior influência sobre a disposição dos indivíduos para compra de sacolas ecológicas;
- b) Exemplificar a utilização do modelo como ferramenta auxiliar à segmentação do mercado consumidor;
- c) Verificar se a disposição para compra da sacola ecológica, por parte dos indivíduos, depende sempre das mesmas variáveis, independente da mudança do preço ou do modelo do produto.

2 ASPECTOS TEÓRICOS

A compreensão deste trabalho requer tanto alguns conhecimentos relacionados ao estudo do mercado quanto à regressão logística. Por essa razão, serão apresentados inicialmente conceitos importantes relacionados a esses dois campos que compõem o eixo temático desta pesquisa. Por fim, serão apresentados os procedimentos metodológicos da pesquisa.

2.1 Estudo do mercado

Nesta seção são exibidos conceitos importantes para obter o entendimento do porquê da importância de estudar o mercado consumidor e o papel da pesquisa de mercado nesse entendimento.

2.1.1 Mercado

Mankiw (2001) define mercado como um grupo de compradores e vendedores de um produto ou serviço. Pyndick e Rubinfeld (2002) consideram mercado como um grupo de vendedores e compradores que, por meio de interações, determinam o preço de um produto ou de um conjunto de produtos.

Para Czincota (2001), um mercado é qualquer indivíduo, grupo de indivíduos ou organizações que queiram e estejam habilitados a adquirir o produto de uma empresa.

Conhecer o mercado para o qual será vendido o produto requer uma análise dos concorrentes deste produto, do grupo de indivíduos que está habilitado a comprá-lo, bem como da interação entre vendedores do produto e compradores. Desse modo, conhecer o mercado consumidor é parte fundamental do conhecimento do mercado.

2.1.2 Estudo do mercado consumidor

Uma empresa que deseja ter sucesso no lançamento de um produto no mercado precisa conhecê-lo. Não conhecer o seu mercado pode levar a empresa a lançar um produto que não esteja em conformidade com os desejos e necessidades do consumidor. Neto (2005) afirma que conhecer o mercado é fundamental para qualquer empresa vender seu produto. Se o

mercado não é conhecido, o risco das pessoas não serem atraídas pelo produto é normalmente alto.

Para Kotler e Amstrong (1993 *apud* SCHNEIDER et al, 2008) conhecer as preferências e percepções do consumidor em relação ao produto, preço e propagandas, pode fornecer uma importante vantagem sobre os concorrentes.

Segundo Mowen e Minor (2003 *apud* ARAÚJO e SILVA, 2006) o comportamento do consumidor envolve o estudo das unidades compradoras e dos processos de troca para adquirir, consumir e dispor de mercadorias, serviços, experiência e idéias.

Além disso, a falta de conhecimento do mercado pode levar a empresa a lançar um produto, que mesmo satisfazendo as necessidades do cliente, não seja competitivo perante a concorrência ou que se torne economicamente inviável em pouco tempo.

Para Kotler e Amstrong (2003) se uma empresa deseja ser bem sucedida no mercado competitivo atual precisa está voltada para o cliente. Deve conquista-lo do concorrente e mantê-lo através da disponibilização de um valor superior. Para fazer isso, segundo os autores, é necessário que as necessidades e desejos do cliente sejam entendidos. Esse entendimento é alcançado através de uma análise cuidadosa dos consumidores.

Para o SEBRAE-SP (*apud* MAIA, 2006) a análise de mercado é a função que relaciona os consumidores com o analista de mercado por meio de informações. Através dessas informações é possível encontrar oportunidades e problemas de mercado, melhorando a compreensão do mesmo. Assim, pode-se afirmar que o estudo de mercado é um instrumento que permite obter dados que de alguma forma serão analisados, com auxílio de técnicas estatísticas, para a análise do produto no mercado.

Maia (2006) afirma que, na prática, os estudos de mercado são aplicáveis nos seguintes campos:

- Conhecimento do consumidor
- Projeto e aceitação do produto
- Distribuição no mercado
- Estudos sobre publicidade

O bom desempenho de empresas nesses quatro campos é algo primordial para que conquistem clientes, matenha-os e se tornem competitivas. Isso mostra o quanto o estudo de mercado é importante para a gestão das empresas.

2.1.3 Segmentação do mercado

Segundo Souza e Cardozo (2009) qualquer que seja o mercado, sempre apresentará grande heterogeneidade. Por essa razão as empresas procuram identificar segmentos de mercado com características distintas na qual possam concentrar seus esforços de marketing e satisfazê-los de maneira mais fácil.

Ferreira (2002) afirma que não é possível para nenhuma empresa atender todo um mercado consumidor, ou pelo menos, da mesma forma. Por isso, as organizações estão segmentando seus mercados para satisfazê-los de maneira mais eficaz. Elas segmentam o mercado, escolhem um ou mais segmentos e tentam satisfazer esse(s) segmento(s) de maneira melhor que seus concorrentes.

Para Doll (2006 *apud* BRANDSTETTER, 2008) qualquer empresa que atue em um mercado amplo, percebe em algum momento que é impossível investir em todas as possibilidades existentes e alcançar um resultado satisfatório, pois a quantidade de recursos despendida para sustentar a estratégia será muito grande. Para o autor, essa é a razão pela qual as empresas precisam escolher um ou mais segmentos para atuar.

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 seguimentos 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.

Para Ferreira (2002) segmentar um mercado significa escolher um grupo de clientes, com necessidades semelhantes, para o qual a empresa poderá fazer uma oferta mercadológica. A segmentação requer que sejam identificados os fatores que afetam as decisões de compras dos consumidores.

Lopes (2009) afirma que o fundamento da segmentação de mercado é muito simples. Tem como base o fato de um único produto não satisfazer as necessidades e desejos de todos os clientes. Para o autor, o centro de toda a discussão sobre segmentação, está no fato de existir diferenças entre os consumidores, que precisam ser consideradas nas tomadas de decisão da empresa.

Lopes (2009) ainda conclui que o processo de segmentação visa identificar diferenças significativas entre os consumidores, separando-os em grupos diferenciados para que a

empresa possa escolher aqueles segmentos que se mostram mais favoráveis a concentrar esforços.

A segmentação pode, então, ser considerada uma ferramenta fundamental para gestão empresarial uma vez que leva as empresas a agir de maneira eficiente, ao permitir que direcionem seus recursos para um melhor atendimento de um mercado consumidor específico.

2.1.4 Pesquisa de mercado

A pesquisa de mercado é um importante instrumento de coleta de informações junto ao consumidor, concorrente ou fornecedor, que tem por finalidade orientar a tomada de decisões de empresários e empreendedores (AZEVEDO, 2009). Através dessas informações é possível conhecer o mercado onde o negócio está inserido ou pretende se inserir.

Uma pesquisa pode ser qualitativa ou quantitativa. Dependendo do objetivo da pesquisa, poderá se optar por uma ou ambas. Gomes (2005) explica ainda que a pesquisa quantitativa é um método de pesquisa que trabalha com indicadores numéricos e segue critérios estatísticos. Ela é apropriada para medir opiniões, atitudes e preferências, estimar o potencial ou volume de vendas de um negócio e para medir o tamanho e a importância de segmentos de mercado.

De acordo com Samara e Barros (2002 *apud* ARAÚJO e SILVA, 2006) a pesquisa de mercado é um projeto formal que visa obter dados de maneira empírica, sistemática e objetiva para solucionar problemas ou determinadas oportunidades relacionadas ao marketing de produtos e serviços.

Camboim e Lima (1999) pesquisa de mercado é um uma importante ferramenta para determinar as características mais valorizadas pelos consumidores. Para o referido autor, seu uso tem se difundido cada vez mais com a aplicação do conceito de marketing.

A pesquisa de mercado é um dos principais instrumentos disponíveis para obter informações que tornem, possíveis conhecer entender as necessidades e desejos dos clientes. Por essa razão, empresas que investem mais em pesquisas de mercado tendem a conhecer melhor os seus clientes.

2.2 Regressão Logística

Os conceitos exibidos nesta seção são importantes para compreensão da análise de regressão logística.

2.2.1 Classificação das variáveis

Mayorga (1999 *apud* MONTENEGRO, 2009) define variável como sendo uma quantidade ou característica que pode assumir qualquer valor numérico. Para o referido autor, por essa razão podem ser consideradas medidas ou classificações: quantidades que variam; conceitos operacionais que contêm ou apresentam valores; ou aspectos discerníveis em um objeto de estudo e passível de mensuração.

Para Prearo (2008 *apud* MONTENEGRO, 2009) as variáveis podem ser métricas (subdivididas em nominal e ordinal) e não-métricas (subdivididas em intervalar e razão). Segundo Pereira (2005 *apud* MONTENEGRO, 2009) classifica as variáveis em quantitativas ou métricas (subdivididas em contínuas e discretas) e qualitativas ou não-métricas (subdivididas em nominal e ordinal).

Para Agresti (2007) considera variável categórica aquela que tem como escala de medidas um conjunto de categorias. Como já dito, essas variáveis podem ser nominais e ordinais. São consideradas ordinais as variáveis que apresentam categorias em uma ordem natural. Se essa ordem não existir as variáveis são ditas nominais.

Agresti (2007) ainda menciona que na análise com variáveis nominais são obtidos os mesmos resultados, independente da forma como as categorias são listadas. Enquanto na análise com variáveis ordinais as categorias devem ser listadas segundo uma das ordens naturais (crescente ou decrescente). Se estas ordens naturais não forem obedecidas os resultados irão apresentar-se diferentes.

Conforme Larson e Farber (2003 *apud* SANT'ANNA, 2006) há uma relação entre a qualidade da análise estatística e a classificação das variáveis utilizadas. Essa relação existe porque a adequação da técnica estatística utilizada depende da natureza que a variável em estudo apresenta. Para os referidos autores, de acordo com a estrutura numérica, as variáveis são classificadas em:

- Quantitativas – quando os resultados das observações são expressos sempre em números, pertencentes a um conjunto dos números reais, representando contagens ou medidas.
- Qualitativas – quando os resultados das observações são expressos em categorias ou níveis diferenciados por uma característica não numérica, apresentando ou não algum ordenamento.

Magalhães e Lima (2002 *apud* SANT'ANNA, 2006) dividem as variáveis qualitativas (ou categóricas) em nominais e ordinais, e dividem as variáveis quantitativas em contínuas e discretas. Subdividem ainda as variáveis nominais em:

- Dicotômicas ou binárias – quando apresentam duas categorias
- Politômicas ou polinomiais – quando apresentam mais de duas categorias

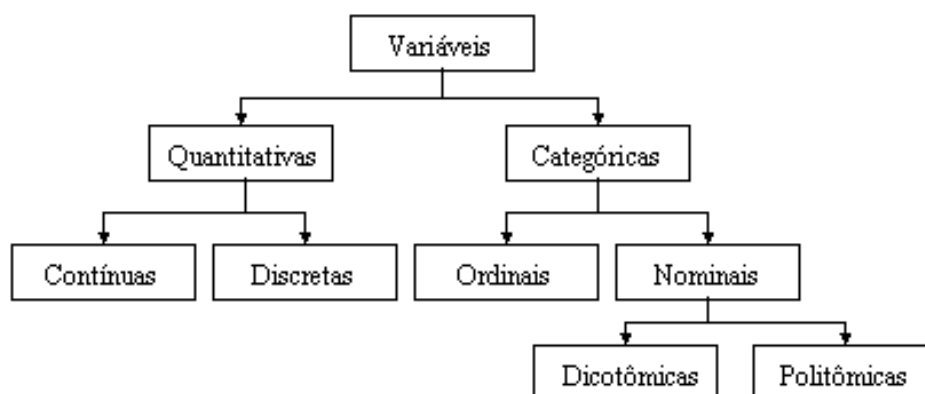
Montenegro (2009) afirma que os métodos utilizados para variáveis ordinais não podem ser utilizados em situações onde estão presentes variáveis nominais. Já os métodos utilizados por variáveis nominais podem ser usados para variáveis ordinais, desde que as categorias obedeçam a uma das ordens naturais.

Agresti (2007) complementa, afirmando que ao usar um método para variáveis nominais em situações com variáveis ordinais estaremos desconsiderando que a ordem das categorias é importante, por essa razão podemos ter uma perda do poder de análise.

Sant'Anna (2006) compartilha a mesma idéia ao mencionar que não é adequado utilizar os mesmos testes estatísticos para diferentes tipos de variáveis. Por isso, é essencial classificar as variáveis do conjunto a ser estudado, pois diferentes classificações podem nos levar a utilizar uma técnica estatística diferente.

A classificação das variáveis é mostrada no diagrama da figura (1).

Figura 1 - Classificação das variáveis



Neste trabalho há apenas variáveis categóricas. Para simplificar a classificação, as variáveis categóricas nominais dicotômicas serão denominadas variáveis dicotômicas e os demais tipos de variáveis categóricas serão chamadas de categóricas. Mas ao utilizar essa simplificação preocupou-se em manter a ordem natural das categorias das variáveis categóricas ordinais.

2.2.2 Análise de Regressão

Segundo Royson e Suerbrei (2008 *apud* MONTENEGRO, 2009) modelos de regressão têm realizado muitas tarefas em diversos campos científicos onde estão presentes dados estatísticos. Entre essas tarefas estão:

- A predição de resultados específicos
- Identificação de preditores importantes
- Compreensão dos efeitos de preditores
- Ajustes para preditores incontroláveis
- Estratificação por risco

Para Devore (2006) o objetivo da análise de regressão é exatamente a exploração da relação entre duas (ou mais) variáveis, de modo a obter informações sobre uma delas, por meio dos valores conhecidos da(s) outra(s). Segundo o referido autor a análise de regressão é a parte da estatística que estuda a relação entre variáveis relacionadas de modo não-determinístico.

Johnson & Wichern (1992 *apud* MONTENEGRO, 2009) consideram a análise de regressão uma metodologia a ser utilizada em predições dos valores de uma variável resposta (dependente) a partir de variáveis explicativas (independentes), podendo também ser usada na avaliação dos efeitos das variáveis explicativas sobre as respostas.

A análise de regressão possibilita, assim, disponibilizar informações importantes acerca da variável dependente e das variáveis independentes, a partir de alguns dados conhecidos sobre as variáveis envolvidas no estudo.

2.2.3 Modelos Lineares Generalizados (MLG's)

Conforme Cox (1996 *apud* SANT'ANNA, 2006) a utilização de um modelo linear normal nem sempre é adequada. Uma das razões para isso é o fato da normalidade ser sempre um requisito para modelos desse tipo. Muitas vezes temos dados que são mensurados como proporções e neste caso dificilmente será possível supor a normalidade. Nessas situações é necessário procurar um outro tipo de modelo para relacionar as variáveis.

Quando o comportamento das variáveis não for normal uma boa alternativa para modelar a relação entre elas é utilizar um Modelo Linear Generalizado. Este tipo de modelo

permite analisar a influência de uma variável sobre a outra e fazer previsões. Uma outra vantagem concebida por esses modelos é a possibilidade de trabalhar com variáveis qualitativas (SANT'ANNA, 2006).

Para Agresti (2002) os modelos lineares generalizados podem ser usados para estender a análise de regressão para inclusão de respostas que não apresentam distribuição normal. Segundo esse autor, os MLG's apresentam três componentes:

- Componente aleatório – que apresenta a distribuição de probabilidade da variável dependente;
- Componente sistemático – que especifica a variável explicativa, entrando do lado direito da equação do modelo;
- Função de ligação – que denota o valor esperado da variável dependente, por meio da função que relaciona a variável resposta as variáveis explicativas.

Para Figueira (2006) existem duas classes de MLG's:

- Modelos logit – onde a variável dependente é associada a uma variável aleatória que se comporta de acordo com uma distribuição binomial;
- Modelos loglinear – onde a variável dependente é associada a uma variável aleatória que se comporta de acordo com distribuição de Poisson.

Segundo Montenegro (2009) os modelos lineares generalizados possuem uma função de densidade de probabilidade com a seguinte forma:

$$f(y_i, \theta_i) = a(\theta_i)b(y_i)\exp[y_i Q(\theta_i)] \quad (1)$$

Nessa equação y_i representa as variáveis aleatórias para a qual procura-se uma estimativa (variável dependente dos modelos). O valor do parâmetro θ_i pode variar de 1 até n, dependendo do valor das variáveis independentes.

2.2.4 Regressão logística

Conforme Hair *et al* (2006 *apud* MINUZZI *et al*, 2007) a regressão logística é um tipo de regressão adequada para explicar o comportamento de uma variável dependente quando essa apresentar um formato binário ou categórico.

Fahat (2003 *apud* CIMROT *et al*, 2009) afirma que a regressão logística é adequada para situações em que se deseja relacionar variáveis independentes a uma variável dependente

dicotômica. Ressalta ainda que mesmo nas situações onde a variável dependente não for dicotômica é possível dicotomizá-la para uma posterior aplicação da regressão logística.

Hosmer e Lemeshow (2000) afirmam que a regressão logística tornou-se um método padrão para realizar a análise de regressão utilizando variáveis binárias. De acordo com os referidos autores, a regressão logística segue os mesmos princípios gerais usados na regressão linear, diferenciando-se, porém, nas respostas e suposições feitas para concepção do modelo.

Segundo Figueira (2006), quando a variável dependente, na análise de regressão, for categórica deve-se utilizar a regressão logística, pois a inadequação do uso do método dos mínimos quadrados para estimação dos parâmetros do modelo não permite que seja utilizada a regressão linear.

Hair (2005 *apud* PACAGNELLA JR. *et al* 2009) declara que o nome dessa técnica é derivado do nome da transformação utilizada na variável resposta e torna possível obter diretamente a probabilidade de um fenômeno estudado vir a ocorrer.

Para Garson (2008 *apud* PACAGNELLA JR. *et al* 2009) deve-se obedecer alguns pressupostos para dedução do modelo de regressão logística:

- A variável dependente deve ser binária ou multinomial (com mais de duas categorias)
- Devem ser incluídas todas as variáveis relevantes no modelo
- Devem ser excluídas as variáveis irrelevantes do modelo
- Não deve existir multicolinearidade
- Deve-se buscar um ajuste adequado

A regressão logística binária apresenta um modelo linear generalizado cuja classe é a *logit*. Como mencionado anteriormente, os modelos *logit* possuem uma variável aleatória que pode ser associada à distribuição binomial. Esses modelos têm uma função de densidade de probabilidade que assumem a seguinte forma:

$$f(y_i, \theta_i) = \pi^{y_i} (1 - \pi)^{1-y_i} \quad (2)$$

Após uma manipulação algébrica chega-se a:

$$f(y_i, \theta_i) = (1 - \pi) \exp \left[y_i \log \left(\frac{\pi}{1 - \pi} \right) \right] \quad (3)$$

É possível perceber que as equações (3) e (1) apresentam uma semelhança. Para que elas sejam iguais basta que consideremos:

$$\theta_i = \pi \quad (4)$$

$$b(y_i) = 1 \quad (5)$$

$$Q(\theta_i) = \log\left(\frac{\pi}{1-\pi}\right) \quad (6)$$

Isso mostra que o modelo de regressão logística pode ser considerado um tipo de modelo linear generalizado.

O parâmetro π do modelo *logit* assume valores entre 0 e 1, e pode ser interpretado como a probabilidade de ocorrência de um evento. O seu valor é dado por:

$$\pi(x_{i1}, x_{i2}, \dots, x_{ip}) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})} \quad (7)$$

ou

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (8)$$

Ao considerar apenas uma das variáveis tendendo a infinito, mantendo todas as outras constantes, é possível concluir que π tende a zero se o parâmetro desta variável for negativo e tende a 1 se o parâmetro for positivo.

Montenegro (2009) ressalta que a variável resposta π apresenta um comportamento aproximadamente linear em valores intermediários e um comportamento curvilíneo nos valores extremos (grandes ou pequenos). Desse modo, a curva apresenta o comportamento indicado na figura 1.

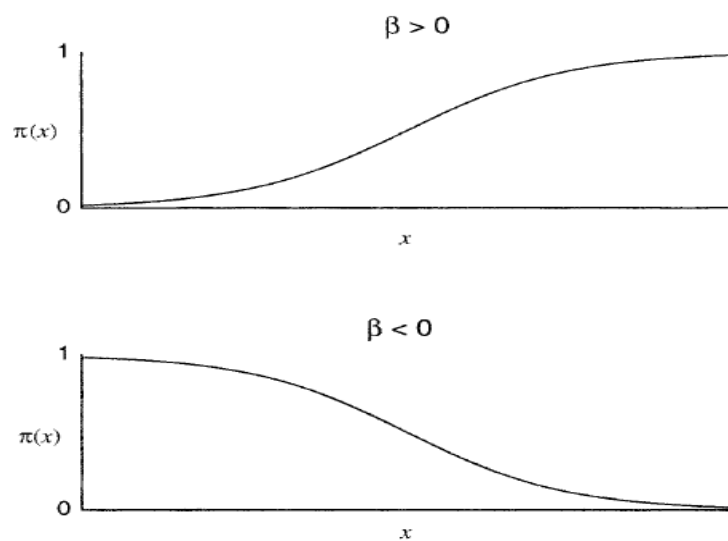


Figura 2 – Curva em S gerada do modelo de regressão logística

Fonte: Agresti (2007)

Para obter os parâmetros do modelo da equação (7) utilizou-se o método da máxima verossimilhança. Para mostrar como esse método é usado é necessário primeiramente exibir a função de verossimilhança que representa o modelo. As observações $y_1, y_2, y_3, \dots, y_p$ podem ser consideradas independentes. Desse modo, utilizando a equação (2) é possível chegar a seguinte função de distribuição de probabilidade conjunta:

$$\prod_{i=1}^n f(y_i, \theta_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (10)$$

Assim a função de verossimilhança será a seguinte:

$$L(\beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (11)$$

Como sabemos os valores máximos de uma função são obtidos onde as derivadas parciais são iguais zero.

Desse modo, o estimador de máxima verossimilhança é a solução para:

$$\sum_{i=1}^n (y_i - \pi_i) = 0 \quad (12)$$

E,

$$\sum_{i=1}^n x_{ij} (y_i - \pi_i) = 0 \quad (13)$$

Em que $j = \{1, 2, \dots, j\}$

Considere as matrizes seguintes:

$$Y = (y_1, y_2, \dots, y_n)'_{1 \times n}$$

$$\Pi = (\pi_1, \pi_2, \dots, \pi_n)'_{1 \times n}$$

$$B = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)'_{1 \times (p+1)}$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{n1} \\ 1 & x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \quad (14)$$

$$V = \begin{bmatrix} \pi_1(1-\pi_1) & 0 & \cdots & 0 \\ 0 & \pi_1(1-\pi_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_1(1-\pi_1) \end{bmatrix} \quad (15)$$

É possível usar essas matrizes para escrever as $p+1$ equações na forma matricial, como:

$$\frac{\partial L(\beta)}{\partial \beta} X'(Y - \Pi) = 0 \quad (16)$$

Um coeficiente importante quando se trabalha com regressão logística é a *odds ratio* (razão da chance) ψ . Esse coeficiente pode informar o efeito da variação em uma determinada variável sobre a chance de ocorrência de um evento. É dado por:

$$\psi = \frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]} \quad (17)$$

Esse coeficiente pode ser obtido algebricamente para cada uma das variáveis substituindo a expressão (9) na expressão (17). Por exemplo, pode-se obter o valor de ψ para a variável x_1 , cujo parâmetro é β_1 . O valor da *odds ratio* nesse caso seria dado por:

$$\psi = \frac{\frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} / \frac{1}{1 + \exp(\beta_0 + \beta_1)}}{\frac{\exp(\beta_0)}{1 + \exp(\beta_0)} / \frac{1}{1 + \exp(\beta_0)}} \quad (18)$$

É fácil mostrar que a expressão (18) é semelhante à expressão (19):

$$\psi = \exp(\beta_1) \quad (19)$$

Utilizando a expressão (19) é possível obter facilmente o valor da *odds ratio* para uma determinada variável.

O significado da *odds ratio* pode ser mais bem compreendido observando a equação (8). Ao aplicar a função exponencial nos dois lados da equação é obtida a equação (20).

$$\left(\frac{\pi}{1 - \pi} \right) = e^{\beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p}} \quad (20)$$

$$\left(\frac{\pi}{1 - \pi} \right) = e^{\beta_0} * (e^{\beta_1})^{x_1} * (e^{\beta_2})^{x_2} * \dots * (e^{\beta_p})^{x_p} \quad (21)$$

Sendo o lado direito da equação (21) a chance de ocorrência de um evento (razão entre a probabilidade de sucesso e a de fracasso de um evento), essa equação indica, por exemplo, que uma variação de uma unidade no valor da variável x_1 resulta em um impacto de e^{β_1} sobre

a chance de o evento representado pelo modelo ocorrer, ou seja, a chance de um evento ocorrer aumentará ou será reduzida em e^{β_1} para cada unidade adicional na variável x_1 .

2.2.5 Métodos de seleção de modelos

Hosmer e Lemeshow (2007) recomendam que sejam utilizados métodos concebidos para selecionar as variáveis que irão fazer parte do modelo. Neste sentido, o autor afirma que o objetivo de qualquer método desse tipo é selecionar os melhores modelos, com um determinado número de variáveis, dentro do contexto científico. Para os referidos autores os melhores modelos são os que apresentam maior simplicidade com maior poder de explicação.

Agresti (2002) cita três métodos de seleção das variáveis:

- Seleção *Forward* - nesse método o processo é iniciado com modelos sem variáveis independentes e são adicionadas as variáveis até que o ajuste não possa mais ser melhorado ou não haja mais variáveis para fazer a análise.
- Seleção *Backward* – nesse método o processo é iniciado com o modelo mais complexo (com o maior número de variáveis) e as variáveis são retiradas até o momento onde o ajuste não possa mais ser melhorado. Nesse caso, deve ser eliminada a variável que ao ser retirada cause o menor impacto no modelo, ou seja, a que apresentar o coeficiente com o maior p-valor.
- Seleção *forward* e *backward* – nesse modelo tanto são acrescentadas variáveis quanto são retiradas.

Draper e Smith (1966 *apud* GUIMARÃES, 2006) compartilham a mesma idéia ao propor alguns métodos de seleção de variáveis, sendo os principais deles os de eliminação *backward*, seleção *forward* e regressão *stepwise* (mistura dos dois outros métodos), sendo o último o mais utilizado. Os autores acrescentam que o uso dessas técnicas é uma maneira eficiente e rápida de selecionar um bom modelo a partir da análise de muitas variáveis.

2.2.6 Parâmetros de avaliação do modelo

a) Critério de Informação de Akaike (AIC)

O Critério de Informação de Akaike (AIC) foi desenvolvido por Hirotugu Akaike em 1971. Essa é uma medida de avaliação de modelos estatísticos e pode ser utilizado para

descrever o equilíbrio entre a complexidade de um modelo e sua precisão ao representar a realidade.

Segundo Agresti (2002), além dos testes de significância outros critérios podem ser utilizados para avaliar o quanto um modelo é bom. Dentre esses critério um muito conhecido é o Critério de Informação de Akaike (AIC).

Em uma estimação por regressão logística, o melhor modelo, segundo esse critério, é o que apresenta o menor AIC, pois esse pode ser considerado, dentre os modelos com maior simplicidade, aquele que mais aproxima o valor ajustado dos valores das probabilidades reais. O autor indica que o valor dessa estatística é dado pela expressão (22).

$$AIC = 2k - 2\ln(L) \quad (22)$$

Nesta expressão k representa o número de parâmetros estimados e L representa o valor do logaritmo da função de verossimilhança.

Pela equação podemos perceber que quanto maior o número de parâmetros estimados para o modelo maior o valor do modelo e, conseqüentemente, pior será esse modelo. Dessa maneira o AIC permite comparar modelos de acordo com a complexidade. Neste sentido, para o critério modelos simples são melhores do que modelos complexos.

Por sua vez, quanto maior o valor assumido pela função de verossimilhança maior o logaritmo desse valor e, conseqüentemente, menor será o valor do AIC, ou seja, melhor será o modelo. Sabe-se que quanto maior for o valor da função de verossimilhança mais verossímil será a estimação, ou seja, mais próximo os valores estimados estarão dos valores reais. Dessa forma o AIC avalia a precisão do modelo.

É possível concluir então que o AIC torna possível escolher o melhor modelo considerando a complexidade e sua precisão e, segundo esse critério, bons modelos são aqueles mais simples e capazes de representar bem a realidade.

Segundo Barbiero (2003) os programas de computadores que calculam o AIC estimam seu valor por meio da expressão (23).

$$AIC \cong n * (1 + \log(2\pi)) + n * \log(\sigma^2) + 2k \quad (23)$$

Nesta expressão n é o número de observações σ^2 é a variância dos resíduos (diferença entre os valores estimados da variável dependente do modelo e os valores reais para dada uma observação).

b) A seleção dos modelos e o p-valor

As informações fornecidas nessa seção são baseadas nas afirmações de Hosmer e Lemeshow (2000).

Para avaliar a significância do modelo por meio do p-valor também se utiliza a função de verossimilhança. O *p-valor* é obtido utilizando-se a estatística *Deviance* que, por sua vez, é calculada usando a função de verossimilhança.

A estatística *Deviance* é calculada pela expressão (24). O modelo saturado é aquele na qual todos os coeficientes das variáveis dadas estão presentes.

$$D = -2 \ln \left(\frac{\text{função de verossimilhança do modelo ajustado}}{\text{função de verossimilhança do modelo saturado}} \right) \quad (24)$$

Se o modelo a ser analisado apresentar apenas uma variável (regressão logística simples), a estatística G será dada pela expressão (25). Caso o modelo apresente mais de uma variável a estatística G é dada pela expressão (26).

$$G_{\text{simples}} = D(\text{modelo com a variável}) - D(\text{modelo sem a variável}) \quad (25)$$

$$G_{\text{multipla}} = D(\text{modelo com todas as variáveis}) - D(\text{modelo sem variáveis}) \quad (26)$$

Sob a hipótese de que o coeficiente da variável considerada na expressão (25) é igual a zero a estatística que foi denominada de G_{simples} assume uma distribuição qui-quadrado (χ^2) com grau de liberdade igual a 1. Isso torna possível que seja testado a significância da variável e calculado o p-valor. Nas expressões (27) e (28) é mostrado a forma de fazer o cálculo do p-valor para um modelo de regressão logística simples.

$$P(\chi^2(1) > G_{\text{simples}}) < p' \quad (27)$$

$$p\text{-valor}_{\text{simples}} = p' \quad (28)$$

Então para encontrar o p-valor deve-se entrar com os valores de G_{simples} e o grau de liberdade 1 em uma tabela de distribuição qui-quadrado e encontrar o valor de p' mostrado na expressão (24).

Sob a hipótese de que os coeficientes de todas as variáveis são nulos a estatística que foi denominada G_{multipla} assumirá uma distribuição qui-quadrado com grau de liberdade igual ao número de variáveis. Por esse meio, é possível encontrar o p-valor e verificar a significância modelo.

Ao considerar o número de variáveis de um modelo de regressão logística igual a k , o p -valor do modelo de regressão logística múltiplo é obtido utilizando as expressões (27) e (28).

$$P(\chi^2(k) > G_{\text{multipla}}) < p'' \quad (29)$$

$$p\text{-valor}_{\text{multipla}} = p'' \quad (30)$$

Ao se estudar um modelo de regressão logística com várias variáveis utiliza-se o p -valor_{simples} para encontrar o p -valor de cada coeficiente e adota-se o p -valor_{multipla} como o p -valor do modelo.

O p -valor do modelo é utilizado para testar a hipótese de que todos os coeficientes do modelo são iguais a zero, ou seja, testar a hipótese de que não existe modelo que relacione a variável dependente as variáveis independentes do modelo analisado. Se o p -valor for maior do que o nível de significância adotado deve-se rejeitar a hipótese de que todos os coeficientes são iguais a zero, o que leva a concluir que existe um modelo que relacione a variável dependente e as variáveis independentes.

O p -valor do coeficiente tem uma função semelhante, porém testa a hipótese de que o coeficiente analisado é igual a zero, ou seja, se a variável desse coeficiente tem alguma influência na variável dependente do modelo. Sendo o p -valor do coeficiente superior ao nível de significância, quanto mais distante for o valor dessas duas estatísticas maior será a tendência de que a variável do coeficiente não influencia na variável dependente do modelo.

c) Erro Padrão e Estatística Wald

Segundo Hosmer e Lemeshow (2000) para os coeficientes β_j de um modelo de regressão múltipla, o erro padrão $SE(\beta_j)$ é obtido a partir da expressão (31).

$$SE(\beta_j) = [\text{var}(\beta_j)]^{1/2} \quad (31)$$

Hosmer e Lemeshow (2000) ainda afirmam que o erro-padrão é utilizado para calcular a estatística *Wald* (W). Essa estatística é um outro meio para analisar a significância dos coeficientes do modelo. A equação (30) é usada para efetuar o cálculo da estatística *Wald*.

$$W = \frac{\beta_j}{SE(\beta_j)} \quad (32)$$

Sob a hipótese de que β_j é igual a zero a estatística W assume uma distribuição normal padrão. Desse modo, é possível utilizar essa distribuição para testar a hipótese de que β_j é igual a zero.

d) Pseudo- R^2

Há, na análise de regressão logística, medidas correspondentes ao coeficiente de determinação presente na análise de regressão linear, que indicam a proporção de variação total ocorrida na variável dependente que em função das independentes. Essas medidas são denominadas de pseudo- R^2 .

Segundo Corrar, Paulo e Dias Filho (2009) embora não se disponha de uma medida rigorosamente idêntica ao R^2 nos modelos de regressão logística há medidas que executam bem essa função. Para os referidos autores, as mais conhecidas são as seguintes:

- R^2 -logit: expressa a variação percentual do valor de verossimilhança do modelo que considera apenas a constante em relação ao modelo que considera as variáveis explicativas. Este coeficiente apenas indica se o acréscimo das variáveis ao modelo irá melhorá-lo em relação ao modelo composto apenas por uma constante.
- Cox-Snell R^2 : esse coeficiente apresenta maiores semelhanças com o coeficiente de determinação da análise de regressão linear. Entre duas equações igualmente válidas deve-se preferir o que apresentar o maior Cox-Snell R^2 . Porém esse coeficiente também não é capaz de indicar a proporção das variações da variável dependente que são explicadas pelas variáveis independentes. Possui uma escala que se inicia em 0, mas não chega ao valor 1 no seu limite superior.
- Nagelkerke R^2 : assim como o Cox-Snell R^2 é capaz de indicar que no julgamento de modelos igualmente válidos, aquele que apresenta o maior valor desse coeficiente é o melhor. Apesar da semelhança com o Cox-Snell R^2 , o Nagelkerke R^2 se apresenta mais compreensível. Uma vez que sua escala varia de 0 até 1, é possível indicar a variação da chance de ocorrência de um evento que pode ser explicada pelas variáveis independentes do modelo.

No software R, utilizado na dedução dos modelos desse trabalho, o coeficiente pseudo R^2 é o Nagelkerke R^2 .

3 ASPECTOS METODOLÓGICOS

Tomando como base Gil (1994), pode-se classificar esta pesquisa como bibliográfica e descritiva. Bibliográfica porque faz uso das contribuições dos diversos autores sobre o tema pesquisado. Descritiva por ter como uma de suas finalidades a descrição de características do mercado consumidor de sacolas ecológicas do município de João Pessoa.

Além disso, a pesquisa pode ser classificada como quantitativa uma vez que se utilizaram indicadores numéricos e segue critérios estatísticos ao atribuir valores numéricos a opiniões, características e preferências de determinada amostra.

Quanto aos meios de investigação, foram utilizados materiais como livros, artigos, dissertações, bem como textos disponíveis na internet, relacionados aos conceitos de pesquisa de mercado e regressão logística.

O ambiente da pesquisa foi o município de João Pessoa, tendo como unidade espacial de referência o mercado de sacolas ecológicas do município.

Com o intuito de alcançar os objetivos propostos, foram adotados os seguintes procedimentos:

Inicialmente, levantou-se na literatura conceitos relacionados à pesquisa de mercado e regressão logística. Em seguida, foram elaboradas 12 questões referentes às preferências, opiniões e características dos indivíduos entrevistados. Os questionários foram aplicados junto a 57 pessoas de 21 bairros de João Pessoa.¹

O maior número de entrevistados (74%) estão em um grupo de sete bairros: Bancários, Bessa, Cabo Branco, Cristo Redentor, Jaguaribe e Manaíra. Além desses, foram entrevistados um menor grupo de indivíduos (26%) dos seguintes bairros: Torre, Mangabeira, Jardim Cidade Universitária, Água Fria, Bairro dos Novaes, Rangel, Bairro dos Estados, Valentina, Intermares, Cruz das Armas, Castelo Branco e Geisel.

Quanto ao perfil dos entrevistados, 49% são do sexo masculino e 51% do sexo feminino. A maioria (44%) tem entre 20 e 30 anos. A faixa etária de indivíduos com idades entre 31 a 40 anos é de cerca de 25%. O grupo de indivíduos com idades acima de 40 anos correspondem a 31% dos entrevistados.

Sobre o estado civil dos participantes da pesquisa, 49% se declararam casados, 46% solteiros e 5% separados. O maior percentual de escolaridade dos entrevistados é o de

¹ Os questionários foram aplicados entre os dias 13 e 29 de julho de 2009.

indivíduos com ensino superior completo (78%), seguido de superior incompleto (16%). Os demais (6%) possuem até o ensino médio.

A maior parte dos entrevistados (29%) declarou renda familiar de três a seis salários mínimos. Logo em seguida, têm-se os indivíduos com renda entre seis a nove salários mínimos (25%). Os indivíduos com renda inferior a três salários correspondem a 14% dos entrevistados e os que possuem renda acima de seis salários mínimos são 32%.

A partir do questionário utilizado na pesquisa de mercado, foram definidas 18 variáveis dicotômicas e seis variáveis categóricas para o estudo da composição do modelo. No entanto, duas das variáveis dicotômicas (não fizeram parte da análise pelo fato de praticamente terem assumido o mesmo valor em todas as observações da amostra).

As variáveis dependentes do estudo representam a disposição dos indivíduos para compra. Para representar essas variáveis podiam ser escolhidos nove conjuntos de dados (cada um correspondente a um tamanho de sacola e um preço específico). Para esse estudo foram escolhidas quatro dessas opções.

Essas variáveis podem assumir o valor 1 para respostas que indicam que o entrevistado declarou que está disposto a pagar, pela sacola, o valor fixado e pode assumir o valor 0 para o entrevistado que declarou que não pagaria o valor fixado. Essas variáveis são as seguintes:

- 1) *compram7* – que será utilizada apenas nos modelos de regressão para sacolas de tamanho médio a R\$ 7,00;
- 2) *compram5* – que será utilizada apenas nos modelos para sacolas de tamanho médio ao preço de R\$ 5,00;
- 3) *comprap4* – que será usada nos modelos para sacolas de tamanho pequeno ao preço de R\$ 4,00;
- 4) *comprag10* – que será usada somente nos modelos para tamanho grande ao preço de R\$ 10,00;

Os fatores que irão formar a matriz de variáveis explicativas estão enumerados a seguir:

- 1) variável dicotômica *cons_amb* que representa a opinião dos participantes da pesquisa sobre a influência da consciência ambiental na compra da sacola ecológica. Para construção do modelo, considerou-se *cons_amb*=1 para os que responderam que a consciência ambiental influenciaria na compra e *cons_amb*=0 para os que afirmaram que não influenciaria.

2) Para representar a opinião dos entrevistados sobre a influência dos descontos nas compras com sacolas ecológicas criou-se a variável dicotômica *descontos*, que assume o valor 1 para os que declararam que isto influenciaria na compra e assume o valor 0 para os que afirmaram que esses descontos não influenciariam.

3) A variável que representaria a influência, na compra, do fato das sacolas se tornarem moda foi denominada *moda*. Essa variável também é dicotômica. Devido ao baixo percentual de pessoas que declararam que esse fator exerceria influência na compra (1,8% dos entrevistados) não foi considerado na avaliação do modelo.

4) Para representar a influência que uma futura lei proibindo as sacolas plásticas teria sobre a compra das sacolas ecológicas incluiu-se a variável dicotômica *obrig* que assume o valor 1 quando a obrigatoriedade da abolição das sacolas plásticas exerce influência e 0 quando esta não exerce nenhuma influência.

5) Os entrevistados foram também questionados sobre a importância de quatro aspectos do produto: resistência, modelo, praticidade e espaço interno das sacolas. Esses quatro aspectos foram representados pelas variáveis dicotômicas *resistência*, *modelo*, *praticidade* e *tamanho* (capacidade interna), respectivamente. Essas variáveis assumem o valor 0 para os que afirmaram que o aspecto é importante, ou muito importante, e assumem o valor 1 para os que afirmaram que o aspecto tinha pouca importância para eles.

6) A cor do produto foi representada pela variável categórica *cor* e assume o valor 1 quando a preferência for pela cor branca, 2 quando for pela cor bege, 3 quando for pela cor preta e 4 quando for pelo modelo com estampas.

7) Os indivíduos que participaram da pesquisa também indicaram o número médio de produtos que costumam comprar. O número de itens que os indivíduos costumam comprar é representada, no modelo, pela variável categórica *num_itens*, que assumirá os valores 1 para compras com 10 ou menos itens, 2 para compras de mais de 10 itens e 3 para as feiras mensais.

8) Perguntou-se também se os entrevistados utilizariam as sacolas ecológicas em quatro locais específicos: supermercados, livrarias, shoppings e padarias. Cada um desses locais é representado por uma variável dicotômica (*supermercados*, *livrarias*, *shopping* e *padarias*, respectivamente) que assumem o valor 0 quando declara-se que não utilizaria e 1 quando declara-se que utilizaria.

9) Os entrevistados declararam a frequência com que vão as compras. Essas categorias foram agrupadas na variável categórica *frequencia*. Essa variável assume o valor 1 quando a

freqüência for uma vez no mês, o valor 2 para uma freqüência quinzenal, o valor 3 para uma freqüência semanal e o valor 4 quando a freqüência for quase todos os dias.

10) Uma outra questão presente na pesquisa foi sobre a preferência de tamanho da sacola. Essa questão foi adicionada ao estudo por meio da variável categórica *tam_ideal* que tem valor 1 para sacolas de tamanho pequeno, 2 para as sacolas de tamanho médio e 3 para as sacolas de tamanho grande.

11) O fator gênero foi definido pela variável dicotômica *sexo*, que assume o valor 1 para o sexo masculino e 0 para o sexo feminino.

12) A faixa etária foi representada pela variável categórica *fx_etária* e assume o valor 1 para idades de 20 a 30 anos, 2 para idades de 31 a 40 anos e 3 para idades superiores a 40 anos.

13) O estado civil dos indivíduos entrevistados também foi questionado. Essa característica foi representada pela variável dicotômica *est_civil* que assume o valor 1 para os que declararam serem casados e 0 para os que declararam não serem casados.

14) A variável usada para representar a escolaridade foi *escol* (variável dicotômica) e tem valor 1 para os que possuem nível superior completo e 0 para os que não possuem.

15) A renda familiar foi incluída no modelo por meio da variável categórica *renda*, que tem valor 1 para uma renda até três salários mínimos, valor 2 para renda de 3 a 6 salários mínimos, valor 3 para uma renda de 6 a 9 salários mínimos e assume valor 4 para uma renda superior a 9 salários mínimos.

A partir da equação (8), estimou-se um modelo de regressão logística, através do software R, que relaciona a variável resposta (*compram7*, *compram5*, *comprag10*, *comprap4*) às variáveis explicativas aqui descritas.

Devido a limitações de tempo e de recursos não foi possível obter dados de uma amostra representativa da população dos bairros de João Pessoa.

Deve-se ressaltar que importância deste trabalho está no fato de servir como base para outros trabalhos no mesmo sentido. Porém, dado o tamanho da amostra, as conclusões obtidas para este conjunto de observações não devem ser utilizadas para inferir sobre o mercado de sacolas ecológicas de João Pessoa. Caso haja interesse apenas nessa inferência recomenda-se que seja feito um trabalho semelhante com uma amostra representativa.

Para julgar a aceitabilidade de um modelo adotou-se alguns critérios. Julgou-se como aceitável os modelos que se mostraram estatisticamente significativos para um nível de significância de 0,2 (p-valor do modelo superior a 0,2 e o p-valor de cada coeficiente também superior a 0,2) e pseudo- R^2 igual a 0,450.

4 RESULTADOS

4.1 Análise descritiva dos dados

Com relação a disposição dos indivíduos a pagarem pelas sacolas ecológicas a pesquisa mostrou que:

1. Para sacola de tamanho pequeno ao preço de R\$ 4,00, 28% afirmaram que estariam dispostos a pagar o valor, enquanto 72% afirmaram que não comprariam.
2. Para sacola de tamanho médio a R\$ 5,00, 16% afirmaram que estariam dispostos a pagar o valor, enquanto 84% afirmaram que não comprariam.
3. Para sacola de tamanho médio a R\$7,00, 73% afirmaram que não estariam dispostos a pagar enquanto 27% afirmaram que estariam dispostos a pagar.
4. Para sacola de tamanho grande a R\$ 10,00, 28% afirmou que não comprariam enquanto 72% afirmaram que comprariam.

Na pesquisa, 77,2% dos entrevistados afirmaram que a consciência ambiental contribuiria para que comprassem o produto. Para 38,6% dos entrevistados, os descontos estimulariam a compra.

Ao serem questionados sobre a influência que uma futura lei proibindo as sacolas plásticas teria sobre a compra das sacolas ecológicas, 7% declaram que seriam influenciados a comprá-las.

Na pesquisa 97% dos entrevistados afirmaram que considerava a resistência um fator importante. Com relação ao modelo esse percentual foi de 61%, enquanto para o espaço interno foi de 91% e para a praticidade foi de 75%.

No que diz respeito a cor da sacola ecológica que mais agradaria os entrevistados 49% afirmaram preferir bege, 19% preferiram a cor branca, 19 % preferiam com estampas e 13% tinham preferência pela cor preta.

Constatou-se que 39% dos pesquisados fazem compras com 10 ou menos itens, 53% fazem compras com mais de 10 itens e 12% apenas compram uma grande quantidade de itens em feiras mensais.

Para 75,4% dos participantes da pesquisa a sacola ecológica seria usada em supermercados. Um percentual de 36,8% afirmou que utilizaria a sacola em padarias. Para shoppings e livrarias os percentuais foram, respectivamente, 7% e 8,8%.

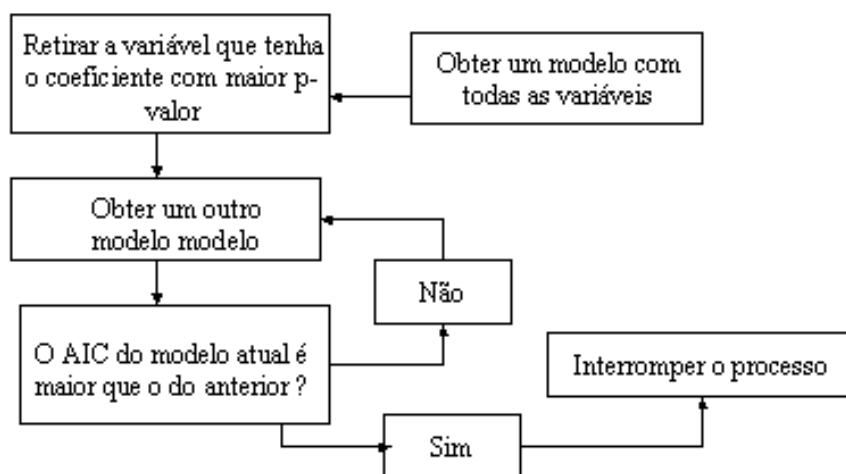
Dos entrevistados, 5% afirmaram ir quase diariamente às compras, 56% afirmaram ir semanalmente, 25% afirmaram ir quinzenalmente e 15% declararam ir uma vez no mês.

No tocante à preferência do tamanho da sacola, 10,5 % declararam a preferência pela a sacola de tamanho pequeno, 73,7% declararam preferir a sacola de tamanho médio e 15,8% informaram que preferiam as de tamanho grande.

4.2 Método utilizado e o resultado da estimação

Na busca de um bom modelo de regressão logística para representar o comportamento dos dados foi usado o método *backward* (Figura 3). Esse método consiste em definir um modelo, retirar a variável com o maior p-valor e observar o valor do AIC. Se o AIC do novo modelo for inferior ao modelo anterior o processo continua. Caso o AIC do novo modelo tenha um valor superior ao do modelo anterior o processo é interrompido e o modelo que procuramos é o que teve a última redução do AIC.

Figura 3 - Método backward



No estudo para a sacola de tamanho médio a um preço de R\$ 7,00, foram gerados 12 modelos até encontrar o mais indicado dentre os que foram obtidos. A tabela 1 mostra informações sobre cada um desses modelos. Nela encontramos os valores do AIC, do pseudo- R^2 e do p-valor do modelo, bem como a variável que apresentou o coeficiente com maior p-valor e o p-valor desse coeficiente.

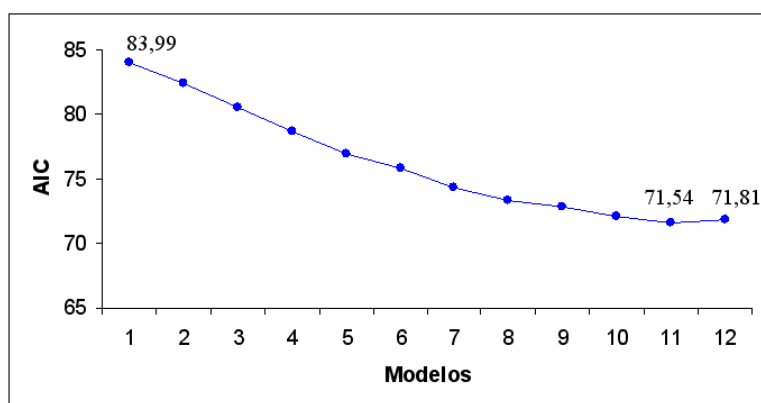
Tabela 1 – Informações sobre os modelos gerados para sacolas ecológicas de tamanho médio ao preço de R\$ 7,00

Modelos	AIC	Variação do AIC	pseudo-R ²	p-valor do modelo	variável a ser excluída	Maior p-valor dos coeficientes das variáveis
Modelo 1	83,99	-	0,599	0,0205	resistência	0,8352
Modelo 2	82,45	-1,54	0,593	0,016	livraria	0,8022
Modelo 3	80,51	-1,94	0,592	0,011	renda	0,7227
Modelo 4	78,64	-1,87	0,590	0,008	cons_amb	0,5822
Modelo 5	76,94	-1,7	0,586	0,0052	supermercados	0,3612
Modelo 6	75,85	-1,09	0,574	0,0044	modelo	0,5016
Modelo 7	74,32	-1,53	0,568	0,0031	shopping	0,3207
Modelo 8	73,35	-0,97	0,554	0,0026	num_médio	0,2474
Modelo 9	72,77	-0,58	0,534	0,0024	est_civil	0,2494
Modelo 10	72,12	-0,65	0,514	0,0022	tam_ideal	0,2427
Modelo 11	71,54	-0,58	0,493	0,002	obrig	0,1553
Modelo 12	71,81	0,27	0,459	0,0025	-	-

Fonte: estimação do autor.

Como o modelo 12 apresentou um AIC superior ao do modelo 11 o melhor modelo pelo método stepwise backward foi o modelo 11. Na figura 4 é possível perceber que a medida que as variáveis menos significativas são retiradas obtém-se modelos melhores.

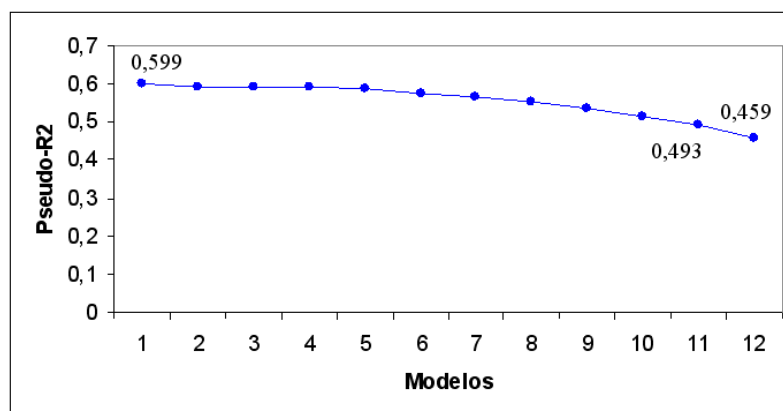
Figura 4 - Comportamento do AIC com a retirada das variáveis menos significativas nos modelos para sacolas de tamanho médio a um preço de R\$ 7,00



O modelo 11 possui nove variáveis e um pseudo-R² igual a 0,493, o que indica que o modelo é capaz de explicar 49,3% da variação da chance relacionada a variável dependente *compram7*. O p-valor do modelo é de 0,002, o que mostra que há um modelo capaz de relacionar a variável *compram7* com as variáveis independentes do modelo 11.

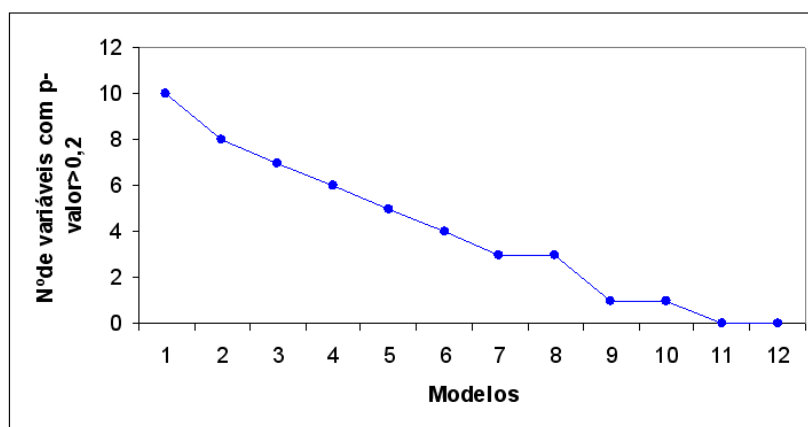
A medida que as variáveis menos significativas de cada modelo são retiradas o pseudo- R^2 é reduzido, o que significa que o poder explicativo é reduzido. A figura 5 mostra como varia o pseudo- R^2 .

Figura 5 - Comportamento do pseudo- R^2 com a retirada das variáveis menos significativas dos modelos para sacolas de tamanho médio a R\$ 7,00



Apesar da redução do poder explicativo, a complexidade dos modelos também é reduzida uma vez que o número de variáveis é reduzido. Além disso, obtém-se modelos com um menor número de variáveis cujos coeficientes não são estatisticamente significativos, uma vez que apresentam p-valor superior a 0,2. É possível visualizar essa observação no gráfico da figura 6.

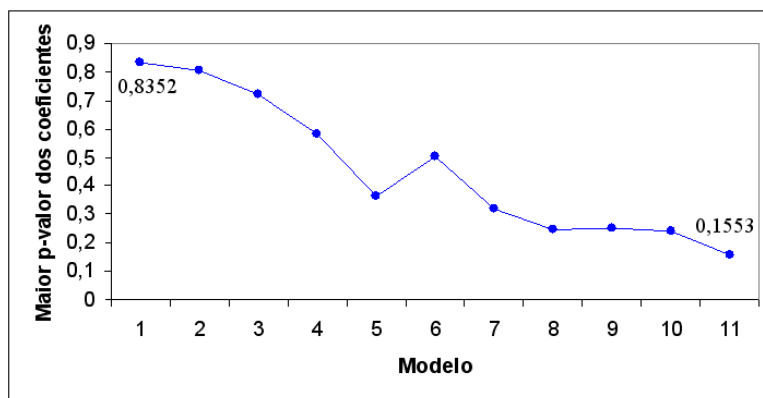
Figura 6 - Comportamento do número de variáveis com p-valor > 0,2 com a retirada das variáveis nos modelos de sacolas de tamanho médio a um preço de R\$ 7,00



À medida que é retirada uma variável a tendência é que sejam obtidos modelos que exigem um nível de significância cada vez menor para julgar todos coeficientes dos modelos

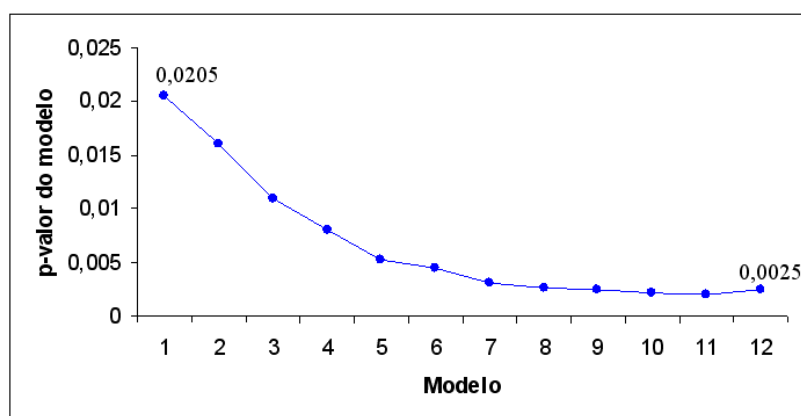
como significativos, ou seja, o maior p-valor dos coeficientes de cada modelo tende a ser cada vez menor. Isso não ocorreu apenas ao retirar a variável *supermercados* do modelo 5. Pode-se visualizar essas observações na figura 7.

Figura 7 - Comportamento do maior p-valor dos coeficientes com a retirada das variáveis nos modelos para sacolas de tamanho médio a um preço de R\$ 7,00



Quando se retira a variável menos significativa do modelo é possível observar uma tendência de redução do p-valor do modelo, ou seja, à medida que a variável menos significativa de cada modelo é retirada obtém-se modelos onde é cada vez mais difícil negar a relação entre a chance, relacionada a variável dependente, e as variáveis independentes. O gráfico presente na figura 8 mostra essa observação.

Figura 8 - Comportamento do p-valor com a retirada das variáveis nos modelos de sacolas de tamanho médio a um preço de R\$ 7,00



O número total de variáveis e o número de variáveis cujos coeficientes apresentam p-valor superior a 0,2 são mostradas, para cada um dos 12 modelos, na tabela 2. Nessa tabela é também mostrado o menor p-valor dos coeficientes dos modelos e o p-valor do intercepto.

Tabela 2 – Outras informações sobre os modelos gerados para sacolas ecológicas de tamanho médio ao preço de R\$ 7,00

Modelos	Número de variáveis	Nº de variáveis com p-valor do coeficiente maior que 0,20	Menor p-valor dos coeficientes das variáveis	p-valor do intercepto
Modelo 1	19	10	0,0034	0,9139
Modelo 2	18	8	0,0071	0,058
Modelo 3	17	7	0,0072	0,0523
Modelo 4	16	6	0,0064	0,0532
Modelo 5	15	5	0,006	0,0561
Modelo 6	14	4	0,0074	0,0513
Modelo 7	13	3	0,008	0,0662
Modelo 8	12	3	0,006	0,0706
Modelo 9	11	1	0,0052	0,0848
Modelo 10	10	1	0,0048	0,0855
Modelo 11	9	0	0,0034	0,1551
Modelo 12	8	0	0,0034	0,2726

Na tabela 2 podemos perceber que nos modelos de 1 a 10 há variáveis que não apresentam coeficientes estatisticamente significativos para um nível de significância de 0,2, uma vez que estas variáveis apresentam um p-valor maior que 0,2. Apenas os modelos 11 e 12 não apresentam coeficientes de variáveis não significativos. No entanto, o coeficiente do intercepto do modelo 12 não é estatisticamente significativo (para esse modelo $p\text{-valor} > 0,2$).

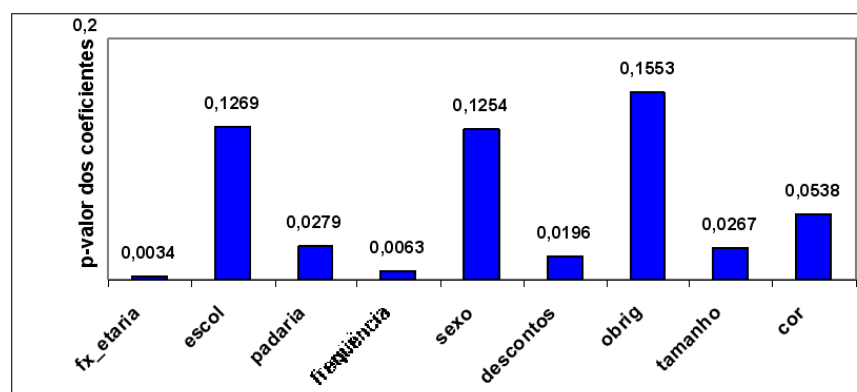
Um dos problemas encontrados no modelo 11 foi o valor relativamente alto do p-valor de alguns coeficientes. Esse problema pode ser identificado ao observarmos o quadro 1, que mostra o p-valor dos coeficientes, o erro padrão, a estatística Wald e os valores dos coeficientes.

Quadro 1 – Informações a cerca dos coeficientes das variáveis independentes e do intercepto do modelo obtido para sacolas de tamanho médio a R\$ 7,00

	Coeficiente	Erro padrão	Wald Z	P-valor
Intercepto	3,8411	2,7016	1,42	0,1551
fx_etaria	-1,9426	0,6634	-2,93	0,0034
escol	-1,589	1,0409	-1,53	0,1269
padaria	-1,8843	0,8569	-2,2	0,0279
frequência	1,6047	0,5877	2,73	0,0063
sexo	-1,3675	0,8924	-1,53	0,1254
descontos	2,8628	1,227	2,33	0,0196
obrig	-2,1579	1,5185	-1,42	0,1553
tamanho	-4,9024	2,2119	-2,22	0,0267
cor	0,7842	0,4067	1,93	0,0538

Perceba que os coeficientes das variáveis *escol*, *sexo* e *obrig*, bem como o coeficiente do intercepto somente podem ser considerados significativos se fosse adotado nesse estudo um nível de significância superior a 0,1553. Como foi adotado um nível de significância de 0,2, o modelo pode ser considerado aceitável. Essa significância dos coeficientes também é mostrada na figura 9.

Figura 9 - p-valor dos coeficientes do modelo para sacolas de tamanho médio a um preço de R\$ 7,00



As expressões (33), (34), (35) detalham a equação do modelo:

$$\ln\left(\frac{p}{1-p}\right) = 3,8411 - 1,9426 * fx_etaria - 1,5890 * escol - 1,8843 * padaria - \\ + 1,6047 * frequência - 1,3675 * sexo + 2,8628 * descontos - 2,1579 * obrig - \\ - 4,9024 * tamanho + 0,7872 * cor = \rho_1 \quad (33)$$

$$\frac{p}{1-p} = e^{\rho_1} \quad (34)$$

$$p = \frac{1}{1 + e^{-\rho_1}} \quad (35)$$

Através do estudo de modelos semelhantes aos obtidos aqui é possível identificar, para um determinado produto, se as opiniões e características dos indivíduos que estão dispostos a pagar o valor fixado, podem estatisticamente ser distinguidas daquelas dos indivíduos que não estão dispostos a pagar o valor fixado.

Ao confirmar que as características e opiniões dos “compradores” são bem definidas e são distintas daquelas dos “não compradores”, é possível saber se um determinado grupo de pessoas, com determinadas opiniões e características, apresentam ou não uma probabilidade alta de serem “compradores”, ou seja, é possível afirmar se um determinado grupo de pessoas forma um bom segmento de mercado ou não, considerando sua disposição para comprar.

Neste caso, a empresa que vende sacolas ecológicas pode utilizar o modelo obtido para tomar decisões no sentido de satisfazer os indivíduos que se declaram dispostos a comprar. Dessa maneira, essa empresa pode evitar despende esforços que surtirão pouco efeito sobre esse grupo de pessoas e evitar despende esforços com pessoas que tem pouca probabilidade de pagar o valor do seu produto.

A equação (34) mostra a chance de um indivíduo está disposto a comprar as sacolas em função das variáveis do modelo. Essa chance é considerada a razão entre a probabilidade de sucesso e a probabilidade de fracasso de um dado evento. Esse evento no nosso caso é alguém pertencer ao grupo de pessoas que pagaria o preço do produto.

Como já foi mencionado é possível encontrar o impacto da variação de uma unidade em uma das variáveis sobre a chance do evento estudado ocorrer. Essa informação é dada pela *odds ratio* (razão da chance) das variáveis. Fazendo isso, é possível dizer se uma mudança no valor de uma variável eleva a chance ou a reduz. A tabela 3 indica o valor da razão da chance para cada uma das variáveis.

Tabela 3 – Valores da *odds ratio* para cada variável do modelo para sacolas ecológicas de tamanho médio ao preço de R\$ 7,00

Variáveis	<i>odds ratio</i>
fx_etaria	0,14
escol	0,20
padaria	0,15
frequência	4,98
sexo	0,25
descontos	17,51
obrig	0,12
tamanho	0,01
cor	2,19

A equação 34 pode ser reescrita de uma maneira que fique fácil visualizar na equação a *odds ratio*. Essa forma de reescrever a equação 34 está presente na equação 36. As *odds ratio* se encontram entre parênteses.

$$\frac{p}{1-p} = e^{3,8411 * (e^{-1,9426})^{fx_etaria} * (e^{-1,5890})^{escol} * (e^{-1,8843})^{padaria} * (e^{1,6047})^{frequência} * (e^{-1,3675})^{sexo} * (e^{2,8628})^{descontos} * (e^{-2,1579})^{obrig} * (e^{-4,9024})^{tamanho} * (e^{0,7872})^{cor}} \quad (36)$$

Baseando-se nos valores da tabela 3, por exemplo, é possível afirmar que uma opinião favorável aos descontos aumenta a razão da chance do modelo. Isto indica que achar os descontos importantes aumenta a probabilidade de alguém pertencer ao grupo de pessoas que tem disposição para pagar o preço das sacolas.

De maneira mais clara, é possível, com base na expressão (36), afirmar que uma opinião favorável ao desconto aumenta a chance de um indivíduo está disposto a comprar em $e^{2,8628}=17,51$ vezes, ou seja, em dois segmentos com características iguais, distinguindo-se apenas pela opinião sobre o desconto, aquele que considera o desconto um importante incentivo a compra do produto possui 17,51 vezes mais chance de pagar R\$ 7,00 pela sacola de tamanho médio do que aquele que considera o desconto pouco importante.

Da mesma forma, também é possível saber que ir às compras com maior frequência eleva a probabilidade de alguém ser um “comprador”. Neste mesmo, sentido uma outra informação útil é saber quais aspectos melhor definem se alguém será um comprador.

A tabela 3 mostra que os aspectos que melhor definem se os indivíduos pertencerão ao grupo de compradores é sua opinião sobre a importância dos descontos nas compras com sacolas ecológicas, a frequência com que vão às compras e a cor do produto. A tabela 4 mostra o uso do método *backward* no desenvolvimento de um modelo para a sacola ecológica de tamanho pequeno ao preço de R\$ 4,00.

Tabela 4 – Informações sobre os modelos gerados para sacolas ecológicas de tamanho pequeno ao preço de R\$ 4,00

Modelos	AIC	Redução do AIC	Pseudo-R ²	p-valor do modelo	Variável a ser excluída	Maior p-valor dos coeficientes das variáveis
Modelo 1	90,26	-	0,379	0,5621	escol	0,9736
Modelo 2	88,26	-2	0,379	0,4952	padaria	0,8596
Modelo 3	86,29	-1,97	0,378	0,4291	tamanho	0,813
Modelo 4	87,13	0,84	0,324	0,558	-	-

Pode-se perceber que o p-valor do modelo foi alto. Nesse caso, somente não rejeitaríamos a hipótese de que há um modelo que relacione a variável dependente com as variáveis independentes se considerássemos um nível de significância superior a 0,4291. Ou seja, ao afirmarmos que existe um modelo a probabilidade de errarmos é de 42,91%. Segundo o critério adotado (nível de significância igual a 0,2) não poderíamos considerar esse modelo aceitável.

Para um nível de significância de 0,16 rejeitaríamos a hipótese de que há um modelo. Isso pode ter ocorrido porque as opiniões e características das pessoas que pagariam R\$ 4,00 pelas sacolas de modelo pequeno não são significativamente distintas daquelas pessoas que não pagariam o preço do produto. Uma outra causa possível seria o fato de as características e opiniões dos “compradores” não serem bem definidas, ou seja, apesar de pertencerem ao grupo de pessoas que pagariam o valor do produto os indivíduos apresentam opiniões e características muito distintas.

A tabela 5 mostra algumas informações relacionadas à aplicação do método *stepwise backward* para uma sacola de tamanho médio ao preço de R\$ 5,00.

Tabela 5 – Informações sobre os modelos gerados para sacolas ecológicas de tamanho médio a R\$ 5,00

Modelos	AIC	Redução do AIC	pseudo-R ²	p-valor do modelo	variável a ser excluída	p-valor da variável a ser excluída
Modelo 1	60,5	-	0,689	0,0626	resistência	1
Modelo 2	58,5	-2	0,689	0,0458	cor	0,9311
Modelo 3	56,51	-1,99	0,689	0,0326	shopping	0,8773
Modelo 4	54,54	-1,97	0,688	0,023	tamanho	0,7563
Modelo 5	52,67	-1,87	0,686	0,0158	consc_amb	0,6203
Modelo 6	50,93	-1,74	0,681	0,0112	supermercados	0,5863
Modelo 7	60,12	9,19	0,456	0,1733	-	-

Para a sacola de tamanho médio a R\$ 5,00 obteve-se um modelo com um pseudo-R² igual a 0,681. O fato do p-valor ter sido igual a 0,0112 mostra que podemos rejeitar a hipótese que todos coeficientes do modelo são iguais a zero para um nível de significância de 0,16, ou seja, existe um modelo que relaciona as variáveis independentes com a variável dependente. O quadro 2 mostra mais informações sobre esse modelo.

Quadro 2 – Informações a cerca dos coeficientes das variáveis independentes e do intercepto do modelo obtido para sacolas de tamanho médio a R\$ 5,00

	Coeficiente	Erro padrão	Wald Z	P-valor
Intercepto	-1,1775	30,0855	-0,04	0,9688
fx_etaria	-3,1973	1,8999	-1,68	0,0924
est_civil	2,1262	2,8536	0,75	0,4562
escol	3,7921	2,9742	1,28	0,2023
renda	1,649	0,9933	1,66	0,0969
num_médio	4,8911	2,2813	2,14	0,032
supermercados	-16,0927	29,5747	-0,54	0,5863
padaria	3,4314	2,4589	1,4	0,1629
livraria	4,3506	4,1194	1,06	0,2909
frequência	-0,8898	0,8895	-1	0,3171
tam_ideal	3,2867	2,7457	1,2	0,2313
sexo	1,4748	1,638	0,9	0,368
descontos	2,8929	3,1798	0,91	0,3629
obrig	-4,1718	3,873	-1,08	0,2814
modelo	4,1578	2,132	1,95	0,0512

Para um nível de significância de 0,16, apenas os coeficientes das variáveis *fx_etária*, *renda*, *num_médio* e modelo poderiam ser considerados significativamente diferentes de zero. Somente poderíamos considerar todos os coeficientes estatisticamente significativos se adotássemos um nível de significância superior a 0,5863. Para o nível de significância adotado neste trabalho o modelo não pode ser considerado aceitável.

O p-valor alto para maior parte dos coeficientes pode ter ocorrido porque as opiniões e características podem ter sido suficientemente bem definidas e distintas entre o grupo de “compradores” e o grupo de “não compradores” apenas para as variáveis que apresentaram coeficientes significativos.

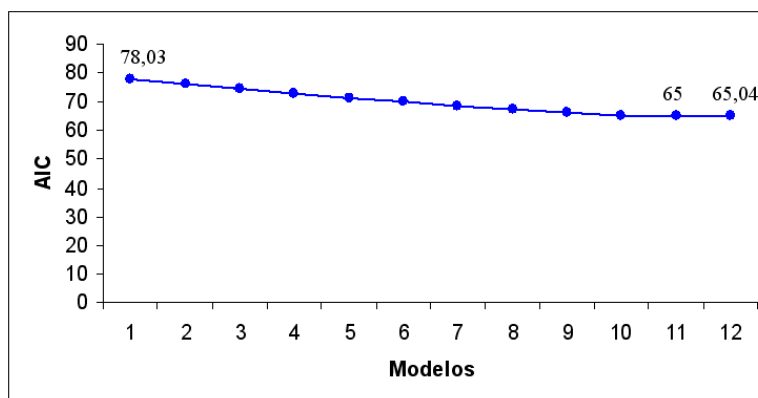
Ao utilizar o método *backward* para o modelo de sacolas de tamanho grande a um preço de R\$ 10,00 obteve-se os resultados mostrados na tabela 6.

Tabela 6 – Informações sobre os modelos gerados para sacolas ecológicas de tamanho grande ao preço de R\$ 10,00

Modelos	AIC	Redução do AIC	pseudo-R ²	p-valor do modelo	variável a ser excluída	p-valor da variável a ser excluída
Modelo 1	78,03	-	0,584	0,0565	est_civil	0,7592
Modelo 2	76,12	-1,91	0,582	0,0421	renda	0,5096
Modelo 3	74,57	-1,55	0,575	0,0336	freqüência	0,4807
Modelo 4	73,08	-1,49	0,568	0,027	cor	0,5609
Modelo 5	71,42	-1,66	0,562	0,0201	tamanho	0,4056
Modelo 6	70,15	-1,27	0,551	0,0165	supermercados	0,4966
Modelo 7	68,65	-1,5	0,543	0,0124	descontos	0,486
Modelo 8	67,17	-1,48	0,535	0,0091	fx_etária	0,3031
Modelo 9	66,29	-0,88	0,517	0,008	obrig	0,351
Modelo 10	65,26	-1,03	0,501	0,0066	sexo	0,2036
Modelo 11	65	-0,26	0,472	0,007	tam_ideal	0,1675
Modelo 12	65,04	0,04	0,437	0,0082	-	-

Como o modelo 12 apresentou um AIC superior ao do modelo 11 o melhor modelo pelo método *backward* foi o modelo 11. Na figura 10 é possível perceber que à medida que as variáveis menos significativas são retiradas obtém-se modelos melhores, de acordo com o AIC.

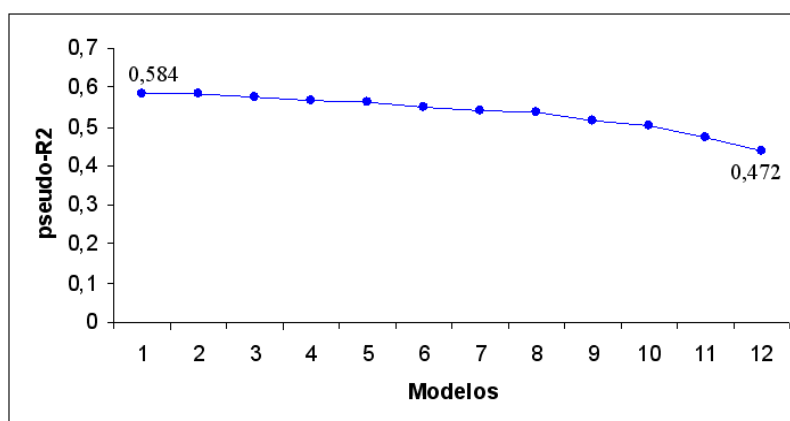
Figura 10 - Comportamento do AIC com a retirada de variáveis menos significativas nos modelos para sacolas de tamanho grande a um preço de R\$ 10,00



Para as sacolas ecológicas de tamanho grande ao preço de R\$ 10,00 é possível obter um modelo (modelo11) com um pseudo- R^2 igual a 0,472, o que significa que o modelo é capaz de explicar 47,2% da variação da chance relacionada a variável *comprag10*. O p-valor é igual a 0,007, o que indica que existe um modelo que relaciona a variável dependente *comprag10* com as variáveis independentes.

A figura 11 mostra como varia o pseudo- R^2 à medida que a variável menos significativa de cada modelo é retirada. Perceba que o pseudo- R^2 é reduzido, ou seja, o poder explicativo diminui com a retirada das variáveis.

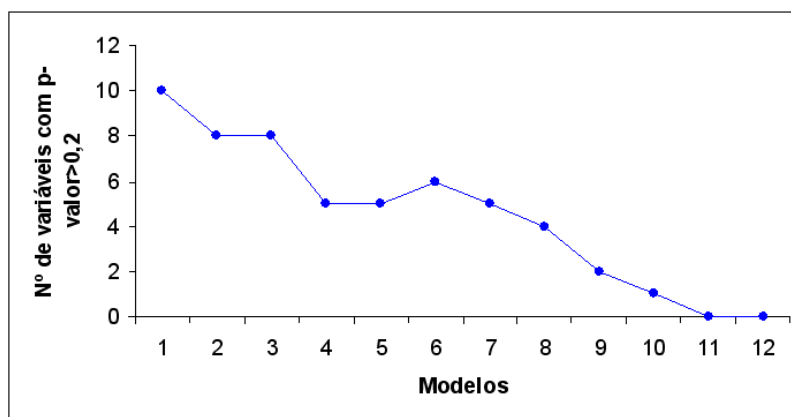
Figura 11 - Comportamento do pseudo-R2 com a retirada das variáveis menos significativas dos modelos para sacolas de tamanho grande a um preço de R\$10,00



Assim como foi observada no modelo selecionada para sacola de tamanho médio ao preço de R\$ 7,00, apesar da redução do poder explicativo, a complexidade dos modelos também é reduzida uma vez que o número de variáveis diminui.

Observa-se também uma tendência de obter modelos com um menor número de variáveis cujos coeficientes não são estatisticamente significativos, uma vez que apresentam p-valor superior a 0,2. A redução somente não foi observada ao retirar a variável *tamanho* do modelo 5. Pode-se verificar essa observação no gráfico da figura 12.

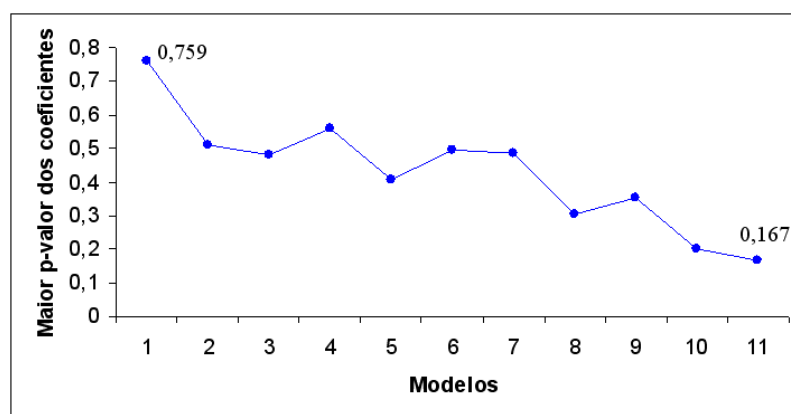
Figura 12 - Comportamento do número de variáveis com p-valor > 0,2 com a retirada de variáveis dos modelos para sacolas de tamanho grande a um preço de R\$ 10,00



Foi verificado também, até chegar ao modelo selecionado, que, em geral, à medida que são retiradas as variáveis menos significativas a tendência é que sejam obtidos modelos que exigem um nível de significância cada vez menor para julgar todos coeficientes dos modelos como significativos, ou seja, o maior p-valor dos coeficientes de cada modelo tende a ser cada vez menor.

No entanto, a redução do maior p-valor não ocorreu ao retirar as variáveis *frequência*, *tamanho* e *fx_etária*, respectivamente, dos modelos 3, 5 e 8. Pode-se visualizar essas observações e as do parágrafo anterior na figura 13.

Figura 13 - Comportamento do maior p-valor dos coeficientes com a retirada de variáveis dos modelos para sacolas de tamanho médio a um preço de R\$ 10,00

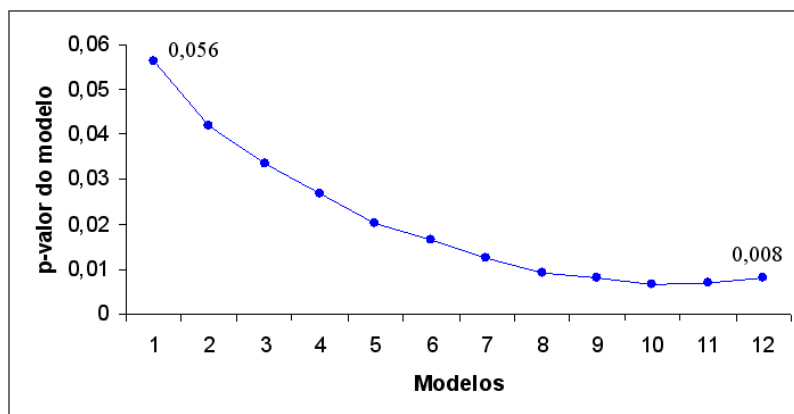


Outras informações acerca dos outros 12 modelos obtidos para as sacolas de tamanho grande ao preço de R\$ 10,00 são mostradas na tabela 7. É possível perceber que os modelos 11 e 12 não apresentam variáveis com p-valor superior a 0,2, ou variáveis que não são estatisticamente significativas. Os coeficientes do intercepto desses modelos também podem ser considerados significativos para o nível de significância adotado.

Há também uma tendência de redução do p-valor do modelo, quando as variáveis menos significativas são retiradas dos modelos. Como foi declarado anteriormente, isso equivale a afirmar que à medida que a variável menos significativa de cada modelo é retirada obtém-se modelos onde é cada vez mais difícil negar a relação entre a chance de ocorrência do evento estudado e as variáveis independentes.

O gráfico presente na figura 14 mostra a tendência de redução do p-valor.

Figura 14 - Comportamento do p-valor do modelo com a retirada de variáveis nos modelos para sacolas de tamanho grande a um preço de R\$ 10,00



A tabela 7 mostra o número de variáveis o número total de variáveis e o número de variáveis cujos coeficientes apresentam p-valor superior a 0,2 para cada um dos 12 modelos. Nessa tabela é também mostrado o menor p-valor dos coeficientes dos modelos e o p-valor do intercepto.

O p-valor do intercepto é importante quando se utiliza o modelo para fazer previsões de maneira mais rigorosa. Valores maiores do p-valor do intercepto indicam que se deve ter mais cautela ao dá atenção excessiva aos números obtidos a partir do modelo.

O menor p-valor dos coeficientes pode ser utilizado juntamente com o maior p-valor destes coeficientes para formar um intervalo onde estão os p-valores de todos os coeficientes do modelo. Além disso, indica o p-valor da variável com maior significância estatística do modelo.

Tabela 7 – Outras informações sobre os modelos gerados para sacolas ecológicas de tamanho grande ao preço de R\$ 10,00

Modelos	Número de variáveis	Nº de variáveis com p-valor do coeficiente maior que 0,20	Menor p-valor dos coeficientes das variáveis	p-valor do intercepto
Modelo 1	19	10	0,0279	0,0988
Modelo 2	18	8	0,0279	0,075
Modelo 3	17	8	0,0223	0,0648
Modelo 4	16	5	0,0239	0,0478
Modelo 5	15	5	0,0268	0,0455
Modelo 6	14	6	0,0298	0,0576
Modelo 7	13	5	0,0326	0,0382
Modelo 8	12	4	0,0161	0,0316
Modelo 9	11	2	0,0161	0,0211
Modelo 10	10	1	0,0179	0,0151
Modelo 11	9	0	0,0244	0,0201
Modelo 12	8	0	0,0271	0,0127

O quadro 3 exibe mais informações sobre o modelo selecionado (modelo 11) para as sacolas grandes ao preço indicado.

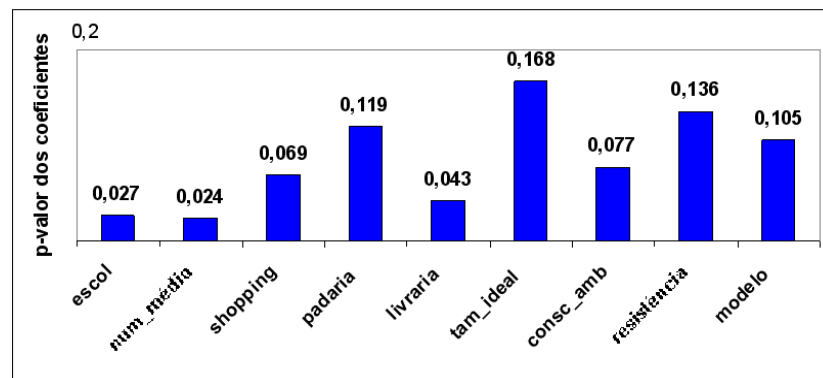
Quadro 3 – Informações a cerca dos coeficientes das variáveis independentes e do intercepto do modelo obtido para sacolas de tamanho grande a R\$ 10,00

	Coeficiente	Erro padrão	Wald Z	P-valor
Intercepto	-15,02	64,608	-2,32	0,0201
escol	2,349	1,0649	2,21	0,0274
num_médio	1,674	0,7436	2,25	0,0244
shopping	-2,514	1,3821	-1,82	0,0689
padaria	1,404	0,9016	1,56	0,1193
livraria	3,911	1,9358	2,02	0,0433
tam_ideal	1,342	0,9722	1,38	0,1675
consc_amb	1,638	0,9246	1,77	0,0765
resistência	6,534	4,3804	1,49	0,1358
modelo	1,521	0,9381	1,62	0,1048

O quadro 3 mostra que se adotássemos um nível de significância de 0,16 não poderíamos rejeitar a hipótese de que o coeficiente da variável *tam_ideal* é igual a zero, ou seja, não poderíamos declarar que o coeficiente dessa variável é significativamente diferente de zero. Para fazer essa afirmação precisaríamos adotar um nível de significância igual a 0,17. Como o critério adotado aqui considera um nível de significância de 0,2, o modelo pode ser julgado como aceitável.

O p-valor do coeficiente de cada uma das variáveis pode ser encontrado também no gráfico da figura *14. A linha superior do gráfico indica o limite 0,2 para que os coeficientes sejam considerados significativos.

Figura 15 - p-valor dos coeficientes do modelo para sacolas de tamanho grande a um preço de R\$ 10,00



As expressões (37), (38) e (39) mostram formas diferentes de escrever a equação do modelo selecionado (modelo 11).

$$\ln\left(\frac{p}{1-p}\right) = -15,02 + 2,349 * escol + 1,627 * num_médio - 2,514 * shopping + 1,404 * padaria + 3,911 * livraria + 1,342 * tam_ideal + 1,638 * consc_amb + 6,534 * resistência + 1,521 * modelo = \rho^2 \quad (37)$$

$$\frac{p}{1-p} = e^{\rho^2} \quad (38)$$

$$p = \frac{1}{1 + e^{-\rho^2}} \quad (39)$$

A tabela 8 mostra o valor incremental da *odds ratio* para cada variável do modelo.

Tabela 8 – Valores da *odds ratio* para cada variável do modelo para as sacolas de tamanho grande ao preço de R\$ 10,00

Variáveis	odds ratio
escol	10,48
num_médio	5,33
shopping	0,08
padaria	4,07
livraria	49,95
tam_ideal	3,83
consc_amb	5,16
resistência	688,15
modelo	4,58

Novamente, a equação do modelo será escrita de maneira que seja possível visualizar nela o valor da *odds ratio*. A *odds ratio* é mostrada entre parênteses na equação (40).

$$\frac{p}{1-p} = e^{-15,02} * (e^{2,349})^{escol} * (e^{1,627})^{num_medio} * (e^{-2,514})^{shopping} * (e^{1,404})^{padaria} * (e^{3,911})^{livraria} * (e^{1,342})^{tam_ideal} * (e^{1,638})^{cons_amb} * (e^{6,534})^{resistência} * (e^{1,521})^{modelo} \quad (40)$$

Baseando-se nos valores da *odds ratio*, mostrados de maneira mais clara na tabela 8, é possível afirmar que ter nível superior completo (observar a *odds ratio* da variável *escol*) eleva a chance de um indivíduo pertencer ao grupo de pessoas que está disposto a comprar a sacola grande por R\$10,00 em $e^{2,349} = 10,48$ vezes.

Observando os valores das *odds ratio* pode-se concluir que uma das variáveis que melhor distinguem os “compradores” dos “não compradores”, considerando a sacola grande a R\$ 10,00, é a variável *resistência*, uma vez que esta apresenta a maior *odds ratio*. Outras duas variáveis importantes para fazer essa distinção são *livraria* e *escol*, que também apresentaram uma *odds ratio* relativamente alta.

O que é possível perceber diante dos modelos selecionados a partir da aplicação do método *backward* é que os modelos a serem obtidos dependem do preço e da modalidade do produto (neste caso o tamanho). Entre as várias diferenças entre os modelos obtidos foi possível observar que no primeiro modelo selecionado a variável *resistência* não estava presente, enquanto neste ultimo modelo selecionado esta era uma das variáveis mais

importantes. No primeiro modelo escolhido a variável *descontos* era uma das mais importantes enquanto neste ultimo essa variável não estava presente.

Essa dependência do modelo de regressão em relação ao preço e a modalidade da sacola ocorre provavelmente porque quando a empresa vende uma modalidade de produto diferente ou vende um mesmo produto numa faixa de preço distinta ela estará agradando a um público diferente. Este grupo apresenta preferências, opiniões e características diferentes daquele que é mais bem satisfeito por uma sacola de tamanho e faixa de preço diferente.

A dependência do modelo de regressão em relação ao preço e a modalidade também é verificada quando não se encontra um modelo significativo estatisticamente para distinguir segmentos para um determinado modelo de sacola a um preço particular.

Neste trabalho identificaram-se dois casos correspondentes a “modalidade de produto/preço”, onde não foi possível obter um modelo estatisticamente significativo. Isso pode ter ocorrido porque as opiniões, preferências e características do grupo de “compradores” e “não compradores” não são suficientemente distintas e/ou bem definidas para distinguir os grupos baseando-se nelas.

4.3 Informações sobre o mercado consumidor de sacolas ecológicas

Pode-se utilizar os dois melhores modelos obtidos neste trabalho para identificar as preferências das pessoas que apresentam maior probabilidade de serem compradoras do produto. Fazendo isso é possível orientar quem deseja vender o produto a tomar decisões que tenham maior probabilidade de satisfazer as necessidades das pessoas que realmente estão interessados em comprar as sacolas.

Se a opção de uma empresa de sacolas ecológicas é vender uma sacola de tamanho médio a R\$ 7,00 então deve saber que:

- Incentivar a garantia de desconto aos que utilizam sacolas ecológicas nas compras é uma forma de agir favoravelmente a uma das opiniões que melhor caracterizam os “compradores”. Essa afirmação é baseada no fato da variável *descontos* ter apresentado a maior *odds ratio*.
- Indivíduos que vão com maior frequência às compras têm maior probabilidade de comprarem o produto. Dessa maneira, ao colocar as sacolas ecológicas em locais onde as pessoas vão com frequência as compras é uma outra maneira de buscar pessoas com uma probabilidade maior de comprar o produto. Isso pode ser declarado com base no coeficiente da variável *frequência* superior a 1.

- Fabricar sacolas de cor preta e com estampas também contribui para satisfazer mais a pessoas que têm uma maior probabilidade pagar o valor do produto. Vender sacolas brancas é uma forma de agradar mais a indivíduos com uma probabilidade maior de não pagarem o preço fixado.
- Sendo a *odds ratio* da variável *fx_etária* inferior a 1, pode-se afirmar que pessoas na faixa etária entre 20 e 30 anos têm uma probabilidade maior de pagarem o valor fixado. Então, modelos de sacolas mais voltados para essa faixa etária tende a agradar mais a indivíduos com maior probabilidade de pagar o valor fixado.
- A *odds ratio* da variável *escol* também foi menor que 1. Desse modo, pessoas com nível superior completo têm uma probabilidade menor de serem compradoras do que aquelas com nível superior incompleto. Como a pesquisa foi feita principalmente com pessoas com nível superior completo e estudantes universitários, isso mostra que os estudantes têm maior probabilidade de comprarem.
- Se a *odds ratio* da variável *sexo* é inferior a 1, podemos afirmar que indivíduos do sexo feminino tem maior probabilidade de pagar o valor da sacola do que as do sexo masculino.
- As pessoas que declaram que a proibição da legislação seria um importante incentivo a compra de sacolas ecológicas tem uma probabilidade menor de serem “compradores” do que aqueles que consideram que essa proibição é pouco importante.
- Fabricar modelos de sacolas com espaço interno grande é gastar mais recursos com algo que tende a agradar mais a pessoas com uma probabilidade baixa de pagar o valor das sacolas ecológicas.
- As pessoas que declaram que usariam a sacola ecológica em panificadoras têm uma probabilidade menor de comprar as sacolas ecológicas do que aquelas que declaram que não utilizariam.

A sacola ecológica de tamanho grande a R\$ 10,00 alcança um público com características e opiniões diferentes daquele da sacola de tamanho médio a R\$ 7,00. Algumas informações sobre esse público, proporcionado pelo modelo de regressão logística, são listadas abaixo:

- Pessoas com nível superior de escolaridade têm maior probabilidade de pagar o valor fixado para esse tipo de sacola ecológica do que pessoas sem nível superior.
- Indivíduos que fazem apenas compras mensais grandes têm uma probabilidade maior de pagar o valor dessa sacola do que aqueles que fazem pequenas compras.

- Pessoas que costumam fazer compras com um maior número de itens têm maior probabilidade de comprar esse tipo de sacola ecológica
- As pessoas que acham que utilizariam sacolas em um shopping tem menor probabilidade de pagarem o valor fixado para esse tipo de sacola.
- Ao contrário do que ocorre com o público das sacolas de tamanho médio a R\$ 7,00, a pessoa que usaria as sacolas grandes em padarias tem maior probabilidade de pagar o valor fixado do que aquelas que declaram que não utilizariam em compras nesse local. O mesmo é válido para o uso nas compras em livrarias.
- Aqueles que declaram que a sua consciência ambiental seria um fator que influenciaria a comprar o produto tem uma probabilidade maior de pagar o preço da sacola do que aqueles que declaram o contrário.
- Aqueles que declaram que a boa resistência da sacola ecológica é importante, tem uma probabilidade maior de comprar a sacola do que aqueles que declaram que este fator é pouco importante. O mesmo é válido para o design da sacola. Mas, a importância dada à resistência é o fator que mais influencia na probabilidade de alguém ser um “comprador”, uma vez que possui a maior *odds ratio* dentre as variáveis do modelo.

4.4 Um exemplo da utilização do modelo como auxílio à segmentação

Vamos considerar que uma empresa de sacolas ecológicas deseja vender a sacola de tamanho médio a R\$ 7,00. Uma atitude importante que deve ser tomada antes da venda do produto é buscar conhecer os consumidores desse produto, pois os conhecendo melhor a empresa poderá planejar de maneira mais eficiente formas de satisfazê-los.

A empresa não conseguirá agradar a todos da região onde pretende vender o produto por isso, é racional que ela tente atender apenas uma parcela do público que possua algumas características e opiniões semelhantes. Ao fazer isso, a empresa estará fazendo uma segmentação do mercado.

A empresa deverá escolher um grupo de consumidores para atender e que posteriormente deverá conhecê-lo. Mas, de que maneira a empresa deverá escolher esse grupo de consumidores? Uma forma de fazer isso seria vender o produto para pessoas que estão, de fato, interessadas nele e, por isso, dispostas a pagar o preço fixado pela empresa.

Para conhecer as pessoas que estão interessadas no produto, a empresa poderia aplicar a regressão logística aos dados obtidos por meio de uma pesquisa de mercado semelhante à utilizada neste trabalho. Fazendo isso ela poderia identificar características e opiniões que estão relacionadas com o desejo de alguém comprar ou não as sacolas. Vamos utilizar o modelo obtido neste trabalho para exemplificar esse auxílio à segmentação do mercado.

O modelo encontrado tem nove variáveis relacionadas com a intenção da compra. Vamos utilizar apenas cinco delas (*freqüência*, *fx_etária*, *sexo*, *descontos* e *cor*) para exemplificar a escolha de um grupo de consumidores que tem uma probabilidade baixa de pagar o valor da sacola ecológica e um grupo que tem uma probabilidade alta.

Dentre aqueles com alta probabilidade de pagarem o valor do produto, é preferível que deixemos de atender a anseios de grupos com probabilidade maior de pagar do que os anseios daqueles com menor probabilidade de pagar, uma vez que é mais fácil perder um cliente que tenha uma menor probabilidade de comprar. Assim, para as demais variáveis que não serão manipuladas devemos trabalhar com valores mais desfavoráveis possíveis (*padaria*=1, *obrig*=1, *escol*=1 e *tamanho*=1).

Vamos considerar que ações da empresa sejam voltadas para satisfazer a um público jovem (entre 20 e 30 anos), do sexo feminino, que tem preferência pelas sacolas com estampas, faz compras semanalmente e considera os descontos importantes (*fx_etária*=1, *sexo*=0, *freqüência*=3 e *descontos*=1). Nesse caso, a empresa estará voltada para pessoas que tem uma probabilidade de pagar o valor das sacolas igual a aproximadamente 90%.

Consideremos agora que a empresa opta por atender a um público entre 30 e 40 anos, do sexo masculino, que faz compras uma vez no mês, tem preferência por sacolas de cor preta e também considera o desconto um incentivo importante (*fx_etária*=2, *sexo*=1, *freqüência*=1 e *descontos*=1). Esse público tem uma probabilidade de menos de 1% de comprar a sacola.

Os resultados do exemplo descrito nos dois últimos parágrafos são mostrados no quadro 4.

Quadro 4 - Informações acerca dos fatores e valores da probabilidade dos segmentos A e B estarem dispostos a comprar as sacolas de tamanho médio a R\$ 7,00

Grupos	faixa etária	sexo	freqüência	descontos	cor da sacola	probabilidade
A	20 a 30	feminino	semanalmente	influenciaria	estampado	0,9
B	31 a 40	masculino	mensalmente	influenciaria	preto	0,01

Se a empresa direcionar seus esforços para atender a um público que apresenta uma probabilidade inferior a 1% de pagarem o preço da sacola estará assumindo um risco alto de

seu produto não ser vendido. A empresa não estará agradando ao público mais interessado no produto. Estará agradando apenas aqueles pouco interessados. Por isso, segundo o modelo é mais aconselhável que a empresa opte pelo segmento A.

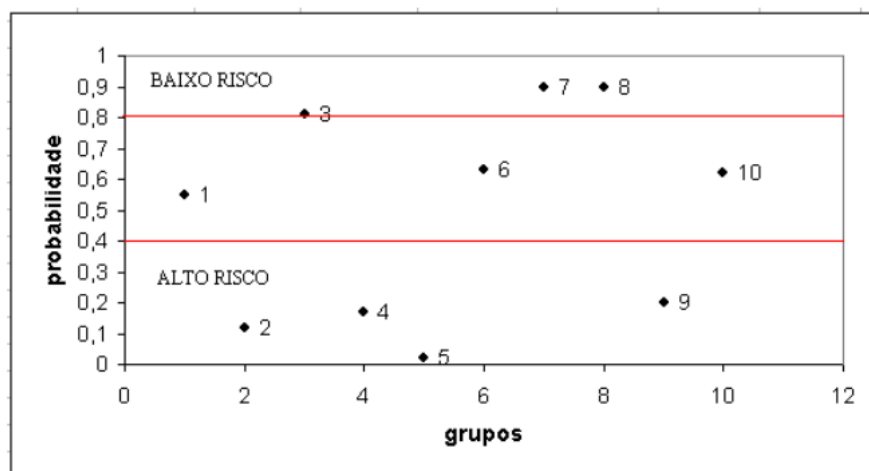
Como um segundo exemplo, considere que a empresa, que fabrica sacolas ecológicas, lista 10 segmentos, com características específicas relacionadas as quatro variáveis presentes no exemplo anterior. Esses segmentos são mostrados no quadro 5 juntamente com suas características.

Quadro 5 – Informações acerca dos fatores para os segmentos enumerados de 1 a 10

Grupos	faixa etária	sexo	freqüência	descontos	cor da sacola
1	31 a 40	feminino	semanalmente	influenciaria	estampado
2	31 a 40	masculino	semanalmente	influenciaria	preto
3	20 a 30	feminino	semanalmente	não influenciaria	preto
4	20 a 30	masculino	semanalmente	não influenciaria	branco
5	acima de 40	masculino	diariamente	influenciaria	branco
6	20 a 30	feminino	semanalmente	influenciaria	estampado
7	20 a 30	masculino	quinzenalmente	influenciaria	estampado
8	20 a 30	feminino	diariamente	influenciaria	bege
9	20 a 30	feminino	diariamente	não influenciaria	branco
10	31 a 40	masculino	diariamente	influenciaria	estampado

As probabilidades dos 10 grupos estarem dispostos a pagar o valor de R\$ 7,00 pela sacola de tamanho médio foram colocadas no gráfico presente na figura 16.

Figura 16 – Gráfico com os valores das probabilidades dos segmentos de 1 a 10 estarem dispostos a pagar R\$ 7,00 pelas sacolas de tamanho médio



O gráfico da figura 16 é dividido em três regiões onde se destacou a área de alto risco de escolher um segmento que não comprará a sacola média por R\$ 7,00, e a área de baixo risco. Baseando-se no modelo, seria mais aconselhável optar por atender aos segmentos 3, 7 e 8, e evitar escolher os segmentos 2, 4, 5 e 9.

O modelo poderia auxiliar nas decisões relacionadas à segmentação de duas formas. Poderia ser útil para fazer uma segmentação de acordo com o risco de escolher um segmento que não estaria disposto a pagar o valor do produto, ou poderia ser utilizado após escolher os segmentos segundo um critério qualquer para avaliá-los.

5 CONSIDERAÇÕES FINAIS

O presente trabalho foi baseado na aplicação da regressão logística, usando o método *stepwise backward*, a dados de uma pesquisa de mercado sobre sacolas ecológicas realizada na cidade de João Pessoa. A pesquisa contou com a participação de 57 indivíduos que opinaram sobre várias questões relacionadas ao produto, citaram suas características e declararam o valor que estariam dispostos a pagar por cada modelo da sacola ecológica.

O objetivo era então obter modelos de regressão logística, para diferentes tamanhos de sacola e para diferentes preços, que relacionassem a intenção de compra das pessoas às suas características e opiniões. Obtendo esse modelo seria possível afirmar quais características e opiniões são importantes para diferenciar as pessoas que têm uma alta probabilidade de pagar pela sacola daquelas que têm uma probabilidade baixa. Além disso, seria possível indicar a influência de cada uma dessas opiniões e características, sobre a probabilidade de alguém está disposto a aceitar o preço fixado para sacola ecológica.

Com a pesquisa percebeu-se que as opiniões e características que distinguem as pessoas que pagariam o valor das sacolas daquelas que não o pagariam mudam para diferentes preços e modelos. Assim, na região estudada, ao ser oferecido um determinado modelo de sacola ecológica numa faixa de preço específica satisfaz-se a um público específico. Quando um outro modelo numa outra faixa de preço é oferecido satisfaz-se a um outro público, com opiniões, características e preferências diferentes.

Tentamos identificar modelos relativamente consistentes para quatro situações com modelos de sacola e preços específicos. Em duas dessas situações (sacola de tamanho médio a R\$ 5,00 e sacola de tamanho pequeno a R\$ 4,00) não foram obtidos modelos consistentes. Para a sacola de tamanho médio a R\$ 7,00 foi obtido um modelo com pseudo- R^2 igual a 0,493 e p-valor igual a 0,002, enquanto para a sacola de tamanho grande a R\$ 10,00 foi obtido um modelo com o pseudo- R^2 igual a 0,472 e p-valor de 0,007.

Os modelos ainda foram utilizados para mostrar as opiniões e características que aumentam e as que diminuem a probabilidade de alguém está disposto a comprar as sacolas ecológicas. Por fim, foi mostrado um exemplo do uso do modelo como ferramenta auxiliar a segmentação de mercado.

Vale chamar atenção para o fato de os modelos encontrados apresentarem problemas. Um deles é os valores baixos do pseudo- R^2 que mostram que os modelos não são capazes de explicar uma parcela grande da variação das suas variáveis dependentes. Um outro problema

encontrado foi o p-valor relativamente alto dos coeficientes de algumas variáveis que para um dos modelos chegou próximo de 0,17.

O modelo também não é capaz de indicar algumas variáveis que são fundamentais para decisão da compra das sacolas ecológicas. Se uma opinião é igualmente importante, tanto para as pessoas que estão dispostas a pagar o preço fixado quanto para aquelas que não estão dispostas a pagar esse preço, o modelo não irá identificar essa opinião. Desse modo, é essencial que também seja feita uma análise explanatória dos dados para identificar opiniões e características que são importantes para compra.

Apesar dos problemas identificados, o baixo p-valor dos modelos mostra que há uma relação entre as variáveis independentes e as variáveis dependentes. Talvez, fazendo outras pesquisas, no mesmo sentido, com amostras maiores obtenha-se resultados mais promissores.

6 REFERÊNCIAS

- AGRESTI, A. **Categorical Data analysis**. Nova Jersey, 2 ed. 2002. Editora John Wiley and Sons.
- ARAÚJO, M.V. P.; SILVA, S. M. **Pesquisa de mercado do comportamento do consumidor de produtos alimentícios de marcas próprias em uma empresa varejista mundial em Natal/RN**. Disponível em:
http://www.simpep.feb.unesp.br/anais/anais_13/artigos/911.pdf. Acesso em: 11 dez.2009.
- AZEVEDO, G. C. I. **Saiba mais**. Disponível em:
<http://www.scribd.com/doc/939526/Pesquisa-de-Mercado?autodown=pdf>. Acesso em: 22 mai.2009.
- BARBIERO, C. C. de M. **Séries temporais: um estudo de previsão para a receita operacional da ect – empresa brasileira de correios e telégrafos**. 2003. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina. 2003.
- 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.
- CAMBOIM, L. G.; LIMA, A. S. **Pesquisa de mercado para empresas do sub-setor de edificações**. Disponível em:
http://www.abepro.org.br/biblioteca/ENEGEP1999_A0632.PDF. Acesso em: 10 dez.2009.
- CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. **Análise multivariada: para cursos de administração, ciências contábeis e economia**. São Paulo:Atlas, 2009.
- CYMROT, R.; OLIVEIRA, Y. M. B. M.; YAMADA, A. M.; AQUINO, R. B. **Relatando a experiência de aplicação de um software estatístico em pesquisa interdisciplinar realizada por alunos de um Curso de Engenharia Elétrica**. Disponível em:
http://www.iiis2009.org/cisci/program/reg_waiting.asp?id_program=2&tx_login=msanchez&vc=2. Acesso em:5 dez.2009.
- CZINKOTA, M.R *et al.* **Marketing: as Melhores Práticas**. Porto Alegre: Ed. Bookman, 2001.
- DEVORE, J. L.; **Probabilidade e estatística: para engenharia e ciências**. São Paulo: Pioneira Thomson Learnig, 2006.

FERREIRA, F. H. G. **Segmentação de mercado**. Disponível em:

[http://www.biblioteca.sebrae.com.br/bds/BDS.nsf/D96040554BF9B03256D520059AE80/\\$File/310_1_Arquivos_segmercado.pdf](http://www.biblioteca.sebrae.com.br/bds/BDS.nsf/D96040554BF9B03256D520059AE80/$File/310_1_Arquivos_segmercado.pdf). Acesso em: 20 dez.2009

FIGUEIRA, C. V. **Modelos de Regressão Logística**. 2006. Dissertação (Mestrado em Matemática) – Instituto de Matemática, Universidade Federal do Rio Grande do Sul. 2006.

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

GUIMARÃES, R. V. **Uso de regressão logística para previsão de fechamento de operações financeiras: termo de moedas**. 2006. Dissertação (Mestrado em Engenharia de Produção) – Universidade de São Paulo. 2006.

HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**. 2 ed. New York: John Wiley & Sons, 2000.

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

LEITÃO, Miriam. **A queda do dólar e o problema das sacolas plásticas**. Disponível em: http://oglobo.globo.com/economia/miriam/post.asp?t=a_queda_do_dolar_o_problema_das_sacolas_plasticas&cod_Post=91791&a=73. Acesso em: 13 mai. 2009.

LOPES, I. J. **A segmentação de mercado como estratégia de marketing**. Disponível em: http://www.alealmeida.com/facul/felix_segmentacao.pdf. Acesso em: 10 dez.2009.

MAIA, M. L. S. **Análise de mercado para polvilho azedo produzido no APL de mandiocultura no município de oliveira – MG**. Disponível: <http://www.ufv.br/dep/engprod/TRABALHOS%20DE%20GRADUACAO/MARCIA%20LUIZA%20SOARES%20MAIA/AN%C3%81LISE%20DE%20MERCADO%20PARA%20POLVILHO%20AZEDO%20PRODUZIDO%20NO%20APL%20DE%20MANDIOCULTURA%20NO%20MUNIC%C3%8DPIO%20DE%20OLIVEIRA%20%E2%80%93%20MG.pdf>. Acesso: 02 jan.2009.

MANKIW, N. G. **Introdução a Economia**. 2ª ed. Ed.Campus. Rio de Janeiro, 2001.

MINUZZI, J.; SANTOS, P. C. F.; LEZANA, A. G. R.; CASAROTTO FILHO, N. **Intenção empreendedora em alunos de engenharia de produção: uma análise com o uso de regressão logística**. Disponível em:

http://www.abepro.org.br/biblioteca/ENEGEP2007_TR660482_9450.pdf. Acesso em:27 dez. 2009

MONTENEGRO, S. G. **Modelo de regressão em dados categóricos na área de ergonomia experimental**. 2009. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal da Paraíba. 2009.

PACAGNELLA JÚNIOR, A.C ; PORTO, G. S.; KANNEBLEY JÚNIOR, S.; SILVA, S. L.; BONACIM, C. A. G. **Obtenção de patentes na indústria do Estado de São Paulo: uma análise utilizando regressão logística**. Disponível em:
http://www.scielo.br/scielo.php?pid=S0103-65132009000200004&script=sci_arttext. Acesso em: 15 dez. 2009.

PENHA, R. N. **Um estudo sobre regressão logística binária**. Disponível em:
<http://www.epr.unifei.edu.br/PFG/producao2002/trabalhos/Renata.PDF>. Acesso em: 26 dez. 2009.

PINDYCK, R. S.; RUBINFELD, D. L. **Microeconomia**. 6ª ed. Ed.Pearson Prentice Hall, São Paulo, 2006.

ROCHA, Dêlcio. **Consciência ambiental dá tom no nicho de “ecobags”**. Disponível em:
<http://www.ambienteemfoco.com.br/?p=8104>. Acesso em: 03 mai. 2009.

SANT’ANNA, Â. M. O. **Método de orientação à modelagem de dados mensurados em proporção**. 2006. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal do Rio Grande do Sul. 2006.

SCHNEIDER D.; WUTTKE D.; FÜHR H.; SCHNEIDER, I. **Pesquisa de marketing: ferramenta indispensável para gestão das organizações**. Disponível em:
http://www.seifai.edu.br/artigos/Artigo_pesquisaMKT.pdf .Acesso em: 03 jan. 2010.

SEBRAE. **Fatores condicionantes e taxas de sobrevivência e mortalidade de micro e pequenas empresas do Brasil 2003-2005**. Disponível em:
http://www.sebraepr.com.br/FCKeditor/userfiles/file/BancodePesquisas%20/Mortalidade/Mortalidade_Empresas_2005.pdf. Acesso em: 26 dez.2009

SEBRAE/MG. **Fatores condicionantes e taxas de mortalidade das empresas no país.Relatório de pesquisa, 2004**. Disponível em:
http://www.sebraemg.com.br/arquivos/informativos/relatorio_pesquisa_mortalidade_minas.pdf. Acesso em: 26 dez.2009.

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.

TUSSI, A. C. S. **Pesquisa de mercado:** despesa ou investimento? (17/03/2008). Disponível em:

http://www.justale.com.br/visao/Pesquisa%20de%20Mercado_Investimento%20ou%20Despesa.pdf. Acesso em: 26 dez.2009.

APÊNDICE A - Resultados da pesquisa de mercado sobre sacolas ecológicas -Dados
utilizados na análise

Amostra	Variáveis						
	comprag14	comprag10	comprap4	compram5	compram7	consc_amb	descontos
1	0	1	0	1	0	0	1
2	0	1	0	1	1	1	0
3	0	1	0	1	1	1	1
4	0	1	0	1	1	1	0
5	0	1	0	1	1	0	1
6	0	0	0	1	1	0	0
7	0	1	1	1	1	1	0
8	1	1	1	1	0	1	0
9	0	0	0	0	0	1	0
10	0	1	1	1	0	1	0
11	0	1	0	1	1	1	0
12	0	1	0	1	1	1	0
13	1	1	1	1	0	1	0
14	0	0	0	0	0	1	0
15	0	1	0	1	1	1	0
16	0	1	0	1	1	0	1
17	0	1	0	1	1	1	1
18	0	1	0	1	1	1	0
19	0	0	0	0	0	1	1
20	0	1	1	1	0	1	0
21	0	1	0	1	1	1	0
22	0	0	0	0	0	1	0
23	0	0	0	0	0	1	0
24	1	1	1	1	1	0	0
25	0	1	0	1	1	0	0
26	0	0	0	0	0	0	0
27	0	1	0	1	1	0	0
28	1	1	1	1	0	1	0
29	1	1	1	1	0	1	0
30	0	1	0	1	1	1	0
31	0	0	0	1	1	1	0
32	0	1	0	1	1	1	1
33	0	1	1	1	1	1	1
34	0	0	0	1	1	0	1
35	0	1	0	1	0	1	1
36	1	1	1	1	0	0	1
37	0	0	0	0	0	1	0
38	0	1	0	1	1	1	1
39	1	1	1	1	1	1	0
40	0	1	0	1	1	1	1
41	0	1	0	1	1	1	1
42	0	0	0	1	1	1	0
43	0	0	0	1	1	1	0
44	0	1	0	1	1	1	0
45	1	1	1	1	0	1	0
46	0	1	0	1	1	0	0
47	1	1	0	1	0	0	0

(Cont.)

48	0	0	0	1	1	0	1
49	1	1	1	1	0	0	1
50	0	0	0	0	0	0	1
51	0	0	0	1	1	1	0
52	0	0	0	0	0	0	1
53	0	1	0	1	1	1	0
54	1	1	1	1	0	0	1
55	0	1	0	1	1	1	0
56	0	1	1	1	0	1	1
57	1	1	1	1	0	1	1

APÊNDICE A (cont.)- Resultados da pesquisa de mercado sobre sacolas ecológicas -Dados
utilizados na análise

Amostra	Variável					
	obrig	resistência	tamanho	modelo	cor	num_médio
1	0	1	1	1	2	3
2	0	1	1	0	2	1
3	0	1	1	0	3	1
4	0	1	1	0	2	1
5	0	1	1	0	2	2
6	0	1	1	0	4	2
7	0	1	1	0	2	3
8	0	1	1	0	2	1
9	0	1	1	0	4	1
10	0	1	1	0	3	2
11	0	1	1	0	2	3
12	0	1	1	0	1	3
13	0	1	1	1	4	2
14	0	1	1	0	4	2
15	0	1	1	1	3	2
16	0	1	1	0	4	2
17	1	1	1	1	4	2
18	0	1	1	0	2	1
19	1	1	1	0	2	1
20	0	1	1	0	2	2
21	0	1	1	1	2	2
22	0	1	1	1	1	1
23	0	1	1	0	1	1
24	0	1	1	1	2	1
25	0	1	1	1	2	2
26	0	1	1	1	4	2
27	0	1	1	1	1	2
28	0	1	1	1	1	2
29	0	1	1	1	4	1
30	0	1	1	1	2	1
31	0	1	1	1	2	2
32	0	1	1	1	1	1
33	1	1	1	1	2	1
34	1	1	1	1	1	1
35	0	1	0	1	1	2
36	0	1	1	0	2	1
37	0	1	0	0	2	1
38	0	1	1	0	3	3
39	0	0	1	0	2	3
40	0	1	1	0	2	2
41	0	1	1	0	2	2
42	0	0	0	0	1	3
43	0	1	1	0	4	1
44	0	1	1	0	2	2
45	0	1	1	1	4	2
46	0	1	1	0	3	2
47	0	1	0	0	1	2

(Cont.)

48	0	1	1	0	1	2
49	0	1	1	0	3	2
50	0	1	1	0	3	1
51	0	1	1	1	2	2
52	0	1	1	0	1	2
53	0	1	1	0	4	2
54	0	1	1	0	2	2
55	0	1	1	0	2	2
56	0	1	1	1	2	1
57	0	1	1	1	2	2

APÊNDICE A (cont.)- Resultados da pesquisa de mercado sobre sacolas ecológicas -Dados utilizados na análise

Amostra	Variáveis						
	supermercados	shopping	padaria	livraria	frequência	tam_ideal	sexo
1	1	1	1	1	3	2	1
2	0	0	1	0	4	2	1
3	1	0	0	0	2	3	1
4	0	0	0	0	3	2	1
5	1	0	0	0	3	2	1
6	1	0	0	0	2	2	1
7	1	0	0	0	3	1	1
8	0	0	0	1	1	1	0
9	1	0	0	0	3	2	0
10	1	0	1	0	3	2	0
11	0	0	1	0	3	2	0
12	0	1	1	1	1	3	0
13	1	1	0	0	2	2	0
14	1	0	1	0	1	2	0
15	1	0	0	0	1	2	0
16	1	1	1	1	2	2	0
17	1	1	1	1	2	2	0
18	1	1	1	1	1	1	0
19	1	0	0	0	2	2	1
20	1	0	0	0	3	2	1
21	1	0	0	0	2	2	1
22	1	0	0	0	3	2	1
23	1	1	1	1	4	2	1
24	1	0	1	0	1	2	1
25	1	0	0	1	3	2	1
26	1	0	0	0	2	2	1
27	1	0	1	0	3	2	0
28	1	0	0	0	2	2	0
29	1	0	0	0	3	2	1
30	1	0	0	0	1	1	1
31	1	0	0	0	3	2	0
32	1	0	1	0	2	1	1
33	1	1	1	0	2	2	0
34	1	0	1	0	2	2	0
35	1	0	0	0	3	2	1
36	1	0	0	0	3	3	1
37	1	0	0	0	3	2	1
38	1	0	0	0	3	2	1
39	1	0	1	1	3	3	0
40	1	0	1	0	3	3	0
41	1	0	1	0	2	3	0
42	1	0	0	0	1	2	0
43	0	0	0	0	1	2	0
44	1	0	1	0	3	3	0
45	1	0	0	0	3	2	0
46	1	0	1	0	3	3	1
47	1	0	0	0	2	2	0
48	1	0	0	0	1	2	1

(Cont.)

49	0	0	1	1	3	2	1
50	1	0	0	0	3	2	1
51	0	1	0	0	1	2	0
52	1	0	0	0	3	2	0
53	1	1	1	1	3	2	1
54	1	0	1	0	3	3	1
55	1	0	0	0	3	3	0
56	1	0	1	0	3	2	0
57	1	0	1	0	3	2	0

APÊNDICE A (cont.)- Resultados da pesquisa de mercado sobre sacolas ecológicas -Dados utilizados na análise

Amostra	Variáveis			
	fx_etaria	est_civil	escol	renda
1	1	0	1	3
2	2	1	1	4
3	3	1	1	4
4	2	1	1	1
5	3	1	1	4
6	2	1	0	4
7	3	1	1	4
8	1	0	1	3
9	1	0	1	4
10	3	1	0	4
11	3	1	0	2
12	1	1	0	1
13	1	0	1	2
14	3	0	1	4
15	1	0	1	2
16	3	1	1	3
17	2	1	1	1
18	1	0	1	1
19	1	0	1	3
20	1	0	1	2
21	3	1	0	3
22	1	1	0	2
23	1	0	0	2
24	1	0	1	1
25	2	0	0	3
26	2	1	0	1
27	3	1	1	3
28	1	0	0	3
29	1	0	1	4
30	1	0	1	4
31	2	1	1	2
32	1	0	1	2
33	1	0	1	2
34	1	0	1	2
35	3	1	1	4
36	2	0	1	2
37	3	1	1	3
38	3	1	1	4
39	2	1	1	4
40	3	0	1	2
41	3	1	1	4
42	3	1	1	4
43	2	1	0	1
44	1	0	1	2
45	2	0	1	2
46	1	0	1	4
47	1	0	1	3
48	1	0	1	2

(Cont.)

49	1	0	1	3
50	3	1	1	1
51	3	0	1	4
52	3	1	1	3
53	2	1	0	3
54	2	1	1	2
55	2	0	1	4
56	1	0	1	2
57	1	1	1	3

APÊNDICE B - Questionário utilizado na pesquisa de mercado

1. O que levaria o (a) senhor (a) a utilizar nas suas compras sacolas ecológicas ao invés de sacolas plásticas?

- A. ☐ Futura obrigatoriedade (lei, decreto, etc.)
- B. ☐ Descontos nas compras
- C. ☐ Moda
- D. ☐ Consciência ambiental
- E. ☐ Outro _____

2. Dos atributos listados abaixo, como o (a) senhor (a) os classificaria ao comprar uma sacola ecológica? Utilize o grau de importância: (1) Muito importante; (2) Importante; (3) pouco importante.

- ☐ Resistência do material da sacola
- ☐ Modelo e cor da sacola
- ☐ Tamanho da sacola
- ☐ Conforto e praticidade ao transportar as compras
- ☐

3. Escolha a cor da sua preferência para o corpo e as alças da sacola ecológica:

	Corpo da sacola	Alças
A	☐ Bege	☐ Da mesma cor do corpo
B	☐ Branca	☐ Verde
C	☐ Preta	☐ Vermelha
D	☐ Estampado (ex: floral)	☐ Azul
E	☐ Outro _____	☐ Outro _____

4. Em média, quantos itens o (a) senhor (a) costuma comprar?

- A. ☐ Até 10 itens (pequenas compras)
- B. ☐ Mais de 10 itens (compras médias)
- C. ☐ Feira do mês (grandes compras)

5. Para que tipo de compras você utilizaria a sacola ecológica?

- A. ☐ Compras de supermercado
- B. ☐ Compras na padaria
- C. ☐ Compras no shopping
- D. ☐ Compras na livraria

6. Com que frequência você costuma fazer compras?

- A. ☐ Uma vez no mês (feira do mês)
- B. ☐ Quinzenalmente
- C. ☐ Semanalmente
- D. ☐ Quase todos os dias
- E. ☐ Raramente
- F. ☐ Outro _____

7. Qual seria o tamanho ideal da sacola ecológica para você?

- A. ☐ Pequeno, modelo vertical (Largura: 30 cm x Altura: 40 cm)

- B. () Médio, modelo horizontal (Largura: 45 cm x Altura: 35 cm x Espessura: 10 cm)
 C. () Grande, modelo horizontal (Largura: 50 cm x Altura: 40 cm x Espessura: 15 cm)

8. Quanto você estaria disposto a pagar por uma sacola ecológica? Obs: Preços em Reais (R\$).

	PEQUENA (R\$)	MÉDIA (R\$)	GRANDE (R\$)
A	2,00 a 4,00	5,00 a 7,00	10,00 a 14,00
B	4,00 a 6,00	7,00 a 9,00	14,00 a 18,00
C	6,00 a 8,00	9,00 a 11,00	18,00 a 22,00
D	Não estou disposto (a) a pagar	Não estou disposto (a) a pagar	Não estou disposto (a) a pagar

9. Sexo: () Feminino () Masculino

13. Faixa etária:

- A. () Menor que 20
 B. () De 20 a 30
 C. () De 30 a 40
 D. () De 40 a 50
 E. () De 50 a 60
 F. () Acima de 60

10. Estado civil: A.()Solteiro(a) B.()Casado(a) C.()Separado(a) D.()Viúvo(a)

11. Escolaridade:

- A. () Ensino fundamental incompleto
 B. () Ensino fundamental completo
 C. () Ensino médio incompleto
 D. () Ensino médio completo
 E. () Ensino superior incompleto
 F. () Ensino superior completo

12. Renda familiar:

- A. () Até 3 salários mínimos
 B. () De 3 a 6 salários mínimos
 C. () De 6 a 9 salários mínimos
 D. () De 9 a 12 salários mínimos
 E. () De 12 a 15 salários mínimos
 F. () Acima de 15 salários mínimos

APÊNDICE C - Rotinas que são requisitos para obter os modelos no software R

`monografia<-read.table("Dados_monografia.txt",head=T)#Comando usado para adicionar os dados presentes no documento "Dados_monografia.txt" presente na pasta "Meus documentos" no data frame "monografia"#`

`attach(monografia)#Comando usado para anexar os dados presentes no data frame "monografia" a pasta de trabalho do R#`

`require(Design) #Comando utilizado para habilitar o uso da função "lrm" do pacote Design#`

ROTINAS UTILIZADAS NA OBTENÇÃO DOS MODELOS NO SOFTWARE R

Sacolas de tamanho médio a R\$ 7,00

`modelo1.1<-`

`glm(compram7~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor,family=binomial(link=logit))`

`modelo1.2<-`

`lrm(compram7~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor)`

`modelo2.1<-`

`glm(compram7~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo+cor,family=binomial(link=logit))`

`modelo2.2<-`

`lrm(compram7~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo+cor)`

`modelo3.1<-`

`glm(compram7~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo+cor,family=binomial(link=logit))`

`modelo3.2<-`

`lrm(compram7~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo+cor)`

`modelo4.1<-`

`glm(compram7~fx_etaria+est_civil+escol+num_médio+supermercados+shopping+padaria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo+cor)`

```
ência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo+cor,family=binomial(link=logit))
```

```
modelo4.2<-
```

```
lrm(compram7~fx_etaria+est_civil+escol+num_médio+supermercados+shopping+padaria+freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo+cor)
```

```
modelo5.1<-
```

```
glm(compram7~fx_etaria+est_civil+escol+num_médio+supermercados+shopping+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+modelo+cor,family=binomial(link=logit))
```

```
modelo5.2<-
```

```
lrm(compram7~fx_etaria+est_civil+escol+num_médio+supermercados+shopping+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+modelo+cor)
```

```
modelo6.1<-glm(compram7~fx_etaria+est_civil+escol+num_médio
```

```
+shopping+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+modelo+cor,family=binomial(link=logit))
```

```
modelo6.2<-lrm(compram7~fx_etaria+est_civil+escol+num_médio
```

```
+shopping+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+modelo+cor)
```

```
modelo7.1<-glm(compram7~fx_etaria+est_civil+escol+num_médio
```

```
+shopping+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+cor,family=binomial(link=logit))
```

```
modelo7.2<-lrm(compram7~fx_etaria+est_civil+escol+num_médio
```

```
+shopping+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+cor)
```

```
modelo8.1<-glm(compram7~fx_etaria+est_civil+escol+num_médio
```

```
+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+cor,family=binomial(link=logit))
```

```
modelo8.2<-lrm(compram7~fx_etaria+est_civil+escol+num_médio
```

```
+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+cor)
```

```
modelo9.1<-
```

```
glm(compram7~fx_etaria+est_civil+escol+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+cor,family=binomial(link=logit))
```

```
modelo9.2<-
```

```
lrm(compram7~fx_etaria+est_civil+escol+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+cor)
```

```
modelo10.1<-
```

```
glm(compram7~fx_etaria+escol+padaria+freqüência+tam_ideal+sexo+descontos+obrig+tamanho+cor,family=binomial(link=logit))
```

```
modelo10.2<-
lrm(compram7~fx_etaria+escol+padaria+frequência+tam_ideal+sexo+descontos+obrig+tamanho
+cor)
```

```
modelo11.1<-
glm(compram7~fx_etaria+escol+padaria+frequência+sexo+descontos+obrig+tamanho
+cor,family=binomial(link=logit))
```

```
modelo11.2<-
lrm(compram7~fx_etaria+escol+padaria+frequência+sexo+descontos+obrig+tamanho+cor)
```

```
modelo12.1<-glm(compram7~fx_etaria+escol+padaria+frequência+sexo+descontos+tamanho
+cor,family=binomial(link=logit))
```

```
modelo12.2<-
lrm(compram7~fx_etaria+escol+padaria+frequência+sexo+descontos+tamanho+cor)
```

Sacolas de tamanho grande a R\$ 10,00

```
modelo1.1<-
glm(comprag10~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padari
a+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+mod
elo+cor,family=binomial(link=logit))
```

```
modelo1.2<-
lrm(comprag10~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padari
a+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+mod
elo+cor)
```

```
modelo2.1<-
glm(comprag10~fx_etaria+escol+renda+num_médio+supermercados+shopping+padaria+livraria
+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor,fa
mily=binomial(link=logit))
```

```
modelo2.2<-
lrm(comprag10~fx_etaria+escol+renda+num_médio+supermercados+shopping+padaria+livraria
+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor)
```

```
modelo3.1<-
glm(comprag10~fx_etaria+escol+num_médio+supermercados+shopping+padaria+livraria+freqü
ência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor,family=
binomial(link=logit))
```

```
modelo3.2<-
lrm(comprag10~fx_etaria+escol+num_médio+supermercados+shopping+padaria+livraria+freqüê
ncia+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor)
```



```
modelo4.1<-
glm(comprag10~fx_etaria+escol+num_médio+supermercados+shopping+padaria+livraria+tam_i
deal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor,family=binomial(link=logit))
```

```
modelo4.2<-
lrm(comprag10~fx_etaria+escol+num_médio+supermercados+shopping+padaria+livraria+tam_i
deal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor)
```

```
modelo5.1<-
glm(comprag10~fx_etaria+escol+num_médio+supermercados+shopping+padaria+livraria+tam_i
deal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo,family=binomial(link=logit))
```

```
modelo5.2<-
lrm(comprag10~fx_etaria+escol+num_médio+supermercados+shopping+padaria+livraria+tam_i
deal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo)
```

```
modelo6.1<-
glm(comprag10~fx_etaria+escol+num_médio+supermercados+shopping+padaria+livraria+tam_i
deal+sexo+consc_amb+descontos+obrig+resistência+modelo,family=binomial(link=logit))
```

```
modelo6.2<-
lrm(comprag10~fx_etaria+escol+num_médio+supermercados+shopping+padaria+livraria+tam_i
deal+sexo+consc_amb+descontos+obrig+resistência+modelo)
```

```
modelo7.1<-
glm(comprag10~fx_etaria+escol+num_médio+shopping+padaria+livraria+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+modelo,family=binomial(link=logit))
```

```
modelo7.2<-
lrm(comprag10~fx_etaria+escol+num_médio+shopping+padaria+livraria+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+modelo)
```

```
modelo8.1<-
glm(comprag10~fx_etaria+escol+num_médio+shopping+padaria+livraria+tam_ideal+sexo+consc_amb+obrig+resistência+modelo,family=binomial(link=logit))
```

```
modelo8.2<-
lrm(comprag10~fx_etaria+escol+num_médio+shopping+padaria+livraria+tam_ideal+sexo+consc_amb+obrig+resistência+modelo)
```

```
modelo9.1<-
glm(comprag10~escol+num_médio+shopping+padaria+livraria+tam_ideal+sexo+consc_amb+obrig+resistência+modelo,family=binomial(link=logit))
```

modelo9.2<-

```
lrm(comprag10~escol+num_médio+shopping+padaria+livraria+tam_ideal+sexo+consc_amb+obrig+resistência+modelo)
```

modelo10.1<-

```
glm(comprag10~escol+num_médio+shopping+padaria+livraria+tam_ideal+sexo+consc_amb+resistência+modelo,family=binomial(link=logit))
```

modelo10.2<-

```
lrm(comprag10~escol+num_médio+shopping+padaria+livraria+tam_ideal+sexo+consc_amb+resistência+modelo)
```

modelo11.1<-

```
glm(comprag10~escol+num_médio+shopping+padaria+livraria+tam_ideal+consc_amb+resistência+modelo,family=binomial(link=logit))
```

modelo11.2<-

```
lrm(comprag10~escol+num_médio+shopping+padaria+livraria+tam_ideal+consc_amb+resistência+modelo)
```

modelo12.1<-

```
glm(comprag10~escol+num_médio+shopping+padaria+livraria+consc_amb+resistência+modelo,family=binomial(link=logit))
```

modelo12.2<-

```
lrm(comprag10~escol+num_médio+shopping+padaria+livraria+consc_amb+resistência+modelo)
```

Sacolas de tamanho grande a R\$ 14,00

modelo1.1<-

```
glm(comprag14~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor,family=binomial(link=logit))
```

modelo1.2<-

```
lrm(comprag14~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor)
```

modelo2.1<-

```
glm(comprag14~fx_etaria+est_civil+escol+num_médio+supermercados+shopping+padaria+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor,family=binomial(link=logit))
```

modelo2.2<-

```
lrm(comprag14~fx_etaria+est_civil+escol+num_médio+supermercados+shopping+padaria+livra
```

```
ria+freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor
)
```

```
modelo3.1<-
glm(comprag14~fx_etaria+est_civil+escol+num_médio+supermercados+shopping+padaria+livra
ria+freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+modelo+cor,family=b
inomial(link=logit))
```

```
modelo3.2<-
lrm(comprag14~fx_etaria+est_civil+escol+num_médio+supermercados+shopping+padaria+livra
ria+freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+modelo+cor)
```

```
modelo4.1<-
glm(comprag14~fx_etaria+est_civil+escol+num_médio+supermercados+shopping+padaria+freq
üência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+modelo+cor,family=binomial
(link=logit))
```

```
modelo4.2<-
lrm(comprag14~fx_etaria+est_civil+escol+num_médio+supermercados+shopping+padaria+freq
üência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+modelo+cor)
```

```
modelo5.1<-
glm(comprag14~fx_etaria+escol+num_médio+supermercados+shopping+padaria+freqüência+ta
m_ideal+sexo+consc_amb+descontos+obrig+resistência+modelo+cor,family=binomial(link=logi
t))
```

```
modelo5.2<-
lrm(comprag14~fx_etaria+escol+num_médio+supermercados+shopping+padaria+freqüência+ta
m_ideal+sexo+consc_amb+descontos+obrig+resistência+modelo+cor)
```

```
modelo6.1<-
glm(comprag14~fx_etaria+escol+num_médio+supermercados+shopping+padaria+freqüência+ta
m_ideal+sexo+consc_amb+descontos+resistência+modelo+cor,family=binomial(link=logit))
```

```
modelo6.2<-
lrm(comprag14~fx_etaria+escol+num_médio+supermercados+shopping+padaria+freqüência+ta
m_ideal+sexo+consc_amb+descontos+resistência+modelo+cor)
```

Sacolas de tamanho pequeno a R\$ 4,00

```
modelop1.1<-
glm(comprap4~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria
+livraria+freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+mode
lo+cor,family=binomial(link=logit))
```

```
modelop1.2<-
```

```
lrm(comprap4~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria  
+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+mode  
lo+cor)
```

```
modelop2.1<-
```

```
glm(comprap4~fx_etaria+est_civil+renda+num_médio+supermercados+shopping+padaria+livrar  
ia+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor,  
family=binomial(link=logit))
```

```
modelop2.2<-
```

```
lrm(comprap4~fx_etaria+est_civil+renda+num_médio+supermercados+shopping+padaria+livrar  
ia+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor)
```

```
modelop3.1<-
```

```
glm(comprap4~fx_etaria+est_civil+renda+num_médio+supermercados+shopping+livraria+freqü  
ência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor,family=  
binomial(link=logit))
```

```
modelop3.2<-
```

```
lrm(comprap4~fx_etaria+est_civil+renda+num_médio+supermercados+shopping+livraria+freqü  
ência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+modelo+cor)
```

```
modelop4.1<-
```

```
glm(comprap4~fx_etaria+est_civil+renda+num_médio+supermercados+shopping+livraria+freqü  
ência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+modelo+cor,family=binomial(l  
ink=logit))
```

```
modelop4.2<-
```

```
lrm(comprap4~fx_etaria+est_civil+renda+num_médio+supermercados+shopping+livraria+freqü  
ência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+modelo+cor)
```

Sacolas de tamanho médio a R\$ 5,00

```
modelom1.1<-
```

```
glm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padari  
a+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+mod  
elo+cor,family=binomial(link=logit))
```

```
modelom1.2<-
```

```
lrm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria  
+livraria+frequência+tam_ideal+sexo+consc_amb+descontos+obrig+resistência+tamanho+mode  
lo+cor)
```

```
modelom2.1<-
glm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+
livraria+freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo+cor,family=binomial(link=logit))
```

```
modelom2.2<-
lrm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+
livraria+freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo+cor)
```

```
modelom3.1<-
glm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+
livraria+freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo,family=binomial(link=logit))
```

```
modelom3.2<-
lrm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+shopping+padaria+
livraria+freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo)
```

```
modelom4.1<-
glm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+padaria+livraria+
freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo,family=binomial(link=logit))
```

```
modelop4.2<-
lrm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+padaria+livraria+
freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+tamanho+modelo)
```

```
modelom5.1<-
glm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+padaria+livraria+
freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+modelo,family=binomial(link=logit))
```

```
modelom5.2<-
lrm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+padaria+livraria+
freqüência+tam_ideal+sexo+consc_amb+descontos+obrig+modelo)
```

```
modelom6.1<-
glm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+padaria+livraria+
freqüência+tam_ideal+sexo+descontos+obrig+modelo,family=binomial(link=logit))
```

```
modelom6.2<-
lrm(compram5~fx_etaria+est_civil+escol+renda+num_médio+supermercados+padaria+livraria+
freqüência+tam_ideal+sexo+descontos+obrig+modelo)
```

```
modelom7.1<-
glm(compram5~fx_etaria+est_civil+escol+renda+num_médio+padaria+livraria+freqüência+tam_ideal+
sexo+descontos+obrig+modelo,family=binomial(link=logit))
```

```
modelom7.2<-  
lm(compram5~fx_etaria+est_civil+escol+renda+num_médio+padaria+livraria+frequência+tam_  
ideal+sexo+descontos+obrig+modelo)
```