

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Estatística

Mário Halfeld Clark Campos

Aplicação do Modelo Binomial aos Dados do Sítio Consumidor.gov.br

Juiz de Fora

2016

Mário Halfeld Clark Campos

Aplicação do Modelo Binomial aos Dados do Sítio Consumidor.gov.br

Monografia apresentada ao Departamento de Estatística da Universidade Federal de Juiz de Fora, na área de concentração em Estatística, como requisito parcial para obtenção do título de Bacharel em Estatística.

Orientador: Clécio da Silva Ferreira

Juiz de Fora

2016

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Campos, Mário Halfeld Clark.

Aplicação do Modelo Binomial aos Dados do Sítio Consumidor.gov.br /
Mário Halfeld Clark Campos. – 2016.
77 f. : il.

Orientador: Clécio da Silva Ferreira

Trabalho de Conclusão de Curso – Universidade Federal de Juiz de Fora,
Instituto de Ciências Exatas. Departamento de Estatística, 2016.

1. Índice de solução. 2. Modelo binomial. 3. Pontos de corte. 4.
Classificação. Ferreira, Clécio da Silva, orient. II. Título.

Mário Halfeld Clark Campos

Aplicação do Modelo Binomial aos Dados do Sítio Consumidor.gov.br

Monografia apresentada ao Departamento de Estatística da Universidade Federal de Juiz de Fora, na área de concentração em Estatística, como requisito parcial para obtenção do título de Bacharel em Estatística.

Aprovada em: 04/03/2016

BANCA EXAMINADORA

Clécio da Silva Ferreira (UFJF) - Orientador
Doutor em Estatística - Universidade de São Paulo

Augusto Carvalho Souza (UFJF)
Doutor em Economia - CEDEPLAR/UFMG

José Humberto Viana Lima Júnior (UFJF)
Doutor em Administração de Empresas – Escola de
Administração de Empresas de São Paulo EAESP/FGV

AGRADECIMENTOS

Primeiramente, a Deus e a seu filho Jesus Cristo, que tanto nos amou e nos ensinou em sua trajetória aqui na terra.

Aos meus pais, meu irmão e minha namorada, pelo apoio em toda minha trajetória até aqui.

Aos meus amigos, ao meu orientador Clécio e demais professores, pela paciência e dedicação.

RESUMO

Nesse trabalho utilizou-se a base de dados públicos do site do consumidor do governo federal (consumidor.gov.br), que é um serviço público para solução alternativa de conflitos de consumo (conflitos entre consumidores e empresas). Essa base de dados contém informações do consumidor e da reclamação. Após cadastrada a reclamação do indivíduo, a empresa terá dez dias para dar resposta a tal reclamação; e posterior a essa resposta o consumidor terá um prazo de vinte dias para classificar sua reclamação como "Resolvida" ou "Não Resolvida". Na prática porém, observa-se que há muitas reclamações que o indivíduo não avalia; isso dificulta a obtenção do real valor do índice de solução calculado pelo site, que é a proporção de reclamações resolvidas em relação ao total de reclamações por empresa. Assim, o governo considera as reclamações não avaliadas como "Resolvidas" para calcular esse índice; o presente trabalho ajustou um modelo binomial conjugado com critérios para obtenção de pontos de corte para classificação das reclamações não avaliadas como: "Resolvidas" ou "Não Resolvidas"; e desse modo pôde calcular o índice de solução. Em seguida, foi realizado um comparativo entre os índices de solução obtidos, ou seja, o índice calculado pelo governo e também, os demais índices obtidos segundo os critérios de classificação utilizados no modelo binomial. Quando comparados, percebeu-se certa heterogeneidade entre os índices de solução.

Palavras-chave: Índice de solução. Modelo binomial. Pontos de corte. Classificação.

ABSTRACT

In this work we used the public database of consumer website federal government (consumidor.gov.br), which is a public service for alternative solution of consumer disputes (conflicts between consumers and businesses). This database contains customer information and complaints. Once registered the individual complaint, the company will have ten days to respond to the claim; and subsequent to that the consumer will have a response within twenty days to sort his complaint as "Resolved" or "Unresolved". In practice however, it is observed that there are many complaints which the individual does not evaluate; Difficult to obtain the actual value of the settlement index calculated by the site, which is the proportion of complaints resolved in relation to the total complaints by company. Thus, the government considers complaints not assessed as "resolved" to calculate this index; this work fits a binomial model in conjunction with Criterion for obtainment cutoff points for classification of not evaluated complaints as "resolved" or "not resolved"; and thus could calculate the solution index. Then, there is a comparison between the obtained solution indexes, the index calculated by the government and also other indexes obtained using the classification used in the binomial model. When compared, certain Heterogeneity between the solution indexes was observed.

Key words: Solution index. Binomial model. Cutoff points. Classification.

LISTA DE ILUSTRAÇÕES

Figura 1 – Formulário Cadastro de Reclamação	15
Figura 2 – Gráfico em barras para variável: Região	22
Figura 3 – Gráfico em barras para variável: Sexo	23
Figura 4 – Gráfico de barras com ordenada em porcentagem para variável: Faixa Etária	24
Figura 5 – Gráfico de barras com ordenada em porcentagem da população residente no Brasil em 2013 por grupos de idade	24
Figura 6 – Gráfico em barras para variável: Mês Abertura	25
Figura 7 – Histograma para variável: Tempo Resposta	25
Figura 8 – Gráfico de barras para variável: Área	26
Figura 9 – Gráfico de barras para variável: Grupo Problema	27
Figura 10 – Gráfico de barras para variável: Como Comprou Contratou	27
Figura 11 – Gráfico de barras para variável: Avaliação Reclamação	28
Figura 12 – Gráfico em barras da distribuição da avaliação da reclamação por Mês de Abertura	29
Figura 13 – Gráfico em barras da distribuição da avaliação da reclamação por Tempo de Resposta	29
Figura 14 – Gráfico em barras da distribuição da avaliação da reclamação por Área	30
Figura 15 – Gráfico em barras da distribuição da avaliação da reclamação por Grupo Problema	31
Figura 16 – Gráfico em barras da distribuição da avaliação da reclamação por Como Comprou Contratou	31
Figura 17 – Funções de ligação do Quadro 2	35
Figura 18 – Exemplo de Curva ROC para um Determinado Modelo de Regressão Logística	39
Figura 19 – Gráfico da Proporção de Predição Correta por Ponto de Corte do Modelo de Regressão Logística	43
Figura 20 – Valores estimados dos β_j 's e $\exp \hat{\beta}_j$	44
Figura 21 – Gráfico da Curva ROC para As Reclamações Avaliadas	46
Figura 22 – Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados	63

LISTA DE QUADROS

Quadro 1 – Quadro Explicativo das Variáveis dos Consumidores	16
Quadro 2 – Quadro da Algumas Funções de Ligações Adequadas para Dados Binários	34
Quadro 3 – Quadro da Algumas Funções de Ligações Adequadas para Dados Binários	36

LISTA DE TABELAS

Tabela 1 – Cálculo do Coeficiente de Contigência para Algumas Variáveis	20
Tabela 2 – Média Aritimética para a Variável Tempo Reposta segundo a Variável Reclamação da Avaliação	21
Tabela 3 – Valor da Razão entre as Porcentagens das Regiões	22
Tabela 4 – Valores dos AIC"s de cada um dos Modelos Finais de Diferentes Funções de Ligação	41
Tabela 5 – Valores Máximos Obtidos para a Proporção de Predições Corretas e os Seus Respective Pontos de Corte para cada um dos Três Modelos . . .	42
Tabela 6 – Índice de Solução de Algumas Empresas Segundo os Quatro Critérios .	47

SUMÁRIO

1	INTRODUÇÃO	11
2	O CONSUMIDOR NO BRASIL	12
2.1	Conceituação	12
2.2	A Defesa do Consumidor no Brasil	12
2.3	Sites Utilizados para Reclamação dos Consumidores	12
3	ORIGEM DOS DADOS	14
3.1	O Site Consumidor.gov.br	14
3.2	Dados Abertos	15
3.2.1	Dados Abertos do Consumidor.gov.br	16
4	ANÁLISE EXPLORATÓRIA DOS DADOS	19
4.1	Análise da Associação da Variável Avaliação da Reclamação com as Demais Variáveis	19
4.2	Análise Exploratória Univariada de Algumas Variáveis	21
4.2.1	Variável Região	21
4.2.2	Variável Sexo	22
4.2.3	Variável Faixa Etária	23
4.2.4	Mês Abertura	24
4.2.5	Variável Tempo Resposta	25
4.2.6	Variável Área	26
4.2.7	Variável Grupo Problema	26
4.2.8	Variável Como Comprou Contratou	27
4.2.9	Variável Avaliação Reclamação	28
4.3	Análise Exploratória Bivariada das Variáveis de Indicativo de Associação com a Variável Avaliação Reclamação	28
4.3.1	Mês Abertura	28
4.3.2	Tempo Resposta	29
4.3.3	Variável Área	30
4.3.4	Variável Grupo Problema	30
4.3.5	Variável Como Comprou Contratou	31
5	MODELOS PARA DESFECHOS CATEGÓRICOS	32
5.1	Síntese Sobre os Modelos Lineares Generalizados (MLG's)	32
5.1.1	Componentes dos Modelos Lineares Generalizados	32
5.2	Modelo Binomial	33

5.2.1	Funções de Ligações Apropriadas para o Modelo Binomial	34
5.2.2	Estimação de β para o Modelo Binomial	36
5.2.3	Alguns Procedimentos para Seleção de Modelos	37
5.2.3.1	Método Forward	37
5.2.3.2	Método backward	37
5.2.3.3	Método Stepwise	38
5.2.3.4	Método Akaike	38
5.3	Curva ROC (<i>Receiver Operating Characteristic</i>)	38
6	RESULTADOS	41
6.1	Ajuste do Modelo Binomial	41
6.1.1	Escolha do Modelo Binomial através do confronto dos AIC's	41
6.1.2	Escolha do Modelo Binomial através do Confronto da Máxima Proporção de Predição Correta	42
6.1.3	Alguns Detalhes do Modelo de Regressão Logística Escolhido	44
6.2	Classificação das Avaliações das Reclamações Não Avaliadas pelo Consumidor	45
6.2.1	Utilizando o Critério Empregado no <i>Default</i> dos <i>Softwares</i>	45
6.2.2	Utilizando o Critério Máxima Proporção de Predição Correta	45
6.2.3	Utilizando o Critério de Classificação da Curva ROC	45
6.2.4	Forma de Classificação do Site do Governo	46
6.3	Construção do Índice de Solução a Partir dos Quatro Critérios de Classificação das Reclamações Não Avaliadas	47
7	CONCLUSÃO	48
	REFERÊNCIAS	50
	APÊNDICE A – Gráfico em barras de algumas das outras va- riáveis não analisadas	51
	APÊNDICE B – Códigos no script do <i>software</i> estatístico R	52
	APÊNDICE C – Índice de Solução de Todas as Empresas Constadas nos Dados pelos quatro critérios	63

1 INTRODUÇÃO

A relação de consumo no Brasil é muito intensa, o que acaba por gerar muitos conflitos. Devido esses conflitos, o governo federal criou um site para auxiliar nesse sentido¹, que tem como objetivo justamente solucionar esse imbróglio entre consumidor e empresa. O presente trabalho, coletou a base de dados aberta desse site, de modo a examiná-lo para aplicação de modelos. Essa base de dados contém informações do consumidor e da reclamação. Nesses dados há uma variável chamada de Avaliação Reclamação, que é a situação final da reclamação; e tem as seguintes classificações: "Resolvida", "Não Resolvida" e "Não Avaliada". As reclamações classificadas como não avaliadas, são aquelas a qual o indivíduo não avaliou; isso dificulta a obtenção do real valor do índice de solução calculado pelo site, que é a proporção de reclamações resolvidas em relação ao total de reclamações por empresa.

Há diversas pesquisas sobre a avaliação dos consumidores como a ELECTRE TRI aplicado à avaliação da satisfação de consumidores, elaborada por COSTA et al (2007), que teve como objetivo conhecer a percepção do consumidor. Um artigo também interessante é: Pesquisa de satisfação de clientes: o estado-da-arte e proposição de um método brasileiro; desenvolvida por ROSSI e SLONGO (1998).

Nesse trabalho, será utilizado o modelo binomial conjugado com critérios para obtenção de pontos de corte para classificação das reclamações não avaliadas como: "Resolvidas" ou "Não Resolvidas". Esses critérios são: A curva ROC, o critério da máxima proporção de predição correta; e o critério empregado no *default* dos *softwares*, que utiliza o ponto de corte igual 50% para uma classificação segundo as probabilidades estimadas obtidas em um modelo binomial.

Assim, o trabalho está organizado em 6 capítulos. O capítulo 1 faz uma breve contextualização e síntese do assunto a ser abordado. Já o Capítulo 2 apresenta o conceito de consumidor conforme a lei brasileira, resume o histórico da defesa do consumidor no Brasil; e introduz alguns dos sites para reclamação dos consumidores. O capítulo 3 faz uma explanação sobre a origem dos dados utilizados nesse trabalho, detalhando o site do consumidor do governo federal (consumidor.gov.br). A análise exploratória univariada e bivariada, além de análise de associação são realizadas no capítulo 4. No capítulo 5 são abordados modelos para desfechos categóricos, além de critérios para obtenção de ponto de corte. Os resultados com o ajuste de modelos binomiais, classificação das reclamações não avaliadas e cálculo dos índices de solução são apresentados no capítulo 6. Finalmente, no capítulo 7 é realizada uma conclusão a cerca dos resultados obtidos.

¹ <http://www.consumidor.gov.br/>

2 O CONSUMIDOR NO BRASIL

2.1 Conceituação

Conforme o artigo 2º da Lei nº 8.078/1990 BRASIL "Consumidor é toda pessoa física ou jurídica que adquire ou utiliza produto ou serviço como destinatário final."

2.2 A Defesa do Consumidor no Brasil

Segundo o site do Ministério da justiça do governo federal¹ a defesa do consumidor no Brasil começou a ser desenvolvida a partir da década de 1960, período esse que foi reconhecida a vulnerabilidade do consumidor e sua importância nas relações comerciais nos Estados Unidos, pois foi em 15 de março de 1962, em mensagem ao Congresso Nacional, que o então presidente John F. Kennedy reconheceu o caráter universal da proteção dos direitos dos consumidores; e conseqüentemente, no dia 15 de março é comemorado o Dia Mundial dos Direitos do Consumidor.

Ainda, de acordo com o esse mesmo site do governo federal, posteriormente, na década de 1970, surgiram os primeiros órgãos de defesa do consumidor, como por exemplo o Grupo Executivo de Proteção ao Consumidor (atual Fundação Procon São Paulo). Anos depois, a Organização das Nações Unidas, por meio da Resolução nº 39-248 de 1985, estabeleceu as Diretrizes para a Proteção do Consumidor, ressaltando a importância da participação dos governos na implantação de políticas de defesa do consumidor. Assim, com o advento da Constituição Federal de 1988, foi consagrado a proteção do consumidor como direito fundamental e princípio da ordem econômica (arts. 5º, XXXII, e 170, V), cabendo ao Estado a promoção da defesa do consumidor, na forma da lei.

Enfim, em 11 de setembro de 1990, por meio da Lei nº 8.078/90, foi criado o Código de Defesa do Consumidor, que assegura o reconhecimento da vulnerabilidade do consumidor e estabelece a boa-fé como princípio basilar das relações de consumo.

O site supracitado relatou também outro ato de suma importância ocorrido em 2012, que foi criação da Secretaria Nacional do Consumidor, que surgiu por meio do Decreto nº 7.738. Ela tem como função exercer as competências estabelecidas na Lei nº 8.078/90 tais como formular, promover, supervisionar e coordenar a Política Nacional de Proteção e Defesa do Consumidor.

2.3 Sites Utilizados para Reclamação dos Consumidores

Dentre os sites mais acessados tem-se o Reclame Aqui, o Consumidor.gov; e o Denuncio. Esses sites tem como objetivo principal sanar conflitos entre consumidores e

¹ <http://www.justica.gov.br/>

empresas.

Segundo o próprio site do reclame aqui ², o mesmo há 13 anos atua como canal de comunicação entre consumidores e empresas de todo o Brasil. Os dados das empresas e dos consumidores que estão cadastrados no site servem apenas para possibilitar o acesso e atendimento entre as duas partes. É importante ressaltar que o banco de dados do reclame aqui é totalmente sigiloso e não comercializável. Assim, ele funciona do seguinte modo: o consumidor digita o nome da empresa a qual quer reclamar e clica em buscar. Caso o indivíduo não encontre a empresa no cadastro do site, não tem problema, ele ainda pode fazer a busca por categorias ou solicitar o cadastramento dessa empresa em um link. Posteriormente, o consumidor preenche todos os campos da reclamação; confere a sua reclamação, se identifica no site fazendo o login ou se cadastrando; ficando pedente apenas a resposta por parte da empresa que será notificada em relação à sua reclamação. Uma observação interessante é que esse site recomenda que o cliente dissemine a sua reclamação pelas redes sociais para que ela ganhe ainda mais visibilidade.

Já em relação ao site do Denuncio³ de acordo com o próprio site ele é o espaço que todo consumidor deseja na rede mundial de computadores, tendo o consumidor brasileiro o poder de exercer seu direito expressando sua reclamação quanto a atendimento, compra, venda, produtos e serviços. Para publicar uma denúncia basta seguir os seguintes passos: Primeiro, o consumidor se cadastra no site; se já for cadastrado, efetua o seu login, em seguida, clica no link e faz a sua reclamação; digita o nome da empresa ou escolhe a empresa pela sua categoria; informa o nome, cidade, estado, e-mail e segmento da empresa; e após selecionar a empresa, preenche o campo título da reclamação e inclui o seu texto. Além disso, o consumidor reclamante tem a opção de preencher os seus telefones de contato, pois muitas empresas contatam nos telefones em anexo para resolver o problema. Assim, um aviso é encaminhado, automaticamente, via RSS e por e-mail à parte reclamada, caso a empresa tenha seu serviço de atendimento ao cliente cadastrado no Denuncio. Os dados de empresas e consumidores que integrarem o sistema do Denuncio são exclusivamente para possibilitar o acesso de ambos ao site e permitir a comunicação do Denuncio com seus cadastrados. O banco de dados do Denuncio é sigiloso e não comercializável.

No próximo capítulo será abordado com detalhes o site do governo federal (consumidor.gov.br), que foi a fonte dos dados desse trabalho.

² <http://www.reclameaqui.com.br/>

³ <http://www.denuncio.com.br/>

3 ORIGEM DOS DADOS

3.1 O Site Consumidor.gov.br

Os dados do trabalho são provenientes do site do Consumidor do governo federal¹. Conforme definições desse site, o mesmo é um novo serviço público para solução alternativa de conflitos de consumo disponibilizado por meio de plataforma tecnológica de informação, interação e compartilhamento de dados monitorada pelos Procons e pela Secretaria Nacional do Consumidor do Ministério da Justiça, com o apoio da sociedade.

Lançado em junho de 2014, o Consumidor.gov.br inaugurou um espaço público de atendimento ao consumidor, promovendo o diálogo efetivo entre consumidores e empresas de diversos segmentos.

A ferramenta, concebida com base em princípios de transparência e controle social, fornece ao Estado informações essenciais à elaboração e implementação de políticas públicas de defesa dos consumidores e incentiva a competitividade no mercado pela melhoria da qualidade e do atendimento ao consumidor.

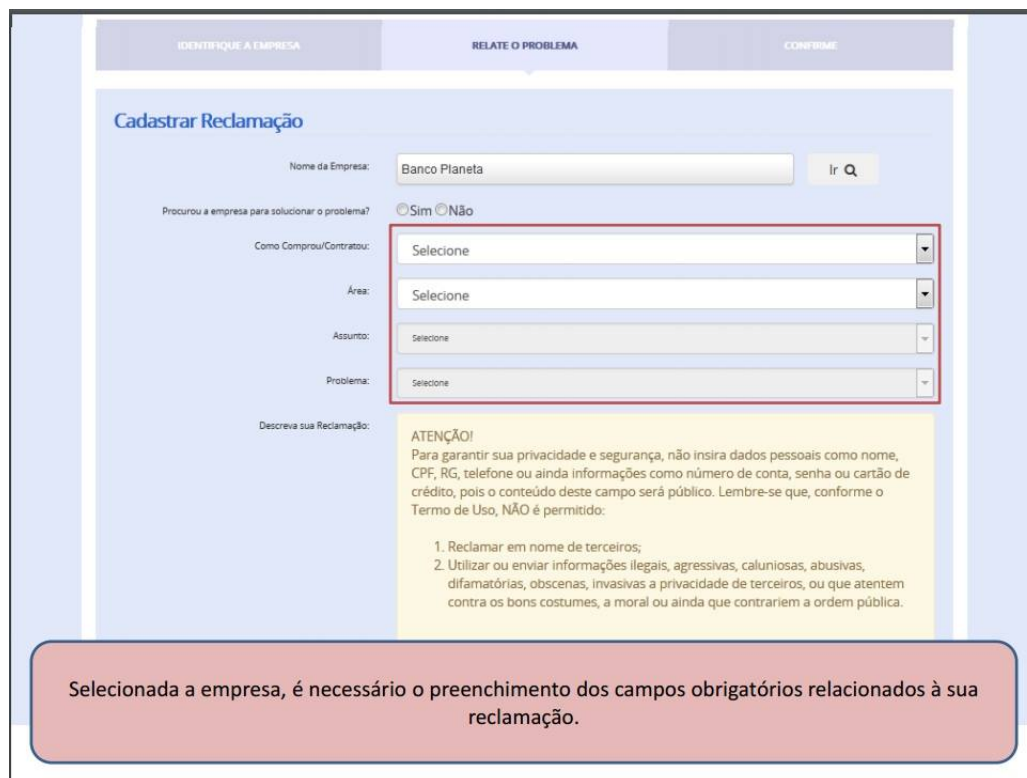
A participação das empresas na plataforma é voluntária, mas é preciso aderir formalmente ao serviço, mediante assinatura de termo no qual se comprometem a conhecer, analisar e investir todos os esforços disponíveis para a solução dos problemas apresentados. O consumidor, por sua vez, deve se identificar adequadamente e comprometer-se a apresentar todos os dados e informações relativas à reclamação relatada.

A Secretaria Nacional do Consumidor é a responsável pela gestão, disponibilização e manutenção do Consumidor.gov.br, bem como pela articulação com demais órgãos e entidades do Sistema Nacional de Defesa do Consumidor que, por meio de cooperação técnica, apoiam e atuam na consecução dos objetivos do serviço.

O funcionamento dessa plataforma é feita da seguinte maneira: Primeiro, o consumidor deve verificar se a empresa da qual quer reclamar está cadastrada no site; posteriormente, o consumidor registra sua reclamação e, a partir daí, é iniciada a contagem do prazo para manifestação da empresa, que tem até 10 dias para analisar e responder a reclamação. Em seguida, é garantida ao consumidor a chance de em até 20 dias, comentar a resposta recebida, classificar a demanda como resolvida ou não resolvida, e ainda indicar seu nível de satisfação com o atendimento recebido que é um escore de 1 a 5. Os dados das reclamações registradas alimentam uma base de dados que é pública. Veja na Figura 1, o formulário do cadastro de reclamação do site:

¹ <http://www.consumidor.gov.br/>

Figura 1 – Formulário Cadastro de Reclamação



Cadastrar Reclamação

Nome da Empresa: Banco Planeta Ir Q

Procurou a empresa para solucionar o problema? Sim Não

Como Comprou/Contratou: Selecione

Área: Selecione

Assunto: Selecione

Problema: Selecione

Descreva sua Reclamação:

ATENÇÃO!
Para garantir sua privacidade e segurança, não insira dados pessoais como nome, CPF, RG, telefone ou ainda informações como número de conta, senha ou cartão de crédito, pois o conteúdo deste campo será público. Lembre-se que, conforme o Termo de Uso, NÃO é permitido:

1. Reclamar em nome de terceiros;
2. Utilizar ou enviar informações ilegais, agressivas, caluniosas, abusivas, difamatórias, obscenas, invasivas a privacidade de terceiros, ou que atentem contra os bons costumes, a moral ou ainda que contrariem a ordem pública.

Selecionada a empresa, é necessário o preenchimento dos campos obrigatórios relacionados à sua reclamação.

Fonte: Site www.consumidor.gov.br.

Consideram-se apenas as reclamações finalizadas aquelas que já tiveram os prazos de resposta da empresa (máximo 10 dias) e de avaliação do consumidor (máximo 20 dias) transcorridos.

De acordo com o site do G1/jornal hoje² edição do dia 18/02/2015 o objetivo do serviço é diminuir o número de casos que vão parar na Justiça. Para se ter uma noção da situação, 70 por cento dos processos que estavam nos juizados especiais na época dessa publicação do G1, são de problemas do consumidor. Cada um deles custava, em média, 1700 reais para o Governo.

3.2 Dados Abertos

Segundo o site do ministério da justiça³ os dados abertos são dados publicados em um formato legível por máquina e sem restrição de licenças, patentes ou mecanismos de controle, de modo a estarem livremente disponíveis para serem utilizados e redistribuídos à vontade. A política de dados abertos governamental, pauta que vem ganhando cada vez mais atenção no âmbito do Poder Público, trata da disseminação de informações públicas na internet, com a finalidade de que possam ser reutilizadas por toda a sociedade.

² <http://http://www.g1.globo.com/jornal-hoje/>

³ <http://www.justica.gov.br/>

Ainda, de acordo com o mesmo, um dos formatos abertos mais comuns de publicação é o CSV (Comma-Separated Values, ou “valores separados por vírgula”), para armazenamento de dados tabulares em texto.

3.2.1 Dados Abertos do Consumidor.gov.br

De posse das informações que já foram explicitadas nesse trabalho sobre o funcionamento da reclamação via site consumidor.gov.br; obtém-se uma base de dados anonimizadas sobre as reclamações cadastradas.

Dessa forma, foram baixados pelo site Consumidor.gov.br o arquivo em formato CSV com os dados de todas as reclamações finalizadas no 1º semestre de 2015. As variáveis desses dados são explicitadas através do Quadro 1 (veja-o abaixo); que é proveniente do dicionário dessa base dados aberto também disponibilizada pelo site.

Quadro 1 – Quadro Explicativo das Variáveis dos Consumidores

(Continua)

Coluna	Tipo/Tamanho	Descrição	Exemplo de Dado
Região	CHAR(2)	Sigla da região geográfica do consumidor reclamante	“CO” (centro-oeste), “N” (norte), “SE” (sudeste), etc.
UF	CHAR(2)	Sigla do estado do consumidor reclamante	“AM”, “DF”, “ES”, “PE”, “SP”, etc.
Cidade	VARCHAR(100)	Município do consumidor reclamante	“Brasília”, “Manaus”, “Recife”, “São Paulo”, etc.
Sexo	CHAR(1)	Sigla do sexo do consumidor reclamante	“F” (feminino) e “M” (masculino)
Faixa Etária	VARCHAR(30)	Faixa etária do consumidor	“até 20 anos”, “entre 21 a 30 anos”, etc.
Ano Abertura	INT	Ano de abertura da reclamação pelo consumidor	“2014”, “2015”, etc.
Mês Abertura	INT	Número do mês de abertura da reclamação pelo consumidor	“5” (maio), “6” (junho), “7” (julho), “8” (agosto), etc.
Data Abertura	CHAR(10)	Data de abertura da reclamação pelo consumidor	“04/11/2014”, etc. (formato dd/mm/aaaa)
Data Resposta	CHAR(10)	Data de resposta da reclamação pela empresa	“04/11/2014”, etc. (formato dd/mm/aaaa)
Data Finalização	CHAR(10)	Data de finalização da reclamação	“04/11/2014”, etc. (formato dd/mm/aaaa)

Quadro 1 - Quadro Explicativo das Variáveis dos Consumidores

(Conclusão)

Coluna	Tipo/Tamanho	Descrição	Exemplo de Dado
Tempo Resposta	CHAR(2)	Número de dias para a resposta da reclamação pela empresa. Diferença, em dias, entre a Data da Resposta e a Data de Abertura	“0”, “1”, “2”, “3”, “4”, “5”, “6”, “7”, “8”, “9” e “10”.
Nome Fantasia	VARCHAR(100)	Nome pelo qual a empresa reclamada é conhecida no mercado	“Claro”, “Oi”, “Tim”, “Vivo”, etc.
Segmento de Mercado	VARCHAR(100)	Principal segmento de mercado da empresa participante	“Administradoras de Consórcios”, etc.
Área	VARCHAR(100)	Área à qual pertence o assunto objeto da reclamação	“Serviços Financeiros”, “Telecomunicações”, etc.
Assunto	VARCHAR(250)	Assunto objeto da reclamação	“Atendimento Bancário”, “Telefonia Fixa”, etc.
Grupo Problema	VARCHAR(150)	Agrupamento do qual faz parte o problema classificado na reclamação	“Atendimento / SAC”, “Contrato / Oferta”, etc.
Problema	VARCHAR(250)	Descrição do problema objeto da reclamação	“Portabilidade não efetivada”, etc.
Como Comprou Contratou	VARCHAR(50)	Descrição do meio utilizado para contratação/aquisição do produto ou serviço reclamado	“Loja física”, “Telefone”, “Internet”, etc.
Procurou Empresa	CHAR(1)	Sigla da resposta do consumidor à pergunta: “Procurou a empresa para solucionar o problema?”	“S” (sim) e “N” (não)
Respondida	CHAR(1)	Sigla que indica se a empresa respondeu à reclamação ou não	“S” (sim) e “N” (não)
Situação	VARCHAR(50)	Situação atual da reclamação no sistema	“Finalizada avaliada”, “Respondida”, etc.
Avaliação Reclamação	VARCHAR(50)	Classificação atribuída pelo consumidor sobre o desfecho da reclamação	“Resolvida”, “Não resolvida” e “Não Avaliada”.
Nota do Consumidor	CHAR(1)	Número da nota de 1 a 5 atribuída pelo consumidor ao atendimento da empresa	“1”, “2”, “3”, “4” e “5”

Será realizada no próximo capítulo, uma análise exploratória de algumas das variáveis do Quadro 1. Precedentemente, serão efetuadas investigações de associação de algumas variáveis com a variável Avaliação Reclamação, considerando apenas as reclamações avaliadas pelo consumidor. Em seguida será executada uma análise univariada de variáveis relevantes; e uma análise exploratória bivariada daquelas variáveis que apresentaram indicativo de uma possível associação com a variável Avaliação Reclamação.

4 ANÁLISE EXPLORATÓRIA DOS DADOS

Conforme as sugestões de BUSSAB e MORETTIN (2010) é importante obter a maior quantidade possível de informações dos dados para que se indique modelos plausíveis a serem utilizados numa fase posterior; assim, faz-se necessário a realização de uma análise exploratória de algumas das variáveis apresentadas no Quadro 1 para que haja uma redução, compreensão e interpretação das mesmas.

Porém, antes de proceder a análise exploratória de algumas variáveis, serão feitas análises de associação de algumas variáveis com a variável avaliação da reclamação, considerando apenas as reclamações avaliadas, ou seja, aquelas que obtiveram avaliação por parte do consumidor como resolvida ou não resolvida, que representa um total de cerca de 62,3%. É importante salientar que essas análises de associação serão realizadas somente com variáveis que façam sentido examinar e que não tiveram excesso de dados faltantes, pois limitaria muito a aplicação de um modelo binomial.

Finalmente, será feita uma análise exploratória univariada de algumas variáveis e uma análise exploratória bivariada daquelas que apresentaram indicativo de uma possível associação com a avaliação da reclamação do consumidor.

Além disso, encontra-se no APÊNDICE A, gráficos exploratórios para outras variáveis que não foram investigadas nessa seção.

4.1 Análise da Associação da Variável Avaliação da Reclamação com as Demais Variáveis

Utilizou-se como medida de associação entre variáveis qualitativas o coeficiente de contingência, dado pela expressão 4.1, que foi proposto por Pearson (1904); e que pode apresentar valores entre 0 e 1, sendo quanto mais próximo de 1 maior a associação entre as variáveis e quanto mais próximo a 0 menor a associação entre as variáveis. Observa-se que, na expressão abaixo que χ^2 representa o valor da estatística qui-quadrado de independência e n o total geral da tabela de contingência de dupla entrada em questão.

$$P = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad (4.1)$$

Dessa forma, foram calculados os valores dos coeficientes de contingência, entre a variável Avaliação da Reclamação (considerando somente as categorias Resolvida; e Não Resolvida) e as variáveis Região, Sexo, Faixa Etária, Mês Abertura, Área, Grupo Problema, Como Comprou/Contratou, Procurou Empresa e Respondida; as demais variáveis não foram utilizadas por terem excesso de dados faltantes ou não fazerem muito sentido no estudo em questão.

A Tabela 1 contém os valores do coeficientes de contingência (calculados através do programa SPSS(2011)) entre a variável Avaliação da Reclamação e as variáveis Região, Sexo, Faixa Etária, Mês abertura, Área, Grupo Problema, Como Comprou/Contratou; e Procurou Empresa. Exposto os valores obtidos, adotou-se $P < 0,07$ como critério de indicação de baixíssima associação com a variável Avaliação da Reclamação, pois se considera um coeficiente de contingência de valor igual 0,15 como de baixa associação, portanto um valor de 0,07 (menor que metade de 0,15) indica um ainda menor indicativo de associação entre as variáveis. Logo as variáveis Região, Sexo, Faixa Etária e Procurou Empresa foram eliminadas da análise bivariada.

Tabela 1 – Cálculo do Coeficiente de Contingência para Algumas Variáveis

Variável	Coeficiente de Contingência
Região	0,042
Sexo	0,026
Faixa Etária	0,048
Mês Abertura	0,079
Área	0,196
Grupo Problema	0,137
Como Comprou/Contratou	0,121
Procurou Empresa	0,000

Fonte: Elaborada pelo Autor, 2016.

No caso da variável Tempo Resposta, por se tratar de uma variável quantitativa discreta, utilizou-se a média aritmética dentro de cada categoria da variável Avaliação da Reclamação para uma investigação inicial do grau de associação entre as mesmas. Essas medidas estão na Tabela 2.

Na Tabela 2, observa-se que há uma certa diferença entre as médias dos tempos de resposta das duas categorias. Logo, por essa análise prévia, percebe-se que a princípio não há motivo para eliminação da variável Tempo Resposta no que tange a associação com a variável Avaliação da Reclamação.

Tabela 2 – Média Aritimética para a Variável Tempo Reposta segundo a Variável Reclamação da Avaliação

Avaliação da Reclamação	\bar{x}
Resolvida	6,839
Não Resolvida	7,501
Todas (incluindo Não Avaliadas)	7,182

Fonte: Elaborada pelo Autor, 2016.

Assim, nas próximas seções será realizada a análise exploratória para examinar melhor algumas das variáveis.

4.2 Análise Exploratória Univariada de Algumas Variáveis

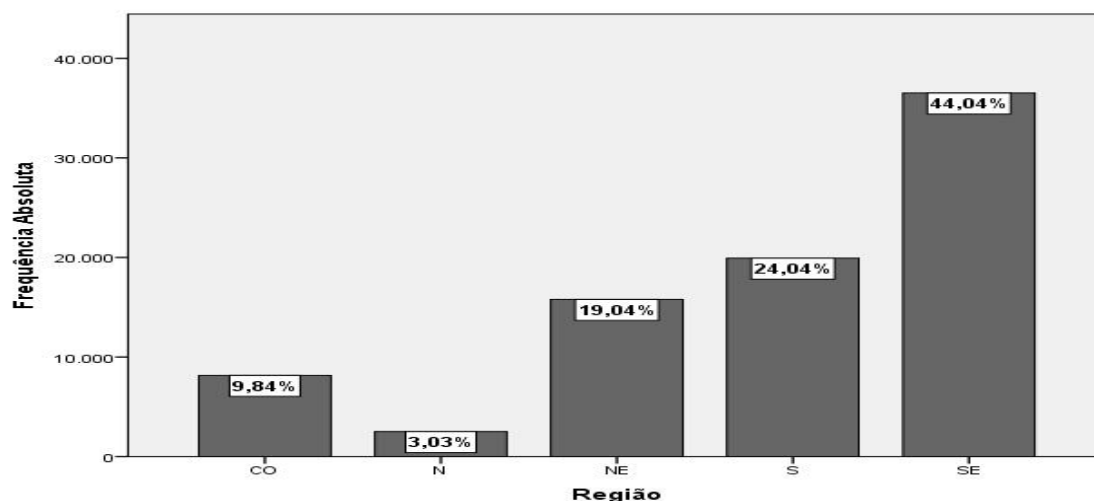
Nessa seção será realizada uma análise exploratória univariada representada através de gráficos de barras de algumas variáveis julgadas interessantes pelo autor. Também será feito um histograma para variável Tempo Resposta.

4.2.1 Variável Região

Sendo Região uma variável qualitativa nominal, apresenta-se na Figura 2 o gráfico de barras que destaca as regiões Sudeste, Sul, e Nordeste; respectivamente como as regiões onde houve mais registros de reclamações na plataforma do site; nele contém também, as porcentagens das reclamações por região. Nota-se que a região Sudeste com 44,04% do total das reclamações supera as regiões Nordeste, Norte e Centro-Oeste; que juntas perfazem um total de 31,91%. Através de dados do site do IBGE¹ da estimativa da população residente no Brasil e unidades da federação com data de referência em 1º de julho de 2014. Tem-se a região Sudeste com aproximadamente 41,97% da população total do período de referência, já a regiões Sul, Nordeste, Centro-oeste e Norte tiveram 14,31%, 27,71%, 7,506% e 8,50%; respectivamente. Diante desses dados, foi realizado um comparativo rápido através da razão entre da porcentagem de reclamação por região e porcentagem da estimativa da população por região mencionadas acima, veja Tabela 3. Houve uma heterogeneidade nesses valores, fato que sugere uma investigação mais apurada quanto a possíveis causas, como por exemplo, falta de uma melhor divulgação da plataforma do site consumidor.gov em algumas regiões; entre outras.

¹ <http://www.ibge.gov.br/>

Figura 2 – Gráfico em barras para variável: Região



Fonte: Elaborado pelo Autor, 2016.

Tabela 3 – Valor da Razão entre as Porcentagens das Regiões

Regiões	Valor da Razão
Sudeste	1,05
Sul	1,68
Nordeste	0,69
Centro-oeste	1,31
Norte	0,36

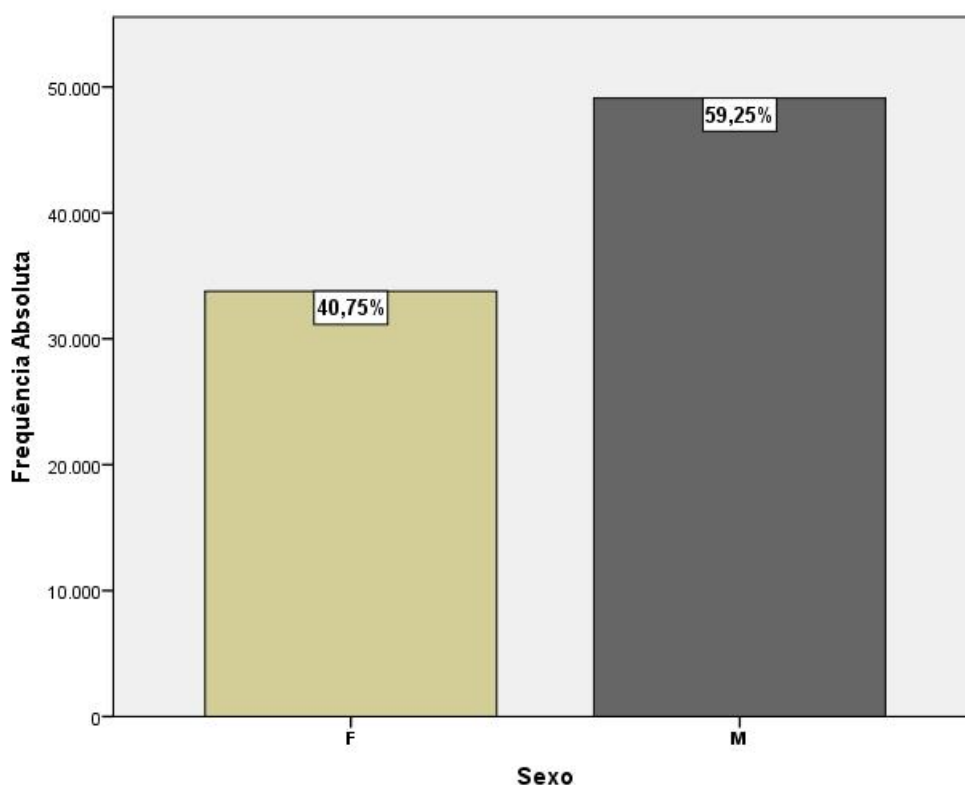
Fonte: Elaborada pelo Autor, 2016.

4.2.2 Variável Sexo

Observando a variável Sexo, a qual é classificada como Qualitativa nominal, apresenta-se na Figura 3 o gráfico de barras, onde percebe-se que em geral há mais reclamação dos homens com um percentual igual 59,25%. Segundo o site do governo² dados da última Pesquisa Nacional por Amostra de Domicílio (PNAD), divulgada pelo IBGE em 2013, indicaram que viviam no Brasil 103,5 milhões de mulheres, o equivalente a 51,4% da população. Isso, novamente sugere uma investigação mais apurada quanto a possíveis causas dessa heterogeneidade da reclamações dos consumidores por sexo no portal.

² <http://www.brasil.gov.br/cidadania-e-justica/2015/03/mulheres-sao-maioria-da-populacao-e-ocupam-mais-espaco-no-mercado-de-trabalho/>

Figura 3 – Gráfico em barras para variável: Sexo



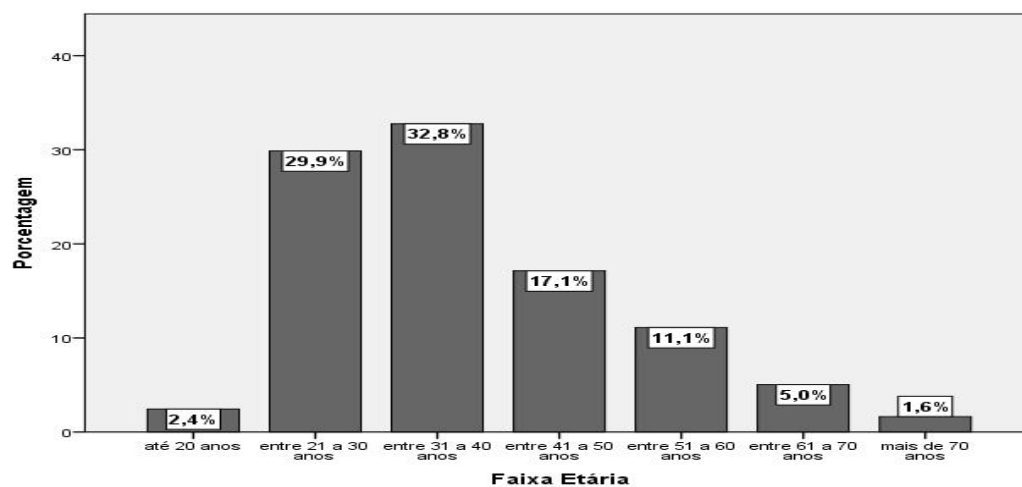
Fonte: Elaborado pelo Autor, 2016.

4.2.3 Variável Faixa Etária

A fim de analisar a variável Faixa Etária de classificação qualitativa ordinal, elaborou-se um gráfico de barras com ordenada y em porcentagem representada pela Figura 4, que destaca as faixas etárias de 31 a 40; e 21 a 30 como as que mais reclamam na plataforma do site com respectivamente; 32,8% (equivale a 27.160 reclamações) e 29,9% (equivale a 24.771) do total das reclamações. Também foi obtido através de pesquisa do autor no site do IBGE³, dados de indicadores de aspectos demográficos do Brasil, que foram elaborados, principalmente, a partir da Pesquisa Nacional por Amostra de Domicílios - PNAD 2013, realizada pelo IBGE. De posse desses dados o autor desse trabalho elaborou a Figura 5, onde mostra o gráfico da distribuição da população brasileira segundo diferentes faixas etárias no ano 2013. Analisando as duas Figuras (4 e 5) percebe-se certo contraste entre as porcentagens da população brasileira; e a porcentagem das reclamações no referido site do governo segundo as faixas etárias; embora, claro não há como fazer um comparativo robusto, pois as faixas etárias de ambos os gráficos são ligeiramente distintas; e os anos de ambos os gráficos também distintos.

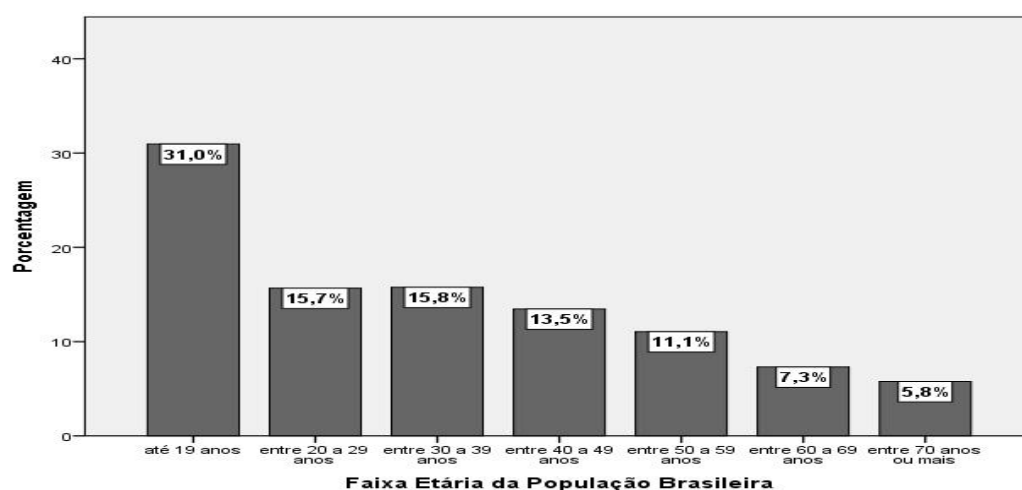
³ <http://www.ibge.gov.br/>

Figura 4 – Gráfico de barras com ordenada em porcentagem para variável: Faixa Etária



Fonte: Elaborado pelo Autor, 2016.

Figura 5 – Gráfico de barras com ordenada em porcentagem da população residente no Brasil em 2013 por grupos de idade

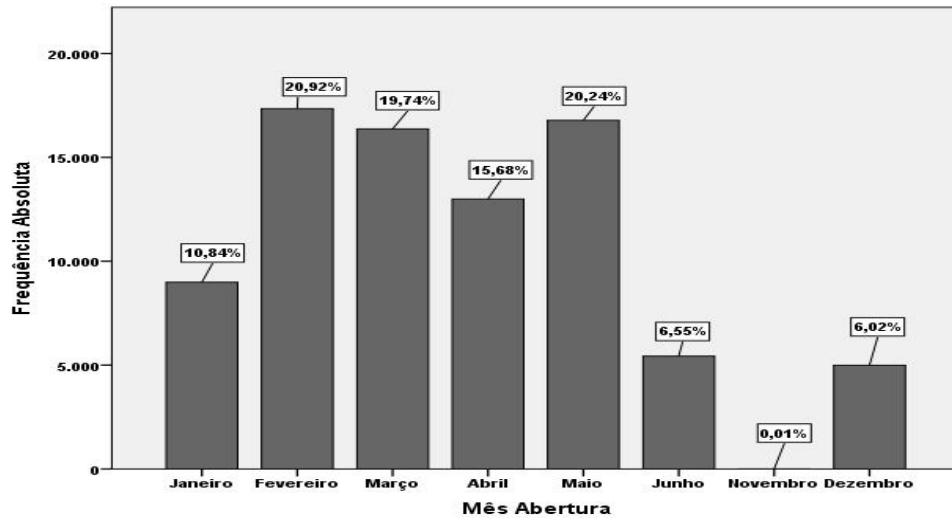


Fonte: Elaborado pelo Autor, 2016.

4.2.4 Mês Abertura

Fazendo a análise exploratória desta vez da variável Mês Abertura, classificada como qualitativa ordinal, observa-se na Figura 6 um gráfico em barras dos meses de abertura das reclamações dos clientes, a qual se destacam os meses de Fevereiro, Março; e Maio como aqueles que obtiveram maior frequência absoluta de reclamação, representando, respectivamente, 20,92%; 19,74% e 20,24% do total geral de Reclamações registradas pelo site do governo. É importante salientar, que por se trata de dados de reclamações finalizadas no 1º semestre de 2015, não há frequência absoluta de alguns meses do ano, como por exemplo os meses de agosto, setembro e outubro.

Figura 6 – Gráfico em barras para variável: Mês Abertura

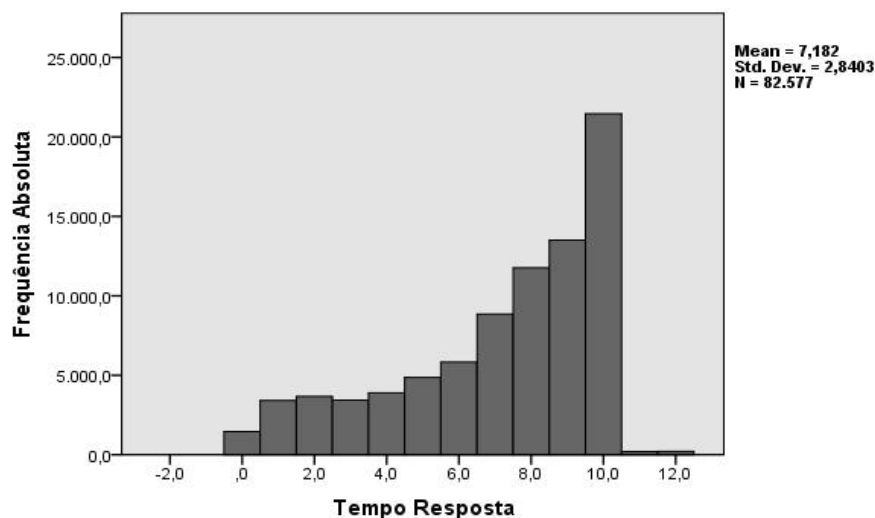


Fonte: Elaborado pelo Autor, 2016.

4.2.5 Variável Tempo Resposta

Já em relação a variável Tempo Resposta, que é classificada como quantitativa discreta, optou-se por construir o histograma desses tempos de resposta das empresas cadastradas no site, representado na Figura 7. Observa-se, nesse histograma, que a maior parte das empresas respondem as reclamações entre seis (incluso) e dez dias (incluso), sendo juntas responsável por um total de 74,1% do total das reclamações. Destaca-se que 25,9% das empresas respondem em 10 dias, fato esse que demonstra a dificuldade dessas em ser mais ágeis nesse quesito.

Figura 7 – Histograma para variável: Tempo Resposta

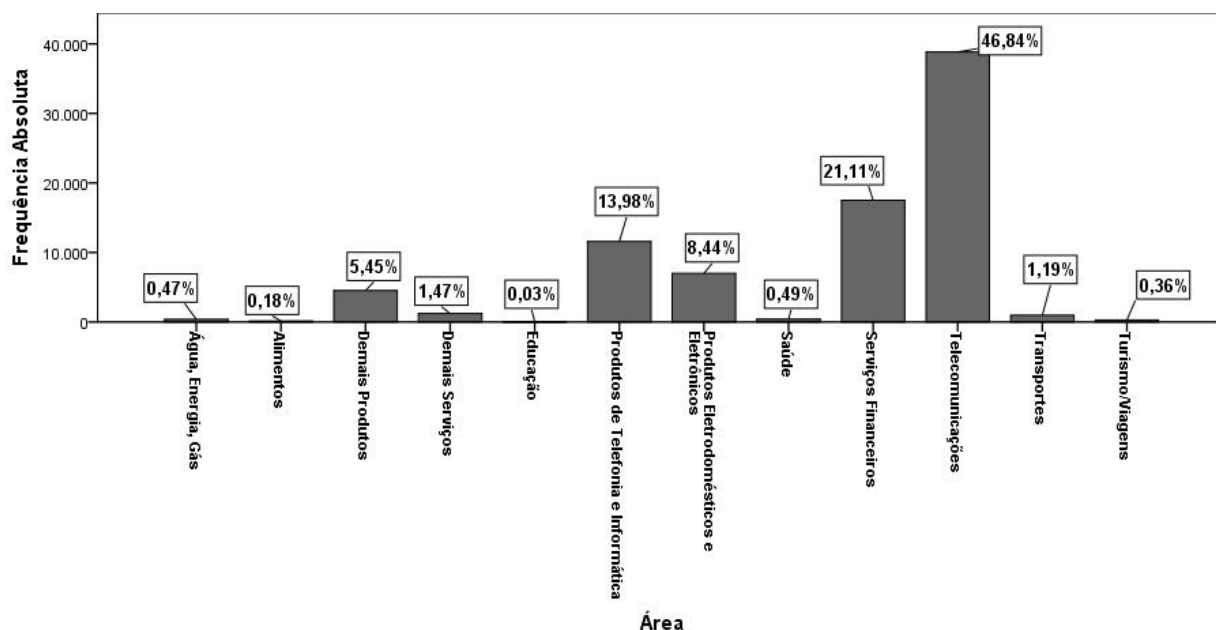


Fonte: Elaborado pelo Autor, 2016.

4.2.6 Variável Área

Analisando desta vez a variável Área, classificada como qualitativa nominal, tem-se na Figura 8 o gráfico em barras univariado, o qual apresenta como destaque as áreas de telecomunicações, serviços financeiros; e produtos de telefonia e informática com a maior frequência entre o total geral das reclamações registradas no site do governo, essas representam respectivamente: 46,84%; 21,11% e 13,98%.

Figura 8 – Gráfico de barras para variável: Área

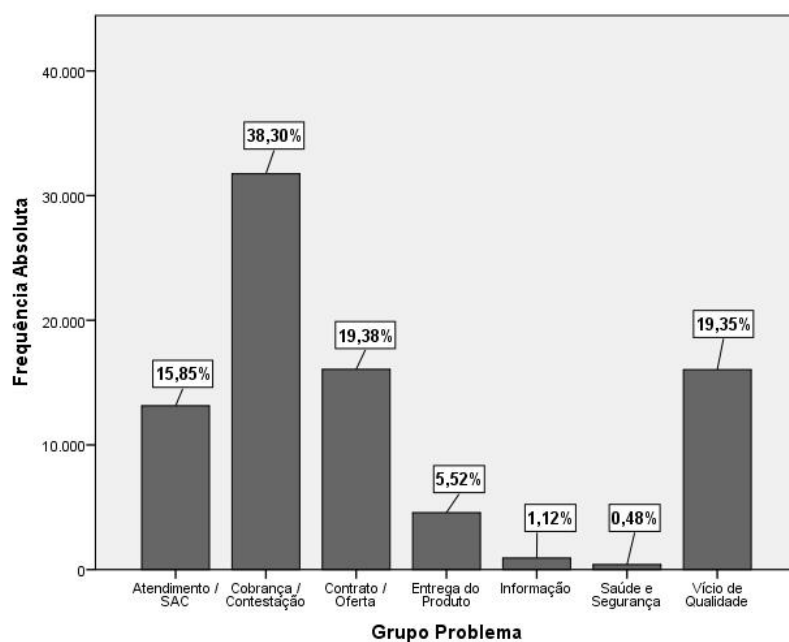


Fonte: Elaborado pelo Autor, 2016.

4.2.7 Variável Grupo Problema

Na análise da variável Grupo Problema, também qualificada como qualitativa nominal, observa-se na Figura 9 o gráfico em barras em que os agrupamentos: Cobrança/Contestação, Contrato/Oferta; Vício de Qualidade; e Atendimento/ SAC são os "vilões" com as maiores frequências do total geral das reclamações no site. Esse tipos de agrupamentos citados acima, tem respectivamente, as seguintes porcentagens: 38,30%; 19,38%; 19,35% e 15,85%.

Figura 9 – Gráfico de barras para variável: Grupo Problema

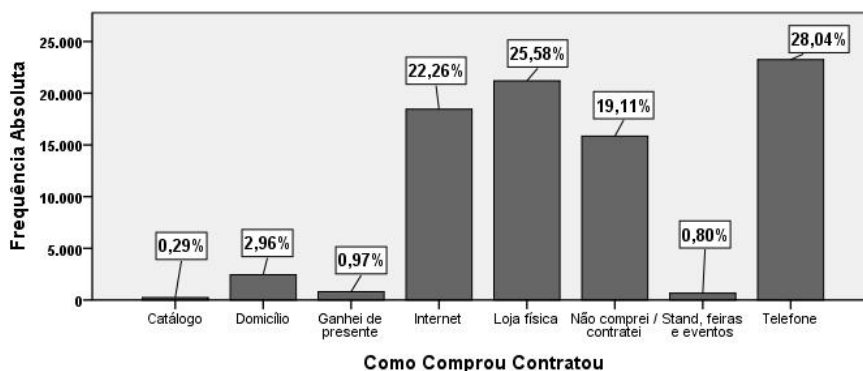


Fonte: Elaborado pelo Autor, 2016.

4.2.8 Variável Como Comprou Contratou

Explorando agora a variável Como Comprou/Contratou, que é qualificada como qualitativa nominal, observa-se na Figura 10 o gráfico em barras, a qual se destaca os seguintes meios de contratação ou aquisição: Telefone representando 28,04% do total de reclamações no site, loja física com 25,58%, internet com 22,26%; e não comprei/ contratei (quando o consumidor contratou determinado serviço ou produto, ao invés de comprar) representando 19,11%.

Figura 10 – Gráfico de barras para variável: Como Comprou Contratou

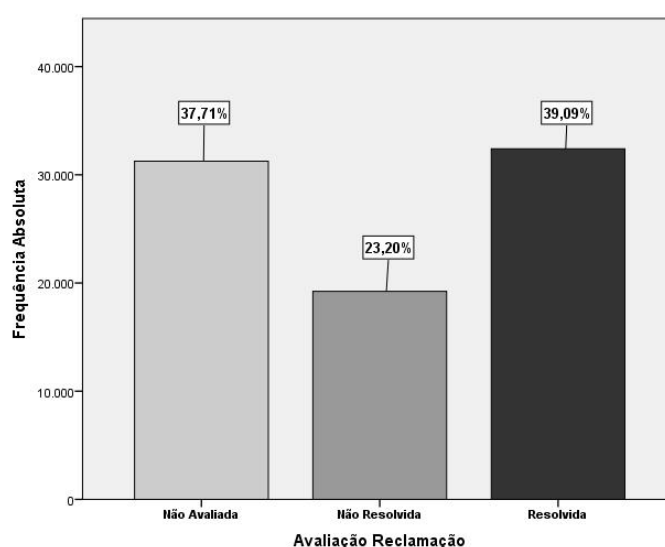


Fonte: Elaborado pelo Autor, 2016.

4.2.9 Variável Avaliação Reclamação

Realizando a análise univariada para variável Avaliação Reclamação, representa-se na Figura 11 um gráfico em barras com a frequência absoluta das três diferentes classificações das reclamações atribuídas pelos consumidores. Nesse gráfico, destaca-se uma maior proporção de avaliações resolvidas com 39,09%, ao passo que há 23,20% de avaliações classificadas como não resolvida. Sendo assim, infelizmente há 37,71% das reclamações que não obtiveram avaliação por parte do consumidor; e que portanto deixam indefinido o real desfecho da reclamação.

Figura 11 – Gráfico de barras para variável: Avaliação Reclamação



Fonte: Elaborado pelo Autor, 2016.

4.3 Análise Exploratória Bivariada das Variáveis de Indicativo de Associação com a Variável Avaliação Reclamação

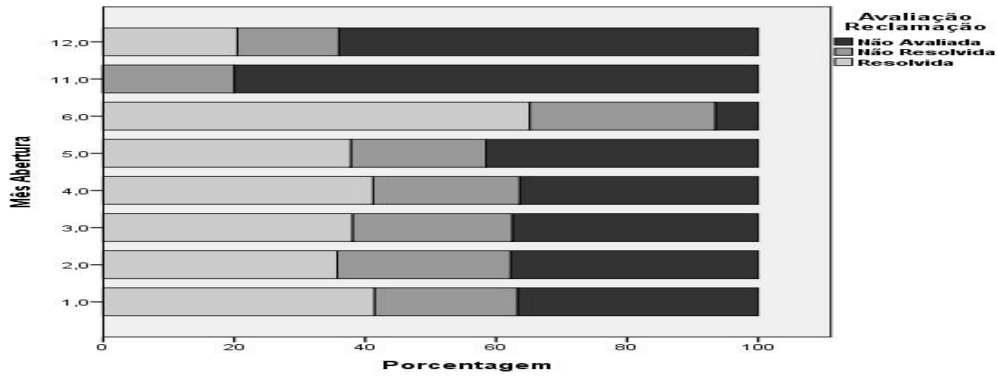
Nessa seção será realizada uma análise exploratória bivariada representada através de gráficos de barras da distribuição da avaliação da reclamação pelas categorias daquelas variáveis que apresentaram coeficiente de contingência P maiores ou iguais a 0,07. Veja novamente Tabela 1. Também será feito o mesmo gráfico para variável Tempo Resposta, que através da verificação da Tabela 2 obteve-se indícios de associação com a variável Avaliação Reclamação.

4.3.1 Mês Abertura

Dando início a análise bivariada, na Figura 12 é representado um gráfico com as proporções dos tipos de avaliação final da reclamação em relação ao total (100%) de cada um dos meses de abertura, representados pela variável Mês Abertura. Percebe-se,

nesse gráfico, que há diferenças nas proporções de Avaliações classificadas como Resolvida em cada Mês de abertura; isso também ocorre nas avaliações classificadas como Não Resolvidas. Assim, tornando a suspeita inicial de associação entre as variáveis Avaliação Reclamação e Mês Abertura ainda mais provável.

Figura 12 – Gráfico em barras da distribuição da avaliação da reclamação por Mês de Abertura

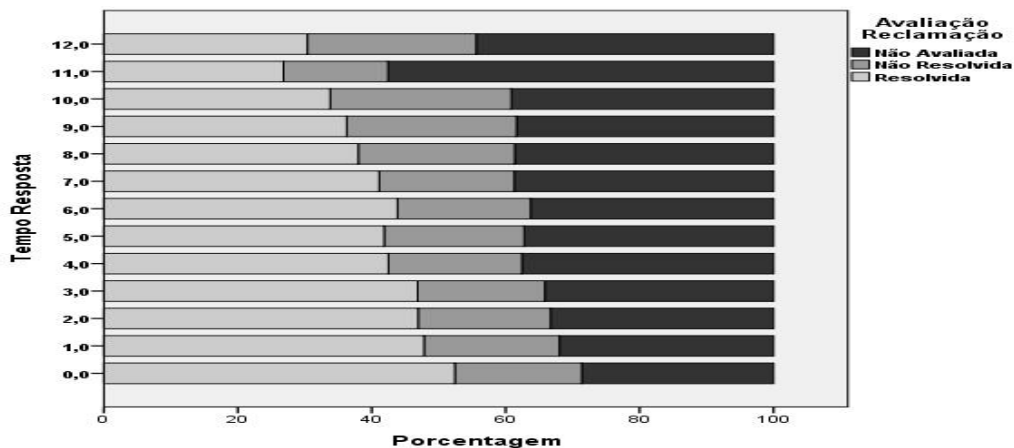


Fonte: Elaborado pelo Autor, 2016.

4.3.2 Tempo Resposta

Examinando desta vez o tempo de resposta em dias da reclamação do consumidor, tem-se na Figura 13, o gráfico com as proporções dos tipos de avaliação final da reclamação em relação ao total (100%) de cada um dos tempos de resposta, representados pela variável Tempo Resposta. Observa-se que, nos diversos tempos de resposta, que há diferenças visíveis das proporções dos tipos de avaliações finais Resolvidas, principalmente, nas que foram respondidas com 0 dias (em menos de 24 horas após o registro da reclamação) e com 11 dias. Essas diferenças de proporções também ocorrem nas avaliações Não Resolvidas.

Figura 13 – Gráfico em barras da distribuição da avaliação da reclamação por Tempo de Resposta

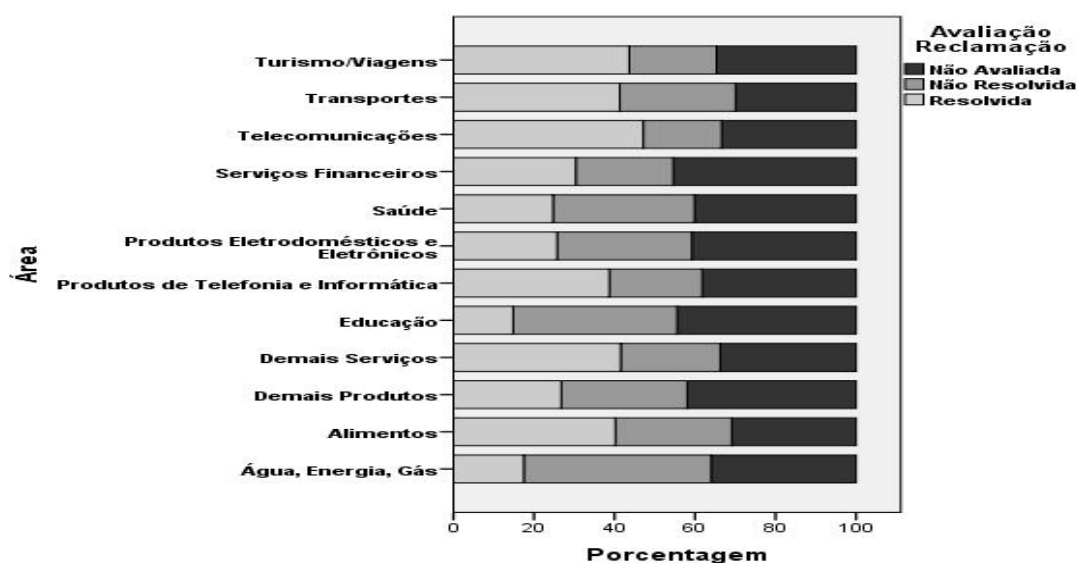


Fonte: Elaborado pelo Autor, 2016.

4.3.3 Variável Área

Na Figura 14 é representado um gráfico com as proporções dos tipos de avaliação final da reclamação em relação ao total (100%) de cada uma das áreas comerciais, representadas na variável Área. Nota-se nesse gráfico, há uma certa heterogeneidade das proporções dos tipos de avaliações finais Resolvidas; e Não Resolvidas nas diversas áreas comerciais, isso ratifica a suspeita provável de um determinado grau de associação entre essas duas variáveis.

Figura 14 – Gráfico em barras da distribuição da avaliação da reclamação por Área

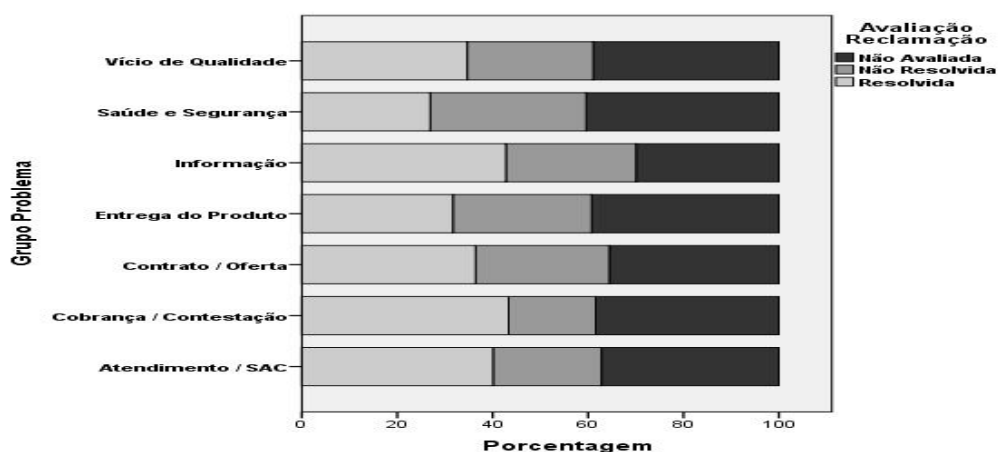


Fonte: Elaborado pelo Autor, 2016.

4.3.4 Variável Grupo Problema

Fazendo a análise bivariada da variável grupo problema, apresenta-se na Figura 15 um gráfico em barras com as proporções dos tipos de avaliação final da reclamação em relação ao total (100%) de cada um dos agrupamentos de problemas dos consumidores que reclamaram pelo site, representadas na variável grupo problema. Nesse gráfico, há uma diferença das proporções dos tipos de avaliações Resolvidas; e Não Resolvidas nos diversos agrupamentos de problemas dos consumidores, isso novamente, ratifica a suspeita provável de um determinado grau de associação entre essas duas variáveis que foram citadas.

Figura 15 – Gráfico em barras da distribuição da avaliação da reclamação por Grupo Problema

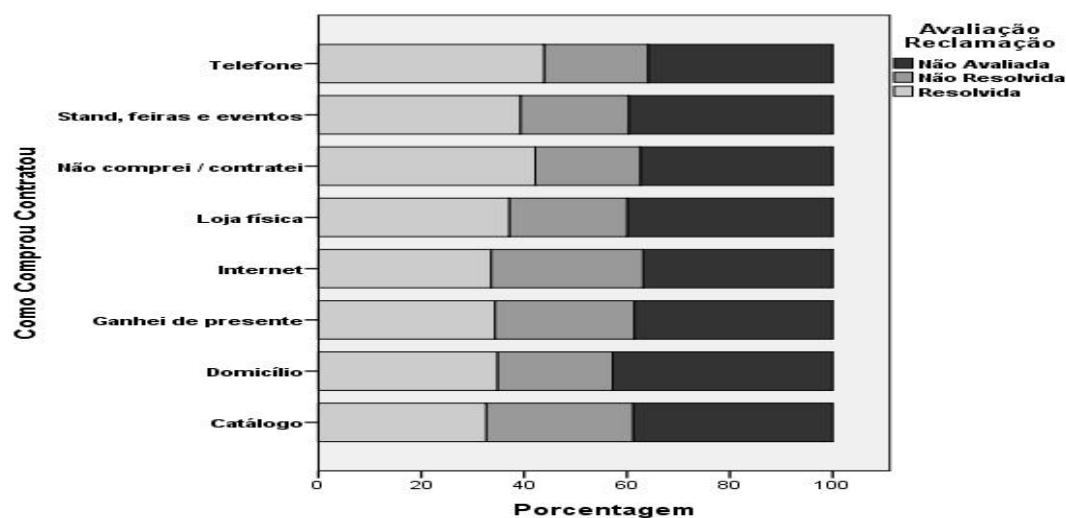


Fonte: Elaborado pelo Autor, 2016.

4.3.5 Variável Como Comprou Contratou

Por último na variável Como Comprou Contratou, apresenta-se na Figura 16 um gráfico em barras com as proporções dos tipos de avaliação final da reclamação em relação ao total (100%) de cada um dos meios de contratação ou aquisição dos consumidores que reclamaram pelo site, representadas na variável Como Comprou Contratou. Nesse gráfico, há uma diferença das proporções dos tipos de avaliações finais Resolvidas; e Não Resolvidas nos diversos meios de contratação ou aquisição dos consumidores, assim, ratifica a suspeita de um determinado grau de associação entre essas duas variáveis mencionadas.

Figura 16 – Gráfico em barras da distribuição da avaliação da reclamação por Como Comprou Contratou



Fonte: Elaborado pelo Autor, 2016.

5 MODELOS PARA DESFECHOS CATEGÓRICOS

Segue-se nas seções seguintes desse capítulo a metodologia utilizada para obtenção dos resultados obtidos pelo autor.

5.1 Síntese Sobre os Modelos Lineares Generalizados (MLG's)

Conforme AGRESTI (1996) os modelos lineares generalizados estendem o modelo de regressão simples abrangendo para distribuições não-normais de resposta e funções de modelagem da média. Três componentes especificam um modelo linear generalizado: Um componente aleatório composto de uma variável Y e sua distribuição de probabilidade; um componente sistemático composto pelas variáveis explicativas tais que produzem uma função linear preditora; e uma função de ligação que especifica uma função de $E(Y)$ igualando ao componente sistemático. Os MLG's foram introduzidos por Nelder e Wedderburn (1972), embora muitos modelos dessa classe só serem bem estabelecidos alguns anos depois.

5.1.1 Componentes dos Modelos Lineares Generalizados

Ainda, conforme preceitos de AGRESTI (1996) o componente aleatório de um MLG consiste de uma variável resposta Y com observações independentes (Y_1, \dots, Y_N) que segue uma distribuição da família exponencial dada pela forma abaixo:

$$f(y_i; \theta_i) = a(\theta_i)b(y_i) \exp[y_i Q(\theta_i)] \quad (5.1)$$

O valor do parametro θ_i pode variar para $i = 1, \dots, N$ dependendo dos valores da variável explicativa. O termo $Q(\theta_i)$ é chamado de parâmetro natural.

Já o componente sistemático relaciona um vetor (η_1, \dots, η_N) de uma variável explicativa através de um modelo linear. Com x_{ij} o valor do preditor j ($j = 1, 2, \dots, p$) para a observação i . Assim, η_i é definido por:

$$\eta_i = \sum_j \beta_j x_{ij}$$

O terceiro componente de um MLG é a função de ligação que conecta os componentes aleatório e sistemático. dessa forma, tem-se $\mu_i = E(Y_i)$. No modelo essa ligação de μ_i com η_i ocorre através de $\eta_i = g(\mu_i)$, onde g é a função de ligação, que é uma função monotônica e diferenciável. Portanto, g está ligando $E(Y_i)$ com o componente sistemático através da

fórmula abaixo:

$$g(\mu_i) = \sum_j \beta_j x_{ij} \quad (5.2)$$

A função de ligação é chamada de identidade quando $\eta_i = \mu_i$. Assim, ela especifica um modelo linear para a própria média, sendo a função de ligação para o modelo de regressão com distribuição normal de Y . A função de ligação que transforma a média no parâmetro natural é chamada de ligação canônica.

Denotando $L(\beta)$ como o logaritmo de um MLG com respostas independentes PAULA(2004, p.8) compreende que:

Uma das vantagens de usarmos ligações canônicas é que as mesmas garantem a concavidade de $L(\beta)$ e conseqüentemente muitos resultados assintóticos são obtidos mais facilmente. Por exemplo, a concavidade de $L(\beta)$ garante a unicidade da estimativa de máxima verossimilhança de β , quando essa existe. Para ligações não canônicas Wedderburn (1976) discute condições para a existência da concavidade de $L(\beta)$.

De acordo com AGRESTI (1996) em síntese, um MLG é um modelo linear que transforma a média de uma variável resposta que tem distribuição natural de uma família exponencial.

Para mais detalhes sobre os modelos lineares generalizados veja seções 1.2 e 1.3 de PAULA (2004).

5.2 Modelo Binomial

Segundo CORDEIRO e NETO (2004) o modelo binomial é uma das mais antigas distriuições de probabilidade e foi desenvolvida por James Bernoulli em seu tratado *Ars Conjectand*, publicado em 1713. A distribuição binomial surge naturalmente em um grande número de situações, a qual as observações Y são contagens não negativas limitadas por um valor fixo.

Uma forma de deduzir essa distribuição é através da distriuição de Bernoulli, expressa na função de probabilidade 5.3. Na distribuição Bernoulli Y_i assume dois valores:

$$Y_i = \begin{cases} 1 & \text{se o evento de interesse ocorre na repetição } i \\ 0 & \text{caso contrário,} \end{cases}$$

tal que

$$P(Y_i = k) = \pi^k (1 - \pi)^{1-k}, \quad k = 0, 1, \quad (5.3)$$

onde π representa a probabilidade de um evento de interesse ocorrer.

Assim, fazendo $S_n = \sum_{i=1}^m Y_i$ de m variáveis aleatórias Y_1, \dots, Y_m de Bernoulli independentes e identicamente distribuídas conforme 5.3, tem-se que S_n segue uma distribuição binomial de função probabilidade dada por 5.4.

$$P(S_n = y) = \binom{m}{y} \pi^y (1 - \pi)^{m-y}, \quad y = 0, 1, 2, \dots, m. \quad (5.4)$$

É importante salientar que a função de probabilidade S_n/m (proporção de sucessos) pertence à família exponencial 5.1 com parâmetro natural $Q(\theta_i) = \log \frac{\mu}{1-\mu}$, a qual $\mu = E(S_n/m)$. Assim, a ligação canônica, representa o logaritmo da razão de chances.

5.2.1 Funções de Ligações Apropriadas para o Modelo Binomial

Ainda, conforme CORDEIRO e NETO (2004) no intuito de investigar a relação entre a probabilidade de sucesso π da variável resposta Y e o vetor de covariáveis (x_1, \dots, x_p) assume-se que a dependência entre π e (x_1, \dots, x_p) ocorre através da combinação linear $\eta_i = \sum_j \beta_j x_{ij}$. Entretanto, como $-\infty < \eta_i < \infty$, expressar π através de uma função linear de η_i seria equivocado do ponto de vista probabilístico, pois sabe-se que π não se encontraria restrito ao intervalo $(0,1)$. Dessa forma, uma maneira mais simples e eficaz para solucionar este problema é a utilização de uma transformação $g(\pi)$ que relacione o intervalo unitário a reta real, de tal maneira que:

$$g(\pi_i) = \eta_i = \sum_j \beta_j x_{ij} \quad i = 1, 2, \dots, n. \quad (5.5)$$

No Quadro 2 abaixo apresenta-se algumas funções de ligação que são adequadas para dados binários, pois elas preservam as restrições sobre a probabilidade π_i .

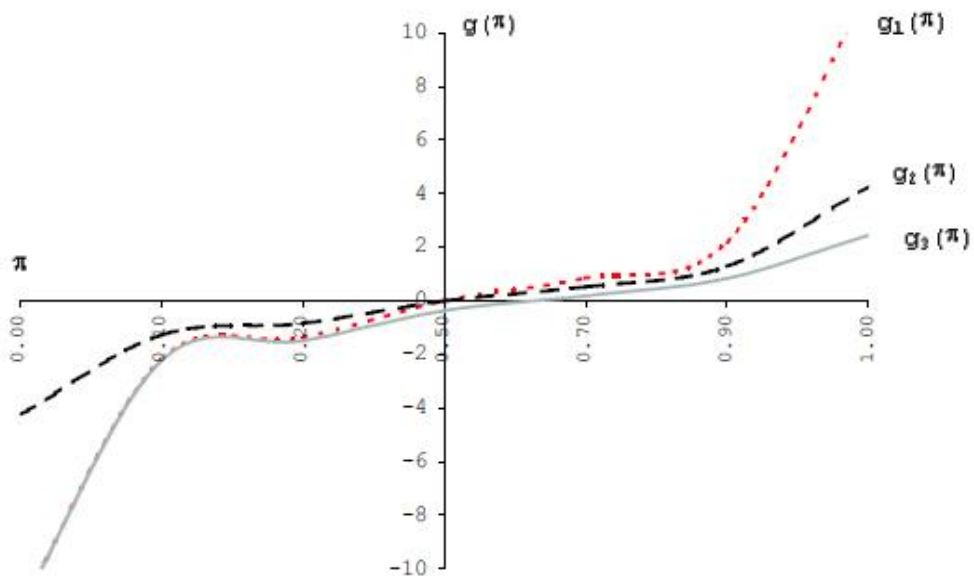
Quadro 2 – Quadro da Algumas Funções de Ligações Adequadas para Dados Binários

Nome atribuído à função de ligação	Função de ligação
Logit ou função logística	$g_1(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$
Função probit ou inversa da distribuição acumulada da normal reduzida	$g_2(\pi) = \Phi(\pi)^{-1}$
Complemento log-log	$g_3(\pi) = \log(-\log(1-\pi))$

Todas as três funções de ligação possuem inversa, são contínuas e crescentes no intervalo (0,1).

Na Figura 17, pode-se examinar o comportamento das três ligações usualmente empregadas no modelo binomial. Essas, apresentam um comportamento praticamente linear no intervalo $0,1 \leq \pi \leq 0,9$. Para baixos valores de π , as ligações logística e complemento log-log encontram-se muito próximas, decaído mais rapidamente que a probit. Mas quando π se aproxima de 1, a ligação complemento log-log cresce mais lentamente do que a probit e a logit. Uma característica da ligação logística é que ela decresce quando π vai para 0 e cresce quando π vai para 1 de forma bastante acelerada.

Figura 17 – Funções de ligação do Quadro 2



Fonte: Cordeiro e Neto (2004).

A função logística apresenta algumas vantagens em relação às outras ligações para análise de dados binários: A primeira é que ela pode ser interpretada como o logaritmo da razão de chances; a segunda é que a mesma apresenta propriedades teóricas mais simples; e por fim ela é mais conveniente para análise de dados coletados de forma retrospectiva. É importante salientar, que isto não significa que as outras transformações não sejam utilizadas na prática.

A fim de compreender melhor o ajuste obtido é necessário a utilização da relação entre π e o preditor linear $\eta = X\beta$. O Quadro 3 estabelece a relação 5.5 das três funções de ligação para dados binários já mencionadas, além da relação de π_i em função do preditor linear η_i .

Quadro 3 – Quadro da Algumas Funções de Ligações Adequadas para Dados Binários

Nome atribuído à função de ligação	Ligação dos componentes do MLG	π_i em função de η_i
Logit ou função logística	$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \eta_i$	$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$
Função probit ou inversa da distribuição acumulada da normal reduzida	$\Phi(\pi_i)^{-1} = \eta_i$	$\pi_i = \Phi(\eta_i)$
Complemento log-log	$\log(-\log(1 - \pi_i)) = \eta_i$	$\pi_i = 1 - \exp(-\exp(\eta_i))$

Fonte: Cordeiro e Neto (2004).

Examinando a relação de π_i em função do preditor linear η_i das três funções de ligação, percebe-se que se a parte sistemática do modelo para uma determinada observação tende para um valor muito negativo, sua probabilidade de sucesso tende para zero. Por outro lado, se a mesma tende para um valor muito grande, esta probabilidade tende para um.

Para obtenção de mais detalhes de funções de ligações utilizadas para modelos binários consultar seção 2.8 de CORDEIRO e NETO (2004).

5.2.2 Estimação de β para o Modelo Binomial

Segundo CORDEIRO e NETO (2004) há diversos métodos de estimar os parâmetros β , dos quais podemos citar: a estimação-M, Bayesiano, qui-quadrado mínimo; e o método da máxima verossimilhança, que frequentemente é utilizados nos programas computacionais.

Um algoritmo de máxima verossimilhança de estimação dos parâmetros β 's, foi desenvolvido por Nelder e Wedderburn (1972) e baseia-se em um método semelhante ao de Newton-Raphson, conhecido como método Escore de Fisher. A principal diferença em relação ao modelo clássico de regressão é que as equações de máxima verossimilhança são não-lineares.

É importante destacar que os *softwares* computacionais de ajuste de MLG sempre utilizam o método escore de Fisher para calcular as estimativas dos β 's. Isso acontece

porque no método de Newton-Raphson há uma maior probabilidade do algoritmo não convergir.

Para entender detalhes sobre o método e o algoritmo de estimação veja seções 2.4 e 2.8.5 de CORDEIRO e NETO (2004).

5.2.3 Alguns Procedimentos para Seleção de Modelos

De acordo com PAULA (2004) existe uma diversidade de procedimentos para a seleção de modelos de regressão, embora nenhum deles sejam consistente, ou seja, que mesmo com amostras grandes selecione com probabilidade de 100% as variáveis explicativas com coeficiente de regressão não nulo. Assim, alguns desses procedimentos serão descritos brevemente a seguir.

5.2.3.1 Método Forward

Iniciamos o método pelo modelo $g(\mu) = \beta_0$. Ajustamos então para cada variável explicativa o modelo:

$$g(\mu) = \beta_0 + \sum_j \beta_j x_j \quad (j = 1, \dots, q)$$

Assim, testa-se $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$. Seja P o menor nível descritivo dentre os q testes. Se $P \leq P_E$, a variável correspondente entra no modelo.

Por exemplo, supondo que X_1 tenha sido escolhida. Então, no passo seguinte ajusta-se os modelos

$$g(\mu) = \beta_0 + \beta_1 x_1 + \beta_j x_j \quad (j = 2, \dots, q)$$

Então, testa-se $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$. Seja P o menor nível descritivo dentre os $(q-1)$ testes. Se $P \leq P_E$, a variável correspondente entra no modelo. Repete-se procedimento até que ocorra $P > P_E$.

5.2.3.2 Método backward

Inicia-se o procedimento pelo modelo:

$$g(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q$$

Testa-se $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$ para $j=1, \dots, q$. Seja P o maior nível descritivo dentre os q testes. Se $P > P_S$, a variável correspondente sai do modelo. Supondo que X_1 tenha saído do modelo. Então, ajusta-se o modelo

$$g(\mu) = \beta_0 + \beta_2x_2 + \dots + \beta_qx_q$$

Assim, testa-se $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$ para $j=1,\dots,q$. Seja P o maior nível descritivo dentre os $(q-1)$ testes. Se $P > P_S$, a variável correspondente sai do modelo. Repete-se o procedimento até que ocorra $P \leq P_E$.

5.2.3.3 Método Stepwise

É uma mescla dos dois procedimentos anteriores. Inicia-se o processo com o modelo $g(\mu) = \beta_0$. Após inclusão de duas variáveis, é verificado se a primeira variável não sai do modelo. O processo continua até que nenhuma variável seja incluída ou seja retirada no modelo. Geralmente $0,15 \leq P_E, P_S \leq 0,25$. Uma sugestão seria utilizar $P_E = P_S = 0,20$.

Ainda segundo PAULA (2004) o método de seleção stepwise é um dos mais aplicados em regressão logística.

5.2.3.4 Método Akaike

O método proposto por Akaike (1974) basicamente se diferencia dos procedimentos anteriores por um processo de minimização que não envolve testes estatísticos. A idéia basilar é selecionar um modelo que seja parcimonioso, ou em outras palavras, que seja bem ajustado e tenha um número reduzido de parâmetros. Como o logaritmo da função de verossimilhança $L(\beta)$ cresce com o aumento do número de parâmetros do modelo, uma proposta razoável seria encontrar o modelo com o menor valor para a função abaixo:

$$AIC = -L(\hat{\beta}) + p$$

Onde, p denota o número de parâmetros do modelo. Para utilização do critério de Akaike no programa estatístico R (2015) é necessário usar antes o comando `require(MASS)` de VENABLES e RIPLEY (2002). Uma maneira de aplicarmos o critério de Akaike é partindo do maior modelo cujos resultados são guardados no objeto `fit.model`. Daí, então, devemos usar o comando `stepAIC(fit.model)`.

Segundo VENABLES e RIPLEY (2002) o comando `stepAIC` realiza seleção de modelos passo a passo pelo método Akaike (AIC), sendo o padrão da função ambos, ou seja, `forward` e `backward`, sendo assim o método stepwise.

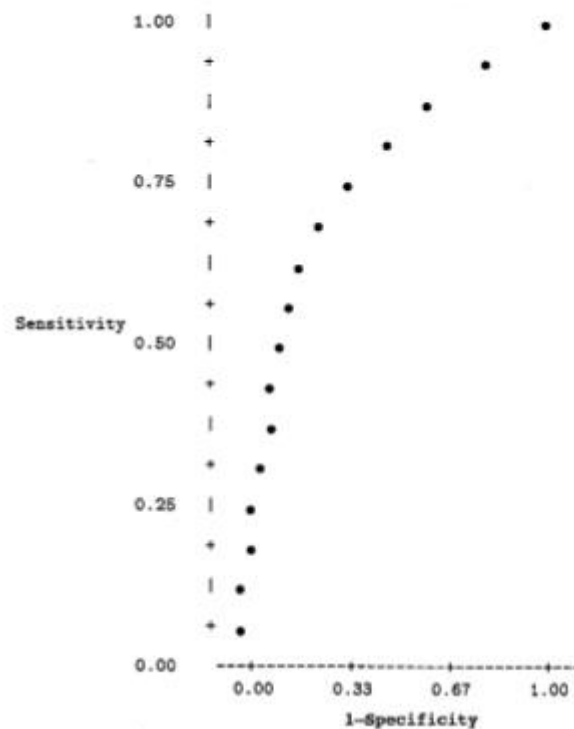
5.3 Curva ROC (*Receiver Operating Characteristic*)

Segundo AGRETI (1996) uma tabela de classificação cruzada de resposta binária tem uma predição de $y = 0$ ou $y = 1$. Assim, o preditor é $\hat{y} = 1$ quando $\hat{\pi}_i > \pi_0$ e $\hat{y} = 0$ quando $\hat{\pi}_i \leq \pi_0$, para algum ponto de corte π_0 . Muitas classificações de tabelas utilizam

$\pi_0 = 0,5$ e sintetizam o poder preditivo pela sensibilidade = $P(\hat{y} = 1|y = 1)$ e pela especificidade = $P(\hat{y} = 0|y = 0)$. É importante ressaltar que a escolha de π_0 é arbitrária.

A curva de características de operação do receptor (curva ROC) é elaborada através do gráfico da sensibilidade como função de (1-especificidade) para as possibilidades de ponto de corte π_0 . Essa curva geralmente possui formato côncavo que liga os pontos (0,0) e (1,1). Na Figura 18 abaixo, segue-se um exemplo de uma curva ROC para um determinado modelo de regressão logística.

Figura 18 – Exemplo de Curva ROC para um Determinado Modelo de Regressão Logística



Fonte: Agresti (1996)

Segundo site do portal action¹ (elaborado em conjunto pelas empresas Estatcamp e DigUp, compostas por profissionais com mestrado e doutorado em estatística e computação), a escolha do ponto de corte deve ser baseada em uma combinação ótima tanto da sensibilidade quanto da especificidade, pois partimos do suposto que classificar o indivíduo como evento dado que ele é não evento (falso positivo) e classificar o indivíduo como não evento dado que ele é evento (falso negativo) traz prejuízos equivalentes para o pesquisador. Pela análise da curva ROC, escolhemos o ponto de corte referente a combinação da sensibilidade e 1-especificidade que mais se aproxima do canto superior esquerdo do gráfico.

¹ <http://www.portaction.com.br/>

Sintetizando o que virá nas seções do próximo capítulo, serão ajustados modelos binomiais para os dados que foram avaliados pelos consumidores, que registraram suas reclamações no site do governo; tudo isso conforme seção 5.2 deste trabalho. Assim, a variável resposta Y será a variável Avaliação da Reclamação; e as variáveis explicativas: Mês Abertura, Grupo Problema, Como Comprou/ Contratou, Área; e Tempo Resposta.

Posterior a esses ajustes, será escolhido o modelo mais parcimonioso, segundo o critério conjugado de Stepwise e AIC, reveja subseção 5.2.3. De posse da Matriz $Beta$ dos estimadores dos parâmetros β_j ; e de posse da Matriz X correspondente a matriz das variáveis explicativas; serão estimados os $\hat{\pi}_i$'s, onde $\hat{\pi}_i$ corresponde a probabilidade de sucesso (probabilidade de a reclamação ser resolvida) das respectivas avaliações das reclamações classificadas como não avaliadas. Em seguida, através de quatro metodologias diferentes; as reclamações não avaliadas por parte do consumidor, serão classificadas como Resolvida ou Não Resolvidas. Uma dessas metodologias é do governo que segundo site do Consumidor do governo federal² caso o consumidor não avalie, sua reclamação é contabilizada como “resolvida”.

De posse dessas classificações, será elaborado o índice de solução que segundo definição do site do consumidor do governo apresenta índices de solução das reclamações por empresa (variável nome fantasia); ou seja, a proporção de avaliações de reclamações resolvidas em relação ao total de reclamações registradas. Assim, será calculado esse índice conforme as quatro metodologias de classificações diferentes. Por fim, serão confrontados os valores dos respectivos índices de solução das empresas, segundo esses quatro métodos.

² <http://www.consumidor.gov.br/>

6 RESULTADOS

Nas seções seguintes desse capítulo, serão expostos os resultados obtidos do trabalho, baseando-se nos conceitos do capítulo anterior.

6.1 Ajuste do Modelo Binomial

Foram realizados ajustes de modelos binomiais no programa computacional estatístico R (2015) para os dados avaliados pelos consumidores, que registraram suas reclamações no site do governo; reveja seção 5.2 deste trabalho. Assim, a variável resposta Y será a variável Avaliação da Reclamação; e as variáveis explicativas: Mês Abertura, Grupo Problema, Como Comprou/ Contratou, Área; e Tempo Resposta. A escolha inicial dessas variáveis explicativas foram baseados na análise exploratória de dados realizada no capítulo 4 desse trabalho, a qual ao final indicou que essas variáveis podem ter associação com a variável Avaliação da Reclamação.

Dessa forma, ajustou-se inicialmente três modelos cada um com uma função de ligação diferente. O primeiro modelo utilizou a função de ligação logit, já o segundo modelo a função de ligação probit; e o terceiro a função de ligação complemento log-log. Após esse ajuste inicial, procedeu-se a seleção de variáveis para cada um dos três ajustes, utilizado o comando `stepAIC` do pacote MASS, que segundo VENABLES e RIPLEY (2002) realiza seleção de modelos por *stepwise* utilizando metodologia Akaike (AIC). Procedendo dessa maneira, obteve-se um modelo final para cada um dos modelos iniciais.

Curiosamente, os respectivos modelos binomiais finais de cada um dos três modelos foram exatamente os modelos iniciais com as cinco variáveis explicativas.

6.1.1 Escolha do Modelo Binomial através do confronto dos AIC's

obteve-se na Tabela 4 os respectivos valores de AIC's para cada um desses três modelos finais.

Tabela 4 – Valores dos AIC's de cada um dos Modelos Finais de Diferentes Funções de Ligação

Função de Ligação	AIC
Logit	64059,31
Probit	64041,30
Complemento Log-log	64052,41

Fonte: Elaborada pelo Autor, 2016.

Percebe-se fazendo um comparativo do AIC dos três modelos que a diferença é irrisória entre os mesmos, logo utilizando-se esse critério é preferível adotar como modelo binomial para se estimar os π_i 's das reclamações não avaliadas, o modelo de função logística, pois como dito na subseção 5.2.1, tal modelo tem três vantagens em relação aos demais: primeiro a função logística pode ser interpretada como o logaritmo da razão de chances; a segunda é que a mesma apresenta propriedades teóricas mais simples; e terceiro ela é mais conveniente para análise de dados coletados de forma retrospectiva.

6.1.2 Escolha do Modelo Binomial através do Confronto da Máxima Proporção de Predição Correta

Quando obtém-se os valores estimados π_i 's para a predição de um modelo binário é necessário escolher um ponto de corte π_0 ; que conforme seção 5.3: o preditor é $\hat{y} = 1$ quando $\hat{\pi}_i > \pi_0$ e $\hat{y} = 0$ quando $\hat{\pi}_i \leq \pi_0$. Assim, definiu-se o ponto de corte π_0 que maximize a proporção de predições corretas, ou seja, o ponto de corte que maximize a expressão 6.1. Esse procedimento foi realizado para os três modelos finais obtidos na seção 6.1, ou seja, os três modelos finais com as diferentes funções de ligação, obtidos através do comando stepAIC. Dessa forma, a Tabela 5 contém os valores máximos obtidos da proporção de predições corretas e os seus respectivos pontos de cortes para cada um dos três modelos. Para mais detalhes do procedimento usado no R (2015), veja Apêndice B.

$$\frac{(N^\circ \text{ de Acertos de "Resolvida"} + N^\circ \text{ de Acertos de "Não Resolvida"})}{(\text{Total de Predições realizadas})} \quad (6.1)$$

Tabela 5 – Valores Máximos Obtidos para a Proporção de Predições Corretas e os Seus Respetivos Pontos de Corte para cada um dos Três Modelos

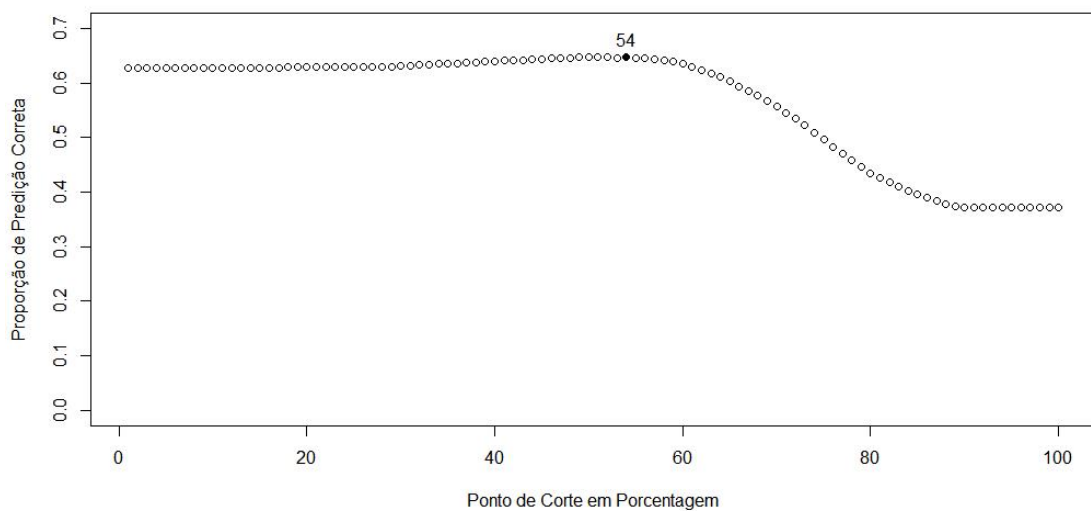
Função de Ligação	Máxima Proporção de Predição Correta	π_0
Logit	64, 77%	0,54
Probit	64, 77%	0,51
Complemento Log-log	64, 70%	0,50

Fonte: Elaborada pelo Autor, 2016.

Percebe-se que o modelo binomial logit e o probit obtiveram uma proporção maior da máxima predição correta, ambas com 64, 77%; assim, pela igualdade mais um vez, opta-se pelo ajuste com o modelo de regressão logística pelas vantagens já relatadas.

Na Figura 19 apresenta-se um gráfico da proporção de predição correta por ponto de corte do modelo com função de ligação logit, elaborado pelo autor desse trabalho através do R (2015). Nesse gráfico foi destacado o ponto de corte (54%) que maximizou a predição correta (64,77%). Além disso, percebe-se pelo formato obtido pelos pontos do gráfico que as melhores predições corretas estão em torno de $\pi_0 = 50\%$; o que faz sentido se for considerado que um modelo de previsão interessante tende a considerar $\hat{\pi}_i > 50\%$ como Resolvida e $\hat{\pi}_i \leq 50\%$ como Não Resolvida.

Figura 19 – Gráfico da Proporção de Predição Correta por Ponto de Corte do Modelo de Regressão Logística



Fonte: Elaborado pelo Autor, 2016.

Ao final, nota-se que a escolha do modelo binomial tanto pelo confronto dos AIC's quanto pelo confronto da máxima predição correta; indicaram o modelo de regressão logística como o mais adequado para essa situação. Portanto, esse será o modelo a ser utilizado para a classificação das reclamações não avaliadas, dando-lhes através do mesmo avaliações como Resolvida; e Não Resolvida.

6.1.3 Alguns Detalhes do Modelo de Regressão Logística Escolhido

Através das análises anteriores, chegou-se em um modelo de regressão logística final com a variável resposta Avaliação da reclamação ("Resolvida", "Não Resolvida") e com variáveis explicativas: Mês Abertura, Grupo Problema, Como Comprou/Contratou, Área; e Tempo Resposta.

Assim, na Figura 20 encontra-se uma tabela com os valores estimados $\hat{\beta}_i$'s e os valores $\exp \hat{\beta}_j$. Categorias com o parâmetro $\hat{\beta}_i = 0$, são categorias consideradas como de referência; dessa maneira, $\hat{\beta}_i$'s positivos aumentam a probabilidade de a reclamação do consumidor ser Resolvida, ao passo que de $\hat{\beta}_i$'s negativos diminuem essa probabilidade. Como já mencionado na subseção 5.2.1 a função de ligação logit pode ser interpretada como o logaritmo da razão de chances, então, aplicando-se a função exponencial natural no modelo de regressão logística, obtém-se a razão de chances. Dessa forma, examinado por exemplo os coeficientes da variável explicativa Grupo Problema, tem-se que $\exp \hat{\beta}$ do coeficiente da cobrança e contestação é $\exp 0,30862 = 1,36155$; onde mostra que se determinada reclamação for de cobrança/ contestação, essa tem aproximadamente 36% a mais de chances de ser resolvida do que se a reclamação for de atendimento/SAC, que é a categoria de referência.

Cumprе salientar também que foram necessárias 8 interações escore de Fisher para a estimação dos parâmetros $\hat{\beta}_i$'s.

Figura 20 – Valores estimados dos β_j 's e $\exp \hat{\beta}_j$

		$\hat{\beta}_j$	$\exp(\hat{\beta}_j)$
Mês Abertura	1	0,00000	
	2	-0,30928	0,73397
	3	-0,16104	0,85126
	4	-0,10625	0,89920
	5	-0,09261	0,91155
	6	0,06895	1,07138
	11	-9,61071	0,00000
	12	-0,26924	0,76396
Grupo Problema	Atendimento/ Sac	0,00000	
	Cobrança/ Contestação	0,30862	1,36155
	Contrato/ Oferta	-0,28704	0,75048
	Entrega do Produto	0,12634	1,13466
	Informação	-0,06813	0,93413
	Saúde e Segurança	-0,36108	0,69692
Como Comprou/ Contratou	Vício de Qualidade	-0,23710	0,78891
	Catálogo	0,00000	
	Domicílio	0,04027	1,04110
	Ganhei de Presente	-0,13028	0,87785
	Internet	-0,02315	0,97712
	Loja física	0,15894	1,17227
	Não comprei/ Contratei	0,28019	1,32338
	Stand, feiras e eventos	0,10182	1,10718
	Telefone	0,22136	1,24760
	Área	Água, Energia, Gás	0,00000
Alimento		1,77195	5,88232
Demais Produtos		1,13609	3,11456
Demais Serviços		1,63821	5,14594
Educação		-0,13117	0,87707
Produto Eléctrodomésticos e		1,04566	2,84528
Produto de Telefonia e Informática		1,71902	5,57905
Saúde		0,80934	2,24644
Serviços Financeiros		1,20533	3,33786
Telecomunicações		2,04706	7,74509
Transportes		1,35053	3,85949
Turismo/ Viagens		1,82546	6,20566
Tempo Resposta		-0,09371	0,90510
Constante		-0,44895	0,63830

Fonte: Elaborada pelo Autor, 2016.

6.2 Classificação das Avaliações das Reclamações Não Avaliadas pelo Consumidor

Como a partir de agora, tem-se a matriz β dos coeficientes β_j 's; e também matriz X correspondente a matriz das variáveis explicativas das reclamações não avaliadas, serão então estimados, os $\hat{\pi}_i$'s, onde $\hat{\pi}_i$ corresponde a probabilidade estimada de sucesso (probabilidade estimada de a reclamação ser resolvida) das respectivas avaliações das reclamações classificadas como não avaliadas. Dessa forma, obteve-se os respectivos η_i 's através da multiplicação dessas duas matrizes mencionadas; e posteriormente os $\hat{\pi}_i$'s das reclamações não avaliadas, através da fórmula do cálculo de π_i para ligação logit do Quadro 3 da subseção 5.2.1. Esses cálculos foram realizados por intermédio de códigos no programa computacional estatístico R (2015), elaborados pelo autor deste trabalho (veja Apêndice B).

Esses $\hat{\pi}_i$'s foram calculados com sucesso, assim, restará apenas um critério para a classificação das reclamações não avaliadas. Nas próximas subseções serão utilizados três critérios para a obtenção do ponto de corte π_0 : Critério empregado no *default* dos *softwares*, o critério da máxima proporção de predição correta; e o critério da curva ROC.

Também será apresentado nessa seção, a forma a qual o site do governo classifica as reclamações não avaliadas.

6.2.1 Utilizando o Critério Empregado no *Default* dos *Softwares*

O Critério aplicado pelo *Default* dos *Softwares* é muito simples: ele considera o ponto de corte $\pi_0 = 50\%$; assim, conseqüentemente, se $\hat{\pi}_i > 50\%$ a reclamação é classificada como Resolvida e se $\hat{\pi}_i \leq 50\%$ como Não Resolvida.

Então, foi feita a classificação das reclamações não avaliadas seguindo esse critério; e armazenou-se.

6.2.2 Utilizando o Critério Máxima Proporção de Predição Correta

Esse critério já foi explicitado e realizado na subseção 6.1.2 (veja-a), sendo definido como ponto de corte $\pi_0 = 54\%$, que resultou na melhor predição correta.

Novamente, foi realizada a classificação das reclamações não avaliadas e; assim, classificadas como Resolvidas se $\hat{\pi}_i > 54\%$, ou Não Resolvidas se $\hat{\pi}_i \leq 54\%$.

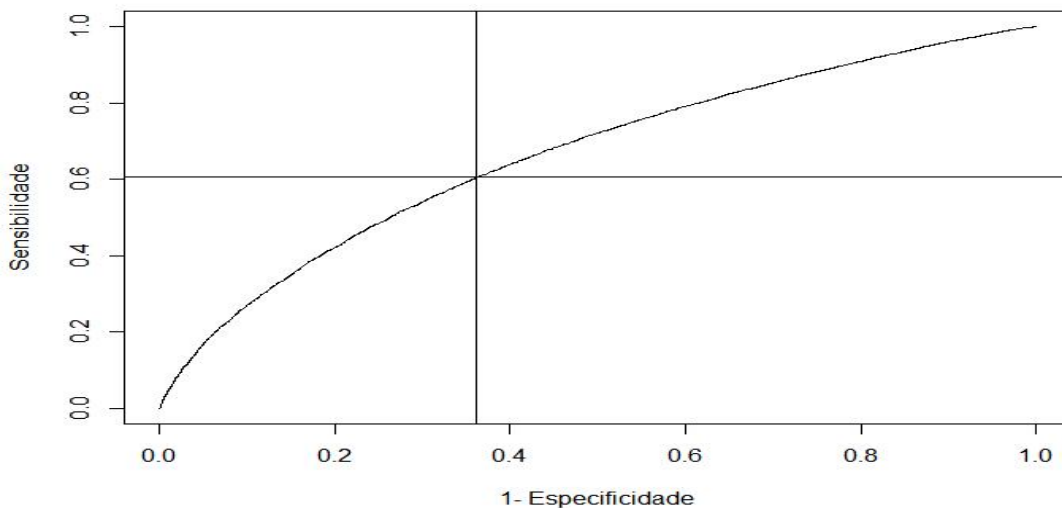
6.2.3 Utilizando o Critério de Classificação da Curva ROC

Esse critério já foi mencionado na seção 5.3 (veja novamente). Como o critério de escolha de π_0 é arbitrário, então neste trabalho optou-se pelo critério de maximização simultânea da sensibilidade e da especificidade para obtenção do ponto de corte. Assim,

obteve-se $\pi_0 = 63,1351\%$; portanto, se $\hat{\pi}_i > 63,1351\%$ a reclamação é classificada como Resolvida e se $\hat{\pi}_i \leq 63,1351\%$ como Não Resolvida. Os valores obtidos da sensibilidade e especificidade na predição do modelo são respectivamente: 0,605351 e 0,638439. Cumpre salientar que a sensibilidade representa a proporção de predição correta de reclamações Resolvidas; e especificidade representa a proporção de predição correta das avaliações Não Resolvidas.

Na Figura 21 apresenta-se o gráfico da curva ROC para as reclamações avaliadas, a qual percebe-se o formato côncavo mencionado na seção 5.3. Além, disso, o autor fez duas retas perpendiculares entre si, a qual o ponto de interseção entre as mesmas, representa exatamente o ponto de corte π_0 . Na interseção dessas duas retas com os eixos coordenados, tem-se os valores da sensibilidade e o valor de (1-especificidade).

Figura 21 – Gráfico da Curva ROC para As Reclamações Avaliadas



Fonte: Elaborado pelo Autor, 2016.

Após todas essas verificações, procedeu-se a classificação das reclamações não avaliadas, através do ponto de corte $\pi_0 = 63,1351\%$.

6.2.4 Forma de Classificação do Site do Governo

Como já foi dito na introdução do capítulo 6, o índice de solução elaborado pelo site do governo é a proporção de avaliações de reclamações resolvidas, em relação ao total de reclamações registradas por nome fantasia das empresas. Para o cálculo desse índice o site do governo classifica as reclamações não avaliadas pelo consumidores como Resolvidas.

6.3 Construção do Índice de Solução a Partir dos Quatro Critérios de Classificação das Reclamações Não Avaliadas

Há um total de 82.896 dados avaliados e não avaliados, desses há um total de 194 dados não avaliados, onde não consta a variável Tempo Resposta. Portanto para esses 194 dados não foi possível estimar os π_i 's, logo essas observações tiveram que ser excluídas do cálculo do índice de solução. Então o total de observação para construção desse índice é igual a 82.702 observações.

Dessa forma, apresenta-se na Tabela 6 o índice de solução de algumas empresas (nome fantasia das empresas) segundo cada um dos quatro critérios anteriormente mencionados. Nota-se obviamente que o índice de solução segundo o critério do site do governo é sempre maior, seguido em geral por ordem decrescente pelos critérios: *Default* dos *softwares*, máxima proporção de predição correta; e o da curva ROC. Percebe-se também, que a diferença desses índices comparando esses critérios, pode ser bem grande. A tabela completa do índice de solução de todas as empresas constante nos dados está no Apêndice C.

Tabela 6 – Índice de Solução de Algumas Empresas Segundo os Quatro Critérios

Nome Fantasia	C. Governo	C. Máxima Predição	C. <i>Default</i>	C. Curva ROC
Brastemp	61,8%	19,1%	20,4%	17,6%
Carrefour	74,7%	35,8%	46,3%	27,4%
Ricardo Eletro	60,5%	34,2%	43,4%	27,6%
Samsung	66,8%	43,9%	50,9%	33,9%
GVT	86,0%	85,7%	85,9%	80,1%

Fonte: Elaborada pelo Autor, 2016.

No próximo capítulo será apresentada a conclusão desse trabalho confrontando a forma de classificação das reclamações não avaliadas do site do governo; e os demais critérios de classificação demonstrados até aqui.

7 CONCLUSÃO

Após a análise do índice de solução obtido pelos quatro critérios, compreende-se que pode haver grandes diferenças na proporção de reclamações solucionadas por empresas por diferentes critérios. Desta forma, embora seja louvável a intenção do site do governo em elaborar tal índice, seria imprescindível a busca incessante por critérios que classifiquem melhor as reclamações não avaliadas, pois como o próprio inciso 3 do artigo 6º. da Lei nº 8.078/1990 BRASIL (Código de Defesa do Consumidor) diz: é direito básico do consumidor a informação adequada e clara sobre os diferentes produtos e serviços. Assim, seria possível aferir melhor o índice de solução; e conseqüentemente levar uma informação mais clara aos consumidores no que tange a solução das empresas. Uma outra maneira que talvez melhore a aferição de tal índice, poderia ser a elaboração de políticas que elevasse a proporção de reclamações avaliadas.

A forma de classificação das reclamações não avaliadas adotadas pelo site do governo federal para o cálculo do índice de solução contém uma imperfeição. Essa imperfeição é justamente fato de a probabilidade de se classificar corretamente uma reclamação não avaliada que na realidade não foi resolvida ser zero. Essa probabilidade ser zero implica que tal forma de classificação do governo, tende a superestimar o índice de solução.

Considerando a análise dos dados desse trabalho, observou-se por exemplo, que o modelo de regressão logístico aliado ao critério da máxima proporção de predição correta, obteve aproximadamente 64,77% de predição correta. Isso demonstra um bom nível de predição, reconhecendo, claro, que aproximadamente 35,23% da variabilidade da avaliação do consumidor não é explicada pelo modelo em questão. Nesse modelo o ponto de corte é 0,54; assim, tem-se uma probabilidade diferente de zero de classificar uma reclamação não avaliada que na realidade não foi resolvida. Analisando através da curva ROC esse ponto de corte, obtém-se uma especificidade de aproximadamente 34,8% e sensibilidade de 82,5%; demonstrando-se que existem probabilidades razoáveis adotando o critério da máxima proporção da predição de se classificar uma reclamação não avaliada como resolvida ou não resolvida corretamente.

Dessa forma, o site do governo deveria apresentar uma justificativa com base científica para classificar todas as reclamações não avaliadas como resolvidas. Essa justificativa deveria estar bem explícita em seu site, para que haja um melhor esclarecimento para o consumidor. Não havendo tal justificativa seria interessante a busca por métodos calcados na ciência para uma estimação mais acurada do índice de solução.

Cumprе salientar também, que essa tendência de reclamações digitais está em crescente, visto que recentemente o endereço eletrônico da prefeitura de São Paulo¹ noticiou o lançamento no dia 11 de fevereiro de 2016 do decreto que institui o Procon

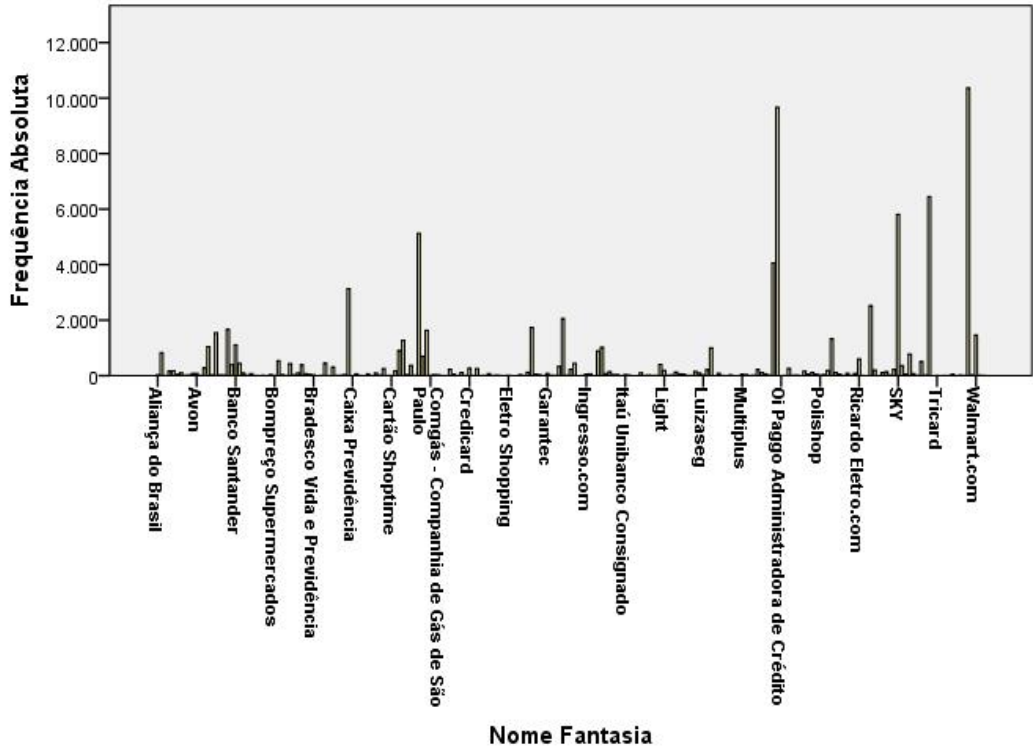
¹ <http://www.capital.sp.gov.br/portal/noticia/9537ad-image-0>

paulistano digital para consulta pública. Ainda, de acordo com esse referido site, haverá acompanhamento e análise de dados na cidade; logo através desse trabalho percebe-se que a utilização de modelos binomiais poderá ser uma ferramenta bastante útil para essas análises de dados.

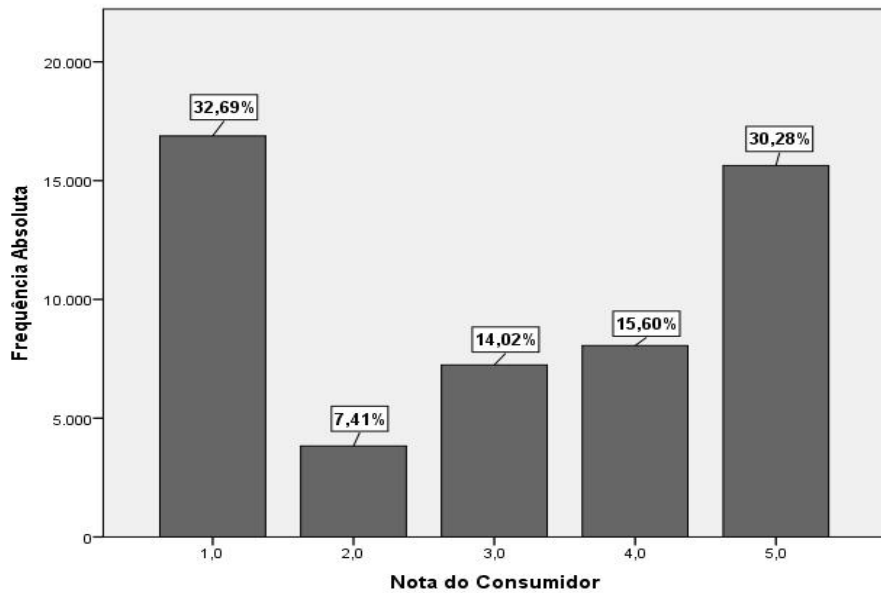
REFERÊNCIAS

- [1] AGRESTI, Alan. **Categorical data analysis**. New York: John Wiley Sons, 1996.
- [2] BRASIL, Constituição Federal de. Lei nº. 8.078, de 11 de setembro de 1990: Dispõe sobre a proteção do consumidor e dá outras providências. **Diário Oficial da União**, 1990.
- [3] BUSSAB, Wilton de O.; MORETTIN, Pedro A. **Estatística básica**. 6. ed. São Paulo: Saraiva, 2010.
- [4] CORDEIRO, Gauss Moutinho; NETO, Eufrásio de AL. Modelos paramétricos. **Associação Brasileira de Estatística**, v. 246, 2004.
- [5] COSTA, Helder Gomes et al. ELECTRE TRI aplicado a avaliação da satisfação de consumidores. **Produção**, v. 17, n. 2, p. 230-245, 2007.
- [6] PAULA, Gilberto Alvarenga. **Modelos de regressão: com apoio computacional**. São Paulo: IME-USP, 2004.
- [7] R Core Team (2015). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [8] ROSSI, Carlos Alberto Vargas; SLONGO, Luiz Antonio. Pesquisa de satisfação de clientes: o estado-da-arte e proposição de um método brasileiro. **Revista de Administração contemporânea**, v. 2, n. 1, p. 101-125, 1998.
- [9] SPSS, IBM. IBM SPSS statistics for Windows, version 20.0. **New York: IBM Corp**, 2011.
- [10] Venables, W. N. Ripley, B. D. (2002) **Modern Applied Statistics with S**. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

APÊNDICE A – Gráfico em barras de algumas das outras variáveis não analisadas



Fonte: Autor



Fonte: Autor

APÊNDICE B – Códigos no script do *software* estatístico R

```
#####
##### Achando o ponto de corte #####

# Leitura dos dados da probabilidade encontrada no modelo e o valor bin da
# observa sendo 1 Resolvida e 0 NResolvida
PI_predicao=fit1$fitted.values
OBSERVACAO=scan("observacao.txt")

# Encontrando o ponto de corte que maximiza a propor de predi

TESTE=Prop.acertos.predicao(fit1$fitted.values)
names(TESTE)

Prop.acertos.predicao=function(PI_predicao)
{
  classificacao=vector()
  RESULTADO=vector()
  ponto_corte= 0
  for(i in 1:100){
    ponto_corte= ponto_corte + 0.01
    for(j in 1:length(PI_predicao)){
      if(PI_predicao[j]>ponto_corte){
        classificacao[j]=1
      }else{
        classificacao[j]=0
      }
    }
  }

  RESULTADO[i]= (1-(sum(abs(classificacao-OBSERVACAO))/51510))

}

max.proporcao=max(RESULTADO)

alpha=list(Prop.pontos.de.corte=RESULTADO,
max.proporcao=max.proporcao)
```

```

}

plot(RERESULTADO,ylim=c(0,0.7))
identify(RERESULTADO,n=1)

## A function to use identify to select points, and overplot the
## points with another symbol as they are selected
identifyPch <- function(x, y = NULL, n = length(x), pch = 19, ...)
{
  xy <- xy.coords(x, y); x <- xy$x; y <- xy$y
  sel <- rep(FALSE, length(x)); res <- integer(0)
  while(sum(sel) < n) {
    ans <- identify(x[!sel], y[!sel], n = 1, plot = FALSE, ...)
    if(!length(ans)) break
    ans <- which(!sel)[ans]
    points(x[ans], y[ans], pch = pch)
    sel[ans] <- TRUE
    res <- c(res, ans)
  }
  res
}

identifyPch(RERESULTADO,n=1)
identify(RERESULTADO,n=1)
abline(h=0.6448651)
abline(v=51)
#####
#####
##### Calculando as Probabilidades pelo modelo encontrado#####

tempo_resposta=scan("tempo_resposta.txt")
mes_abertura= scan("mes_abertura.txt")
mes_abertura= factor(mes_abertura)
mes_abertura= C(mes_abertura,treatment)

area=scan("area_mod.txt")
area= factor(area)
area= C(area,treatment)

```

```
como_comprou_contratou= scan("como_comprou.txt")
como_comprou_contratou= factor(como_comprou_contratou)
como_comprou_contratou= C(como_comprou_contratou,treatment)

grupo_problema =scan("grupo_problema.txt")
grupo_problema= factor(grupo_problema)
grupo_problema= C(grupo_problema,treatment)
avaliacao_reclamacao= scan("avaliacao_reclamacao.txt")
length(avaliacao_reclamacao)
tempo=as.vector(tempo_resposta)

length(tempo_resposta)
length(mes_abertura)
length(area)
length(como_comprou_contratou)
length(grupo_problema)

library(MASS)
# ajustando o modelo com fun de liga logit

fit1 <- glm( avaliacao_reclamacao~ mes_abertura + grupo_problema +
  como_comprou_contratou + area + tempo_resposta ,
  family=binomial(link="logit"))
fit.AIC1=stepAIC(fit1)
summary(fit.AIC1)

# ajustando o modelo com fun de liga probit
### MELHOR MODELO pelo AIC#####

fit2 <- glm( avaliacao_reclamacao~ mes_abertura + grupo_problema +
  como_comprou_contratou + area + tempo_resposta ,
  family=binomial(link="probit"))
fit.AIC2=stepAIC(fit2)
summary(fit.AIC2)

# ajustando o modelo com fun de liga Complementary log-log
# o melhor AIC

fit3 <- glm( avaliacao_reclamacao~ mes_abertura + grupo_problema
```



```

+ area + tempo_resposta ,
family=binomial(link="cloglog"))

fit.AIC3=stepAIC(fit3)
summary(fit.AIC3)

# comparando os AIC's de cada um dos modelos de diferentes funs de liga

fit.AIC1$aic
fit.AIC2$aic # melhor AIC
fit.AIC3$aic
fit.AIC4$aic
(64041/64059)
# comparando os modelos quanto melhores Propor de acertos na Predi
# 0 Melhor pelas vantagens na pequena diferen

TESTE_modelo1=Prop.acertos.predicao(fit1$fitted.values)
TESTE_modelo1$max.proporcao
TESTE_modelo1$Prop.pontos.de.corte
which.max(TESTE_modelo1$Prop.pontos.de.corte)
# ponto de corte 0,54 ( prop = 64,76995 %)

TESTE_modelo2=Prop.acertos.predicao(fit2$fitted.values)
TESTE_modelo2$max.proporcao
which.max(TESTE_modelo2$Prop.pontos.de.corte)
TESTE_modelo2$Prop.pontos.de.corte
# ponto de corte 0,51 ( prop = 64,76995 %)

TESTE_modelo3=Prop.acertos.predicao(fit3$fitted.values)
TESTE_modelo3$max.proporcao
which.max(TESTE_modelo3$Prop.pontos.de.corte)
TESTE_modelo3$Prop.pontos.de.corte
# ponto de corte 0,50 ( prop = 64,698125 %)

# grco dos ponstos de corte do modelo 1 (logit)
plot(TESTE_modelo1$Prop.pontos.de.corte,ylim=c(0,0.7),
ylab="Propor de Predi Correta", xlab= "Ponto de Corte em Porcentagem")

```

```

identify(TESTE_modelo1$Prop.pontos.de.corte,n=1)

## A function to use identify to select points, and overplot the
## points with another symbol as they are selected
identifyPch <- function(x, y = NULL, n = length(x), pch = 19, ...)
{
  xy <- xy.coords(x, y); x <- xy$x; y <- xy$y
  sel <- rep(FALSE, length(x)); res <- integer(0)
  while(sum(sel) < n) {
    ans <- identify(x[!sel], y[!sel], n = 1, plot = FALSE, ...)
    if(!length(ans)) break
    ans <- which(!sel)[ans]
    points(x[ans], y[ans], pch = pch)
    sel[ans] <- TRUE
    res <- c(res, ans)
  }
  res
}
identifyPch(TESTE_modelo1$Prop.pontos.de.corte,n=1)
identify(TESTE_modelo1$Prop.pontos.de.corte,n=1)
abline(h=0.6476995)
abline(v=54)

#####
##### Estabelecendo o ponto de corte pela curva ROC #####

library(ROCR)

pred <- prediction(fit1$fitted.values,
avaliacao_reclamacao)
length(fit1$fitted.values)
length(OBSERVACAO)
perf <- performance(pred,"tpr","fpr")
length(pred)
plot(perf, xlab="1- Especificidade",ylab="Sensibilidade")

```

```

opt.cut = function(perf, pred){
  cut.ind = mapply(FUN=function(x, y, p){
    d = (x - 0)^2 + (y-1)^2
    ind = which(d == min(d))
    c(sensitivity = y[[ind]], specificity = 1-x[[ind]],
      cutoff = p[[ind]])
  }, perf@x.values, perf@y.values, pred@cutoffs)
}

print(opt.cut(perf, pred))

# Ponto de corte pela curva ROC ,631351
#sensitivity 0.605351
#specificity 0.638439

abline(h=0.605351)
abline(v=1-0.638439)

#####
##### Calculando as Probabilidades Estimadas #####

matriz.coeficientes=fit.AIC1$coefficients
matriz.coeficientes=as.matrix(matriz.coeficientes)
dim(matriz.coeficientes)

matriz.x = model.matrix(fit.AIC1)
dim(matriz.x)
matriz.beta.x = matriz.x%%matriz.coeficientes
dim(matriz.beta.x)

# fun para o culo das probabilidades estimadas

calculo.prob=function(mat){
  Prob = vector()
  for(i in 1 : nrow(mat)){

    Prob[i]= exp(mat[i,])/(1+exp(mat[i,]))
  }
}

```

```

    }
    alpha=list(probabilidade=Prob)

}

PI=calculo.prob(matriz.beta.x)
PI$probabilidade[51510]
fit.AIC1$fitted.values[51510]
length(fit.AIC3$fitted.values)

# verificando novamente Propor de Acertos

OBSERVACAO=scan("observacao.txt")

Prop.acertos.predicao=function(PI_predicao)
{
  classificacao=vector()
  RESULTADO=vector()
  ponto_corte= 0
  for(i in 1:100){
    ponto_corte= ponto_corte + 0.01
    for(j in 1:length(PI_predicao)){
      if(PI_predicao[j]>ponto_corte){
        classificacao[j]=1
      }else{
        classificacao[j]=0
      }
    }
  }

  RESULTADO[i]= (1-(sum(abs(classificacao-OBSERVACAO))/51510))

}

max.proporcao=max(RESULTADO)

alpha=list(Prop.pontos.de.corte=RESULTADO,

```

```

max.proporcao=max.proporcao)

}

TESTE=Prop.acertos.predicao(PI$probabilidade)
names(TESTE)
TESTE$max.proporcao
which.max(TESTE$Prop.pontos.de.corte)
TESTE$Prop.pontos.de.corte

# OK ponto de corte 0,54 ( prop = 64,76995 %)
Estorreto
#####XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX#####
#####
##### Ajustando o modelo para os Dados Navaliados #####

tempo_resposta.nao.Av=scan("tempo_resposta.nao.Av.txt")
mes_abertura.nao.Av= scan("mes_abertura.nao.Av.txt")
mes_abertura.nao.Av= factor(mes_abertura.nao.Av)
mes_abertura.nao.Av= C(mes_abertura.nao.Av,treatment)

area.nao.Av=scan("area.nao.Av.txt")
area.nao.Av= factor(area.nao.Av)
area.nao.Av= C(area.nao.Av,treatment)

como_comprou_contratou.nao.Av= scan("como_comprou_contratou.nao.Av.txt")
como_comprou_contratou.nao.Av= factor(como_comprou_contratou.nao.Av)
como_comprou_contratou= C(como_comprou_contratou.nao.Av,treatment)

grupo_problema.nao.Av =scan("grupo_problema.nao.Av.txt")
grupo_problema.nao.Av= factor(grupo_problema.nao.Av)
grupo_problema.nao.Av= C(grupo_problema.nao.Av,treatment)

tempo=as.vector(tempo_resposta.nao.Av)

length(tempo_resposta.nao.Av)
length(mes_abertura.nao.Av)

```

```
length(area.nao.Av)
length(como_comprou_contratou.nao.Av)
length(grupo_problema.nao.Av)

variavel.dummy= c(rep(1,31067))
length(variavel.dummy)

fit.dummy = glm( variavel.dummy ~ mes_abertura.nao.Av + grupo_problema.nao.Av
+ como_comprou_contratou.nao.Av + area.nao.Av + tempo_resposta.nao.Av ,
family=binomial(link="logit"))

matriz.x.nao.Av= model.matrix(fit.dummy)
matriz.x.nao.Av[1,]
dim(matriz.x.nao.Av)

matriz.coeficientes=fit.AIC1$coefficients

matriz.beta.x.nao.Av = matriz.x.nao.Av%*%matriz.coeficientes
dim(matriz.beta.x.nao.Av)

calculo.prob=function(mat){
  Prob = vector()
  for(i in 1 : nrow(mat)){

    Prob[i]= exp(mat[i,])/(1+exp(mat[i,]))

  }
  alpha=list(probabilidade=Prob)
}

PI.nao.Av=calculo.prob(matriz.beta.x.nao.Av)
PI.nao.Av$probabilidade

# classificando as Reclamas NAvaliadas
```

```

classificacao.das.nao.Av=function(PI_previsao,ponto.corte)
{
classificacao=vector()
RESULTADO=vector()
ponto_corte= ponto.corte
  for(j in 1:length(PI_previsao)){
    if(PI_previsao[j]>ponto_corte){
      classificacao[j]=1
    }else{
      classificacao[j]=0
    }
  }
}

alpha=list(classificacao=classificacao)
}

```

```

##### Classificando com o ponto de corte do crito Mma Predi#####
##### MX PREDIO #####
previsao=classificacao.das.nao.Av(PI.nao.Av$probabilidade,0.54)
previsao$classificacao
edit(previsao$classificacao)

```

```
length(previsao$classificacao)
```

```
Total_RESOLVIDAS=sum(previsao$classificacao)
```

```
Total_RESOLVIDAS
```

```
Total.NAO.RESOLVIDA=length(previsao$classificacao)-21706
```

```
Total.NAO.RESOLVIDA
```

```
PROP.RESOLVIDA=Total_RESOLVIDAS/(length(previsao$classificacao))
```

```
PROP.RESOLVIDA
```

```

##### Classificando com o ponto de corte do crito 0,5 Default Software####
##### Default Software #####
previsao.default=classificacao.das.nao.Av(PI.nao.Av$probabilidade,0.5)
previsao.default$classificacao
edit(previsao.default$classificacao)

```

```
length(previsao.default$classificacao)
```

```
Total_RESOLVIDAS.default=sum(previsao.default$classificacao)
```

```
Total_RESOLVIDAS.default
```

```
Total.NAO.RESOLVIDA.default=length(previsao.default$classificacao)-Total_RESOLVIDAS
```

```
Total.NAO.RESOLVIDA.default
```

```
PROP.RESOLVIDA.default=Total_RESOLVIDAS.default/(length(previsao.default$classifica
```

```
PROP.RESOLVIDA.default
```

```
#####
```

```
##### Classificando Conforme curva ROC #####
```

```
# Classificando com o ponto de corte do crito do ponto de corte o
```

```
# estabelecido na curva ROC ponto de corte (0,631351)
```

```
previsao.roc=classificacao.das.nao.Av(PI.nao.Av$probabilidade,0.631351)
```

```
previsao.roc$classificacao
```

```
edit(previsao.roc$classificacao)
```

```
length(previsao.roc$classificacao)
```

```
Total_RESOLVIDAS.roc=sum(previsao.roc$classificacao)
```

```
Total_RESOLVIDAS.roc
```

```
Total.NAO.RESOLVIDA.roc=length(previsao.roc$classificacao)-Total_RESOLVIDAS.roc
```

```
Total.NAO.RESOLVIDA.roc
```

```
PROP.RESOLVIDA.roc=Total_RESOLVIDAS.roc/(length(previsao.roc$classificacao))
```

```
PROP.RESOLVIDA.roc
```


**APÊNDICE C – Índice de Solução de Todas as Empresas Constadas nos
Dados pelos quatro critérios**

Figura 22 – Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continua)

Nome Fantasia das Empresas	GOVERNO	Máx Predição	Default softwares	Curva ROC
	Resolvida	Resolvida	Resolvida	Resolvida
Aliança do Brasil	14	8	10	6
	51,9%	29,6%	37,0%	22,2%
American Express - Amex	139	120	129	88
	84,2%	72,7%	78,2%	53,3%
Americanas Viagens	3	3	3	3
	75,0%	75,0%	75,0%	75,0%
Americanas.com	571	345	391	256
	69,3%	41,9%	47,5%	31,1%
Amil	115	46	51	41
	66,5%	26,6%	29,5%	23,7%
AOC	31	25	25	21
	73,8%	59,5%	59,5%	50,0%
Ativos S.A	79	79	79	77
	77,5%	77,5%	77,5%	75,5%
Atlas Eletrodomésticos	9	5	7	4
	81,8%	45,5%	63,6%	36,4%
Avianca - Oceanair	61	38	47	29
	78,2%	48,7%	60,3%	37,2%
Avianca (Vôos Internacionais)	11	5	6	4
	55,0%	25,0%	30,0%	20,0%
Avon	50	30	32	20
	63,3%	38,0%	40,5%	25,3%
Banco BMG	191	99	121	62
	68,2%	35,4%	43,2%	22,1%
Banco Bradesco	752	560	645	340
	72,2%	53,8%	62,0%	32,7%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Banco Cifra	19	11	14	8
	70,4%	40,7%	51,9%	29,6%
Banco do Brasil	1181	918	1020	658
	76,5%	59,5%	66,1%	42,6%
Banco Fiat	26	15	22	7
	83,9%	48,4%	71,0%	22,6%
Banco Intermedium	9	5	7	1
	42,9%	23,8%	33,3%	4,8%
Banco Itaú Unibanco	1301	1033	1147	753
	78,0%	61,9%	68,7%	45,1%
Banco Pan (Panamericano)	271	179	199	78
	69,7%	46,0%	51,2%	20,1%
Banco Santander	779	545	653	280
	70,6%	49,4%	59,2%	25,4%
Banco Santander Cartões	316	242	274	169
	71,8%	55,0%	62,3%	38,4%
Banco Votorantim	75	54	59	40
	78,1%	56,3%	61,5%	41,7%
Banco24Horas (TecBan)	5	5	5	4
	83,3%	83,3%	83,3%	66,7%
BB Consórcios	10	9	10	7
	76,9%	69,2%	76,9%	53,8%
BB Seguro Auto	40	22	28	22
	58,0%	31,9%	40,6%	31,9%
BCV	5	2	3	1
	100,0%	40,0%	60,0%	20,0%
BeeFitness.com.br	1	1	1	1
	100,0%	100,0%	100,0%	100,0%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

BIG	15	10	10	8
	83,3%	55,6%	55,6%	44,4%
Blockbuster	2	2	2	2
	100,0%	100,0%	100,0%	100,0%
Boa Vista Serviços - SCPC (Serviço Central de Proteção ao Crédito)	26	24	25	23
	68,4%	63,2%	65,8%	60,5%
Bompreço Supermercados	6	2	3	2
	60,0%	20,0%	30,0%	20,0%
Bradescard	483	363	426	260
	89,8%	67,5%	79,2%	48,3%
Bradesco Auto/RE	24	15	17	9
	63,2%	39,5%	44,7%	23,7%
Bradesco Capitalização	7	5	6	4
	87,5%	62,5%	75,0%	50,0%
Bradesco Cartões	377	325	342	237
	86,1%	74,2%	78,1%	54,1%
Bradesco Consórcio	19	15	17	10
	65,5%	51,7%	58,6%	34,5%
Bradesco Financiamentos	78	62	73	40
	83,9%	66,7%	78,5%	43,0%
Bradesco Promotora	291	151	180	67
	74,4%	38,6%	46,0%	17,1%
Bradesco Saúde	45	25	26	24
	62,5%	34,7%	36,1%	33,3%
Bradesco Seguros	29	14	17	8
	72,5%	35,0%	42,5%	20,0%
Bradesco Vida e Previdência	22	16	19	11
	71,0%	51,6%	61,3%	35,5%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Brasilcap	1	1	1	1
	100,0%	100,0%	100,0%	100,0%
Brasilprev Seguros e Previdência	4	4	4	4
	50,0%	50,0%	50,0%	50,0%
Brastemp	278	86	92	79
	61,8%	19,1%	20,4%	17,6%
BuscaPê	0	0	0	0
	0,0%	0,0%	0,0%	0,0%
BV Financeira	240	222	229	171
	76,9%	71,2%	73,4%	54,8%
BV Leasing	3	3	3	1
	37,5%	37,5%	37,5%	12,5%
Caixa Capitalização	12	11	11	8
	80,0%	73,3%	73,3%	53,3%
Caixa Consórcios	23	18	20	16
	65,7%	51,4%	57,1%	45,7%
Caixa Econômica Federal	2308	1592	1880	1017
	73,7%	50,8%	60,0%	32,5%
Caixa Previdência	2	2	2	1
	66,7%	66,7%	66,7%	33,3%
Caixa Seguros	53	45	46	36
	80,3%	68,2%	69,7%	54,5%
Caixa Seguros Saúde	6	3	4	3
	100,0%	50,0%	66,7%	50,0%
Cardif Cap	2	1	1	1
	100,0%	50,0%	50,0%	50,0%
Cardif Seguros e Garantias	51	33	37	28
	76,1%	49,3%	55,2%	41,8%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Cardif Vida	1	1	1	0
	100,0%	100,0%	100,0%	0,0%
Carrefour	71	34	44	26
	74,7%	35,8%	46,3%	27,4%
Cartão Americanas.com	10	7	9	6
	52,6%	36,8%	47,4%	31,6%
Cartão Carrefour	193	147	165	103
	75,1%	57,2%	64,2%	40,1%
Cartão Quero-Quero VerdeCard	7	5	5	4
	100,0%	71,4%	71,4%	57,1%
Cartão Shoptime	6	4	4	4
	75,0%	50,0%	50,0%	50,0%
Cartão Submarino	126	108	114	96
	74,6%	63,9%	67,5%	56,8%
Casas Bahia	728	432	491	359
	80,4%	47,7%	54,3%	39,7%
Casasbahia.com	833	382	440	305
	65,6%	30,1%	34,7%	24,0%
Centauro	21	16	18	13
	67,7%	51,6%	58,1%	41,9%
Centauro.com.br	238	113	113	107
	64,7%	30,7%	30,7%	29,1%
City Lar	9	6	6	5
	69,2%	46,2%	46,2%	38,5%
Claro Celular	4333	4210	4281	3785
	84,4%	82,1%	83,4%	73,8%
Claro Fixo - Embratel	590	574	582	513
	84,0%	81,8%	82,9%	73,1%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Claro TV	1467	1451	1465	1308
	90,1%	89,1%	89,9%	80,3%
Comgás - Companhia de Gás de São Paulo	15	7	7	7
	83,3%	38,9%	38,9%	38,9%
Compra Certa	29	21	23	14
	82,9%	60,0%	65,7%	40,0%
Consórcio Fiat	20	14	17	11
	76,9%	53,8%	65,4%	42,3%
Consórcio Luiza	4	3	3	2
	100,0%	75,0%	75,0%	50,0%
Consórcio Realiza	1	0	0	0
	100,0%	0,0%	0,0%	0,0%
Consul	144	59	67	56
	64,0%	26,2%	29,8%	24,9%
Copel - Companhia Paranaense de Energia	32	8	8	8
	62,7%	15,7%	15,7%	15,7%
Copel Telecom	9	9	9	9
	90,0%	90,0%	90,0%	90,0%
CPFL Energia	56	19	19	19
	53,3%	18,1%	18,1%	18,1%
Crediare	6	5	6	5
	100,0%	83,3%	100,0%	83,3%
Credicard	233	191	213	157
	86,6%	71,0%	79,2%	58,4%
Credipar	7	5	6	4
	87,5%	62,5%	75,0%	50,0%
Decolar.com	217	201	209	182
	86,5%	80,1%	83,3%	72,5%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Dental Uni	2	1	1	1
	50,0%	25,0%	25,0%	25,0%
DIX Amico	5	0	1	0
	55,6%	0,0%	11,1%	0,0%
Dotz	43	41	42	36
	64,2%	61,2%	62,7%	53,7%
Dpaschoal	3	3	3	3
	75,0%	75,0%	75,0%	75,0%
Droga Raia	15	11	12	11
	68,2%	50,0%	54,5%	50,0%
Drogasil	6	3	3	3
	75,0%	37,5%	37,5%	37,5%
Eletro Shopping	8	7	7	7
	61,5%	53,8%	53,8%	53,8%
Época Cosméticos	2	2	2	2
	100,0%	100,0%	100,0%	100,0%
Esmaltec	24	12	13	10
	61,5%	30,8%	33,3%	25,6%
Expresso Itamarati	1	1	1	1
	33,3%	33,3%	33,3%	33,3%
Extra	92	64	72	53
	74,8%	52,0%	58,5%	43,1%
Extra.com	1052	540	598	449
	60,8%	31,2%	34,6%	26,0%
Fast Shop	45	40	43	38
	88,2%	78,4%	84,3%	74,5%
Financeira Itaú Americanas	23	16	21	7
	69,7%	48,5%	63,6%	21,2%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Garantec	80	60	62	55
	97,6%	73,2%	75,6%	67,1%
Girafa.com.br	7	4	5	4
	53,8%	30,8%	38,5%	30,8%
Gol Linhas Aéreas	229	209	212	182
	69,6%	63,5%	64,4%	55,3%
Golden Cross	6	2	2	2
	75,0%	25,0%	25,0%	25,0%
GVT	1770	1764	1768	1650
	86,0%	85,7%	85,9%	80,1%
Hiper Bompreço	11	6	8	6
	68,8%	37,5%	50,0%	37,5%
Hipercard	182	148	167	101
	78,4%	63,8%	72,0%	43,5%
HSBC	286	240	261	178
	64,3%	53,9%	58,7%	40,0%
Icatu Capitalização	1	0	0	0
	100,0%	0,0%	0,0%	0,0%
Icatu Seguros	6	5	6	4
	100,0%	83,3%	100,0%	66,7%
Ingresso.com	43	36	37	29
	71,7%	60,0%	61,7%	48,3%
Insinuante	38	16	21	11
	61,3%	25,8%	33,9%	17,7%
Intelig	2	2	2	2
	66,7%	66,7%	66,7%	66,7%
Itaú BMG Consignado	734	348	438	193
	82,4%	39,1%	49,2%	21,7%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Itaú Consórcio	50	26	32	16
	70,4%	36,6%	45,1%	22,5%
Itaú Seguros	30	18	19	14
	83,3%	50,0%	52,8%	38,9%
Itaú Seguros Auto e Residência	24	17	19	14
	77,4%	54,8%	61,3%	45,2%
Itaú Unibanco Capitalização	2	2	2	2
	66,7%	66,7%	66,7%	66,7%
Itaú Unibanco Consignado	26	13	16	7
	76,5%	38,2%	47,1%	20,6%
Itaú Unibanco Crédito Imobiliário	15	10	12	6
	75,0%	50,0%	60,0%	30,0%
Itaucard	849	714	779	575
	83,3%	70,1%	76,4%	56,4%
Itauleasing	113	72	96	39
	76,4%	48,6%	64,9%	26,4%
Jocar	1	0	0	0
	100,0%	0,0%	0,0%	0,0%
Jocar.com.br	2	1	1	1
	100,0%	50,0%	50,0%	50,0%
KaBuM!	64	54	55	38
	64,6%	54,5%	55,6%	38,4%
Komeco	0	0	0	0
	0,0%	0,0%	0,0%	0,0%
Latina Eletrodomésticos	10	7	7	6
	50,0%	35,0%	35,0%	30,0%
Lebes.com.br	1	1	1	1
	25,0%	25,0%	25,0%	25,0%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Lexmark	0	0	0	0
	0,0%	0,0%	0,0%	0,0%
LG Electronics	233	183	201	147
	57,4%	45,1%	49,5%	36,2%
Light	87	31	31	31
	46,5%	16,6%	16,6%	16,6%
Loja Brastemp	8	4	4	4
	53,3%	26,7%	26,7%	26,7%
Loja Consul	15	10	10	8
	83,3%	55,6%	55,6%	44,4%
Lojas Americanas	110	100	107	90
	84,6%	76,9%	82,3%	69,2%
Lojas Colombo	39	20	24	15
	62,9%	32,3%	38,7%	24,2%
Lojas Colombo.com	29	20	23	18
	55,8%	38,5%	44,2%	34,6%
Lojas Lebes	7	7	7	7
	77,8%	77,8%	77,8%	77,8%
Lojas Quero-Quero	5	2	2	1
	71,4%	28,6%	28,6%	14,3%
Losango	124	108	117	82
	82,1%	71,5%	77,5%	54,3%
Luizacred	74	57	71	38
	79,6%	61,3%	76,3%	40,9%
Luizaseg	20	14	16	11
	71,4%	50,0%	57,1%	39,3%
Magazine Luiza	172	131	141	103
	80,0%	60,9%	65,6%	47,9%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Magazineluiza.com	806	531	582	404
	80,4%	53,0%	58,1%	40,3%
Mapfre Consórcios	0	0	0	0
	0,0%	0,0%	0,0%	0,0%
Mapfre Seguros	54	41	42	32
	63,5%	48,2%	49,4%	37,6%
Marítima Saúde	2	1	1	1
	50,0%	25,0%	25,0%	25,0%
Mercadorama	2	1	1	1
	100,0%	50,0%	50,0%	50,0%
Moip	20	16	18	13
	80,0%	64,0%	72,0%	52,0%
Mueller Eletrodomésticos	3	1	2	1
	100,0%	33,3%	66,7%	33,3%
Mueller Fogões	1	0	0	0
	50,0%	0,0%	0,0%	0,0%
Multiplus	31	31	31	29
	58,5%	58,5%	58,5%	54,7%
Mundial Editora	23	7	10	5
	62,2%	18,9%	27,0%	13,5%
Nacional	2	2	2	2
	33,3%	33,3%	33,3%	33,3%
Netfarma	13	11	11	8
	81,3%	68,8%	68,8%	50,0%
Nextel	194	194	194	189
	86,2%	86,2%	86,2%	84,0%
Novo Mundo	64	35	40	28
	61,5%	33,7%	38,5%	26,9%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

O Boticário	37	25	30	22
	59,7%	40,3%	48,4%	35,5%
Odontoprev	13	4	4	4
	56,5%	17,4%	17,4%	17,4%
Oi Celular	3275	3208	3252	2920
	80,7%	79,1%	80,1%	72,0%
Oi Fixo	7440	7224	7369	6140
	77,0%	74,8%	76,3%	63,5%
Oi Paggo Administradora de Crédito	0	0	0	0
	0,0%	0,0%	0,0%	0,0%
Omni Financeira	3	3	3	2
	75,0%	75,0%	75,0%	50,0%
Ourocard	189	172	180	153
	71,6%	65,2%	68,2%	58,0%
Panvel Farmácias	8	7	7	7
	72,7%	63,6%	63,6%	63,6%
Pão de Açúcar	14	12	13	11
	70,0%	60,0%	65,0%	55,0%
ParPerfeito	3	3	3	3
	100,0%	100,0%	100,0%	100,0%
Pernambucanas	114	78	90	56
	71,3%	48,8%	56,3%	35,0%
Philips Audio e Vídeo	35	21	21	18
	61,4%	36,8%	36,8%	31,6%
Philips TV e Monitores	85	50	53	49
	76,6%	45,0%	47,7%	44,1%
Philips Walita, Avent, Saeco e Sonicare	41	21	27	17
	70,7%	36,2%	46,6%	29,3%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Polishop	17	10	11	7
	85,0%	50,0%	55,0%	35,0%
Polishop.com.br	32	27	29	24
	82,1%	69,2%	74,4%	61,5%
Ponto Frio	154	93	101	73
	80,6%	48,7%	52,9%	38,2%
Pontofrio.com	854	409	462	332
	64,3%	30,8%	34,8%	25,0%
Portal Terra	89	88	88	86
	79,5%	78,6%	78,6%	76,8%
Purificador de Agua Brastemp	40	35	38	32
	78,4%	68,6%	74,5%	62,7%
QBE Brasil Seguros	5	5	5	2
	71,4%	71,4%	71,4%	28,6%
Recovery do Brasil Consultoria	61	61	61	58
	68,5%	68,5%	68,5%	65,2%
RGE - Rio Grande Energia	8	2	2	2
	57,1%	14,3%	14,3%	14,3%
Ricardo Eletro	46	26	33	21
	60,5%	34,2%	43,4%	27,6%
Ricardo Eletro.com	309	147	164	119
	53,9%	25,7%	28,6%	20,8%
SAAE Sorocaba	2	1	1	1
	28,6%	14,3%	14,3%	14,3%
Sam's Club	3	2	2	2
	100,0%	66,7%	66,7%	66,7%
Samsung	1687	1109	1284	856
	66,8%	43,9%	50,9%	33,9%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Continuação)

Santander Financiamentos	126	81	114	35
	61,5%	39,5%	55,6%	17,1%
Senff	71	69	70	60
	79,8%	77,5%	78,7%	67,4%
Serasa Experian	87	61	72	56
	60,4%	42,4%	50,0%	38,9%
Shopfato.com	4	4	4	4
	28,6%	28,6%	28,6%	28,6%
Shoptime	159	107	122	77
	71,0%	47,8%	54,5%	34,4%
SKY	4683	4645	4674	4258
	80,6%	80,0%	80,5%	73,3%
Smiles	259	222	239	186
	71,5%	61,3%	66,0%	51,4%
Sou barato	31	27	28	21
	62,0%	54,0%	56,0%	42,0%
Submarino	493	362	400	263
	63,7%	46,8%	51,7%	34,0%
Submarino Viagens	51	46	49	36
	68,9%	62,2%	66,2%	48,6%
SuperMuffato.com	1	0	1	0
	100,0%	0,0%	100,0%	0,0%
Tam	360	292	309	237
	70,9%	57,5%	60,8%	46,7%
TAM Viagens	14	12	13	10
	58,3%	50,0%	54,2%	41,7%
Tim	5002	4906	4961	4517
	77,6%	76,1%	77,0%	70,1%

Fonte: Autor

Figura 22 - Tabela Completa do Índice de Solução de Todas as Empresas Constadas nos Dados

(Conclusão)

Todo Dia Supermercados	2	1	2	1
	66,7%	33,3%	66,7%	33,3%
Tricard	5	5	5	5
	100,0%	100,0%	100,0%	100,0%
Unimed Belém	2	2	2	2
	100,0%	100,0%	100,0%	100,0%
Unimed Fortaleza	4	3	3	3
	80,0%	60,0%	60,0%	60,0%
Unimed Paraná	5	2	3	2
	71,4%	28,6%	42,9%	28,6%
Unimed Paulistana	22	12	14	8
	45,8%	25,0%	29,2%	16,7%
Valejet.com	1	1	1	1
	50,0%	50,0%	50,0%	50,0%
ViajaNet	15	12	12	12
	71,4%	57,1%	57,1%	57,1%
Virginia Surety Seguros	5	3	4	3
	83,3%	50,0%	66,7%	50,0%
Vivo - Telefônica	8478	8246	8385	7099
	81,8%	79,6%	80,9%	68,5%
Walmart	21	16	16	10
	67,7%	51,6%	51,6%	32,3%
Walmart.com	1073	561	595	467
	73,2%	38,3%	40,6%	31,9%
Wine	10	8	8	7
	71,4%	57,1%	57,1%	50,0%
Yasuda Marítima Seguros	5	1	2	1
	71,4%	14,3%	28,6%	14,3%
TOTAL	63471	54110	56782	45582
	76,7%	65,4%	68,7%	55,1%

Fonte: Autor