



Karine de Almeida Karam

**Regressão Logística: Um modelo de
Risco de Cancelamento de Clientes**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Administração de Empresas do Departamento de Administração da PUC-Rio.

Orientador: Prof. Jorge Ferreira da Silva

Rio de Janeiro
Março de 2006



Karine de Almeida Karan

**Regressão Logística: Um modelo de
Risco de Cancelamento de Clientes**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Administração de Empresas da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Jorge Ferreira da Silva

Orientador

Departamento de Administração – PUC-Rio

Prof. Marie Agnes Chauvel

Departamento de Administração - PUC-Rio

Prof. Leticia Moreira Casotti

UFRJ

Prof. João Pontes Nogueira

Vice-Decano de Pós-Graduação do CCS

Rio de Janeiro, 06 de março de 2006

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Karine de Almeida Karam

Graduou-se em Estatística pela UERJ.

Ficha Catalográfica

Karam, Karine de Almeida

Regressão logística: um modelo de risco de cancelamento de clientes / Karine de Almeida Karam ; orientador: Jorge Ferreira da Silva. – Rio de Janeiro : PUC, Departamento de Administração, 2006.

98 f. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Administração.

Inclui referências bibliográficas.

1. Administração – Teses. 2. Retenção. 3. Lealdade. 4. Satisfação. 5. Marketing de relacionamento. 6. Regressão logística. 7. Segmentação. I. Silva, Jorge Ferreira da. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Administração. III. Título.

CDD: 658

Para minha mãe, Maria José e minha irmã, Mônica pelo apoio e confiança.

Para Maurício, que me acompanhou neste ano de inteira dedicação.

Para os amigos, Flávia, Elisa e Sandro, pela inspiração.

Para Fontinha, pelos conselhos no momento certo.

Para Duda, pela orientação profissional.

Agradecimentos

Ao meu orientador Professor Jorge Ferreira da Silva pelo estímulo e parceria na realização deste trabalho e pelo aprendizado durante todo o curso. Agradeço, principalmente, pelas dicas preciosas e inteligentes nos momentos mais difíceis da construção deste projeto.

À PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

A Sonia da Biblioteca da PUC, pela incansável dedicação nas pesquisas deste tema.

Aos meus colegas da PUC-Rio.

Aos professores que participaram da Comissão examinadora.

A todos os professores e funcionários do Departamento pelos ensinamentos e pela ajuda.

A todos os amigos e familiares que de uma forma ou de outra me estimularam ou ajudaram-me.

Resumo

Karam, Karine de Almeida; da Silva, Jorge Ferreira; (Orientador). **Regressão Logística: Um Modelo de Risco de Cancelamento de Clientes.** Rio de Janeiro, 2006. 98p. Dissertação de Mestrado – Departamento de Administração de Empresas, Pontifícia Universidade Católica do Rio de Janeiro.

O tema central deste projeto é a retenção de clientes como estratégia competitiva para aumentar a lucratividade da empresa. O objetivo é desenvolver um modelo estatístico que relacione variáveis transacionais, demográficas e dados sobre o histórico de eventos com a probabilidade de cancelamento dos clientes assinantes de jornal e definir o perfil dos clientes com maior risco de desligamento. Em uma primeira etapa, este estudo fornece uma revisão teórica sobre lealdade, satisfação e marketing de relacionamento, a fim de buscar uma relação com a retenção de clientes. Em seguida, a revisão de literatura levantou as variáveis mais usadas na segmentação de clientes tais como: variáveis transacionais, geográficas, demográficas, psicográficas e comportamentais para definir o perfil dos clientes que cancelam e dos que não cancelam sua assinatura. Depois de construir um modelo teórico, a regressão logística foi utilizada como técnica estatística para desenvolver um modelo de previsão de cancelamento. Os resultados foram analisados com o auxílio do programa estatístico SPSS e conclui-se que o perfil do cliente que cancela a assinatura do jornal é o jovem de até 30 anos; com baixo nível sócio-demográfico; morador da baixada, subúrbio e outros estados que não o Rio de Janeiro; que tenha adquirido sua assinatura através do canal telemarketing ativo; com a assinatura da modalidade anual e forma de pagamento em boleto ou débito em conta corrente; clientes que adquiriram sua assinatura mais recentemente; que compreendem menos de 3 produtos da empresa e que não tenham feito reclamações através da central de atendimento. O modelo final de previsão de cancelamento contou com 11 variáveis e a tabela de classificação mostrou uma taxa de acerto geral de 75,3%. A última etapa apresenta algumas conclusões, implicações e sugestões para pesquisas futuras.

Palavras-chave

Retenção, lealdade, satisfação, marketing de relacionamento, regressão logística, segmentação e cancelamento.

Abstract

Karam, Karine de Almeida; da Silva, Jorge Ferreira; (Advisor). **Logistic Regression: A Model to Measure Signature's Cancellation Risk**. Rio de Janeiro, 2006. 98p. M.Sc. Dissertation – Departamento de Administração de Empresas, Pontifícia Universidade Católica do Rio de Janeiro.

The core subject of this project is the customers' retention as a competitive strategy to increase the company's profitability. The goal is to develop a statistical model that links transactional and demographic variables and customer's history data with the subscribers' churn of a certain publication. In the first part, this study provides a revision on loyalty, satisfaction and relationship marketing theory in order to find a relation with customers' retention. After that, the literature revision raised the most used variables for the segmentation of customers, such as: transactional, geographic, demographic, psychological and behavior variables to define the profile of the customer who churns and the profile of that one who doesn't. After constructing a theoretical model, the logistic regression was used as a statistical technique to develop a model of cancellation forecasting. The results has been analyzed with the aid of statistical program SPSS and conclude that the profile of the customer who cancels the subscription of the publication is young up to 30 years old; with low social-demographic level; living at Baixada, Suburb, and other states than Rio De Janeiro; that bought the subscription through the outbound telemarketing sales channel; with one year subscription and payment through invoice or direct debit in current account; customers who has bought its signature more recently; that do not buy less than 3 other products of the company and that have not made complaints through the customer service. The final model of churn forecasting uses 11 variables and the classification table showed an accuracy of 75,3%. The last part presents some conclusions, implications and suggestions for future research.

Keywords

Retention, loyalty, satisfaction, relationship marketing, logistic regression, segmentation and churn.

Sumário

1. Introdução	10
1.1. Objetivos do Estudo	13
1.2. Relevância do Estudo	14
1.3. Delimitação do Estudo	15
1.4. A Indústria Jornalística	15
2. Revisão de Literatura	19
2.1. O Conceito de Retenção	19
2.1.1. Retenção e MKt de Relacionamento	21
2.1.2. Retenção e Lealdade	25
2.1.3. Retenção e Satisfação	28
2.1.4. A Relação entre Retenção, Marketing de Relacionamento, Lealdade e Satisfação	30
2.2. Segmentação do Mercado Consumidor	32
2.2.1. Variáveis Geográficas	36
2.2.2. Variáveis Demográficas	37
2.2.3. Variáveis Psicográficas	39
2.2.4. Variáveis Comportamentais	41
2.3. Variáveis Transacionais	43
3.1. Tipo de Pesquisa	47
3.2. Universo e Amostra	48
3.3. Coleta dos Dados	49
3.4. Limitações do Método	49
3.5. Tratamento dos Dados	50
3.5.1. Problema e Objetivo	50
3.5.2. O Método de Regressão Logística	51
3.5.3. Coleta e Análise dos Dados	60
3.5.4. Definição das Variáveis	60
3.5.5. Descrição da Base de Dados	67
4. Análise dos Resultados	69
4.1. Construção do Modelo de Regressão Logística	69
4.2. Avaliação do Ajuste Geral	71
4.3. Precisão da Estimativa	73
4.4. Validação do Modelo (<i>holdout sample</i>)	75
4.5. Interpretação dos Resultados	76
5. Conclusões e Recomendações	84
5.1. Conclusões	84
5.2. Recomendações.	85
5.3. Recomendação para futuras pesquisas	87
6. Referências Bibliográficas	89
Anexo A – Estatísticas Descritivas	93
Anexo B – Distribuição Geográfica	97

Lista de Gráficos

Gráfico 1: Circulação Média Mensal / Ano – Mercado do Rio de Janeiro	16
Gráfico 2: Circulação Média Mensal / Ano – Mercado do Rio de Janeiro	17
Gráfico 3: VPL do Cliente x Risco de Desligamento	20
Gráfico 4: Curva Logística	54
Gráfico 5: Total de Assinaturas x Score de Risco de Cancelamento.	77
Gráfico 6: Score de Risco de Cancelamento x Gênero	78
Gráfico 7: Score de Risco de Cancelamento x Idade	78
Gráfico 8: Score de Risco de Cancelamento x Tipo De Assinatura	79
Gráfico 9: Score de Risco de Cancelamento x Região	79
Gráfico 10: Score de Risco de Cancelamento x Região	80
Gráfico 11: Score de Risco de Cancelamento x Participação em Ações de Fidelização	81
Gráfico 12: Score de Risco De Cancelamento x Canal de Venda	81
Gráfico 13: Score de Risco de Cancelamento x qtde de Produtos Agregados Comprados	82
Gráfico 14: Score de Risco de Cancelamento x Tempo de Permanência	82
Gráfico 15: Score de Risco de Cancelamento x SD &W	83
Gráfico 16: Score de Risco de Cancelamento x Indicador de Reclamação	83