

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Aplicações de Estatística em Marketing

Thiago Daniel Carvalho

Orientador: Prof. Dr. Carlos Alberto Ribeiro Diniz

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

São Carlos

2008

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

C331ae

Carvalho, Thiago Daniel.

Aplicações de estatística em marketing / Thiago Daniel
Carvalho. -- São Carlos : UFSCar, 2008.
97 f.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2007.

1. Marketing. 2. Estatística aplicada. I. Título.

CDD: 519.5 (20^a)

Resumo

Neste trabalho, procuramos analisar como a estatística pode melhorar algumas atitudes de uma empresa com relação a estratégia de marketing. Abordamos métodos os quais permitem identificar os melhores consumidores e a preferência dos mesmos por características de produtos/serviços de forma que a empresa possa se relacionar melhor com seus consumidores.

Abstract :

In this work, we analyse the use of the statistics in the improvement of some attitudes of a company regarding marketing strategy. We deal with methods which permit the identification of the best costumers and the preferences of those costumers for product/service characteristics, so that a company can relate itself better to its costumers.

Sumário

1	INTRODUÇÃO	2
2	Marketing de Produto e Relacionamento Estruturado com Clientes	7
2.1	Introdução	7
2.2	Marketing de Produto	8
2.2.1	Segmentação para marketing de produto ou serviço	9
2.3	Modelos preditivos	10
2.4	Regressão Logística	11
2.4.1	Modelo de Regressão Logística com várias covariáveis	12
2.5	Aplicação	12
2.6	Identificação dos melhores clientes via R/F/M (Recenticidade/Frequência/Valor monetário)	16
2.6.1	Exemplo da utilização da fórmula R/F/M.	16
2.6.2	Utilização da Recenticidade como covariável no modelo logístico	18
2.6.3	Aplicação	19
2.7	Relacionamento Estruturado com Clientes	20
2.7.1	Relacionamento com Clientes	20
2.7.2	A base analítica	22
2.7.3	Construção da base analítica	23
2.7.4	Aplicação	25

3	VALIDAÇÃO DE MODELOS DE REGRESSÃO LOGÍSTICA VIA BOOTSTRAP	30
3.1	Análise decil e cum lift para o modelo de regressão logística	32
3.1.1	Exemplo 1	32
3.2	Metodologia Bootstrap	34
3.3	Construção do intervalo de confiança via Bootstrap	35
3.4	Análise decil de validação Bootstrap	36
3.5	Outras questões à serem analisadas	37
3.6	Exemplo 2	39
3.6.1	Conclusão:	42
3.7	Análise Gráfica dos Decis	43
3.7.1	Construção dos gráficos	45
3.7.2	Utilização dos Gráficos	46
4	ANÁLISE CONJUNTA - <i>CONJOINT ANALYSIS</i>	50
4.1	Introdução	50
4.2	Uso de variáveis dummy	52
4.3	Principais etapas da análise conjunta	53
4.3.1	1º Etapa: Caracterização do Produto	53
4.3.2	2º Etapa: Planejamento da pesquisa	53
4.3.3	3º Etapa: Análise dos dados	55
4.4	Exemplo	56
4.5	Análise Conjunta para marketing de clientes	86
4.5.1	1ª Etapa: Caracterização do Cliente	86
4.5.2	2ª Etapa: Planejamento da pesquisa	87
4.5.3	3º Etapa: Análise dos dados	87
4.5.4	Aplicação da análise conjunta em marketing de clientes	88
5	BIBLIOGRAFIA	98

Capítulo 1

INTRODUÇÃO

Em marketing buscamos conhecer os pensamentos e atitudes de clientes com o intuito de melhorar o relacionamento com os mesmos. O relacionamento com o cliente é definido como o conjunto de contatos que a empresa realiza com seus clientes por meio de vendas, entrega de produtos/serviços, cobrança, ações de comunicação em geral e ações voltadas à exploração do potencial de compra dos clientes.

Em nosso estudo, mostramos como a estatística pode auxiliar no desenvolvimento de estratégias de marketing, ao longo dos capítulos está apresentado técnicas utilizadas em marketing, tanto de produtos, quanto de clientes, e buscamos aplicar técnicas estatísticas para auxiliar na identificação de bons clientes e conhecer características relevantes de produtos/serviços para os consumidores.

Para auxiliar o marketing de produtos abordamos métodos que permitem segmentar os consumidores, e utilizamos modelos logísticos para obter a propensão do cliente vir a abandonar ou comprar um determinado produto. O modelo logístico é utilizado também para fornecer a propensão do cliente voltar a comprar o produto, segmentação em marketing se resume na divisão do mercado em grupos de indivíduos com características, necessidades e modos de atuação semelhantes, segundo seu perfil financeiro, psicológico etc.

Visando estruturar o relacionamento com os clientes, abordamos métodos que permitem identificar segmentos de clientes com alta propensão a dar uma resposta positiva (compra do produto/serviço) e segmentos de clientes com alta propensão a abandonar a empresa.

Com este conhecimento, a empresa pode, então, desenvolver ações de *marketing* para retenção de clientes.

É possível, ainda, a empresa calcular o **valor do cliente**, que é determinado a partir das receitas que o cliente gera em um determinado período de tempo menos o custo administrativo e de serviço, e com isto obter conhecimento sobre o quanto se pode gastar para reter um específico cliente ou um grupo de clientes.

Neste estudo utilizamos uma medida chamada *cum lift* (cumulative lift) para avaliar a qualidade de modelos preditivos de respostas. O *cum lift* indica o ganho esperado de uma **solicitação** implementada utilizando um modelo sobre uma **solicitação** sem utilizar o modelo.

Uma **solicitação** significa um contato com o cliente buscando uma resposta positiva ou negativa. Por exemplo, realizar uma oferta e observar uma resposta positiva (aquisição de um determinado produto).

A medida *cum lift* é obtida a partir de um método denominado análise decil, que fornece uma tabela contendo o desempenho do modelo considerado (modelo logístico). Esta tabela consiste dos decis, número de indivíduos, número de respondentes, taxa de resposta decil, taxa de resposta cumulativa e *cum lift*.

A validação do modelo é feita dividindo aleatoriamente a amostra disponível em duas novas amostras, uma amostra para a construção do modelo e uma para validação, em seguida realiza-se a análise decil nestas duas novas amostras.

Após obtidas as análises decil os *cum lifts* são comparados e espera-se observar uma diminuição dos mesmos da análise decil de validação em relação aos *cum lifts* da análise decil de treinamento. Se a diminuição dos *cum lifts* e os valores dos *cum lifts* são aceitáveis então o modelo é considerado válido.

O maior problema nesta abordagem é que uma única amostra não fornece uma medida de variabilidade que permita uma afirmação sobre o nível de confiança em relação a validação. Desta forma, surgem questões do tipo: Tendo disponível um único conjunto de dados, é possível obter uma estimativa acurada do *cum lift* e seu erro padrão?

Se amostras adicionais estiverem disponíveis, novas análises decil de validação podem ser realizadas e com isto a comparação dos *cum lifts*. Com várias análises de validação é possível determinar a variação do *cum lift* dentro do decil, além de um *cum lift* médio.

Caso não possuímos novas amostras, as amostras adicionais podem ser criadas via técnicas bootstrap. O modelo utilizado pode ser validado e um intervalo de confiança para cada *cum lift* pode ser criado. Estes intervalos podem ser avaliados tomando como base suas extensões (comprimento). Intervalos extensos possuem margens de erro elevadas, o que não é desejado pela empresa.

Assim, quando os intervalos são considerados grandes, a medida a ser tomada é trabalhar com os tamanhos amostrais para tentar obter margens de erro menores e com isto diminuir a extensão dos intervalos para os *cum lifts*.

Para auxiliar na procura do tamanho amostral, que forneça margens de erro pré-fixadas pela empresa, foram contruídos gráficos contendo o tamanho amostral versus margens de erro.

Com o modelo válido, pode-se desenvolver estratégias de marketing utilizando as medidas *cum lift* para alcançar melhores resultados nas atitudes da empresa, como, por exemplo aumentar o número de respostas positivas para uma determinada oferta.

Para responder quais características de um produto/serviço são mais importantes e estabelecer a preferência dos clientes ou respondentes por produtos/serviços, utilizamos a *análise conjunta*. Nesta técnica apresentamos aos respondentes várias produtos, os quais são descritos através de suas características (denominadas atributos).

A *análise conjunta* permite que os respondentes façam suas escolhas de maneira similar a uma situação real de compra, isto é, comparem as várias opções de produtos com base em critérios subjetivos de valor dos atributos mais importantes do produto.

Desta forma, uma empresa conhecendo os atributos mais importantes de um produto/serviço e conhecendo os perfis mais preferidos pelos respondentes, pode desenvolver seu produto/serviço levando em conta os resultados da *análise conjunta* e obter um produto com melhor nível de aceitação no mercado.

Esta técnica pode ser aplicada no desenvolvimento de novos produtos em todos os setores industriais, seleção de segmento-alvo de mercado entre outras aplicações.

A *análise conjunta*, é uma técnica utilizada em marketing de produtos, em nosso trabalho buscando obter novos resultados, aplicamos esta técnica em marketing de clientes onde, ao invés de analisarmos quais características de um produto são mais importantes para o cliente, analisamos quais características dos clientes são mais relevantes para a

empresa e qual o perfil de cliente que é de maior importância para a empresa.

Tendo em mãos este conhecimento, a empresa pode direcionar seus esforços de marketing para seus clientes ativos, de maior importância, e desenvolver ações de marketing tomando como base as características relevantes de seus clientes, para identificar em seu mercado alvo possíveis clientes potenciais (prospects), identificados os possíveis clientes, pode-se direcionar estratégias de marketing à este grupos específico de consumidores não ativos, para conquistar novos bons clientes.

Mais especificamente, este trabalho é organizado como se segue:

O Capítulo 2, abrange a utilização de técnicas estatísticas como auxílio ao marketing de produtos e de clientes, técnicas as quais permitem segmentar os consumidores e identificar bons clientes, modelos logísticos foram utilizados, para obter a propensão do cliente vir a abandonar, comprar um determinado produto e para fornecer a propensão do cliente voltar a comprar o produto, estas informações servem de apoio ao desenvolvimento de ações de marketing de produto.

O modelo logístico, foi também utilizado, como modelos de abandono e modelos preditivos de respostas, que estabelecem respectivamente, o grau de propensão que cada cliente tem a abandonar a empresa e prediz a propensão de cada cliente dar respostas positivas a ofertas.

Tendo conhecimento da propensão ao abandono, a empresa pode ainda, calcular o valor do cliente, e com estas informações, desenvolver ações de retenção, podendo limitar o gasto de retenção.

O Capítulo 3 é dedicado a validação de modelos de regressão logística via bootstrap, onde é utilizado uma medida chamada *cum lift*, obtida a partir da análise decil, para validar o modelo logístico utilizado.

A validação do modelo é feita dividindo aleatoriamente a amostra disponível em duas novas amostras, uma para treinamento e uma para validação, os *cum lifts* são comparados e espera-se observar uma diminuição dos mesmos da análise decil de validação em relação aos *cum lifts* da análise decil de treinamento.

Neste ponto do trabalho deparamos com um problema, sendo que, uma única amostra não fornece uma medida de variabilidade que permita uma afirmação sobre o nível de confiança em relação a validação.

Conciderando que, não possuímos amostras adicionais, novas amostras foram criadas via técnicas bootstrap, com isso, o modelo utilizado pôde ser validado e um intervalo de confiança para cada *cum lift* pôde ser criado, estes intervalos, foram avaliados tomando como base suas extensões. Intervalos extensos possuem margens de erro elevadas, o que não é viável à empresa, uma medida a ser tomada é trabalhar com os tamanhos amostrais para tentar obter margens de erro menores e com isto diminuir a extensão dos intervalos para os *cum lifts*.

Para auxiliar na procura do tamanho amostral, que forneça margens de erro pré-fixadas pela empresa, foram contruídos gráficos contendo o tamanho amostral versus margens de erro.

O foco do Capítulo 4 é a análise conjunta, técnica utilizada para fornecer quais características de um produto/serviço são mais importantes e estabelecer a preferência dos clientes ou respondentes por produtos/serviços, para isso, utilizamos, dois métodos, um tradicional, que envolvem cálculo de médias e outro que, faz uso de um modelo de regressão com variáveis dummy.

A contribuição deste capítulo, está na aplicação de análise conjunta, que é, uma técnica até então voltada para o marketing de produto, em marketing de cliente. Neste ponto ao invés de analisarmos a preferência de clientes por determinados produtos, analisamos a preferência da empresa por determinados perfis de clientes, e quais características dos clientes são mais relevantes para a empresa.

Capítulo 2

Marketing de Produto e Relacionamento Estruturado com Clientes

2.1 Introdução

No cenário empresarial encontramos empresas que não diferenciam clientes, e com isto mantém um relacionamento não customizado com os mesmos. Encontramos também aquelas que buscam conhecer os pensamentos de seus clientes para saber seus pontos de vista, e o que eles necessitam. Este conhecimento é usado para diferenciar o relacionamento com os diversos tipos de clientes da empresa e para criar uma vantagem competitiva.

Segundo Costa (2004), o número de empresas que buscam ouvir seus clientes aumenta a medida que cresce a concorrência. Estas empresas podem ser divididas em dois grupos, um grupo composto das empresas que utilizam marketing de produto e outro grupo composto das que estruturam o relacionamento com seus clientes a partir de dados individuais.

As ferramentas utilizadas pelas empresas que adotam o marketing de produto segmentam os clientes para conhecer o perfil de cada um deles. Na seqüência pesquisam as necessidades destes grupos e os produtos necessários para satisfazê-los e calculam o potencial de compra de cada grupo de cliente, para concentrar os esforços de marketing nos grupos mais atrativos. O potencial de compra de cada grupo é calculado utilizando

as receitas que cada grupo gera em um determinado período de tempo menos o custo de servir este grupo. O custo de servir é o custo de produzir e entregar os produtos comprados e o potencial de compra de um determinado grupo é o valor do grupo para a empresa.

Baseando-se nas necessidades e potencial de compra de cada grupo são desenvolvidos produtos e formas de comunicação para atingir parcialmente ou totalmente os segmentos de forma diferenciada em relação a concorrência.

Na estruturação do relacionamento com clientes, em primeiro lugar determina-se o valor de cada cliente existente (valor do cliente, que é formado a partir das receitas que o cliente gera em um determinado período de tempo menos o custo de servir o cliente), o próximo passo é identificar comportamentos diferentes de grupos de clientes por meio de análise de dados concretos de venda (que produtos/serviços eles compram, com que frequência etc), com isto, busca-se esboçar os perfis dos diferentes grupos de clientes identificados pela análise de dados de compra. Traçado o perfil de cada grupo são desenvolvidas estratégias de comunicação direcionada a cada um deles e desenvolvidas formas de buscar no mercado novos clientes com o perfil semelhante aos perfis mais atrativos.

2.2 Marketing de Produto

De acordo com a American Marketing Association; "Marketing é o processo de planejar a concepção de idéias, produtos e serviços definindo preço, promovendo e distribuindo-os para criar trocas que atendam aos objetivos de satisfação de necessidades e a realização de desejos individuais dos consumidores".

A essência do marketing é o processo de trocas, nas quais uma ou mais partes oferecem algo de valor para o outro, com o objetivo de satisfazer necessidades e desejos. Em síntese como o objetivo do marketing é identificar necessidades não satisfeitas, devemos procurar desenvolver produtos e serviços que atendam as necessidades de cada público. Deve ainda buscar melhorar a qualidade de vida das pessoas em geral.

Portanto, o papel do marketing é utilizar recursos de conhecimento (dados concretos dos clientes, tal como, dados da última compra) que possibilitem uma melhor compreensão do impacto do esforço a ser realizado na administração da demanda de produtos e serviços,

isto é, o quanto de determinado produto ou serviço que uma coletividade está necessitando e procurando.

Empresas que realizam o marketing de produto ou serviço utilizam métodos que permitem segmentar seus clientes e conhecer o perfil de cada grupo, em seguida calculam o potencial de compra dos segmentos de clientes, a partir das receitas geradas em um determinado período de tempo menos o custo de servir estes segmentos. Na sequência pesquisam as necessidades dos grupos e os produtos para satisfazê-los.

2.2.1 Segmentação para marketing de produto ou serviço

A segmentação para o marketing de produtos ou serviços é realizada com o objetivo de identificar grupos de consumidores de um determinado produto ou serviço oferecido pela empresa, que possuem características em comum, como por exemplo, clientes que utilizam um mesmo veículo de comunicação com a empresa, clientes que apresentam um poder aquisitivo semelhante, entre outras.

Segmentação do público alvo do produto ou serviço

A segmentação do público alvo é feita agrupando os consumidores de um determinado produto, que apresentam características em comum, tais como, a frequência com que os consumidores utilizam ou adquirem o produto ou serviço, tempo de fidelidade ao produto ou potencial de compra do cliente.

O potencial de compra do cliente e o tempo de fidelidade ao produto podem ser utilizados em conjunto para realizar a segmentação dos clientes.

Após a segmentação é realizada a seleção do público alvo, onde é importante medir o potencial de compra de cada grupo de clientes (atratividade do grupo). Isto permite selecionar os segmentos mais atrativos, nos quais a empresa deverá concentrar seu esforço de marketing. Em seguida são pesquisadas as necessidades de cada grupo selecionado e maneiras para satisfazê-las, também são elaboradas estratégias de marketing diferenciadas para cada grupo de clientes com o objetivo de atingí-los parcialmente ou totalmente.

As estratégias de marketing para produtos e serviços podem ser apoiadas por resultados de estudos estatísticos, como a construção de modelos preditivos, que fornecem informações relevantes sobre os consumidores do produto.

2.3 Modelos preditivos

Os modelos que podem ser utilizados para auxiliar no desenvolvimento de estratégias de marketing de produto são os modelos de abandono de um determinado produto ou serviço e modelos de resposta positiva a compra do produto.

Modelos de abandono estabelecem o grau de propensão que cada cliente, ou grupos de clientes, tem a abandonar um produto ou serviço (isto é, a probabilidade de cada cliente deixar de comprar produtos/serviços da empresa). Conhecendo a propensão ao abandono do produto, a empresa pode calcular o potencial de compra de um grupo para que possa limitar seus gastos de retenção.

Os modelos preditivos de respostas positivas a compra predizem a propensão de cada cliente comprar um determinado produto. Tendo conhecimento da propensão do cliente dar uma resposta positiva a compra é possível direcionar ofertas relacionadas ao produto com maior probabilidade de aceitação pelo cliente.

Com estes modelos a empresa tem conhecimento da quantidade de clientes que é propenso a comprar e o número de clientes que devem ser abordados com ações de retenção.

Para obter a propensão de abandono e a propensão a dar uma resposta positiva podemos utilizar modelos de regressão logística. O modelo logístico prediz a probabilidade de um determinado cliente, cujos dados não foram usados na construção do modelo, de vir a abandonar a empresa ou de um determinado cliente adquirir um determinado produto. As probabilidades, para este novo grupo de clientes, podem ser ordenadas para melhor identificação de bons clientes com relação a não abandonar ou com relação a aquisição de produtos.

2.4 Regressão Logística

A regressão logística é utilizada nos casos onde a variável aleatória de interesse é uma variável dicotômica tomando valor 1 com probabilidade π e valor 0 com probabilidade $(1-\pi)$. Desta forma a variável aleatória é chamada de ponto binomial ou variável Bernoulli e tem uma distribuição de probabilidade discreta dada por:

$$Pr(Y = y) = \pi^y(1 - \pi)^{1-y}; \quad y = 0, 1.$$

E a resposta para o i^{th} indivíduo de uma amostra, Y_i , $i = 1, 2, 3, \dots, n$, é uma variável bernoulli com,

$$Pr(Y_i = y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}; \quad y_i = 0, 1.$$

A análise de regressão logística assume que a relação entre π_i e o valor de uma covariável x_i de um mesmo individuo é descrito pela função logística

$$\pi_i = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_i)]}, \quad i = 1, 2, \dots, n.$$

A transformação de π_i que é o centro do estudo de regressão logística, é a transformação *logit*. Esta transformação é definida em termos de π_i como:

$$\pi_i = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_i))},$$

$$\pi_i = \frac{1}{1 + \exp(-(y_i))},$$

$$y_i = \ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_i.$$

onde, neste simples modelo de regressão, $y_i = \ln \frac{\pi_i}{1 - \pi_i} = \beta_0 + \beta_1 x_i$, (função de ligação).

Em um modelo de regressão logística simples, a função de verossimilhança é dada por:

$$L = \prod_{i=1}^n Pr(Y_i = y_i) = \prod_{i=1}^n \frac{\{\exp[\beta_0 + \beta_1 x_i]\}^{y_i}}{1 + \exp[\beta_0 + \beta_1 x_i]}; \quad y_i = 0, 1.$$

As estimativas de máxima verossimilhança de β_0 e β_1 podem ser obtidas utilizando um processo iterativo de Newton-Raphson para otimizar a função de verossimilhança.

2.4.1 Modelo de Regressão Logística com várias covariáveis

Considere K covariáveis $x_1, x_2, x_3, \dots, x_k$, a relação entre π_i e o valor de uma covariável x_i de um mesmo indivíduo, que é descrita pela função logística apresentada acima, pode ser, facilmente generalizado como:

$$\pi_i = \frac{1}{1 + \exp[-(\beta_0 + \sum_{j=1}^k \beta_j x_{ji})]}; \quad i = 1, 2, \dots, n,$$

conseqüentemente,

$$y_i = \ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \sum_{j=1}^k \beta_j x_{ji}$$

Isto nos leva a função de verossimilhança dada por:

$$L = \prod_{i=1}^n \frac{\{\exp[\beta_0 + \sum_{j=1}^k \beta_j x_{ji}]\}^{y_i}}{1 + \exp[\beta_0 + \sum_{j=1}^k \beta_j x_{ji}]}; \quad y_i = 0, 1.$$

As estimativas de máxima verossimilhança para os parâmetros de $\beta_0, \beta_1, \dots, \beta_{k+1}$ podem ser obtidas utilizando um processo iterativo de Newton-Raphson para otimizar a função de verossimilhança.

A seção 2.5 apresenta dois casos de utilização de modelos logísticos para obter a propensão de resposta positiva e a propensão de abandono dos clientes.

2.5 Aplicação

· Caso 1

Suponha que possuímos uma amostra com $n = 1000$ clientes de uma determinada empresa, e a empresa deseja conhecer clientes com maior probabilidade a dar uma resposta positiva com relação a aquisição de um determinado produto. Para cada um dos n clientes calculamos o potencial de compra considerando os últimos 6 meses, representada pela

variável X_1 , digamos, e observamos a variável resposta Y , com o valor 1 denotando que o cliente é respondente e o valor 0 denotando que o cliente é não respondente.

O objetivo neste caso é obter a probabilidade de um grupo de clientes, não utilizados na construção do modelo, a dar uma resposta positiva a aquisição de um determinado produto. Estas probabilidades são previstas via modelo de regressão logística dado por

$$\ln(\pi_i/(1 - \pi_i)) = \beta_0 + \beta_1 X_{1i} + \varepsilon_i, i = 1, 2, \dots, 1000;$$

No ajuste do modelo aos dados,

Y	X1
1	420,13
1	310,75
1	374,29
...
1	334,89
1	311,25

obtemos as estimativas para os parâmetros β_0 e β_1 , dados por $\hat{\beta}_0 = -0,4202$ e $\hat{\beta}_1 = 0,0014$ e a análise dos parâmetros estimados mostrada na tabela 1.

Tabela 1 - Análise dos parâmetros estimados.

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence	Chi-Square	Pr > ChiSq
Intercept	1	-0.4202	0.2048	-0.8216	-0.0187	4.21	0.0402
x1	1	0.0014	0.0007	0.0002	0.0027	4.83	0.0279

A partir da tabela 1, temos que, a variável preditora x1 é significativa para o modelo, a um nível de 5%, e obtemos os intervalos de 95% de confiança para os parâmetros estimados.

Para $\hat{\beta}_0$ o intervalo de 95% de confiança é [-0.8216 ; -0.0187].

Para $\hat{\beta}_1$ o intervalo de 95% de confiança é [0.0002 ; 0.0027].

Com o modelo ajustado obtemos as probabilidades estimadas $\hat{\pi}_i, i = 1, 2, \dots, n$, que cor-

responderá a probabilidade do i -ésimo consumidor a apresentar uma resposta positiva a aquisição do produto em estudo.

As probabilidades $\hat{\pi}_i, i = 1, 2, \dots, n$, são obtidas da seguinte forma:

Temos que,

$$\ln(\pi_i/(1 - \pi_i)) = \beta_0 + \beta_1 X_{1i}, i = 1, 2, \dots, n;$$

logo,

$$\ln(\hat{\pi}_i/(1 - \hat{\pi}_i)) = \hat{\beta}_0 + \hat{\beta}_1 X_{1i}, i = 1, 2, \dots, n;$$

de onde,

$$\hat{\pi}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 X_{1i})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_{1i})}.$$

Obtidas as probabilidades $\hat{\pi}_i, i = 1, 2, \dots, n$, a empresa pode identificar consumidores com alta probabilidade de apresentar uma resposta positiva a aquisição do produto e desenvolver estratégias de marketing visando o aumento de respostas positivas. Isto acarretará, certamente, um aumento nas vendas do produto em estudo.

· Caso 2

Suponha que possuímos uma amostra com $n = 1000$ clientes de uma determinada empresa e a empresa deseja conhecer clientes com maior probabilidade a abandonar (deixar de comprar) um determinado produto oferecido pela empresa. Para cada um dos n clientes calculamos o potencial de compra considerando os últimos 6 meses, representada pela variável X_1 , digamos. Além disso, buscamos nos registros da empresa a frequência que cada cliente consumiu o produto, representada pela variável X_2 . Observamos a variável binária Y , com o valor 1 denotando que o cliente abandonou o produto e o valor 0 denotando que o cliente não abandonou o produto.

O objetivo é obter a probabilidade de um grupo de clientes, não utilizados na construção do modelo, a abandonar o produto. Estas probabilidades são preditas via modelo de regressão logística dado por

$$\ln(\pi_i/(1 - \pi_i)) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, i = 1, 2, \dots, 300;$$

Ajustando o modelo aos dados,

Y	X1	X2
1	420,13	11
1	310,75	8
1	374,29	13
...
1	334,89	6
1	311,25	8

obtemos as estimativas para os parâmetros β_0 , β_1 e β_2 , dados por $\hat{\beta}_0 = -0,9169$, $\hat{\beta}_1 = 0,0015$ e $\hat{\beta}_2 = 0,0605$, e a análise dos parâmetros estimados mostrada na tabela 2.

Tabela 2 - Análise dos parâmetros estimados.

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence	Chi-Square	Pr > ChiSq
Intercept	1	-0.9169	0.2814	-1.4684	-0.3654	10.62	0.0402
x1	1	0.0015	0.0007	0.0002	0.0028	5.17	0.0279
x2	1	0.0605	0.0233	0.0149	0.1061	6.76	0.0093

A partir da tabela2, temos que, as variáveis preditoras x1 e x2 são significativas para o modelo a um nível de 5%, e obtemos os intervalos de 95% de confiança para os parâmetros estimados.

Para $\hat{\beta}_0$ o intervalo de 95% de confiança é [-1.4684 ; -0.3654].

Para $\hat{\beta}_1$ o intervalo de 95% de confiança é [0.0002 ; 0.0028].

Para $\hat{\beta}_2$ o intervalo de 95% de confiança é [0.0149 ; 0.1061].

Com o ajuste do modelo podemos obter as probabilidades de abandono para um determinado grupo de clientes. Com o conhecimento destas probabilidades de abandono, a empresa deve desenvolver ações de retenção. Estas ações de retenção podem ser elaboradas considerando a velha regra 80/20, onde temos que 80% dos negócios da empresa

devem-se a 20% de seus clientes, que é uma generalização muito acurada segundo Aaker (2004)

Assim, a prioridade é a retenção dos 20% melhores clientes, além de um trabalho em relação aos 25% seguintes para aperfeiçoá-los, e pode-se dedicar algum esforço aos 55% restantes.

Outro método a ser utilizado para a identificação dos melhores clientes é a fórmula R/F/M, apresentado na seção a seguir.

2.6 Identificação dos melhores clientes via R/F/M (Recenticidade/Frequência/Valor monetário)

Uma ferramenta essencial para a identificação de melhores clientes da empresa é a fórmula de recenticidade/frequência/valor monetário (R/F/M).

- Recenticidade da compra – Quanto tempo faz que este cliente fez seu último pedido?
- Frequência da compra – Com que frequência este cliente costuma comprar da empresa?
- Valor monetário da compra – Quanto este cliente costuma gastar em uma transação típica?

2.6.1 Exemplo da utilização da fórmula R/F/M.

Neste exemplo três clientes tem seus históricos de transação calculados em 12 meses.

A tabela 3 abaixo ilustra a utilização da fórmula R/F/M na avaliação de clientes de uma empresa, que utilizam um determinado produto para escritório,

Tabela 3 -histórico de transação calculado em 12 meses.

<i>Cliente</i>	C_n	R_m	P_a	P_p ($\times 5$)	F	P_a <i>at.</i>	P_p ($\times 2$)	$V.M.$	P_a	P_p ($\times 3$)	T_{pp}	<i>Pontos</i> <i>Cumulativos</i>
1	1	2	20	100	1	3	6	R\$40,00	4	12	118	118
	2	4	10	50	1	3	6	R\$120,00	12	36	92	210
	3	9	3	15	1	3	6	R\$60,00	6	18	39	249
2	1	6	5	25	2	6	12	R\$400,00	25	75	112	112
3	1	2	20	100	1	3	6	R\$90,00	9	27	133	133
	2	4	10	50	1	3	6	R\$70,00	7	21	77	210
	3	6	5	25	2	6	12	R\$80,00	8	24	61	271
	4	9	3	15	1	3	6	R\$40,00	4	12	37	308

Onde:

$C.(n^\circ)$ = Número da compra, F = Frequencia, $R.(m)$ = Recenticidade em meses, $V.$ = Valor monetário, $P.at.$ = Pontos atribuídos, $P.p$ = Pontos ponderados, $Total\ p.p.$ = Total pontos ponderados.

É elementar destacar que cada organização deve determinar, por meio de sua própria análise, quais os fatores que mais influenciam nas compras. Uma regra prática, é que o cliente que comprou mais recentemente é aquele com maior probabilidade de comprar novamente.

Neste exemplo os pontos a serem atribuídos a recenticidade, frequência e valor monetário da transação, adotados pela empresa obedecem a seguinte sequência.

Recenticidade da transação: 20 pontos, se nos últimos 2 meses;
 10 pontos, se nos últimos 4 meses;
 5 pontos, se nos últimos 6 meses;
 3 pontos, se nos últimos 9 meses;

1 ponto, se nos últimos 12 meses.

Frequência da transação: 3 pontos para cada compra dentro dos 12 meses; Máximo = 15 pontos.

Valor monetária da transação: 10% do Volume Monetário da Compra dentro dos 12 meses; Maximo = 25 pontos.

Os pesos relativos para a recenticidade, frequência e valor monetário da transação, são respectivamente 5, 2 e 3.

Desta forma foram atribuídos os pesos de 50%, 20% e 30%, respectivamente a R/F/M.

Na tabela obtida, observamos que escores cumulativos resultantes para os consumidores 1, 2 e 3 são 249,112 e 308 respectivamente, o que indica que a empresa deve ter um preferencial em relação ao cliente 3.

Com base no histórico R/F/M do cliente 3, um montante maior de investimento promocional poderia ser justificável (como o envio de um catálogo sazonal), o cliente 1 pode ser visto como um bom potencial, enquanto enviar material promocional para o cliente 2 seria um esforço de marketing equivocado.

2.6.2 Utilização da Recenticidade como covariável no modelo logístico

De acordo com Aaker (2004), uma regra utilizada em marketing de produto ou serviço é que o cliente que comprou mais recentemente é aquele com maior probabilidade de comprar novamente, com isto podemos utilizar um modelo logístico para prever a probabilidade de clientes virem a comprar um determinado produto novamente.

O modelo logístico estima a probabilidade de clientes virem a comprar o produto novamente utilizando dados da última transação dos clientes, ou seja, utilizando o tempo desde a última compra (Recenticidade). Obtidas as probabilidades, a empresa pode ordená-las para identificar seus clientes com maior propensão a uma nova aquisição do produto e direcionar novas ofertas a estes consumidores.

2.6.3 Aplicação

Suponha que possuímos uma amostra com $n = 1500$ clientes de uma empresa, para cada um dos n clientes buscamos nos históricos da empresa o tempo desde a última compra, representada pela variável X_1 , digamos. O valor de X_1 é igual a 2, se o cliente comprou o produto pela última vez há 2 meses. Além disso observamos a variável binária Y , com o valor 1 denotando que o cliente respondeu a última oferta do produto e o valor 0 denotando que o cliente não respondeu a última oferta do produto.

O desejo da empresa neste caso é obter a probabilidade de um grupo de clientes, comprar o produto novamente. Estas probabilidades são preditas via modelo de regressão logística dado por,

$$\ln(\pi_i/(1 - \pi_i)) = \beta_0 + \beta_1 X_{1i} + \varepsilon_i, i = 1, 2, \dots, 1500;$$

Ajustando o modelo aos dados disponíveis,

Y	X1
0	9
0	4
0	4
...
0	4
1	5

obtemos as estimativas para os parâmetros β_0 e β_1 , dados por $\hat{\beta}_0 = 0,3372$ e $\hat{\beta}_1 = -0,0650$ e a análise dos parâmetros estimados mostrada na tabela 4.

Tabela 2 - Análise dos parâmetros estimados.

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence	Chi-Square	Pr > ChiSq
Intercept	1	0.3372	0.1356	0.0715	0.6029	6.19	0.0129
x1	1	-0.0650	0.0209	-0.1059	-0.0241	9.69	0.0019

A partir da tabela 4, temos que, a variável preditora x_1 é significativa para o modelo a um nível de 1%, e obtemos os intervalos de 95% de confiança para os parâmetros estimados.

Para $\hat{\beta}_0$ o intervalo de 95% de confiança é [0.0715 ; 0.6029].

Para $\hat{\beta}_1$ o intervalo de 95% de confiança é [-0.1059 ; -0.0241].

Com o modelo ajustado obtemos as probabilidades estimadas $\hat{\pi}_i, i = 1, 2, \dots, n$, que corresponderá a probabilidade do i -ésimo consumidor vir a comprar o produto novamente.

Com as probabilidades calculadas podemos ordená-las e identificar os clientes mais prováveis a adquirir o produto novamente. Identificando estes clientes podemos desenvolver estratégias de marketing e direcionar ofertas aos melhores clientes, para melhorar as vendas do produto.

2.7 Relacionamento Estruturado com Clientes

2.7.1 Relacionamento com Clientes

Podemos definir, segundo Costa (2004), que o relacionamento com clientes é o conjunto de contatos que a empresa realiza com seus clientes, por exemplo, vendas, entrega de produtos/serviços, cobrança, ações de comunicação em geral e ações voltadas à exploração do potencial de compra dos clientes.

Para que o relacionamento com clientes tenha importância estratégica, o relacionamento deve ser estruturado e ter a capacidade de fornecer elementos importantes para a definição da estratégia.

Estrutura do Relacionamento

De acordo com Costa (2004), a estruturação do relacionamento com clientes se inicia com foco em três elementos:

- O valor do cliente;
- As formas de geração de valores;
- A posição de cada um no ciclo de relacionamento.

O valor do cliente é formado a partir das receitas que o cliente gera em um determinado período de tempo menos o custo de servir o cliente, este custo é composto pelos custos de produzir e entregar os produtos comprados.

Este valor pode ser ajustado pelo tempo esperado de retenção do cliente, pelas vendas adicionais que se espera fazer ao cliente e pelo custo esperado para servir o cliente neste tempo.

O valor do cliente é um dos primeiros dados a ser calculado, pois permite realizar a segmentação dos clientes, em clientes que geram pouco e muito valor, esta segmentação pode se tornar mais apurada se levar em consideração as formas de geração deste valor, isto é, maneiras com que os clientes utilizam os produtos/serviços da empresa, por exemplo (que produto este cliente compra, quais serviços o cliente utiliza).

É possível, também, identificar clientes que proporcionam prejuízo, valor do cliente negativo, mais que no passado apresentavam valor positivo e reduziram seu relacionamento com a empresa. Com posse desta informação, pode-se elaborar estratégias para tornar o valor destes clientes novamente positivo.

O valor do cliente é, então, a primeira referência que podemos utilizar para tomar decisões sobre quanto devemos investir em cada grupo de clientes para retê-los, se é viável realizar novas vendas para os mesmos, como criar ações para buscar novos clientes no mercado com o mesmo perfil.

Ciclo do Relacionamento

O ciclo de relacionamento é composto de três fases principais no relacionamento :

- Aquisição do cliente;
- Retenção do cliente;
- Venda adicional.

O processo de aquisição, onde é feita a primeira venda, inclui atividades que vão da pré-venda até a primeira entrega. O processo de retenção envolve ações de prevenção ao abandono da empresa pelo cliente e o processo de venda adicional é onde buscamos o aumento das compras dos clientes quando estes já estão estabilizados, ou seja, tem um risco pequeno de abandonar a empresa.

O ciclo de relacionamento precisa ser levado em conta nas ações de relacionamento,

pois os clientes que se encontram em cada um destas fases devem ser tratados de forma diferenciada.

Outro processo importante é o de servir o cliente, que começa com a entrada de cada pedido de compra.

Este processo de servir compreende verificar se o pedido está de acordo com as políticas da empresa e da existência dos produtos e serviços comprados, a entrega com sucesso dentro do prazo combinado, a cobrança do preço combinado e as atividades do pós-venda. Estas medidas devem ser ajustadas a medida do possível às necessidades de cada cliente, mas não deixando de focar a eficiência e baixo custo.

Portanto, é preciso ter um bom conhecimento do ciclo de relacionamento, as ações de aquisição, retenção e venda adicional precisam ser planejadas para cada perfil de cliente, para que a empresa possa explorar ao máximo o valor dos seus clientes com relação às vendas e aos custos envolvidos.

Para que as estratégias de relacionamento (ações de aquisição, retenção e vendas adicionais) tenham a eficácia desejada, é construída uma base analítica que fornece informações relevantes sobre os clientes, para apoiar o desenvolvimento destas estratégias.

2.7.2 A base analítica

A base analítica tem como objetivo fornecer informações necessários para o desenvolvimento de estratégias de relacionamento com clientes, bem como entender o comportamento passado de clientes, avaliar a capacidade da empresa atuar em seus mercados-alvo vendendo seus produtos/serviços e estimar comportamentos futuros de clientes. Costa (2004).

Para fornecer estas informações são contruídos modelos que utilizam dados, disponíveis na empresa, que são úteis para compreender o comportamento dos clientes, além de dados de mercado, que não se encontram nos sistemas operacionais da empresa.

No decorrer do desenvolvimento da base analítica são construídos três tipos de modelos: os modelos de primeira, de segunda e os de terceira geração.

Para a construção dos modelos de primeira geração são necessários dados de transações de cada cliente com a empresa e os custos necessários para atendê-los. Estes dados podem ser adquiridos a partir dos dados individuais de cada cliente, os quais estão disponíveis

no sistema de faturamento e cadastro da empresa, e dados gerais sobre custos que se encontram nos livros contábeis.

Os modelos de primeira geração, fornecem informações indispensáveis sobre os clientes, como a propensão ao abandono da empresa pelos clientes e a propensão a uma resposta positiva (compra de um produto).

Os modelos de segunda geração, têm a função de aprimorar os modelos estatísticos utilizando dados de relacionamento, que podem ser as respostas às ações desenvolvidas com o uso dos modelos de primeira geração.

Os modelos de terceira ordem são capazes de prever comportamentos e medir o potencial de compra, utilizando dados pessoais de clientes nas modelagens estatísticas.

A seguir mostramos como é construída a base analítica.

2.7.3 Construção da base analítica

Dois elementos essenciais para a análise são o valor de cada cliente e o tempo de relacionamento.

O valor do cliente pode ser usado como uma referência da importância do cliente para a empresa. Com base no valor do cliente a empresa pode saber o quanto pode investir no mesmo.

O tempo de relacionamento inicia com o ano em que o cliente foi "adquirido", ou seja, o ano em que ele realizou a primeira compra na empresa. O tempo de relacionamento é uma referência importante, pois permite analisar o comportamento dos clientes ao longo do ciclo de relacionamento e definir as ações adequadas a cada momento, além de permitir a criação de segmentos de clientes.

Análise preliminar dos dados

Após ter definido os segmentos dos clientes é feito um estudo dos dados disponíveis para identificar as variáveis mais importantes, o comportamento das variáveis e identificar as deficiências dos dados se este as possui.

Se os dados possuírem algum tipo de deficiência, analisamos se estas podem ser toleradas, corrigidas ou causar a retirada dos dados das análises.

Os dados podem trazer conhecimento das atividades e formas de relacionamento utilizadas no passado, mostrar as experiências que foram desenvolvidas com sucesso e as que fracassaram.

Modelos de Primeira Geração

Os modelos de primeira geração mais utilizados são os modelos preditivos de abandono e os modelos de resposta positiva a ofertas.

Modelos de abandono estabelecem o grau de propensão que cada cliente tem a abandonar a empresa (isto é, a probabilidade de cada cliente deixar de comprar produtos/serviços da empresa).

Tendo conhecimento da propensão ao abandono, a empresa pode desenvolver ações de retenção. E com o valor do cliente, a empresa pode limitar o gasto de retenção.

Os modelos preditivos de respostas prediz a propensão de cada cliente dar respostas positivas a ofertas. Tendo conhecimento da propensão do cliente dar uma resposta positiva a uma oferta é possível direcionar tais ofertas com maior probabilidade de aceitação pelo cliente.

Com estes modelos a empresa tem conhecimento da quantidade de clientes que é propenso a ofertas e o número de clientes que devem ser abordados com ações de retenção.

Para obter a propensão de abandono e a propensão a dar uma resposta positiva podemos utilizar modelos de regressão logística. O modelo logístico prediz a probabilidade de um determinado cliente vir a abandonar a empresa ou de um determinado cliente adquirir um produto. As probabilidades, para este novo grupo de clientes, podem ser ordenadas para melhor identificação de bons clientes com relação a não abandonar ou com relação a adquirir produtos.

A utilização de modelos logísticos para obter a propensão de resposta positiva e a de abandono dos clientes é apresentada nos dois casos a seguir.

2.7.4 Aplicação

· Caso 1

Suponha que possuímos uma amostra com $n = 300$ clientes de uma determinada empresa e a empresa deseja conhecer clientes com maior probabilidade a dar uma resposta positiva a uma oferta. Para cada um dos n clientes calculamos o valor do cliente considerando os últimos 4 meses, representada pela variável X_1 , digamos. Buscamos nos registros da empresa o salário mensal de cada cliente, representada pela variável X_2 , e observamos a variável resposta Y , com o valor 1 denotando que o cliente é respondente e o valor 0 denotando que o cliente é não respondente.

O objetivo do estudo é obter a probabilidade de clientes, não utilizados na construção do modelo, a dar uma resposta positiva a uma oferta. Para obter as probabilidades utilizamos um modelo de regressão logística representado por

$$\ln(\pi_i/(1 - \pi_i)) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, i = 1, 2, \dots, 300.$$

Neste caso, ajustando o modelo aos dados,

Y	X1	X2
1	376,07	1468,08
0	297,29	1484,68
1	418,62	1883,26
...
0	217,36	1675,59
1	286,81	1400,01

as estimativas para os parâmetros de regressão β_0 , β_1 e β_2 , são, $\hat{\beta}_0 = 2,8973$, $\hat{\beta}_1 = -0.0024$

e $\hat{\beta}_2 = -0.0011$, e a análise dos parâmetros estimados mostrada na tabela 5.

Tabela 5 - Análise dos parâmetros estimados.

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence	Chi-Square	Pr > ChiSq
Intercept	1	2.8973	1.0337	0.8712	4.9234	7.86	0.0051
x1	1	-0.0024	0.0011	-0.0046	-0.0002	4.70	0.0302
x2	1	-0.0011	0.0006	-0.0022	0.0000	3.76	0.0426

A partir da tabela 5, temos que, as variáveis preditoras x1 e x2, são significativas para o modelo a um nível de 5%, e obtemos os intervalos de 95% de confiança para os parâmetros estimados.

Para $\hat{\beta}_0$ o intervalo de 95% de confiança é [0.8712 ; 4.9234].

Para $\hat{\beta}_1$ o intervalo de 95% de confiança é [-0.0046 ; -0.0002].

Para $\hat{\beta}_2$ o intervalo de 95% de confiança é [-0.0022 ; 0.0000].

Com o modelo ajustado obtemos as probabilidades estimadas $\hat{\pi}_i, i = 1, 2, \dots, n$, que correspondem as probabilidades dos consumidores a dar respostas positivas.

Com as probabilidades calculadas podemos ordená-las e identificar os clientes mais prováveis a dar respostas positivas. Identificando estes clientes podemos direcionar ofertas, melhorando a probabilidade de aceitação pelos clientes.

· Caso 2

Suponha que uma empresa deseja conhecer os clientes com maior probabilidade ao abandono. Considerando uma amostra de $n = 300$ clientes calculamos os valores dos clientes levando em conta os últimos 4 meses, representada pela variável X_1 , digamos. Além disso, buscamos nos registros da empresa o salário mensal de cada cliente, representada pela variável X_2 . Observamos a variável binária Y , com o valor 1 denotando que o cliente abandonou a empresa e o valor 0 denotando que o cliente não abandonou a empresa.

O objetivo é obter a probabilidade de um grupo de clientes, não utilizados na construção do modelo, a abandonar a empresa. Estas probabilidades são preditas via modelo

de regressão logística dado por

$$\ln(\pi_i/(1 - \pi_i)) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, i = 1, 2, \dots, 300;$$

No ajuste do modelo aos dados,

Y	X1	X2
1	333,97	1650,41
1	486,14	1418,03
0	325,44	1434,64
...
0	276,11	1597,76
1	209,35	1475,23

obtemos as estimativas para os parâmetros β_0 , β_1 e β_2 , dados por $\hat{\beta}_0 = 3.4882$, $\hat{\beta}_1 = -0.0022$ e $\hat{\beta}_2 = -0.0014$, e a análise dos parâmetros estimados mostrada na tabela 6.

Tabela 6 - Análise dos parâmetros estimados.

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence	Chi-Square	Pr > ChiSq
Intercept	1	3.4882	1.1240	1.2852	5.6912	9.63	0.0019
x1	1	-0.0022	0.0012	-0.0044	0.0001	3.54	0.0500
x2	1	-0.0014	0.0006	-0.0025	-0.0002	5.03	0.0249

A partir da tabela 6, temos que, as variáveis predictoras x1 e x2, são significativas para o modelo a um nível de 5%, e obtemos os intervalos de 95% de confiança para os parâmetros estimados.

Para $\hat{\beta}_0$ o intervalo de 95% de confiança é [1.2852 ; 5.6912].

Para $\hat{\beta}_1$ o intervalo de 95% de confiança é [-0.0044 ; 0.0001].

Para $\hat{\beta}_2$ o intervalo de 95% de confiança é [-0.0025 ; -0.0002].

Com o ajuste do modelo podemos obter as probabilidades de abandono para um determinado grupo de clientes. Com o conhecimento destas probabilidades de abandono,

a empresa deve desenvolver ações de retenção.

· Modelos de Segunda Geração

Os modelos de segunda geração em geral apresentam aprimoramentos em relação aos modelos de primeira ordem.

Enquanto que nos modelos de primeira geração, o valor do cliente é calculado de forma simples, agora o valor do cliente é substituído pelo “life time value” (LTV) dos clientes, que é obtido pela diferença entre as receitas futuras esperadas e os respectivos custos para servir, durante o tempo esperado de retenção do cliente, tudo isso trazido para o valor presente.

O LTV não é utilizado nos modelos de primeira geração, por motivos de ocorrer muitos valores de clientes distorcidos por políticas e práticas inadequadas no passado. Quando estas distorções são identificadas com a utilização de modelos de primeira geração, estas devem ser corrigidas, para tornar possível o cálculo do LTV para sua utilização em modelos de segunda geração.

Os modelos de segunda geração são aperfeiçoados pelo uso de dados de relacionamento, como os dados de respostas a ações geradas pelos modelos de primeira ordem.

· Modelos de Terceira Geração

Os modelos de terceira geração utilizam dados individuais dos clientes. Com isto a empresa passa a conhecer os clientes de forma individual para prover a potencialidade de cada cliente.

Neste ponto é necessário modelos estatísticos avançados, que fazem uso do valor do cliente, das formas que ele gera este valor, e também da expectativa de valor do cliente no futuro, baseado nas suas características atuais.

Neste trabalho não desenvolvemos modelos de segundo e terceira geração, pelo motivo de estarmos interessados na análise dos clientes até a presente data e não em ocasiões futuras.

Os dados fornecidos pelos modelos descritos acima servem de base para apoiar o desenvolvimento de estratégias de relacionamento, como nos casos a seguir.

i) Com a segmentação dos clientes, com base no valor do cliente e na forma como

este valor é gerado, a empresa pode identificar as transações de um determinado cliente e identificar a forma de servir que melhor se adequa a este cliente,

ii) É possível também, considerando os segmentos de clientes, determinar o perfil do grupo de clientes dentro do segmento e, com base neste perfil, buscar no mercado potenciais novos clientes, ou seja, clientes que apresentam um perfil semelhante ao perfil dos clientes mais atraentes. Esta ação é denominada aquisição de novos clientes. Realizando as ações de aquisição utilizando informações provenientes do perfil do segmento, o risco de aquisição de clientes que possam apresentar valor negativo para a empresa diminui, e o número de clientes com qualidade aumenta.

iii) As ações de vendas adicionais podem ser realizadas considerando a propensão do cliente dar resposta positiva, para direcionar ofertas específicas com maior chance de serem aceitas.

Capítulo 3

VALIDAÇÃO DE MODELOS DE REGRESSÃO LOGÍSTICA VIA BOOTSTRAP

Na construção de um modelo de regressão logística o passo inicial é dividir a amostra em questão em duas novas amostras distintas, sendo uma amostra de treinamento para a construção do modelo e uma outra para a validação do modelo.

A medida utilizada para avaliar a qualidade do modelo é a *cum lift* (cumulative lift), obtida a partir de uma análise decil, o *cum lift* indica o ganho esperado de uma solicitação implementada utilizando um modelo sobre uma solicitação implementada sem o modelo.

Neste estudo, uma solicitação significa um contato com o cliente buscando uma resposta positiva ou negativa, por exemplo, realizar uma oferta e observar uma resposta positiva, a aquisição do produto pelo cliente, ou uma resposta negativa, não aquisição do produto.

A avaliação é feita através da comparação das medidas *cum lifts* obtidas das análises decil de treinamento e validação. Espera-se observar uma diminuição dos *cum lifts* da análise decil de validação em relação aos *cum lifts* da análise decil de treinamento, se a diminuição dos *cum lifts* e os valores dos *cum lifts* são aceitáveis então o modelo é considerado válido, mas o problema é que uma única amostra não fornece uma medida de variabilidade, que permita uma afirmação sobre o nível de confiança em relação a validação.

Sendo assim, se amostras adicionais estiverem disponíveis, novas análises decis de validação podem ser realizadas e desta forma a comparação das medidas *cum lifts* obtidas da análise decil de treinamento e das análises decil de validação utilizando as amostras adicionais.

Desta forma com várias análises de validação decil tem-se uma variação do *cum lift* dentro do decil, o que conduz a incertezas das estimativas dos *cum lifts*, tendo mais (menos) confiança na estimativa se ocorrer uma pequena (grande) variação, com isso algumas questões podem ser levadas em consideração tais como:

- Como um "*cum lift médio*" pode ser definido através das várias estimativas, para servir como uma estimativa mais confiável do *cum lift* para um determinado decil?

- O valor calculado para uma estatística como o *cum lift*, pode ser considerado como uma estimativa pontual, que nos propicia uma idéia do verdadeiro valor da estatística, mas há a necessidade de quantificar a certeza associada a tais estimativas pontuais, através de um intervalo de confiança.

- Como o desvio padrão para a estimativa da medida *cum lift* pode ser obtido, para a avaliação da variabilidade da estimativa?

- Tendo a disposição um único conjunto de dados, é possível obter uma estimativa honesta do *cum lift* e seu desvio padrão?

Para responder tais questões, pode-se fazer uso da metodologia bootstrap (ver seção 3.2).

Neste trabalho foi utilizado um modelo de regressão logística (ver seção 2.4), que foi considerado como um modelo preditivo de respostas, isto é, o modelo prediz qual é a probabilidade do cliente dar uma resposta positiva a certa oferta. O modelo foi avaliado de acordo com as medidas *cum lift* obtidas de uma análise decil baseada no modelo em questão e utilizamos a metodologia bootstrap para obter uma estimativa para a medida *cum lift* e quantificar a certeza associada a tal estimativa.

3.1 Análise decil e cum lift para o modelo de regressão logística

A análise decil para o modelo de regressão logística é um método que fornece uma tabela contendo o desempenho do modelo considerado. Esta tabela consiste dos decis, número de indivíduos, número de respondentes, taxa de resposta decil, taxa de resposta cumulativa e *cum lift*.

3.1.1 Exemplo 1

Neste exemplo ilustramos a construção e interpretação da análise decil para o modelo de regressão logística.

Considere uma amostra com $n = 1000$ consumidores de uma determinada empresa, para cada um dos n consumidores dispomos de duas variáveis, a variável binária Y , com o valor 1 denotando que o consumidor é respondente e o valor 0 denotando que o consumidor é não respondente, e a variável X representando o lucro (relativa aos últimos três meses) que cada um dos n consumidores propiciou à empresa.

Considerações:

1. A variável resposta $Y_i = y_i, i = 1, 2, 3, \dots, n$ está relacionada com o fato do consumidor ser respondente ou não, onde o valor 1 denota que o consumidor é respondente e o valor 0 denota que o consumidor é não respondente.

2. Distribuição da variável resposta: Binomial

3. Função de Ligação: função logística $\ln\left(\frac{\pi_i}{1-\pi_i}\right), i = 1, 2, 3, \dots, n$

4. Variável explicativa $X_i = x_i, i = 1, 2, 3, \dots, n$ representando o lucro que o i -ésimo consumidor propiciou à empresa.

Objetivo: obter a análise decil para o modelo de regressão logística.

O modelo de regressão logística é dado por:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_i, \quad i = 1, 2, 3, \dots, n$$

A construção da análise decil, mostrada na Tabela 7, consiste de oito passos:

Tabela 7 - Análise decil para o modelo de regressão logística.

<i>Decil</i>	<i>N.I.</i>	<i>N.R.</i>	<i>T.R.D.</i>	<i>T.R.C.</i>	<i>Cum Lift</i>
<i>Top</i>	100	60	60%	60%	123
2	100	48	48%	54%	111
3	100	55	55%	54.33%	111
4	100	48	48%	52.75%	108
5	100	54	54%	53%	109
6	100	50	50%	52.5%	108
7	100	47	47%	51.71%	106
8	100	44	44%	50.75%	104
9	100	42	42%	49.78%	102
<i>Bottom</i>	100	40	40%	48,8%	100
<i>Total</i>	1000	488	48,8%		

Passo 1: Ajuste o modelo logístico aos dados e obtenha a probabilidade de resposta de cada indivíduo.

Passo 2: Ordene as probabilidades de resposta em ordem decrescente.

Passo 3: Divida a amostra de probabilidades em 10 grupos iguais, desta forma a variável decil é criada assumindo 10 valores ordenados *top(1)*, 2, 3, 4, 5, 6, 7, 8, 9 e *bottom(10)*. O top decil contém os 10% dos indivíduos com maior probabilidade de resposta, o decil 2 consiste dos próximos 10% dos indivíduos com maior probabilidade de resposta. E, similarmente para o decil 3, 4, 5, 6, 7, 8, 9 e *bottom(10)*.

Passo 4: Número de Indivíduos (N. I): número de indivíduos em cada decil, ou seja 10% do total de indivíduos.

Passo 5: Número de Respondentes (N. R.): número de respondentes em cada decil. Neste exemplo o modelo identificou 60 respondentes no decil top e 48 respondentes no decil 2. O número de respondentes para os demais decis é analisado da forma similar.

Passo 6: Taxa de Resposta Decil (T. R. D.): taxa de resposta para cada decil, dada pelo número de respondentes dividido pelo número de indivíduos para um determinado decil. Para o decil top a taxa de resposta é $60\% = 60/100$.

Passo 7: Taxa de Resposta Cumulativa (T. R. C.) para uma dada parte da amostra de consumidores (deciles cumulativos): taxa de resposta entre os indivíduos nos decis cumulativos. Para o decil top, ou seja, 10% da amostra de consumidores, a taxa de resposta cumulativa é $60\% (= [60/100] * 100)$, para os dois decis superiores, ou seja, 20% da amostra de consumidores, a taxa de resposta cumulativa é $54\% (= [60 + 48]/[100 + 100] * 100)$.

Passo 8: *Cum Lift* (C. L.) para uma dada parte da amostra de consumidores é dado pela taxa de resposta cumulativa dividido pela taxa de resposta geral da amostra e, então, multiplicado por 100. A taxa de resposta geral da amostra de consumidores é dada pelo número total de respondentes dividido pelo número de indivíduos na amostra. Para o decil top (10% melhores consumidores da amostra) o *cum lift* 123 indica que ao solicitarmos 10% da amostra, baseando-se no modelo, é esperado 1,23 vezes o número total de respondentes encontrados por uma solicitação aleatória. O *cum lift* 111 para os dois decis superiores (20% melhores consumidores da amostra), indica que ao solicitarmos 20% da amostra, baseando-se no modelo, é esperado 1,11 vezes o número total de respondentes encontrados por uma solicitação aleatória.

3.2 Metodologia Bootstrap

O método bootstrap é o método computacional mais popular para reamostrar uma amostra. A reamostragem é realizada através de uma seleção aleatória dos elementos da amostra com reposição da mesma, onde um elemento da amostra pode se repetir na nova amostra ou não aparecer.

As amostras bootstrap de mesmo tamanho são ligeiramente diferentes entre si, ocorrendo uma variação entre as amostras. Com isto uma distribuição empírica amostral da estatística desejada pode ser construída e, assim, estimativas de viés e variabilidade podem ser determinadas.

A técnica bootstrap pode ser utilizada para avaliar a precisão de qualquer estatística, servindo como alternativa para os métodos paramétricos tradicionais, e também pode ser utilizado no estudo de estatísticas como o *cum lift* onde os métodos paramétricos tradicionais não se adequam.

Um algoritmo bootstrap utilizado para encontrar uma estimativa para a medida *cum lift*, seu desvio padrão e um intervalo de confiança para a estatística em estudo é mostrado a seguir.

3.3 Construção do intervalo de confiança via Bootstrap

A metodologia bootstrap considera que a amostra é a melhor estimativa para uma população desconhecida.

Nesta metodologia a amostra é tratada como uma população, com isto gera-se repetidamente amostras aleatórias de mesmo tamanho com reposição da amostra original. A partir destas amostras bootstrap a distribuição amostral da estatística desejada é estimada e, portanto, é possível calcular uma estimativa bootstrap da estatística em estudo e encontrar o intervalo de confiança para tal estatística.

Um algoritmo bootstrap pode ser dado como:

ALGORITIMO:

1. A ponte para a estatística desejada, digamos, Y
2. Trate a amostra como população.
3. Calcule Y na amostra/população; denote-o de SAM_EST .
4. Gere uma amostra bootstrap da população, isto é, uma seleção aleatória com reposição de tamanho n , do tamanho da amostra original.
5. Calcule Y na amostra bootstrap para produzir um pseudo valor, denote-o de BS_1 .
6. Repita os passos 4 e 5 “ m ” vezes.
7. Após repetir os passos 4 e 5 “ m ” vezes teremos: BS_1, BS_2, \dots, BS_m .
8. Calcule a estimativa bootstrap da estatística:

$$BS_{est}(Y) = 2 * SAM_EST - média(BS_i)$$

9. Calcule a estimativa bootstrap do desvio padrão da estatística:

$$SE_{BS}(Y) = desvio\ padrão\ de\ (BS_i)$$

10. O intervalo de confiança bootstrap é

$$BS_{est}(Y) \pm |Z_{0.025}| * SE(Y)$$

Também pode-se estimar intervalos de confiança através dos quantis da amostra gerada BS_1, BS_2, \dots, BS_m .

3.4 Análise decil de validação Bootstrap

No exemplo anterior foi obtida a estimativa pontual da estatística *cum lift* para cada decil, mas há a necessidade de quantificar a certeza associada a tais estimativas pontuais, isto é encontrar um intervalo de confiança para a estatística em estudo.

Utilizando a metodologia bootstrap, podemos construir intervalos de confiança para as estimativas *cum lift* e através das várias estimativas para o *cum lift* fornecida pela utilização do procedimento bootstrap podemos calculada uma estimativa mais confiável para cada decil chamada de *cum lift bootstrap* (C. L. B.).

Considerando o modelo utilizado anteriormente e a mesma amostra de tamanho $n = 1000$, utilizado o procedimento bootstrap de 10 passos para implementar uma "análise decil de validação bootstrap".

Utilizamos 1000 amostras bootstrap, todas com o mesmo tamanho da amostra original ($n = 1000$).

Para o exemplo 1, o *cum lift bootstrap* para o top decil é 122,92 e tem uma margem de erro de 19,02, desta forma para o top decil, o intervalo de confiança bootstrap com 95% de confiança é $122,92 \pm 19,02$, logo o intervalo de confiança bootstrap é (103,90; 141,94)

O decil 2 tem um *cum lift bootstrap* de 110,73 e tem uma margem de erro de 12,97, e o intervalo de confiança bootstrap com 95% de confiança vai de 97,76 à 123,69.

Esta validação bootstrap indica que o *cum lift bootstrap* é 111,50 utilizando o modelo para selecionar 30% dos indivíduos com maior probabilidade de resposta na amostra e

espera-se que o *cum lift bootstrap* fique entre 101,66 e 121,35.

Tabela 8 - Análise decil de validação bootstrap (amostra de tamanho 1000).

<i>Decil</i>	<i>N.I.</i>	<i>N.R.</i>	<i>T.R.D.</i>	<i>T.R.C.</i>	<i>C.L.</i>	<i>C.L</i>	<i>M.E.</i>	<i>I.C.95%</i>
					<i>M.</i>	<i>B.</i>		
<i>Top</i>	100	60	60%	60%	123	122,92	19.02	(103,90; 141,94)
2	100	48	48%	54%	111	110,73	12.97	(97,76; 123,69)
3	100	55	55%	54.33%	111	111,50	9.85	(101,66; 121,35)
4	100	48	48%	52.75%	108	108,11	7.83	(100,29; 115,94)
5	100	54	54%	53%	109	108,58	6.37	(102,21; 114,96)
6	100	50	50%	52.5%	108	107,51	5.14	(102,37; 112,65)
7	100	47	47%	51.71%	106	105,87	4.11	(101,76; 109,98)
8	100	44	44%	50.75%	104	103,98	3.15	(100,86; 107,13)
9	100	42	42%	49.78%	102	101,96	2.05	(99,91; 104,01)
<i>Bottom</i>	100	40	40%	48,8%	100	100	0	(100; 100)
<i>Total</i>	1000	488	48,8%					

onde C.L.M. corresponde ao *cum lift do modelo*.

3.5 Outras questões à serem analisadas

Com o intervalo de confiança já calculado surgem questões do tipo :

Se a margem de erro obtida (consequentemente o intervalo de confiança) é grande para satisfazer os objetivos do negócio em questão, o que pode ser feito?

A resposta para esta questão se encontra no relacionamento fundamental entre o tamanho da amostra e a extensão do intervalo de confiança, que fornece a seguinte relação.

- Aumentando o tamanho da amostra, aumenta-se a confiança na estimativa.
- Diminuindo o tamanho da amostra, diminui-se a confiança na estimativa.

de maneira equivalente,

- Aumentando o tamanho da amostra, diminui-se o erro padrão.
- Diminuindo o tamanho da amostra, aumenta-se o erro padrão.

Esta relação entre tamanho da amostra e a extensão do intervalo de confiança pode

ser utilizada para controlar a confiança na estimativa bootstrap do *cum lift*.

Para aumentar a confiança na estimativa bootstrap, a relação pode ser utilizada da seguinte forma:

- Se um número grande de consumidores está disponível, este conjunto pode ser adicionado na amostra de validação até obter uma amostra com um determinado tamanho que propicie a margem de erro e a extensão do intervalo de confiança desejados.

Com um modelo tido como válido, ou seja pronto para a implementação, a questão que se faz presente é a seguinte:

- Qual é o tamanho necessário da amostra para implementar uma solicitação com um nível aceitável de confiança baseando-se no modelo, e obter um valor desejado da medida *cum lift*, ou seja, maximizar o desempenho para uma dada parte da amostra (10%, 20%, ..., ou 100% da amostra)?

O relacionamento fundamental entre o tamanho da amostra e a extensão do intervalo de confiança também pode ser utilizado na questão do tamanho da amostra necessário para maximizar a quantidade de desempenho. Determinar o tamanho da amostra na verdade é obter a menor amostra que fornece o valor *cum lift* desejado.

Esta questão é trabalhada da seguinte forma.

- Identifique o decil e o intervalo de confiança que contenham o valor *cum lift* o mais próximo possível do valor *cum lift* desejado, baseando-se na análise decil de validação a disposição.

- Se a extensão do intervalo de confiança for aceitável, então a amostra exigida tem o mesmo tamanho da amostra de validação.

- Se o intervalo de confiança é muito grande, então aumenta-se o tamanho da amostra de validação adicionando consumidores e *bootstrapping* esta nova amostra, adiciona-se novos elementos até que a amostra forneça a extensão desejada do intervalo de confiança. A amostra exigida tem o mesmo tamanho da amostra de validação que forneceu a extensão desejada.

- Se o intervalo de confiança for pequeno, isto indica que uma amostra com um tamanho menor que o tamanho da amostra de validação pode ser utilizada, desta forma diminui-se o tamanho da amostra de validação e *bootstrapping* a nova amostra, realiza-se este procedimento até que a amostra forneça a extensão desejada do intervalo. A amostra

exigida tem o mesmo tamanho da amostra de validação que forneceu a extensão desejada do intervalo de confiança.

O exemplo, a seguir ilustra o uso de análise decil de validação bootstrap para avaliar o modelo utilizado no Exemplo 1. Os intervalos de confiança são também obtidos.

3.6 Exemplo 2

Considere o modelo de regressão logística utilizado no exemplo 1, com uma taxa de resposta geral de 48,8%.

A análise decil de validação baseada em uma amostra de tamanho $n = 1000$, juntamente com uma estimativa bootstrap se encontra na Tabela 8.

As margens de erro de 95% de confiança para os quatro decis superiores, e o intervalo de 95% confiança são grandes para utilizar o modelo em questão.

Da Tabela 8 temos:

Para o decil *top* um intervalo com 95% de confiança [103,90 ; 141,94], um *cum lift* bootstrap 122,92 e uma margem de erro 19,02.

Para o *decil 2* um intervalo com 95% de confiança [97,76 ; 123,69], um *cum lift* bootstrap 110,73 e uma margem de erro 12,97.

Para o *decil 3* um intervalo com 95% de confiança [101,66 ; 121,35], um *cum lift* bootstrap 111,50 e uma margem de erro 9,85.

Para o *decil 4* um intervalo com 95% de confiança [100,29 ; 115,94], um *cum lift* bootstrap 108,11 e uma margem de erro 7,83.

Pelo fato dos intervalos serem grandes, é criada uma nova amostra de tamanho 1000, esta amostra é acrescentada na primeira amostra de tamanho $n = 1000$. Desta forma temos, agora, uma amostra $n_1 = 2000$ e um novo modelo de regressão logística, dado por

$$\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_i, \dots, i = 1, 2, 3, \dots, n_1$$

é ajustado a esta nova amostra. Obtemos as estimativas $\hat{\beta}_0 = 0.4526045$ e $\hat{\beta}_1 = -0.0017618$, para β_0 e β_1 , respectivamente.

Considerando este novo modelo e a amostra de tamanho $n_1 = 2000$, implementa-

se o procedimento bootstrap de 10 passos para viabilizar uma "análise decil de validação bootstrap". Os resultados se encontram na Tabela 9. Utilizou-se 1000 amostras bootstrap, todas com o mesmo tamanho da amostra original ($n_1 = 2000$).

As margem de erro e o intervalo de confiança para a *decil top* e para o *decil 2* ainda são inaceitáveis, grandes amplitudes para os nossos objetivos,

Tabela 9 - Análise decil de validação bootstrap (amostra de tamanho 2000).

<i>Decil</i>	<i>N.I.</i>	<i>N.R.</i>	<i>T.R.D.</i>	<i>T.R.C.</i>	<i>C.L.</i>		<i>M.E.</i>	<i>I.C.95%</i>
					<i>M.</i>	<i>B.</i>		
<i>Top</i>	200	121	60.5%	60.5%	126	126,02	13,20	(112,82; 139,22)
2	200	103	51.5%	56%	116	116,60	8,85	(107,75; 125,45)
3	200	99	59.5%	53.83%	112	111,97	6,92	(105,05; 118,89)
4	200	91	45.5%	51.75%	108	107,57	5,46	(102,11; 113,02)
5	200	94	47%	50.8%	106	105,54	4,60	(100,94; 110,14)
6	200	112	56%	51.67%	107	107,38	3,74	(103,63; 111,12)
7	200	92	46%	50.86%	106	105,71	2,97	(102,75; 108,68)
8	200	86	43%	49.88%	104	103,66	2,24	(101,42; 105,90)
9	200	83	41.5%	48.94%	102	101,75	1,45	(100,30; 103,19)
<i>Bottom</i>	200	81	40.5%	48.1%	100	100	0	(100; 100)
<i>Total</i>	2000	962	48.1%					

Cria-se, então, uma amostra de tamanho 3000, esta amostra é incluída na amostra de tamanho n_1 , aumentando o tamanho da amostra para $n_2 = 5000$. Com isto obtemos outro modelo de regressão logística dado por:

$$\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_i, \dots, i = 1, 2, 3, \dots, n_2$$

Este modelo é ajustado considerando a amostra de tamanho n_2 , de onde obtemos as estimativas $\hat{\beta}_0 = 0.1977388$ e $\hat{\beta}_1 = -0.0007915$.

Com a amostra $n_2 = 5000$ e o modelo acima, implementa-se o procedimento bootstrap de 10 passos para viabilizar uma "análise decil de validação bootstrap". Utilizou-se 1000 amostras bootstrap, todas com o mesmo tamanho da amostra original ($n_2 = 5000$). Os

resultados desta análise estão presentes na Tabela 10.

Observa-se que a margem de erro e o intervalo com 95% de confiança para o decil *top*, ainda são inaceitáveis.

Tabela 10 - Análise decil de validação bootstrap (amostra de tamanho 5000).

<i>Decil</i>	<i>N.I.</i>	<i>N.R.</i>	<i>T.R.D.</i>	<i>T.R.C.</i>	<i>C.L.</i>		<i>M.E.</i>	<i>I.C.95%</i>
					<i>M.</i>	<i>B.</i>		
<i>Top</i>	500	281	56.2%	56.2%	115	114,79	8,37	(106,42;123,16)
2	500	257	51.4%	53.8%	110	109,85	5,65	(104,20;115,50)
3	500	242	48.4%	52%	106	106,20	4,27	(101,93;110,47)
4	500	222	44.4%	50.1%	102	102,28	3,45	(98,83;105,72)
5	500	243	48.6%	49.8%	102	101,62	2,75	(98,88;104,37)
6	500	260	52%	50.17%	102	102,34	2,21	(100,13;104,55)
7	500	256	51.2%	50.31%	103	102,67	1,81	(100,87;104,48)
8	500	230	46%	49.78%	102	101,58	1,43	(100,15;103,00)
9	500	220	44%	49.13%	100	100,26	0,95	(99,30;101,21)
<i>Bottom</i>	500	239	47.8%	49%	100	100	0	(100;100)
<i>Total</i>	5000	2450	49%					

O estudo prossegue criando uma amostra de tamanho 5000, que é acrescentada a amostra de tamanho $n_2 = 5000$. Obtemos então uma nova amostra de tamanho $n_3 = 10000$ e um novo modelo de regressão logística, dado por

$$\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_i, \quad i = 1, 2, 3, \dots, n_3$$

é ajustado à amostra de tamanho n_3 . Obtemos as estimativas $\hat{\beta}_0 = 0.1425257$ e $\hat{\beta}_1 = -0.0006115$.

Utilizamos novamente 1000 amostras bootstrap, todas com o mesmo tamanho da

amostra original ($n_3 = 10000$). A Tabela 11 apresenta os resultados para esta análise.

Tabela 11 - Análise decil de validação bootstrap (amostra de tamanho 10000).

<i>Decil</i>	<i>N.I.</i>	<i>N.R.</i>	<i>T.R.D.</i>	<i>T.R.C.</i>	<i>C.L.</i>		<i>M.E.</i>	<i>I.C.95%</i>
					<i>M.</i>	<i>B.</i>		
<i>Top</i>	1000	532	53.2%	53.20%	109	108,59	5,82	(102,78; 114,41)
2	1000	503	50.3%	51.75%	106	105,66	3,88	(101,79; 109,54)
3	1000	496	49.6%	51.03%	104	104,20	2,98	(101,23; 107,18)
4	1000	479	47.9%	50.25%	103	102,62	2,42	(100,21; 105,04)
5	1000	500	50%	50.20%	103	102,50	1,97	(100,53; 104,46)
6	1000	491	49.1%	50.02%	102	102,15	1,61	(100,54; 103,77)
7	1000	509	50.9%	50.14%	102	102,42	1,30	(101,12; 103,73)
8	1000	450	45%	49.5%	101	101,10	0,99	(100,11; 102,09)
9	1000	450	45%	49%	100	100,08	0,67	(99,42; 100,75)
<i>Bottom</i>	1000	487	48.7%	48.97%	100	100	0	(100; 100)
<i>Total</i>	10000	4897	48.97%					

Os intervalos com 95% de confiança e as margens de erro para os quatro primeiros deciles são:

Para o decil *top* o intervalo de 95% de confiança é [102,78 ; 114,41] e margem de erro 5,82.

Para o decil 2 o intervalo de 95% de confiança é [101,79 ; 109,54] e margem de erro 3,88.

Para o decil 3 o intervalo de 95% de confiança é [101,23 ; 107,18] e margem de erro 2,98.

Para o decil 4 o intervalo de 95% de confiança é [100,21 ; 105,04] e margem de erro 2,42.

3.6.1 Conclusão:

Com os intervalos com 95% de confiança e as margens de erro, presentes na Tabela 11, podemos utilizar o modelo de regressão logística

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_i, \quad i = 1, 2, 3, \dots, n_3,$$

com as estimativas $\hat{\beta}_0 = 0.1425257$ e $\hat{\beta}_1 = -0.0006115$, onde $n_3 = 10000$, para implementar uma solicitação tendo mais confiança nas medidas *cum lift* e, com isto, obter um número satisfatório de respostas positivas.

Esta metodologia pode ser utilizada com um número maior de variáveis preditoras.

3.7 Análise Gráfica dos Decis

A análise gráfica dos decis, permite ter uma melhor idéia do tamanho amostral a ser utilizado em um estudo de desempenho para obter as margens de erro desejadas.

Os gráficos são construídos considerando o *tamanho amostral* e a *margem de erro*, e a proporção de respondentes na amostra. Neste estudo são apresentados quatro gráficos, gráfico 1,2,3 e 4, construídos a partir de amostras contendo 5%, 10%, 15% e 20% respondentes respectivamente. Cada curva representa um decil e cada ponto indica o tamanho amostral e a margem de erro obtida, para um determinado decil.

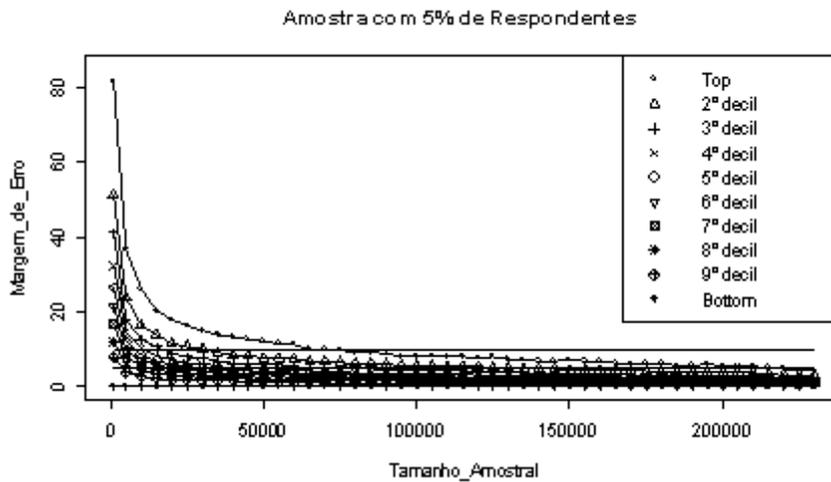


Gráfico 1 - Tamanho Amostral vs Margem de Erro para Amostras com 5% de respondentes

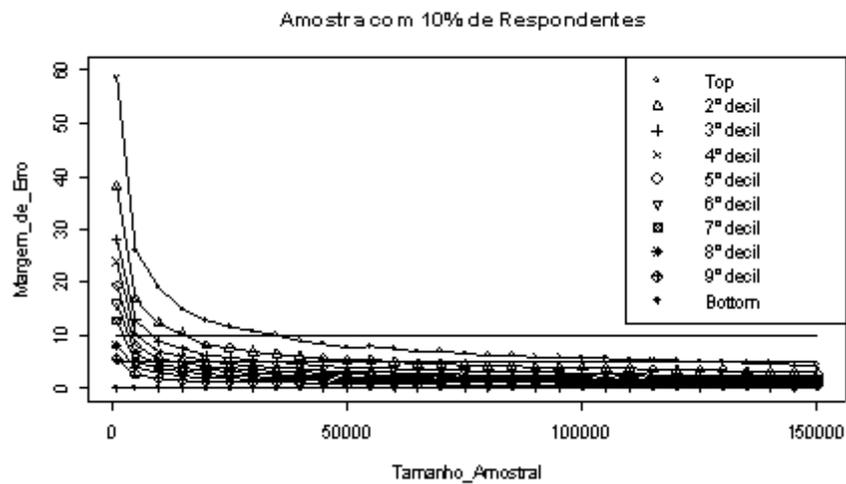


Gráfico 2 - Tamanho Amostral vs Margem de Erro para Amostras com 10% de respondentes

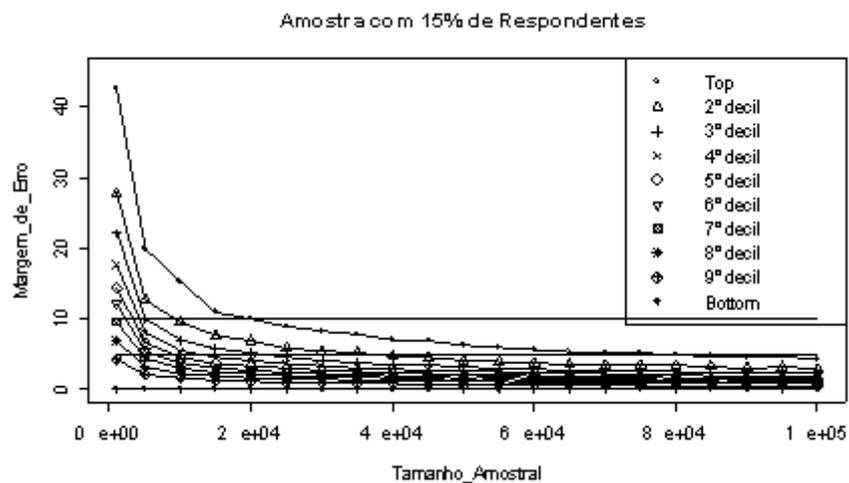


Gráfico 3 - Tamanho Amostral vs Margem de Erro para Amostras com 15% de respondentes

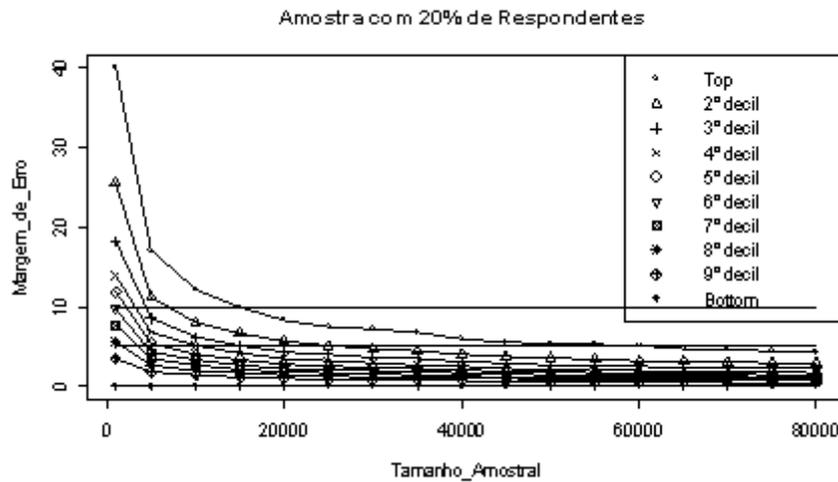


Gráfico 4 - Tamanho Amostral vs Margem de Erro para Amostras com 20% de respondentes

3.7.1 Construção dos gráficos

A partir do Exemplo 2 notamos que é necessário aumentar o tamanho amostral para obter margens de erro satisfatórias. Os gráficos são importantes nesta etapa pois auxiliam na escolha do tamanho amostral para obter determinadas margens de erro desejadas.

Para a construção dos gráficos considere, por exemplo, uma amostra com $n = 1000$ consumidores de uma determinada empresa, com uma proporção de respondentes igual a z , $z = 5\%, 10\%, 15\%$ ou 20% . Para cada um dos n consumidores dispomos de duas variáveis, a variável binária Y , com o valor 1 denotando que o consumidor é respondente e o valor 0 denotando que o consumidor é não respondente, e a variável X representando o lucro (relativa aos últimos três meses) que cada um dos n consumidores propiciou à empresa.

Um modelo de regressão logística é ajustado aos dados para obter a análise decil, após realizada a análise decil o modelo é validado utilizando o procedimento bootstrap de onde obtém-se as margens de erro e o intervalo de confiança para a estimativa *cum lift* de cada decil.

Obtidas as margens de erro, utilizando a amostra de tamanho igual a 1000, estas são

arquivadas.

Em seguida é acrescentada uma nova amostra de tamanho 4000 à amostra original, criando-se desta forma uma amostra de tamanho $n_2 = 5000$, mantendo a proporção de respondentes igual a z , o processo para obter as margens de erro é realizado e estas são novamente arquivadas.

O processo para obter as margens de erro continua, aumentando-se a amostra de 5000 em 5000 clientes.

Após obtidas as margens de erro, é criado o gráfico de margem de erro versus tamanho amostral mostrados nos gráficos 1, 2, 3 e 4.

3.7.2 Utilização dos Gráficos

Como citado acima, os gráficos são utilizados na orientação da escolha do tamanho amostral, necessário para obter uma determinada margem de erro, em uma amostra com proporção z de respondentes.

Para um melhor entendimento da utilização dos gráficos, é apresentado um caso onde é utilizado o gráfico para amostras com 20% de respondentes.

Suponha que uma empresa deseja abordar 20% de seus melhores clientes com uma ação de marketing e esperamos observar margens de erro menores ou iguais a 10.

Neste caso o Gráfico 5, mostrado abaixo, pode ser utilizado para orientar na escolha do tamanho amostral.

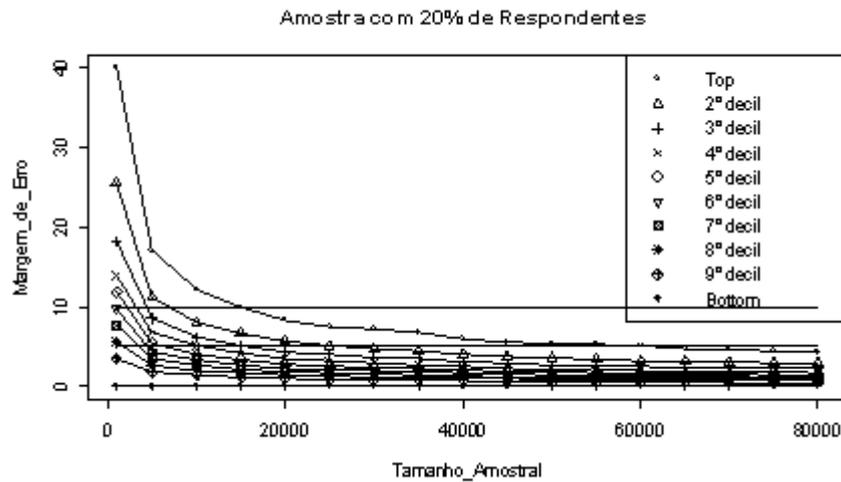


Gráfico 5 - Tamanho Amostral vs Margem de Erro.

Observando no Gráfico 5 a segunda curva de cima para baixo, que representa o 2º decil, concluimos que o tamanho amostral, necessário para obter margens de erro menores ou iguais a 10 é $n \geq 15000$.

Buscando uma amostra com $n = 15000$ consumidores, e, para cada um dos n consumidores, dispomos de duas variáveis, a variável binária Y , com o valor 1 denotando que o consumidor é respondente e o valor 0 denotando que o consumidor é não respondente, e a variável X representando o lucro (relativa aos últimos três meses) que cada um dos n consumidores propiciou à empresa.

Um modelo logístico dado por,

$$\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_i, \dots, i = 1, 2, 3, \dots, n,$$

é ajustado aos dados, considerando a amostra de tamanho n ,

Y	X1
0	243,84
0	374,57
0	277,04
...
0	298,39
0	444,33

de onde obtemos as estimativas $\hat{\beta}_0 = -1,4184931$ e $\hat{\beta}_1 = 0,0001812$.

Com a amostra $n = 15000$ e o modelo acima, implementa-se o procedimento bootstrap de 10 passos para viabilizar uma análise decil de validação bootstrap, mostrada na Tabela 12. Utilizou-se 1000 amostras bootstrap, todas com o mesmo tamanho da amostra original ($n = 15000$).

Na Tabela 12, está apresentada a análise decil de validação bootstrap, onde se encontram as margens de erro e os intervalos com 95% de confiança para os decis.

Tabela 12 - Análise decil de validação bootstrap (amostra de tamanho 15000).

Decil	N.I.	N.R.	T.R.D.	T.R.C.	C.L.	C.L.	M.E.	I.C.95%
					M.	B.		
Top	1500	329	21,93%	21,93%	107.73	107.44	9.59	(97,85 ; 117,04)
2	1500	312	20,80%	21,367%	104.94	104.87	6.14	(98,73 ; 111,01)
3	1500	299	19,93%	20,89%	102.59	102.56	4.66	(97,89 ; 107,22)
4	1500	313	20,87%	20,88%	102.57	102.53	3.91	(98,62 ; 106,44)
5	1500	280	18,67%	20,44%	100.39	100.39	3.25	(97,14 ; 103,64)
6	1500	301	20,07%	20,378%	100.09	100.12	2.74	(97,38 ; 102,86)
7	1500	292	19,47%	20,25%	99.45	99.46	2.15	(97,31 ; 101,61)
8	1500	300	20,00%	20,22%	99.29	99.30	1.59	(97,71 ; 100,89)
9	1500	313	20,87%	20,29%	99.65	99.66	1.04	(98,62 ; 100,69)
Bottom	1500	315	21,00%	20,36%	100	100	0	(100 ; 100)
Total	15000	3054						

Analisando a Tabela 12, observamos que as margens de erro são menores que 10, o que era desejado pela empresa, com estes resultados concluímos que o tamanho amostral a ser escolhido, através do auxílio do gráfico, é $n \geq 15000$.

Capítulo 4

ANÁLISE CONJUNTA - *CONJOINT ANALYSIS*

4.1 Introdução

Em marketing buscamos o conhecimento das características de um determinado produto ou serviço que são mais valorizadas pelos consumidores. A idéia é desenvolver produtos/serviços que contenham essas características, visando uma melhor aceitação dos mesmos por parte dos consumidores.

Uma técnica que podemos utilizar para estabelecer a preferência dos consumidores por certas características de um produto/serviço é a *Análise Conjunta*, desenvolvida por volta de 1970. A *análise conjunta* é uma técnica de pesquisa de mercado que investiga a estrutura da preferência dos clientes.

Em análise conjunta, o produto (ou serviço) é definido através de suas características (denominadas atributos). Supondo que o produto seja um tênis temos, por exemplo, os atributos materiais, preço, marca etc. Os possíveis valores/estados que os atributos podem assumir recebem o nome de níveis de atributos (material pode ter os níveis couro, nylon, couro e nylon, etc). As combinações formadas pelos níveis dos atributos são chamadas de perfis ou estímulos (“tênis de couro, preço R\$ 70,00, marca reconhecida”).

A preferência de um indivíduo pode ser mensurada através de notas, ordenamento, por soma-constante, em que se avalia um conjunto de atributos por pontos/notas, de forma

que, quanto mais pontos maior a preferência, ou ainda, através da escolha do estímulo preferido em um conjunto de estímulos (escolha discreta).

Neste estudo a tarefa dos consumidores, ou respondentes é a de ordenar os perfis da menor à de maior preferência.

A idéia da *análise conjunta* é permitir que os respondentes façam suas escolhas de maneira similar a uma situação real de compra, comparando as varias opções de produtos, com base em critérios subjetivos de valor que atribuem aos níveis considerados mais importantes. A importância ou utilidade de cada nível, também conhecidas como *part-worth utilities*, é inferida a partir da ordenação de preferência dos produtos, utilizando um modelo de regressão com variáveis dummy, ao invés de ser indagada diretamente ao cliente.

O cálculo das utilidades pode ser feito de duas maneiras.

i) A estimação é feita através do cálculo do valor médio da classificação, para cada atributo (variável dependente). Ou obtendo o número de perfis em que um determinado nível está presente e em seguida, buscando a classificação de cada perfil. A utilidade fica representada pela seguinte expressão.

$$U_i = \frac{\sum_{j=1}^n C_j}{n},$$

onde U_i se refere a utilidade do i-ésimo nível, C_j se refere a classificação do j-ésimo perfil que contém o i-ésimo nível e n é o número de perfis que contém o i-ésimo nível.

Depois de calculadas as utilidades obtemos a utilidade total para cada perfil, que é dada pela soma das utilidades dos níveis relacionados com um determinado perfil.

$$Utilidade\ Total_i = \sum_{j=1}^{n_1} U_j,$$

onde $Utilidade\ Total_i$ é a utilidade total do i-ésimo perfil e n_1 é o número de níveis em cada perfil, que é constante para todos os perfis.

ii) A estimação das utilidades é feita utilizando um modelo de regressão com variáveis dummy. Assumimos, então, que a utilidade total de um perfil (preferência por um determinado perfil) é construída através do somatório das utilidades, desta forma podemos escrever a utilidade total através de um modelo de regressão com variáveis dummy, da forma:

$$Y_j = \beta_0 + \sum_j \sum_i X_{ji} \beta_i + \varepsilon$$

onde Y_j é a utilidade total do j -ésimo perfil, β_0 é o intercepto, $X_{ji} \beta_i$ é o escore do nível i de um atributo com relação ao j -ésimo perfil.

Em seguida são obtidas as estimativas dos β 's e dos escores para cada nível. As estimativas dos escores, são transformados em uma específica escala, para obtermos as utilidades para cada nível em uma escala positiva.

Com as utilidades calculadas obtém-se as utilidades totais de cada perfil através do somatório das utilidades dos níveis relacionados com o perfil.

Esta técnica tem sido aplicada com frequência ao desenvolvimento de novos produtos em todos os setores industriais e a seleção de segmento-alvo de mercado, entre outras aplicações.

4.2 Uso de variáveis dummy

Quando trabalhamos com modelos de regressão poderemos lidar com situações em que os inputs são variáveis categóricas (nominais), por exemplo, as características de produtos/serviços.

Para inserirmos as informações destas variáveis no modelo é necessário criar uma ou mais variáveis com valores numéricos que possam representar as características variáveis categóricas consideradas.

As variáveis que serão criadas são chamadas de variáveis dummy.

Uma maneira de construir variáveis dummy que representem uma variável categórica A com k categorias, A_1, A_2, \dots, A_k , é definindo $(k-1)$ variáveis, X_1, X_2, \dots, X_{k-1} , que assumem apenas dois valores, 0 e 1, para $i = 1, 2, 3, \dots, K - 1$, temos:

$$X_i = \left\{ \begin{array}{l} 1, \text{ se a unidade amostral considerada pertence a categoria } A_i, \\ 0, \text{ se a unidade amostral considerada pertence a categoria } A_j, j \neq i \end{array} \right\}$$

Desta forma, a sequência $(1, 0, 0, \dots, 0)$ implica que a unidade amostral em questão estará classificada na categoria A_1 . As outras sequências possíveis são analisadas de forma análoga, até $(0, 0, 0, \dots, 0)$, onde a categoria relacionada é A_k .

Devemos salientar que a forma escolhida para a criação de variáveis dummy é totalmente arbitrária.

4.3 Principais etapas da análise conjunta

4.3.1 1ª Etapa: Caracterização do Produto

Na análise conjunta os produtos/serviços são definidos sobre um número limitado de atributos relevantes (características importantes) cada qual com um número limitado de níveis, desta forma na primeira etapa o produto/serviço é caracterizado através de atributos relevantes que influenciam na preferência ou intenção de compra dos respondentes.

4.3.2 2ª Etapa: Planejamento da pesquisa

Determinação dos atributos relevantes:

É necessário identificar quais atributos são mais importantes no ponto de vista dos clientes, para isto, pode ser necessário a realização de entrevista com um grupo foco, que fornece os atributos que são mais importantes e aqueles que mais influenciaram a preferência dos respondentes em relação à preferência por produtos/serviços.

Escolha dos níveis dos atributos.

Os níveis dos atributos são analisados de maneira que tenha uma amplitude suficiente para abranger as alternativas de interesse. O número de níveis geralmente utilizado é 2 ou 3 níveis, para não sobrecarregar os respondentes na sua tarefa. Um número alto de níveis pode dificultar uma avaliação adequada pelos respondentes.

Determinação das combinações dos níveis dos atributos (perfis que serão avaliados pelos respondentes)

Caso o número de combinações for grande, será necessário selecionar uma fração das opções disponíveis, utilizando experiência a priori na empresa ou técnicas estatísticas de planejamento de experimentos.

Coleta dos dados

A coleta dos dados é realizada através de entrevista pessoal, métodos iterativos com o uso de um computador (os respondentes recebem instruções de um programa computacional), envio de questionários aos respondentes ou entrevista por telefone.

Apresentação dos perfis aos respondentes.

A maneira pela qual os perfis (produtos/serviços) serão apresentados aos respondentes deve ser selecionada dentre as seguintes opções:

i) Forma de parágrafo:

Os níveis de cada perfil é descrito em um parágrafo.

Por exemplo:

“tênis de couro, preço R\$ 70,00, procedência nacional, marca reconhecida, aspecto visual clássico”

ii) Cartões:

Neste método cada perfil é apresentado em um cartão que contenha os atributos e seus níveis relacionados.

Exemplo:

O quadro 1 abaixo representa um cartão, utilizado para representar um perfil.

Quadro 1 - Cartão representando um perfil.

Perfil A	
Atributos	Níveis
Preço	R\$ 70,00
Material	Couro
Procedência	Nacional
Marca	Reconhecida
Aspecto Visual	Clássico

iii) Representação ilustrada:

Os perfís são apresentados aos respondentes em forma de figura contendo as características relacionadas.

iv) Produto real:

Este método utiliza amostras dos possíveis produtos para que sejam avaliadas pelos respondentes.

Tipo de tarefa a ser solicitada aos respondentes para obter uma resposta desejada.

As opções mais utilizadas são: classificação dos perfís por nota em uma determinada escala e ordenação das alternativas por preferência ou intenção de compra. Os respondentes fornecem um escore para um determinado perfil, um escore real na abordagem de classificação e um escore implícito na abordagem de ordenação.

4.3.3 3^o Etapa: Análise dos dados

A análise conjunta permite a obtenção das utilidades ou importâncias de cada nível dos atributos, a partir das preferências dos respondentes pelos vários produtos disponíveis, obtida a partir de uma pesquisa de preferência realizada com uma amostra de respondentes que represente nosso público alvo.

Os cálculos das utilidades para cada nível dos atributos e a utilidade total de cada perfil são feitos como vimos anteriormente na seção 4.1.

Devemos calcular também a importância de um determinado atributo que pode estar em função das utilidades associadas a seus níveis ou seja,

$$\text{Importância do Atributo} = \text{utilidade(máxima)} - \text{utilidade(minima)}$$

A importância relativa de cada atributo é obtida a partir da padronização em relação a importância dos demais atributos como:

$$\text{Importância relativa do atributo} = \frac{\text{Importância do Atributo}}{\text{Soma das importâncias de todos os atributos}} * 100$$

Com a importância e a importância relativa de cada atributo podemos identificar quais atributos são mais importantes para os clientes.

A seguir está ilustrado um exemplo no qual foi obtida as utilidades totais para cada perfil em estudo, a importância de cada atributo e a importância relativa de cada atributo.

4.4 Exemplo

Neste exemplo será analisado quais atributos são mais importantes na escolha de um tênis e qual o perfil preferido pelos clientes.

Esta aplicação foi desenvolvida por Drumond F.B. (2004), com o objetivo de estudar o efeito de alguns atributos na preferência dos respondentes, no momento da compra de um tênis e qual o perfil preferido pelos respondentes, estimando as utilidades dos níveis dos perfis, utilizando o primeiro método de estimação, após apresentar o método utilizado por Drumond F.B. (2004), utilizamos o segundo método de estimação com o objetivo de comparar os resultados obtidos. Os atributos (características do produto) considerados estão mostrados na tabela 13 :

Tabela 13 -Atributos Considerados.

<i>Atributo</i>
<i>Material</i>
<i>Preço</i>
<i>Procedência</i>
<i>Marca</i>
<i>Aspecto Visual</i>

Os níveis dos atributos utilizados neste estudo estão mostrados na tabela 14.

Tabela 14 - Atributos e seus Níveis.

	<i>Atributo</i>	<i>Nível do atributo</i>
<i>Atributos de três níveis</i>	<i>Preço</i>	= R\$70,00
		= R\$100,00
		= R\$130,00
	<i>Material</i>	= <i>Couro</i>
		= <i>Nylon</i>
		= <i>Couro e Nylon</i>
<i>Atributos de dois níveis</i>	<i>Procedência</i>	= <i>Nacional</i>
		= <i>Importada</i>
	<i>Marca</i>	= <i>Reconhecida</i>
		= <i>Não reconhecida</i>
	<i>Aspecto Visual</i>	= <i>Clássico</i>
		= <i>Arrojado</i>

Se calcularmos o número de combinações para os diversos níveis de atributos há $3 * 3 * 2 * 2 * 2 = 72$ possíveis combinações. Como é inviável avaliar todos os possíveis produtos é necessário escolher um subconjunto de combinações que seja representativo dos 72 produtos neste caso foram escolhidas 18 combinações utilizando experiência a priori da empresa, como experiências frustradas com determinadas combinações no passado.

As 18 combinações que forma o subconjunto utilizado estão mostradas na tabela 15.

Tabela 15 - Combinações dos níveis dos atributos.

<i>Produto</i>	<i>Preço R\$</i>	<i>Material</i>	<i>Procedência</i>	<i>Marca</i>	<i>Aspecto Visual</i>
<i>A</i>	70,00	<i>Couro</i>	<i>Nacional</i>	<i>Reconhecida</i>	<i>Clássico</i>
<i>B</i>	70,00	<i>Nylon</i>	<i>Importada</i>	<i>Reconhecida</i>	<i>Arrojado</i>
<i>C</i>	70,00	<i>Couro e Nylon</i>	<i>Nacional</i>	<i>Não Reconhecida</i>	<i>Clássico</i>
<i>D</i>	100,00	<i>Couro</i>	<i>Importada</i>	<i>Não Reconhecida</i>	<i>Arrojado</i>
<i>E</i>	100,00	<i>Nylon</i>	<i>Nacional</i>	<i>Reconhecida</i>	<i>Clássico</i>
<i>F</i>	100,00	<i>Couro e Nylon</i>	<i>Nacional</i>	<i>Reconhecida</i>	<i>Clássico</i>
<i>G</i>	130,00	<i>Couro</i>	<i>Nacional</i>	<i>Reconhecida</i>	<i>Arrojado</i>
<i>H</i>	130,00	<i>Nylon</i>	<i>Nacional</i>	<i>Não Reconhecida</i>	<i>Clássico</i>
<i>I</i>	130,00	<i>Couro e Nylon</i>	<i>Importada</i>	<i>Reconhecida</i>	<i>Clássico</i>
<i>J</i>	70,00	<i>Couro</i>	<i>Nacional</i>	<i>Não Reconhecida</i>	<i>Clássico</i>
<i>K</i>	70,00	<i>Nylon</i>	<i>Nacional</i>	<i>Reconhecida</i>	<i>Arrojado</i>
<i>L</i>	70,00	<i>Couro e Nylon</i>	<i>Importada</i>	<i>Reconhecida</i>	<i>Clássico</i>
<i>M</i>	100,00	<i>Couro</i>	<i>Nacional</i>	<i>Reconhecida</i>	<i>Clássico</i>
<i>N</i>	100,00	<i>Nylon</i>	<i>Importada</i>	<i>Não Reconhecida</i>	<i>Clássico</i>
<i>O</i>	100,00	<i>Couro e Nylon</i>	<i>Nacional</i>	<i>Reconhecida</i>	<i>Arrojado</i>
<i>P</i>	130,00	<i>Couro</i>	<i>Importada</i>	<i>Reconhecida</i>	<i>Clássico</i>
<i>Q</i>	130,00	<i>Nylon</i>	<i>Nacional</i>	<i>Reconhecida</i>	<i>Clássico</i>
<i>R</i>	130,00	<i>Couro e Nylon</i>	<i>Nacional</i>	<i>Não Reconhecida</i>	<i>Arrojado</i>

O próximo passo da *Análise Conjunta* consiste da coleta de dados, onde são obtidas as ordenações de preferência para os produtos. Para isso foram confeccionados cartões com os produtos hipotéticos, como no quadro 2.

Quadro 2 - Cartão contendo o perfil A.

<p><i>Perfil A</i></p> <p><i>Preço : R\$ 70,00</i></p> <p><i>Material : Couro</i></p> <p><i>Procedência : Nacional</i></p> <p><i>Marca : Reconhecida</i></p> <p><i>Aspecto Visual : Clássico</i></p>
--

Cada cartão descreve uma das 18 possíveis combinações e os respondentes os ordenam de acordo com a preferência por cada tênis. O respondente selecionou e ordenou primeiro os de maior preferência, depois os de menor preferência e por último os de preferência intermediária.

Foram abordados 26 respondentes, 13 do sexo masculino e 13 do sexo feminino. Com os dados fornecidos pelos respondentes, obteve-se a ordenação geral (classificação), calculando a média das ordenações feitas pelos respondentes para cada combinação. A ordenação geral foi calculada separadamente, para o sexo feminino e para o sexo masculino e estão mostradas na tabela 16, onde o maior valor indica a maior preferência. As ordenações apresentadas na tabela 16, foram calculadas considerando o seguinte critério:

$Ordenação (Perfil) = \frac{\sum_{i=1}^{13} x_i}{13}$, onde x_i , é a posição que o i -ésimo respondente classificou o perfil em estudo.

Tabela 16 - Preferência pelas combinações dos atributos.

<i>Combinação</i>	<i>Preferência</i>	
	<i>Sexo Feminino</i>	<i>Sexo Masculino</i>
<i>A</i>	17	17
<i>B</i>	7	7
<i>C</i>	11	8
<i>D</i>	6	9
<i>E</i>	9	13
<i>F</i>	15	14
<i>G</i>	4	11
<i>H</i>	1	2
<i>I</i>	13	10
<i>J</i>	12	12
<i>K</i>	5	4
<i>L</i>	18	15
<i>M</i>	16	16
<i>N</i>	8	3
<i>O</i>	10	5
<i>P</i>	14	18
<i>Q</i>	3	6
<i>R</i>	2	1

(Maior Valor Indica Maior Preferência)

O próximo passo é obter as utilidades para cada nível dos atributos, que serão usadas para calcular as utilidades totais de cada perfil, a importância de cada atributo e a importância relativa de cada atributo. Utilizando os dois métodos citados na seção 3.1.

Utilizando o primeiro método, estimamos as utilidades para cada nível dos atributos, calculando o valor médio da variável dependente (ordenação) para cada nível dos atributos, por exemplo:

Vamos calcular a utilidade para o nível R\$ 70,00 do atributo preço, analisando as preferências do sexo feminino, observamos na tabela 15 que o nível R\$ 70,00 está presente em 6 perfis, em seguida buscamos na tabela 16 a classificação de cada um dos 6 perfis, e calculamos o escore médio para o nível em estudo, que é dado por:

$$\frac{(17 + 7 + 11 + 12 + 5 + 18)}{6} = 11,67$$

Realizando os mesmos cálculos para os demais perfis, para o sexo feminino e para o sexo masculino, com as devidas alterações, obtemos as Tabelas 17 e 18.

Para obter as utilidades em uma escala particular, basta converter seus escores médios para uma determinada escala. Para o sexo feminino, as médias que variam de 5,50 à 11,67, foram padronizadas para o intervalo 0,1 à 1, utilizando interpolação linear, onde o menor valor, que neste caso é 5,5 corresponde a 0,1 na nova escala e o maior valor 11,67 corresponde a 1 na nova escala. Similarmente para o sexo masculino, os escores foram convertidos e estão mostrados nas Tabelas 17 e 18.

Tabela 17 - Escores e Utilidades dos níveis dos atributos para o sexo feminino.

<i>Nível do atributo</i>	<i>Escore</i>	<i>Utilidade</i>
<i>Preço</i>		
<i>R\$130,00</i>	6,17	0,20
<i>R\$100,00</i>	10,67	0,85
<i>R\$70,00</i>	11,67	1,00
<i>Material</i>		
<i>Couro e Nylon</i>	11,50	0,98
<i>Nylon</i>	5,50	0,10
<i>Couro</i>	11,50	0,98
<i>Procedência</i>		
<i>Importada</i>	11,00	0,90
<i>Nacional</i>	8,75	0,57
<i>Marca</i>		
<i>Não reconhecida</i>	6,67	0,27
<i>Reconhecida</i>	10,92	0,89
<i>Aspecto Visual</i>		
<i>Arrojado</i>	5,67	0,12
<i>Clássico</i>	11,42	0,96

Tabela 18 - Escores e Utilidades dos níveis dos atributos para o sexo maculino.

<i>Nível do atributo</i>	<i>Escore</i>	<i>Utilidade</i>
<i>Preço</i>		
<i>R\$130,00</i>	8,00	0,35
<i>R\$100,00</i>	10,00	0,57
<i>R\$70,00</i>	10,50	0,63
<i>Material</i>		
<i>CouroeNylon</i>	8,83	0,44
<i>Nylon</i>	5,83	0,10
<i>Couro</i>	13,83	1,00
<i>Procedência</i>		
<i>Importada</i>	10,33	0,61
<i>Nacional</i>	9,08	0,47
<i>Marca</i>		
<i>Nãoreconhecida</i>	5,83	0,10
<i>Reconhecida</i>	11,33	0,72
<i>AspectoVisual</i>		
<i>Arrojado</i>	6,17	0,14
<i>Clássico</i>	11,17	0,70

Com as utilidades de cada nível, podemos calcular as utilidades totais, que são dadas pelo somatório das utilidades dos níveis relacionados com um determinado perfil, por exemplo, a utilidade total do perfil A para o sexo feminino, é dada por

$$UtilidadeTotal(A) = (1,00 + 0,98 + 0,57 + 0,89 + 0,96) = 4,40$$

Fazendo esses cálculos para os demais perfis, utilizando os dados do sexo feminino e do sexo masculino obtemos os resultados contidos na tabela 19.

Tabela 19 - Utilidades Totais para os perfis.

<i>Perfil</i>	<i>Feminino</i>	<i>Masculino</i>
	<i>Utilidade Total</i>	<i>Utilidade Total</i>
<i>A</i>	4,40	3,96
<i>B</i>	3,01	2,87
<i>C</i>	3,78	3,43
<i>D</i>	3,12	2,90
<i>E</i>	3,37	3,10
<i>F</i>	4,25	3,81
<i>G</i>	2,76	2,65
<i>H</i>	2,10	2,06
<i>I</i>	3,93	3,56
<i>J</i>	3,78	3,43
<i>K</i>	2,68	2,61
<i>L</i>	4,73	4,23
<i>M</i>	4,25	3,81
<i>N</i>	3,08	2,84
<i>O</i>	3,41	3,16
<i>P</i>	3,93	3,56
<i>Q</i>	2,72	2,59
<i>R</i>	2,14	2,12

Após o cálculo das utilidades totais os valores foram ordenados e obtemos a classificação dos perfis de acordo com a preferência dos respondentes, para ambos os sexos, as classificações estão mostradas nas tabelas 20 e 21.

Tabela 20- Classificação para o sexo fem. Tabela 21- Classificação para o sexo mas.

<i>Classificação</i>	<i>Perfil</i>
1°	<i>L</i>
2°	<i>A</i>
3°	<i>F</i>
4°	<i>M</i>
5°	<i>I</i>
6°	<i>P</i>
7°	<i>C</i>
8°	<i>J</i>
9°	<i>O</i>
10°	<i>E</i>
11°	<i>D</i>
12°	<i>N</i>
13°	<i>B</i>
14°	<i>G</i>
15°	<i>Q</i>
16°	<i>K</i>
17°	<i>R</i>
18°	<i>H</i>

<i>Classificação</i>	<i>Perfil</i>
1°	<i>L</i>
2°	<i>A</i>
3°	<i>F e M</i>
4°	<i>I e P</i>
5°	<i>J e C</i>
6°	<i>O</i>
7°	<i>E</i>
8°	<i>D</i>
9°	<i>B</i>
10°	<i>N</i>
11°	<i>G</i>
12°	<i>K</i>
13°	<i>Q</i>
14°	<i>R</i>
15°	<i>H</i>

A importância e a importância relativa para o atributo preço, com relação as preferências do sexo feminino são dadas por.

$$\text{Importância do Atributo (Preço)} = 1,00 - 0,20 = 0,80$$

Para os outros atributos temos,

Tabela 22 - Importância de cada atributo para o sexo feminino.

<i>Atributo</i>	<i>Importância</i>
<i>Preço</i>	0,80
<i>Material</i>	0,88
<i>Procedência</i>	0,33
<i>Marca</i>	0,62
<i>Aspecto Visual</i>	0,84

Soma das importâncias de todos os atributos = 3,47

e

$$\text{Importância relativa do atributo (Preço)} = \frac{\text{Importância do Atributo(Preço)}}{\text{Soma das importâncias de todos os atributos}} * 100$$

$$\text{Importância relativa do atributo (Preço)} = 23,05\%$$

Realizando o mesmo processo para os demais atributos, obtemos os resultados mostrados na Tabela 23.

Tabela 23 - Importância relativa de cada atributo para o sexo feminino.

<i>Atributo</i>	<i>Importância Relativa</i>
<i>Preço</i>	23,05%
<i>Material</i>	25,36%
<i>Procedência</i>	9,51%
<i>Marca</i>	17,87%
<i>Aspecto Visual</i>	24,21%

A importância e a importância relativa para cada atributo, foram calculadas em relação a preferência do sexo masculino da mesma maneira que para o sexo feminino, e estão mostradas nas tabelas 24 e 25.

Tabela 24 - Importância de cada atributo para o sexo masculino.

<i>Atributo</i>	<i>Importância</i>
<i>Preço</i>	0,28
<i>Material</i>	0,90
<i>Procedência</i>	0,14
<i>Marca</i>	0,62
<i>Aspecto Visual</i>	0,56

Soma das importâncias de todos os atributos = 2,5

Tabela 25 - Importância relativa de cada atributo para o sexo masculino.

<i>Atributo</i>	<i>Importância Relativa</i>
<i>Preço</i>	11,20%
<i>Material</i>	36,00%
<i>Procedência</i>	5,60%
<i>Marca</i>	24,80%
<i>Aspecto Visual</i>	22,50%

Analisando os resultados obtidos utilizando as preferências do sexo feminino observamos que os atributos mais importantes e que mais interferem na escolha de um tênis é, em primeiro lugar, o material, em segundo lugar o aspecto visual e em seguida o preço.

Da tabela 20 podemos concluir que, para o sexo feminino, os perfis mais preferidos são: o perfil L (Preço = R\$ 70,00, Material = Couro e Nylon, Procedência = Importada, Marca = Reconhecida, Aspecto Visual = Clássico) e o perfil A (Preço = R\$ 70,00, Material = Couro, Procedência = Nacional, Marca = Reconhecida, Aspecto Visual = Clássico).

Para o sexo masculino, os atributos mais importantes e que mais interferem na escolha de um tênis é, em primeiro lugar o material, em segundo lugar a marca e em seguida o aspecto visual, para o sexo masculino, os perfis que apresentaram maior preferência são: o

perfil L (Preço = R\$ 70,00, Material = Couro e Nylon, Procedência = Importada, Marca = Reconhecida, Aspecto Visual = Clássico) e o perfil A (Preço = R\$ 70,00, Material = Couro, Procedência = Nacional, Marca = Reconhecida, Aspecto Visual = Clássico).

Utilizando o segundo método de determinação das estimativas das utilidades, foram consideradas as seguintes codificações;

Tabela 26 - Codificação dos níveis dos atributos.

<i>Atributo</i>	<i>Nível do atributo</i>	<i>Codificação</i>
		(X1, X2)
<i>Preço</i>	= R\$ 70,00	(0, 0)
	= R\$ 100,00	(0, 1)
	= R\$ 130,00	(1, 0)
		(X3, X4)
<i>Material</i>	= <i>Couro</i>	(0, 0)
	= <i>Nylon</i>	(0, 1)
	= <i>Couro e Nylon</i>	(1, 0)
		(X5)
<i>Procedência</i>	= <i>Nacional</i>	(0)
	= <i>Importada</i>	(1)
		(X6)
<i>Marca</i>	= <i>Reconhecida</i>	(0)
	= <i>Não reconhecida</i>	(1)
		(X7)
<i>AspectoVisual</i>	= <i>Clássico</i>	(0)
	= <i>Arrojado</i>	(1)

O modelo de regressão com variável dummy é dado por :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \varepsilon_i, \quad i = 1, \dots, 18$$

que pode ser representado na forma matricial como:

$$Y = X\beta + \varepsilon,$$

onde Y é o vetor coluna de preferências, X é a matriz contendo a codificação dos atributos, β é o vetor de parâmetros e ε é o vetor de erros aleatórios.

Para o sexo feminino a expressão $Y = X\beta + \varepsilon$ é dada por :

$$\begin{array}{c|c|c|c}
 \begin{array}{c} 17 \\ 7 \\ 11 \\ 6 \\ 9 \\ 15 \\ 4 \\ 1 \\ 13 \\ 12 \\ 5 \\ 18 \\ 16 \\ 8 \\ 10 \\ 14 \\ 2 \\ 3 \end{array} & = & \begin{array}{c|c} \begin{array}{cccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \end{array} & * & \begin{array}{c} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \end{array} & + & \begin{array}{c} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{16} \\ \varepsilon_{17} \\ \varepsilon_{18} \end{array}
 \end{array}$$

Podemos observar que o vetor coluna Y representa a ordenação dos perfis para o sexo feminino, obtido através dos dados dos respondentes, e cada linha da matriz X representa um perfil, onde a primeira linha representa o perfil A, a segunda representa o perfil B e assim por diante.

O vetor de parâmetros estimados é dado por :

$$\hat{\beta} = \begin{array}{c|c} \hat{\beta}_0 & 16.25 \\ \hat{\beta}_1 & -5,5 \\ \hat{\beta}_2 & -1 \\ \hat{\beta}_3 & 0 \\ \hat{\beta}_4 & -6 \\ \hat{\beta}_5 & 2.25 \\ \hat{\beta}_6 & -4,25 \\ \hat{\beta}_7 & -5,75 \end{array}$$

O próximo passo do estudo envolve a utilização da soma para zero, para obter as estimativas para os escores de todos os níveis dos atributos.

Exemplo: Suponha que um determinado atributo possui 3 níveis então:

$$\sum_{i=1}^3 \text{escore}_i = 0$$

Os escores para as utilidades níveis dos atributos estão mostradas na tabela 27.

Com o ajuste do modelo obtemos os limites mínimo e máximo para os parâmetros, depois de calculados todos os escores, são utilizados os limites mínimo e máximo, para obter os intervalos para os escores "utilidades", em seguida são obtidos os intervalos para as utilidades totais dos perfis.

Tabela 27 - Escores e Utilidades para o sexo feminino.

<i>Nível do atributo</i>	<i>Utilidade</i>	<i>Intervalo para os Escores</i>
<i>Preço</i>		
R\$ 130,00	-5,5	(-7,25 ; -4,25)
R\$ 100,00	-1	(-2,75 ; 0,25)
R\$ 70,00	6,5	(4,75 ; 7,75)
<i>Material</i>		
<i>Couro e Nylon</i>	0	(-1,75 ; 1,25)
<i>Nylon</i>	-6	(-7,75 ; -4,75)
<i>Couro</i>	6	(4,25 ; 7,25)
<i>Procedência</i>		
<i>Importada</i>	2,25	(0,5 ; 3,5)
<i>Nacional</i>	-2,25	(-4 ; -1)
<i>Marca</i>		
<i>Não reconhecida</i>	-4,25	(-6 ; -3)
<i>Reconhecida</i>	4,25	(2,5 ; 5,5)
<i>Aspecto Visual</i>		
<i>Arrojado</i>	-5,75	(-7,5 ; -4,5)
<i>Clássico</i>	5,75	(4 ; 7)

Depois de obtidas as utilidade para cada nível e seus intervalos, podemos obter as utilidades totais para cada perfil, que é igual a soma das utilidades dos níveis que definem o perfil e os intervalos correspondentes. Por exemplo, a utilidade total para o perfil A é :

$$Utilidade\ Total\ (Perfil\ A) = utilidade(R\$70,00) + utilidade(Couro) + utilidade(Nacional) +$$

$utilidade(Reconhecida) + utilidade(Clássico)$

$$Utilidade Total (Perfil A) = 6,5 + 6 - 2,25 + 4,25 + 5,75 = 20,25$$

Procedendo da mesma maneira para os demais perfis obtemos a tabela 28.

Tabela 28 - Utilidade Total dos Perfis

<i>Perfil</i>	<i>Utilidade Total</i>	<i>Intervalo para as Utilidades totais</i>
<i>A</i>	4,118	(11,5 ; 26,5)
<i>B</i>	2,750	(-7,5 ; 7,5)
<i>C</i>	3,074	(-3 ; 12)
<i>D</i>	2,462	(-11,5 ; 3,5)
<i>E</i>	2,714	(-8 ; 7)
<i>F</i>	3,146	(-2 ; 13)
<i>G</i>	2,426	(-12 ; 3)
<i>H</i>	1,778	(-21 ; -6)
<i>I</i>	3,146	(-2 ; 13)
<i>J</i>	3,506	(3 ; 18)
<i>K</i>	2,426	(-12 ; 3)
<i>L</i>	4,010	(10 ; 25)
<i>M</i>	3,578	(4 ; 19)
<i>N</i>	2,426	(-12 ; 3)
<i>O</i>	2,318	(-13,5 ; 1,5)
<i>P</i>	3,578	(4 ; 19)
<i>Q</i>	2,390	(-12,5 ; 2,5)
<i>R</i>	1,387	(-26,5 ; -11,5)

Tendo calculadas as utilidades totais para cada perfil e os devidos intervalos, podemos obter a classificação dos perfis, que está mostrada na tabela 29.

Tabela 29 - Classificação dos Perfis para o sexo feminino.

<i>Classificação</i>	<i>Perfil</i>
1 ^o	<i>A</i>
2 ^o	<i>L</i>
3 ^o	<i>M e P</i>
4 ^o	<i>J</i>
5 ^o	<i>F e I</i>
6 ^o	<i>C</i>
7 ^o	<i>B</i>
8 ^o	<i>E</i>
9 ^o	<i>D</i>
10 ^o	<i>G, K e N</i>
11 ^o	<i>Q</i>
12 ^o	<i>O</i>
13 ^o	<i>H</i>
14 ^o	<i>R</i>

Uma possibilidade é obter a classificação dos perfis, utilizando os limites dos intervalos para as utilidades totais, obtendo a classificação dos perfis utilizando os limites inferiores ou superiores, obtemos a mesma classificação mostrada na tabela 29.

A importância de cada atributo é apresentada na Tabela 30 e a importância relativa na Tabela 31.

Tabela 30 - Importância de cada atributo para o sexo feminino.

<i>Atributo</i>	<i>Importância do Atributo</i>
<i>Preço</i>	12
<i>Material</i>	12
<i>Procedência</i>	4,5
<i>Marca</i>	8,5
<i>Aspecto Visual</i>	11,5

Soma das importâncias de todos os atributos = 48,5

Tabela 31 - Importância Relativa de cada Atributo.

<i>Atributo</i>	<i>Importância Relativa do Atributo</i>
<i>Preço</i>	24,74%
<i>Material</i>	24,74%
<i>Procedência</i>	9,28%
<i>Marca</i>	17,53%
<i>Aspecto Visual</i>	23,71%

Os intervalos para importância de cada atributo, não é obtido, pois calculando a importância utilizando o limite inferior ou superior das utilidades, obtemos o mesmo resultado, como mostrado a seguir.

Calculando a importância para o atributo preço, utilizando o limite inferior, temos:

$$\text{Importância (Preço)} = 4,75 - (-7,25) = 12$$

Calculando a importância para o atributo preço, utilizando o limite superior, temos:

$$\text{Importância (Preço)} = 7,75 - (-4,25) = 12$$

Observando os resultados obtidos, não são contruídos os intervalos para a importância de cada atributo.

Podemos concluir, que para o sexo feminino os perfis mais preferidos são: o perfil A (Preço = R\$ 70,00, Material = Couro, Procedência = Nacional, Marca = Reconhecida, Aspecto Visual = Clássico) e o perfil L (Preço = R\$ 70,00, Material = Couro e Nylon, Procedência = Importada, Marca = Reconhecida, Aspecto Visual = Clássico).

Efetuada o mesmo processo para o sexo masculino, fazendo as devidas alterações, como alterando Y na expressão $Y = X\beta + \varepsilon$, onde Y será agora a ordenação dos perfis para o sexo masculino, foi obtido o vetor estimado de parâmetros, dado por:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \\ \hat{\beta}_6 \\ \hat{\beta}_7 \end{pmatrix} = \begin{pmatrix} 17,91 \\ -2,5 \\ -5 \\ -8 \\ 1,25 \\ 2,25 \\ -5,5 \\ -5 \end{pmatrix}$$

Utilizando a soma para zero, obtemos as estimativa para as utilidades dos níveis dos atributos e fazendo uso dos limites para os parâmetros obtemos os devidos intervalos, que estão mostradas na Tabela 32.

Tabela 32 - Escores e Utilidades.

<i>Nível do atributo</i>	<i>Utilidades</i>	<i>Intervalo para os Escores</i>
<i>Preço</i>		
<i>R\$130,00</i>	-2,5	(-4,92 ; 1,08)
<i>R\$100,00</i>	-0,5	(-2,92 ; 3,08)
<i>R\$70,00</i>	3	(0,58 ; 6,58)
<i>Material</i>		
<i>Couro e Nylon</i>	-5	(-7,42 ; -1,42)
<i>Nylon</i>	-8	(-10,42 ; -4,42)
<i>Couro</i>	13	(10,58 ; 16,58)
<i>Procedência</i>		
<i>Importada</i>	1,25	(-1,17 ; 4,83)
<i>Nacional</i>	-1,25	(-3,67 ; 2,33)
<i>Marca</i>		
<i>Nãoreconhecida</i>	-5,5	(-7,92 ; -1,92)
<i>Reconhecida</i>	5,5	(3,08 ; 9,08)
<i>Aspecto Visual</i>		
<i>Arrojado</i>	-5	(-7,42 ; -1,42)
<i>Clássico</i>	5	(2,58 ; 8,58)

Conhecidas as utilidades, pode-se calcular, as utilidades totais e seus intervalos e com isto obter a classificação dos perfis e calcular as importâncias relativas de cada atributo, que estão mostradas nas Tabelas 33, 34 e 35.

Tabela 33 - Utilidade Total dos Perfis

<i>Perfil</i>	<i>Utilidade Total</i>	<i>Intervalo para as Utilidades totais</i>
<i>A</i>	25,25	(13,15 ; 43,15)
<i>B</i>	-3,25	(-15,35 ; 14,65)
<i>C</i>	-3,75	(-15,85 ; 14,15)
<i>D</i>	3,25	(-8,85 ; 21,15)
<i>E</i>	0,75	(-11,35 ; 18,65)
<i>F</i>	3,75	(-8,35 ; 21,65)
<i>G</i>	9,75	(-2,35 ; 27,65)
<i>H</i>	-12,25	(-24,35 ; 5,65)
<i>I</i>	4,25	(-7,85 ; 22,15)
<i>J</i>	14,25	(2,15 ; 32,15)
<i>K</i>	-5,75	(-17,85 ; 12,15)
<i>L</i>	9,75	(-2,35 ; 27,65)
<i>M</i>	21,75	(9,65 ; 39,65)
<i>N</i>	-7,75	(-19,85 ; 10,15)
<i>O</i>	-6,25	(-18,35 ; 11,65)
<i>P</i>	22,25	(10,15 ; 40,15)
<i>Q</i>	-1,25	(-13,35 ; 16,65)
<i>R</i>	-19,25	(-31,35 ; -1,35)

Tabela 34 - Classificação para o sexo Masculino.

<i>Classificação</i>	<i>Perfil</i>
1°	<i>A</i>
2°	<i>P</i>
3°	<i>M</i>
4°	<i>J</i>
5°	<i>L – G</i>
6°	<i>I</i>
7°	<i>F</i>
8°	<i>D</i>
9°	<i>E</i>
10°	<i>Q</i>
11°	<i>B</i>
12°	<i>C</i>
13°	<i>K</i>
14°	<i>O</i>
15°	<i>N</i>
16°	<i>H</i>
17°	<i>R</i>

Tabela 35 - Importância Relativa de cada Atributo.

<i>Atributo</i>	<i>Importância Relativa do Atributo</i>
<i>Preço</i>	11%
<i>Material</i>	42%
<i>Procedência</i>	5%
<i>Marca</i>	22%
<i>Aspecto Visual</i>	20%

Podemos concluir, que para o sexo masculino os perfis mais preferidos são: o perfil A (Preço = R\$ 70,00, Material = Couro, Procedência = Nacional, Marca = Reconhecida, Aspecto Visual = Clássico) e o perfil P (Preço = R\$ 130,00, Material = Couro, Procedência = Importada, Marca = Reconhecida, Aspecto Visual = Clássico).

Conclusão geral do exemplo (Comparação dos Resultados das duas Análises)

Com os resultados da análise conjunta realizada utilizando os dois métodos apresentados, os especialistas em marketing podem aprimorar seus produtos ao gosto do consumidor, tendo conhecida a importância de cada atributo.

Após efetuados todos os cálculos, utilizando os dois métodos, podemos compará-los para verificar se ocorre desigualdades nos resultados.

Primeiramente, buscamos comparar os resultados obtidos com relação às preferências do sexo feminino.

Os atributos mais importantes para o sexo feminino utilizando o primeiro método é, em primeiro lugar o material, em segundo lugar o aspecto visual e em seguida o preço, e utilizando o segundo método são, em primeiro lugar o material e o preço, e em seguida o aspecto visual.

Observando os atributos mais importantes para o sexo feminino nos dois métodos, visualizamos que os atributos são os mesmos, somente ocorreu uma mudança na ordem de importância para o preço e aspecto visual.

A seguir na tabela 36 está mostrada as classificações dos perfis obtidos no primeiro e no segundo método e também a classificação fornecida pelos respondentes do sexo feminino.

Tabela 36 - Classificações dos perfis para o sexo Feminino.

<i>Classificação</i>	<i>Classificação dos Resp.</i>	<i>Primeiro Método</i>	<i>Segundo Método</i>
1º	<i>L</i>	<i>L</i>	<i>A</i>
2º	<i>A</i>	<i>A</i>	<i>L</i>
3º	<i>M</i>	<i>F</i>	<i>M e P</i>
4º	<i>F</i>	<i>M</i>	<i>J</i>
5º	<i>P</i>	<i>I</i>	<i>F e I</i>
6º	<i>I</i>	<i>P</i>	<i>C</i>
7º	<i>J</i>	<i>C</i>	<i>B</i>
8º	<i>C</i>	<i>J</i>	<i>E</i>
9º	<i>O</i>	<i>O</i>	<i>D, G, K e N</i>
10º	<i>E</i>	<i>E</i>	<i>Q</i>
11º	<i>N</i>	<i>D</i>	<i>O</i>
12º	<i>B</i>	<i>N</i>	<i>H</i>
13º	<i>D</i>	<i>B</i>	<i>R</i>
14º	<i>K</i>	<i>G</i>	
15º	<i>G</i>	<i>Q</i>	
16º	<i>Q</i>	<i>K</i>	
17º	<i>R</i>	<i>R</i>	
18º	<i>H</i>	<i>H</i>	

Analisando a Tabela 36, observamos que os dois métodos representam bem a classificação realizada pelos respondentes, tendo os perfis L e A em primeiro e segundo lugar respectivamente, no primeiro método e no segundo método, observamos os perfis A e L em primeiro e segundo lugar respectivamente, os dois perfis estão mostrados abaixo,

<i>Perfil</i>	<i>L</i>
<i>Preço</i>	<i>R\$ 70.00</i>
<i>Material</i>	<i>Couro e Nylon</i>
<i>Procedência</i>	<i>Importada</i>
<i>Marca</i>	<i>Reconhecida</i>
<i>Aspecto Visual</i>	<i>clássico</i>

<i>Perfil</i>	<i>A</i>
<i>Preço</i>	<i>R\$ 70.00</i>
<i>Material</i>	<i>Couro</i>
<i>Procedência</i>	<i>Nacional</i>
<i>Marca</i>	<i>Reconhecida</i>
<i>Aspecto Visual</i>	<i>clássico</i>

Os atributos mais importantes para o sexo masculino utilizando o primeiro método são, em primeiro lugar o material, em segundo lugar a marca e em seguida o aspecto visual, utilizando o segundo método são, em primeiro lugar material, em segundo lugar a marca e em seguida o aspecto visual.

Observando os atributos mais importantes para o sexo masculino nos dois métodos, visualizamos que os atributos que mais interferem na compra de um tênis são exatamente os mesmos nos dois métodos.

A seguir na Tabela 37 está mastrada as classificações dos perfis obtidos no primeiro e no segundo método e também a classificação fornecida pelos respondentes do sexo masculino.

Tabela 37 - Classificações dos perfis para o sexo Masculino.

<i>Classificação</i>	<i>Classificação geral dos Resp.</i>	<i>Primeiro Método</i>	<i>Segundo Método</i>
1º	<i>P</i>	<i>L</i>	<i>A</i>
2º	<i>A</i>	<i>A</i>	<i>P</i>
3º	<i>M</i>	<i>F e M</i>	<i>M</i>
4º	<i>L</i>	<i>I e P</i>	<i>J</i>
5º	<i>F</i>	<i>J e C</i>	<i>L</i>
6º	<i>E</i>	<i>O</i>	<i>G</i>
7º	<i>J</i>	<i>E</i>	<i>I</i>
8º	<i>G</i>	<i>D</i>	<i>F</i>
9º	<i>I</i>	<i>B</i>	<i>D</i>
10º	<i>D</i>	<i>N</i>	<i>E</i>
11º	<i>C</i>	<i>G</i>	<i>Q</i>
12º	<i>B</i>	<i>K</i>	<i>B</i>
13º	<i>Q</i>	<i>Q</i>	<i>C</i>
14º	<i>O</i>	<i>R</i>	<i>K</i>
15º	<i>K</i>	<i>H</i>	<i>O</i>
16º	<i>N</i>		<i>N</i>
17º	<i>H</i>		<i>H</i>
18º	<i>R</i>		<i>R</i>

Analisando a Tabela 37, observamos que o segundo métodos representa a classificação geral dos respondentes melhor que o primeiro método, tendo os perfis A e P em primeiro e segundo lugar respectivamente, sendo que o primeiro método mostra a perfil L em primero lugar e o perfil A em segundo lugar, colocando o prefil P em quarto lugar, uma vez que ele ocupa o primeiro lugar na classificação fornecida pelos respondentes do sexo masculino.

Vamos comparar os perfil preferidos pelo sexo masculino obtidos no estudo.

No primeiro método temos o perfil L e o perfil A.

<i>Perfil</i>	<i>L</i>
<i>Preço</i>	<i>R\$ 70.00</i>
<i>Material</i>	<i>Couro e Nylon</i>
<i>Procedência</i>	<i>Importada</i>
<i>Marca</i>	<i>Reconhecida</i>
<i>Aspecto Visual</i>	<i>clássico</i>

<i>Perfil</i>	<i>A</i>
<i>Preço</i>	<i>R\$ 70.00</i>
<i>Material</i>	<i>Couro</i>
<i>Procedência</i>	<i>Nacional</i>
<i>Marca</i>	<i>Reconhecida</i>
<i>Aspecto Visual</i>	<i>clássico</i>

Para o segundo método temos o perfil A e o perfil P.

<i>Perfil</i>	<i>A</i>
<i>Preço</i>	<i>R\$ 70.00</i>
<i>Material</i>	<i>Couro</i>
<i>Procedência</i>	<i>Nacional</i>
<i>Marca</i>	<i>Reconhecida</i>
<i>Aspecto Visual</i>	<i>clássico</i>

<i>Perfil</i>	<i>P</i>
<i>Preço</i>	<i>R\$ 130,00</i>
<i>Material</i>	<i>Couro</i>
<i>Procedência</i>	<i>Importada</i>
<i>Marca</i>	<i>Reconhecida</i>
<i>Aspecto Visual</i>	<i>clássico</i>

Já para os perfis apontados no segundo método como os mais preferidos pelo sexo masculino, observamos que o atributo preço que foi dado como menos importante varia

de R\$ 70,00 para R\$ 130,00.

Com os resultados obtidos no primeiro método, uma empresa pode desenvolver um tênis que pode satisfazer os dois sexos, já com os resultados obtidos com o segundo método a empresa pode desenvolver um tênis para o sexo feminino e um para o sexo masculino, utilizando a informação de que o atributo preço é pouco importante para o sexo masculino.

Através da comparação dos resultados, concluímos que o segundo método onde é utilizado regressão com variáveis dummy, apresenta melhores resultados uma vez que fornece a classificação dos perfis com relação ao sexo masculino, mais relacionada com a classificação dos respondentes do sexo masculino, e observamos a variação do atributo preço, que pode ser uma informação importante para a empresa.

4.5 Análise Conjunta para marketing de clientes

Vimos anteriormente, que a análise conjunta é uma ferramenta poderosa, utilizada em marketing de produtos, que possibilita analisar os pensamentos dos clientes em relação a um determinado produto, utilizando um modelo com variáveis dummy, esta ferramenta permite estabelecer a preferência dos clientes por determinados perfis do produto, e permite conhecer as características do produto, que mais influenciam na hora da compra, a partir da classificação dos possíveis perfis do produto.

Neste ponto do nosso trabalho, vamos utilizar a análise conjunta em marketing de clientes, para estabelecer a preferência da empresa por determinados perfis de clientes, e os atributos que agora serão as características dos clientes, mais valiosos para a empresa, a partir de uma classificação dos clientes, utilizando um modelo de regressão com variáveis dummy.

Para marketing de clientes as etapas da análise conjunta são:

4.5.1 1^a Etapa: Caracterização do Cliente

Neste caso os clientes são definidos sobre um número limitado de atributos relevantes (características importantes) cada qual com um número limitado de níveis, desta forma o

cliente é caracterizado através de atributos relevantes que influenciam na preferência da empresa pelo cliente ou intenção de conquista de novos clientes.

4.5.2 2ª Etapa: Planejamento da pesquisa

- Determinação dos atributos relevantes:

Nesta etapa é necessário identificar os atributos mais importantes para a empresa, neste estudo, utilizamos dois atributos que são de extrema importância em marketing de clientes, que são: o valor do clientes e o tempo de relacionamento, também utilizados para segmentação de clientes.

- Escolha dos níveis dos atributos.

Os níveis dos atributos devem ter uma amplitude suficiente para abranger as alternativas de interesse. O número de níveis utilizados foi 3, para não sobrecarregar a análise de classificação dos clientes.

- Classificação dos clientes

A classificação dos perfis do produto, na análise conjunta é feita a partir da coleta dos dados, realizada através de entrevista pessoal, métodos iterativos com o uso de um computador (os respondentes recebem instruções de um programa computacional), envio de questionários aos respondentes ou entrevista por telefone.

A classificação dos clientes é feita analisando o valor de cada cliente e o tempo de relacionamento em um determinado período. Sendo o melhor cliente, aquele com maior valor e menor tempo de relacionamento.

4.5.3 3ª Etapa: Análise dos dados

Na análise dos dados, é obtida, a utilidade de cada nível dos atributos, a utilidade total de cada cliente e com isso é possível obter a classificação dos clientes, de forma similar a análise conjunta para produtos, utilizando um modelo de regressão com variáveis dummy.

4.5.4 Aplicação da análise conjunta em marketing de clientes

Nesta aplicação, os atributos utilizados foram, o valor do clientes (V.C.) e o tempo de relacionamento (T.R.) e os níveis escolhidos estão mostrados na tabela 38:

Tabela 38- Atributos e seus Níveis.

<i>Atributo</i>	<i>Níveis</i>
<i>Valor do cliente</i>	<i>R\$ 0 – 1000</i>
	<i>R\$ 1000 – 2000</i>
	<i>R\$ 2000 – 3000</i>
<i>Tempo de relacionamento</i>	<i>1 mês</i>
	<i>2 meses</i>
	<i>3 meses</i>

Desta forma, com 2 atributos contendo 3 níveis cada, obtemos $3 * 3 = 9$ possíveis perfís, que não é um número alto que possa trazer complicações na classificação, os perfís estão mostrados na tabela 39.

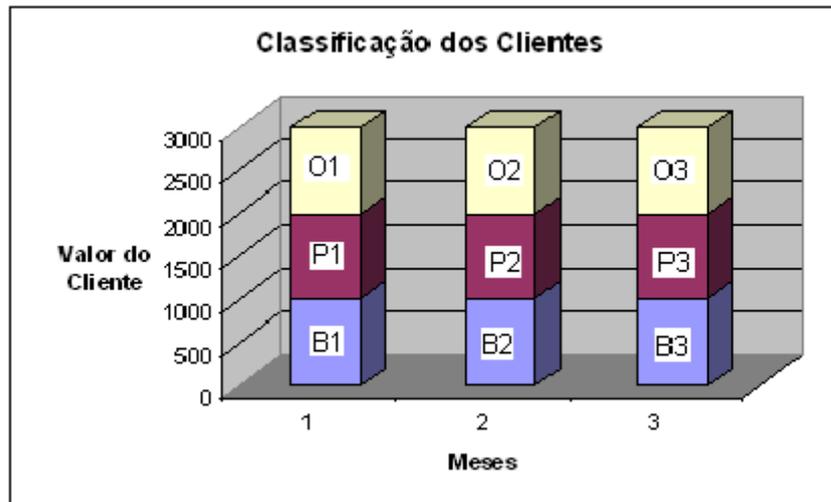
Tabela 39 - Combinações dos níveis dos atributos.

<i>Cliente</i>	<i>Atributo</i>	<i>Atributo</i>
	<i>V.C.(R\$)</i>	<i>T. R.</i>
<i>A</i>	0 – 1000	1 <i>mês</i>
<i>B</i>	0 – 1000	2 <i>meses</i>
<i>C</i>	0 – 1000	3 <i>meses</i>
<i>D</i>	1000 – 2000	1 <i>meses</i>
<i>E</i>	1000 – 2000	2 <i>meses</i>
<i>F</i>	1000 – 2000	3 <i>meses</i>
<i>G</i>	2000 – 3000	1 <i>meses</i>
<i>H</i>	2000 – 3000	2 <i>meses</i>
<i>I</i>	2000 – 3000	3 <i>meses</i>

O próximo passo é a classificação dos perfís (clientes).

A classificação dos perfís, foi realizada utilizando os dois atributos em estudo dispostos no gráfico 6, com o valor do cliente versus tempo de relacionamento, o gráfico 6 e a classificação estão mostrados no gráfico abaixo e na tabela 40.

Gráfico 6 - Classificação dos Clientes.



Os clientes estão classificados como, O1 = ouro 1, O2 = ouro 2, O3 = ouro 3, P1 = prata 1, P2 = prata 2, P3 = prata 3, B1 = bronze 1, B2 = bronze 2, B3 = bronze 3.

Tabela 40 - Classificação dos clientes.

Cliente	Clas. O, P ou B	Classificação
A	B1	4
B	B2	2
C	B3	1
D	P1	7
E	P2	5
F	P3	3
G	O1	9
H	O2	8
I	O3	6

(Maior Valor Indica Maior Preferência)

Obtida a classificação dos clientes, é realizada a codificação dos níveis dos atributos,

para realizar o ajuste do modelo e obter as utilidades dos níveis dos atributos, a codificação está mostrada na tabela 41.

Tabela 41 - Codificação dos níveis dos atributos.

<i>Atributo</i>	<i>Nível do atributo</i>	<i>Codificação</i>
		(X1, X2)
V.C.	R\$ 0 – 1000	(0, 0)
	R\$ 1000 – 2000	(0, 1)
	R\$ 2000 – 3000	(1, 0)
		(X3, X4)
T.R.	1 mês	(0, 0)
	2 meses	(0, 1)
	3 meses	(1, 0)

Desta forma, o modelo de regressão com variável dummy é dado por :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i, \quad i = 1, \dots, 9$$

que pode ser representado na forma matricial como:

$$Y = X\beta + \varepsilon$$

onde Y representa o vetor coluna de preferências da empresa pelos clientes, X a matriz contendo a codificação dos atributos, β o vetor de parâmetros e ε o vetor de erros aleatórios.

A expressão $Y = X\beta + \varepsilon$ é dada por :

$$\begin{array}{c|c|c|c|c|c|c}
 4 & & 1 & 0 & 0 & 0 & 0 & & \varepsilon_1 \\
 2 & & 1 & 0 & 0 & 0 & 1 & & \varepsilon_2 \\
 1 & & 1 & 0 & 0 & 1 & 0 & \beta_0 & \varepsilon_3 \\
 7 & & 1 & 0 & 1 & 0 & 0 & \beta_1 & \varepsilon_4 \\
 5 & = & 1 & 0 & 1 & 0 & 1 & \beta_2 & \varepsilon_5 \\
 3 & & 1 & 0 & 1 & 1 & 0 & \beta_3 & \varepsilon_6 \\
 9 & & 1 & 1 & 0 & 0 & 0 & \beta_4 & \varepsilon_7 \\
 8 & & 1 & 1 & 0 & 0 & 1 & & \varepsilon_8 \\
 6 & & 1 & 1 & 0 & 1 & 0 & & \varepsilon_9
 \end{array}$$

O vetor de parâmetros estimados é dado por :

$$\hat{\beta} = \begin{array}{c|c}
 \hat{\beta}_0 & 4 \\
 \hat{\beta}_1 & 5,33 \\
 \hat{\beta}_2 & 2,67 \\
 \hat{\beta}_3 & -3,33 \\
 \hat{\beta}_4 & -1,67
 \end{array}$$

O próximo passo, envolve a utilização da soma para zero, para obter os escores de todos os níveis dos atributos.

As utilidades para os níveis dos atributos foram consideradas sendo iguais aos escores dos níveis, obtidos aplicando a soma para zero, e estão mostradas na tabela 42.

Tabela 42 - Utilidades dos níveis dos atributos.

<i>Nível do atributo</i>	<i>Utilidades</i>
<i>V. C.</i>	
<i>R\$ 0 – 1000</i>	<i>–8</i>
<i>R\$ 1000 – 2000</i>	<i>2,67</i>
<i>R\$ 2000 – 3000</i>	<i>5,33</i>
<i>T. R.</i>	
<i>1 mês</i>	<i>5</i>
<i>2 meses</i>	<i>–1,67</i>
<i>3 meses</i>	<i>–3,33</i>

Depois de obtidas as utilidade para cada nível, podemos obter as utilidades totais para cada perfil, que é igual a soma das utilidades dos níveis que definem o perfil. Por exemplo, a utilidade total para o perfil A :

$$Utilidade Total (Perfil A) = utilidade(R\$ 0 - 1000) + utilidade(1 mês)$$

$$Utilidade Total (Perfil A) = 4 - 8 + 5 = -3$$

Procedendo da mesma maneira para os demais perfis obtemos a tabela 43.

Tabela 43 - Utilidade Total dos Perfis.

<i>Perfil</i>	<i>Utilidade Total</i>
<i>A</i>	1
<i>B</i>	-5,67
<i>C</i>	-7,33
<i>D</i>	11,67
<i>E</i>	5
<i>F</i>	3,34
<i>G</i>	14,33
<i>H</i>	7,67
<i>I</i>	6

Com o cálculo da utilidade total de cada perfil, podemos obter a classificação dos perfis, que está mostrada na tabela 44.

Tabela 44 - Classificação dos Perfis dos Clientes.

<i>Classificação</i>	<i>Perfil</i>
1º	G
2º	D
3º	H
4º	I
5º	E
6º	F
7º	A
8º	B
9º	C

O próximo passo é o cálculo da importância de cada atributo (I) e a importância relativa (IR) de cada um deles,

$$I = \text{utilidade}(máx) - \text{utilidade}(min)$$

Exemplo, para o atributo valor do cliente (V.C.), temos:

$$I.R.(V.C.) = 5,333 + 8 = 13,33$$

Procedendo da forma similar considerando o atributo tempo de relacionamento (T.R.) obtem-se a tabela 45.

Tabela 45 - Importância de cada atributo (I).

<i>Atributo</i>	<i>I</i>
V.C.	13,33
T.R.	8,33

Cálculo da I. R. de cada atributo.

$$I.R. = \frac{\text{Importância do Atributo}}{\text{Soma das importâncias de todos os atributos}} * 100$$

$$\text{Soma das importâncias de todos os atributos} = 21,66$$

Exemplo, para o V.C.:

$$I.R.(V.C.) = \frac{13,33}{21,66} * 100 = 61,54\%$$

Efetuando o cálculo para o atributo T.R., foi construída a tabela 46:

Tabela 46 - Importância Relativa de cada Atributo (I.R.).

<i>Atributo</i>	<i>I.R.</i>
V.C	61,54%
T.R.	38,46%

Conclusão:

Na conclusão deste exemplo, está confrontada as duas classificações realizadas, para observarmos se elas seguem um padrão semelhante, e estão mostradas na Tabela 47.

Tabela 47 - Classificação dos Perfis dos Clientes.

<i>Classificação</i>	<i>Usual</i>	<i>Modelo</i>
1º	G	G
2º	H	D
3º	D	H
4º	I	I
5º	E	E
6º	A	F
7º	F	A
8º	B	B
9º	C	C

Analisando a tabela 47, observamos que o perfil G, está classificado em primeiro lugar nas duas classificações, e os perfis B, C, I e E se encontram na mesma posição nas duas classificações enquanto os demais perfis sofrem alteração de uma unidade para mais ou para menos em sua classificação.

Da tabela 46, temos que o atributo mais importante é o V.C. com 61,54% de importância.

Com estes resultados concluímos que as duas classificações seguem um padrão semelhante, e o cliente de maior importância para a empresa é o cliente com perfil G, e em

segundo lugar o cliente com perfil D.

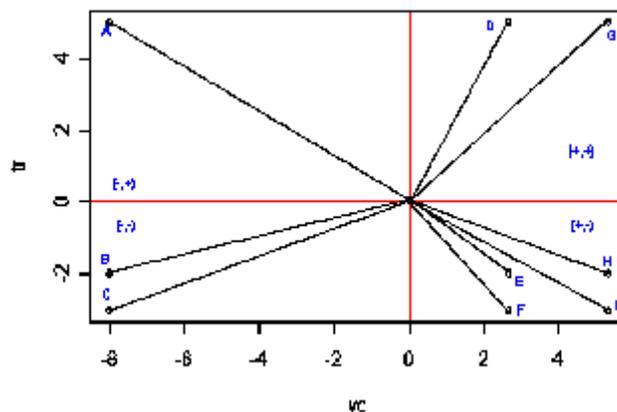
Onde:

$$\text{Perfil (G)} = V.C.(2000 - 3000) \text{ e } T.R.(1\text{mês})$$

$$\text{Perfil (D)} = V.C.(1000 - 2000) \text{ e } T.R.(1\text{mês})$$

Pode-se também analisar graficamente os resultados, construindo um gráfico contendo os atributos em seus eixos, como está apresentado no gráfico 7

Gráfico 7 - Tempo de relacionamento vs Valor do Cliente.



Somando as coordenadas dos pontos os quais representam os clientes, obtemos as utilidades totais dos perfis, efetuando os cálculos encontra-se que o ponto com maior valor da soma é o ponto G, logo este é o perfil de maior preferência pela empresa.

Analisando o gráfico 7 pode-se projetar os pontos sobre os eixos e verificar quais são mais semelhantes ao ponto G considerando um determinado atributo. Por exemplo, observa-se que os perfis G, D e A são idênticos considerando o atributo T.R., e em relação ao atributo V.C. temos que os perfis G, H e I são idênticos.

Com a identificação dos clientes mais importantes e conhecendo o atributo mais importante para a empresa, a empresa deve traçar o perfil completo destes clientes, e desenvolver estratégias de marketing, para buscar novos clientes no mercado, com perfil semelhante aos perfis mais importantes, que neste caso são, os perfis G e D. A empresa pode direcionar uma atenção especial ao atributo valor do cliente, na hora de buscar novos clientes e mesmo para desenvolver estratégias para reter seus clientes, pois este atributo foi dado como o mais importante para a empresa.

Capítulo 5

BIBLIOGRAFIA

[1] Ratner B.”Statistical Modeling and Analysis for Database Marketing: Effective Techniques for Mining Big Data”, Chaoman & Hall/CRC, 2003.

[2] AAKER A. D., KUMAR V., DAY S.G., “Pesquisa de Marketing 2º edição”, editora Atlas S.A., São Paulo, 2004.

[3] Clóvis Corrêa da COSTA, As Novas Fronteiras do Relacionamento com Clientes, São Paulo, Brasil, 2004;

[3] HAIR J.F.JR., ANDERSON R.E., TATHAM, BLACK W.C., “Multivariate Data Analysis 4º edição” Prentice Hall Publishers, New Jersey,1995.

[4] Drumond F.B., “Técnicas estatísticas para o planejamento de produtos” Universidade Federal de M.G. Dep. De Estatística., 2004.

[5] Hosmer David W, Lemeshow Stanley., "Applied logistic regression", 2º edição.