

FABIO HENRIQUE ROCHA SANTOS

**APLICAÇÃO DA METODOLOGIA *LEAN STARTUP* EM EMPRESA DO
MERCADO DE FIDELIZAÇÃO DO BRASIL**

Trabalho de Formatura apresentado à Escola
Politécnica da Universidade de São Paulo para
obtenção do diploma de Engenheiro de Produção.

SÃO PAULO

2014

FABIO HENRIQUE ROCHA SANTOS

**APLICAÇÃO DA METODOLOGIA *LEAN STARTUP* EM EMPRESA DO
MERCADO DE FIDELIZAÇÃO DO BRASIL**

Trabalho de Formatura apresentado à Escola Politécnica da Universidade de São Paulo para obtenção do diploma de Engenheiro de Produção.

Orientador: Prof. Dr. André Leme Fleury

SÃO PAULO

2014

À meus pais Mário e Luzia Santos que tanto lutaram para meu sucesso

AGRADECIMENTOS

Agradeço, em primeiro lugar a Deus e aos meus anjos da guarda, que sempre me protegeram e me auxiliarem em tudo o que faço na vida. Sem eles, com certeza, eu não poderia estar aqui hoje publicando este trabalho.

Em segunda lugar, gostaria de deixar aqui meus mais sinceros agradecimentos a minha família, em especial a meus pais Mário e Luzia, que tanto lutaram a vida toda para que eu e minha irmã Júlia tivéssemos acesso a uma educação de qualidade.

Em terceiro lugar, gostaria de agradecer a minha namorada, Ana Paula, que me ajudou a todo instante na tarefa árdua que foi escrever este trabalho., Se não fosse sua força a meu lado, confesso que não sei se estaria entregando este trabalho de formatura na presente data.

Gostaria de agradecer também aos meus colegas da Poli, em especial ao Paulo Machado, Pedro Joaquim e Rodrigo Bianchini com quem compartilhei grandes momentos dentro desta universidade, os quais ficarão marcados para sempre em minha memória

Gostaria de prestar meus agradecimentos e minhas homenagens ao professor André Fleury, que com sua experiência, sabedoria e bom humor conduziu o andamento deste trabalho de maneira elogiável.

Agradeço ainda a todos os meus professores desde o Ensino Fundamental até a Poli, pois sem dúvida, se hoje estou aqui publicando este trabalho é porque cada um colaborou com um pouquinho de seus conhecimentos e experiências.

Por fim e não menos importantes gostaria de agradecer a toda a equipe da Loyalty S.A. que não só confiaram em mim para a realização deste trabalho como também colaboraram muito com seus conhecimentos, experiências e visão de negócio.

Se não puder voar, corra. Se não puder correr, ande. Se não puder andar, rasteje, mas continue em frente de qualquer jeito.

Martin Luther King

RESUMO

Este trabalho fundamenta-se nos conhecimentos ligados à Metodologia Lean, tal qual concebida por Eric Ries e Steve Blank e à Modelagem Estatística tradicional, aplicando-os de forma integrada no contexto de uma grande empresa do segmento de Fidelidade no Brasil.

O foco principal deste projeto está em desenvolver um processo em prol da gestão do canal de vendas *E-mail Marketing* baseado nos conceitos do ciclo “Construir-Medir-Aprender” e da Aprendizagem Validada. A parte referente ao uso de Modelagem Estatística integra-se a estes conceitos uma vez que fornece métodos de geração de hipóteses acerca do negócio com embasamento teórico-científico.

Trata-se sob uma perspectiva mais ampla, de um exemplo de aplicação em caráter prático organizacional de um Modelo que tem como ponto fundamental a preocupação com o cliente desenvolvido inicialmente para *Startups*. Deste modo, todas as propostas de melhorias geradas neste projeto foram validadas em ambiente de interação com o clientes antes que pudessem ser consideradas ganhos de fato para a organização em estudo.

Na parte final, esboça-se um modelo de organização do trabalho que visa institucionalizar o processo desenvolvido neste projeto, tornando-o parte do dia-a-dia da organização estudada.

Palavras-chave: Metodologia Lean, Regressão Linear, Mercado de Fidelidade, Aprendizagem Validada, Ciclo “Medir-Construir-Aprender”.

ABSTRACT

This paper dialogues directly with the fields of knowledge related to Lean Methodology, which is based on Eric Ries and Steve Blank proposals and traditional Statistical Modeling. Both concepts are applied seamlessly to the context of a large company of Loyalty Market in Brazil.

This project's main focus is to establish a process in order to help the management of the sales channel called E-mail Marketing based on the concepts of "Build- Measure - Learn" cycle and Validated Learning. The part concerning the use of statistical modeling integrates these concepts since it provides methods for generating hypotheses about the business with theoretical and scientific basis.

This is, from a broader perspective, an example of practical application in a company environment of a model, initially developed for Startups, whose fundamental point concerns about its customers. Thus, all proposals were validated through an interaction with customers before being considered improvements for the company that is being studied.

Finally, it outlines a model of job organization which aims to institutionalize the process developed in this project, making it part of the day-to-day operations of the company studied.

Keywords: Lean Methodology, Linear Regression, Loyalty market, Validated learning, cycle "Build- Measure - Learn"

LISTA DE FIGURAS

Figura 1: Ciclo tradicional de desenvolvimento do produto (Blank, 2006)	29
Figura 2: Ciclo de desenvolvimento de clientes de Blank (2006).	30
Figura 3: Ciclo de vida dos clientes extraído de autoria de Croll e Yoskovitz (2013).....	31
Figura 4: Lean Canvas de Croll e Yoskovitz (2013)	33
Figura 5: Ciclo “Construir – Medir- Aprender”. Fonte Ries (2013).....	38
Figura 6: Exemplo de <i>Scatterplot</i> para modelo homocedástico.	50
Figura 7: Interligações das macro-etapas da pesquisa	55
Figura 8: Penetração dos programas de fidelidade na população total (%). Destaque para as brasileiras Multiplus e Smiles. Fonte: Divulgação de Resultados da Multiplus (Ago/2014).....	63
Figura 9: <i>Lean canvas</i> da organização Loyalty S.A.	68
Figura 10: Ciclo de vida dos clientes das Loyalty S.A. baseado no diagrama do ciclo de vida de Croll e Yoskovitz (2013).....	69
Figura 11: Determinação do foco do estudo baseado no diagrama do ciclo de vida de Croll e Yoskovitz (2013).	74
Figura 12: Amostra de dados no software PASW Statistics 17.....	134
Figura 13: Inserindo os parâmetros da regressão no software estatístico	135
Figura 14: Seleção dos parâmetros de controle a ser mostrados na tela de saída da análise multivariada	136
Figura 15: Parâmetros da análise residual a priori.....	136
Figura 16: Parâmetros residuais para a análise de resíduos	137
Figura 17: Critérios utilizados na regressão linear multivariada.	138
Figura 18: Tela inicial da análise de resíduos	139
Figura 19: Selecionando-se os parâmetros e gráficos a ser exibidos na análise de resíduos	140
Figura 20: Normal <i>Plot-test</i> para análise de resíduos.	145
Figura 21: <i>Scatterplot</i> dos resíduos do modelo em análise.....	146
Figura 22: Normal <i>Plot-test</i> para análise de resíduos.	164
Figura 23: <i>Scatterplot</i> dos resíduos do modelo em análise.....	165
Figura 24: Normal <i>Plot-test</i> para análise de resíduos.	183
Figura 25: <i>Scatterplot</i> dos resíduos do modelo em análise.....	184

LISTA DE TABELAS

Tabela 1: Conversão da variável “Satisfação dos Clientes” em variáveis binárias	43
Tabela 2: Análise de Variância com mais de uma Variável independente.....	48
Tabela 3: Transformação dos valores da variável “Dia da Semana”	81
Tabela 4: Tipos das variáveis após a transformação binária das categóricas	82
Tabela 5: Valores de R e R ² no modelo em estudo.....	85
Tabela 6: Coeficiente e sua respectiva significância para o modelo 1.	86
Tabela 7: Correlações de Pearson entre a taxa de Abertura e as variáveis independentes com coeficiente significante	87
Tabela 8: Resultados dos testes de normalidade para o modelo 1	88
Tabela 9: Resumo das hipóteses validadas em ambiente prático	91
Tabela 10: Valores de R e R ² no modelo em estudo.....	94
Tabela 11: Coeficiente e sua respectiva significância para o modelo 2.	95
Tabela 12: Correlações de Pearson entre a taxa de abertura e as variáveis independentes com coeficiente significante	96
Tabela 13: Resultados dos testes de normalidade para o modelo 2.....	97
Tabela 14: Resumo das hipóteses validadas em ambiente prático	101
Tabela 15: Valores de R e R ² no modelo em estudo.....	103
Tabela 16: Coeficiente e sua respectiva significância para o modelo 3.	104
Tabela 17: Correlações de Pearson entre a taxa de Abertura e as variáveis independentes com coeficiente significante	105
Tabela 18: Resultados dos testes de normalidade para o modelo 3	106
Tabela 19: Resumo das hipóteses validadas em ambiente prático	109
Tabela 20: Descrição das variáveis.....	122
Tabela 21: Transformação dos valores assumidos pela variável “Tipo de Comunicação”	123
Tabela 22: Transformação binária dos valores assumidos pela variável categórica “Tipo de Assunto”	124
Tabela 23: Resultados das estatísticas descritivas para regressão com dados padronizados	141
Tabela 24: Correlações de Pearson entre a taxa de Abertura e as variáveis independentes	142
Tabela 25: ANOVA para modelo 1 – após a padronização das variáveis	143

Tabela 26: resumo das informações do modelo 1.....	143
Tabela 27: Coeficiente e sua respectiva significância para o modelo 1.	144
Tabela 28: Resultados dos testes de normalidade para o modelo 1	147
Tabela 29: Resultados das estatísticas descritivas para regressão com dados padronizados	160
Tabela 30: Correlações de Pearson entre a taxa de Abertura e as variáveis independentes	161
Tabela 31: ANOVA para modelo 2	162
Tabela 32: Resumo das informações do modelo 2	162
Tabela 33: Coeficiente e sua respectiva significância para o modelo 2.	163
Tabela 34: Resultados dos testes de normalidade para o modelo 2.....	166
Tabela 35: Resultados das estatísticas descritivas para regressão com dados padronizados	179
Tabela 36: Correlações de Pearson entre a taxa de Abertura e as variáveis independentes	180
Tabela 37: ANOVA para modelo 3	181
Tabela 38: resumo das informações do modelo 3.....	181
Tabela 39: Coeficiente e sua respectiva significância para o modelo 3.	182
Tabela 40: Resultados dos testes de normalidade para o modelo 3	185

LISTA DE QUADROS

Quadro 1: Transformação da variável X1_dia_semana.....	83
Quadro 2: Transformação da variável X2_dia_semana.....	84
Quadro 3: Transformação da variável X3_dia_semana.....	85
Quadro 4: Exemplo de transformação da variável X1_dia_semana	93
Quadro 5: Exemplo de transformação da variável X1_tipo_comunicação	93
Quadro 6: Exemplo de transformação da variável X1_dia_semana	102
Quadro 7: Exemplo de transformação da variável X1_tipo_comunicação	102
Quadro 8: Processo de testes baseado no Ciclo “Construir Medir Aprender” de Ries (2013) aplicado a Loyalty S.A	112
Quadro 9: Transformação da variável X1_tipo_comunicação	125
Quadro 10: Transformação da variável X2_tipo_comunicação	126
Quadro 11: Transformação da variável X1_assunto.....	127
Quadro 12: Transformação da variável X2_assunto.....	128
Quadro 13: Transformação da variável X3_assunto.....	129
Quadro 14: Transformação da variável “saldo médio”.....	130
Quadro 15: Transformação da variável “saldo médio”.....	131
Quadro 16: Transformação da variável “grau transacional”.....	132
Quadro 17: Transformação da variável “taxa de abertura”	133
Quadro 18: Transformação da variável “X1_dia_semana”	148
Quadro 19: Transformação da variável “X2_dia_semana”	149
Quadro 20: Transformação da variável “X3_dia_semana”	150
Quadro 21: Transformação da variável “X1_tipo_comunicação”	151
Quadro 22: Transformação da variável “X2_tipo_comunicação”	152
Quadro 23: Transformação da variável “X1_assunto”	153
Quadro 24: Transformação da variável “X2_assunto”	154
Quadro 25: Transformação da variável “X3_assunto”	155
Quadro 26: Transformação da variável “Saldo Médio”	156
Quadro 27: Transformação da variável “Percentual impactado”	157
Quadro 28: Transformação da variável “Grau transacional”.....	158
Quadro 29: Transformação da variável “Taxa de abertura”	159
Quadro 30: Transformação da variável “X1_dia_semana”	167
Quadro 31: Transformação da variável “X2_dia_semana”	168

Quadro 32: Transformação da variável “X3_dia_taxa de abertura”	169
Quadro 33: Transformação da variável “X2_tipo_comunicação”	170
Quadro 34: Transformação da variável “X2_tipo_comunicação”	171
Quadro 35: Transformação da variável “X1_assunto”	172
Quadro 36: Transformação da variável “X2_assunto”	173
Quadro 37: Transformação da variável “X3_assunto”	174
Quadro 38: Transformação da variável “Saldo Médio”	175
Quadro 39: Transformação da variável “Percentual impactado”	176
Quadro 40: Transformação da variável “Grau transacional”	177
Quadro 41: Transformação da variável “Taxa de abertura”	178

LISTA DE ABREVIATURAS E SIGLAS

BRA – República Federativa do Brasil

BSC – comércio efetuado entre a empresa e o cliente final

CAN: Canadá

CEO – Diretor Executivo

DEU: República Federal da Alemanha

Df – Graus de Liberdade

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária

Et. al.. – entre outros

E-learning – Aprendizado pela internet

E-mail – Correio eletrônico

EUA – Estados Unidos da América

Intuite S.A. – Intuite Sociedade Anônima

IPO – Oferta Pública de Ações

JPN: Japão

Loyalty S.A - Loyalty Sociedade Anônima

MEX: Estados Unidos Mexicanos

PDCA – Planejar, Fazer, Checar, Agir

Return Path S.A. – *Return Path* Sociedade Anônima

Sig. – Significância Estatística

SMS: Serviço de Mensagens de texto via celular

SPAM: Mensagens não solicitadas

Transf. – Transformado

UK – Reino Unido da Grã-Bretanha e Irlanda do Norte

XPTO – Qualquer

SUMÁRIO

1.	Introdução	22
1.1.	Contexto em estudo	22
1.2.	Problemas	23
1.3.	Objetivo	24
1.4.	Justificativa	25
1.5.	Estrutura do Trabalho.....	25
2.	Revisão de Literatura.....	28
2.1.	Metodologia <i>Lean</i>	28
2.2.	<i>E-mail Marketing</i>	39
2.3.	Análise Estatística	41
3.	Metodologia de Pesquisa	54
3.1.	Mapeamento do negócio.....	56
3.1.1.	Compreender o mercado de fidelidade.....	56
3.1.2.	O Modelo de negócio da Loyalty S.A.	56
3.1.3.	Ciclo de Vida dos clientes da Loyalty S.A.	56
3.2.	Planejamento do Experimento	57
3.2.1.	Seleção da abordagem <i>Lean</i>	57
3.2.2.	Segmentação dos Clientes	57
3.2.3.	Seleção do Modelo de Referência	57
3.3.	Modelagem Estatística.....	58
3.3.1.	Construção do Modelo Estatístico	58
3.3.2.	Extração das Hipóteses	59
3.4.	Testes e Validação	59
3.5.	Implementação.....	60
4.	Resultados	62
4.1.	Mapeamento do negócio.....	62
4.1.1.	O Mercado de Fidelidade.....	62
4.1.2.	O Modelo de negócio da Loyalty S.A.	64
4.1.3.	Ciclo de Vida dos Clientes	68

4.1.3.1.	Aquisição	69
4.1.3.2.	Ativação.....	70
4.1.3.3.	Retenção	71
4.1.3.4.	Revenda	71
4.1.3.5.	Referenciamento.....	72
4.1.3.6.	Perda de Clientes	72
4.2.	Planejamento do Experimento.....	73
4.2.1.	Seleção da Abordagem <i>Lean</i>	73
4.2.2.	Segmentação dos clientes	74
4.2.3.	Seleção do Modelo de Referência.....	75
4.3.	Modelagem Estatística	76
4.3.1.	Construção do Modelo Estatístico	76
4.3.1.1.	Mapeamento das variáveis do modelo	76
4.3.1.2.	Os dados.....	80
4.3.1.3.	Transformação das variáveis	80
4.3.1.4.	Padronização dos valores	81
4.3.2.	A regressão linear.....	85
4.3.2.1.	Poder de Explicação do Modelo	85
4.3.2.2.	Coeficientes da Equação de Regressão.....	86
4.3.2.3.	Análise de Resíduos	87
4.3.3.	Extração das hipóteses	88
4.4.	Teste e Validação das Hipóteses	89
4.5.	Refinamento do Modelo	92
4.5.1.	Transformação A	92
4.5.2.	A regressão linear com os dados quadráticos.....	94
4.5.2.1.	Pode de explicação do Modelo.....	94
4.5.2.2.	Coeficientes da Equação de Regressão.....	94
4.5.2.3.	Análise de Resíduos	96
4.5.3.	Extração das hipóteses pós 1 etapa do refino.....	97
4.5.4.	Testes e Validação das Hipóteses.....	98
4.5.5.	Transformação B	101
4.5.6.	A regressão linear pós transformação B	103

4.5.6.1. Poder de explicação do Modelo.....	103
4.5.6.2. Coeficientes da Equação de Regressão	103
4.5.6.3. Análise de Resíduos.....	105
4.5.7. Extração das hipóteses	106
4.5.8. Experimentação e Validação.....	107
4.5.9. Analise dos Resultados	109
4.6. Institucionalizando a Metodologia	110
5. Conclusão	114
6. Referências Bibliográficas.....	118
7. Anexos	122
7.1. Anexo I: Tipos das variáveis utilizadas no modelo.....	122
7.2. Anexo II: Transformação binária das variáveis “Tipo de comunicação” e “Tipo de Assunto”	123
7.3. Anexo III: Padronização das variáveis.....	125
7.4. Anexo IV: Regressão Linear no PASW <i>Statistics</i>	134
7.5. Anexo V: Resultado da Regressão Linear I.....	141
7.6. Anexo VI: Transformação das variáveis para o refinamento do modelo (quadráticos)	148
7.7. Anexo VII: Resultado da Regressão Linear II.....	160
7.8. Anexo VIII: Transformação das variáveis para o refinamento do Modelo (raiz cúbica)	167
7.9. Anexo IX: Resultado da Regressão Linear III.....	179

Capítulo 1: Introdução

1. Introdução

1.1. Contexto em estudo

O mundo globalizado cria cenários em que as organizações cada vez mais veem-se obrigadas a buscar alternativas e formas de dinamizar-se visando garantir sua lucratividade e consequente sobrevivência ao longo prazo.

Serra, Torres e Torres (2004) afirmam que a análise do ambiente externo pode ser realizada por meio do modelo da competitividade, desenvolvido por Michael Porter na década de 1970. Segundo este modelo, um setor da economia, independentemente de qual seja, pode ser analisado levando-se em conta os concorrentes, novos entrantes, poder de barganha dos compradores e fornecedores e surgimento de produtos ou bens substitutos.

Para que obtenham vantagens competitivas e galguem posições de destaque no mercado torna-se imprescindível que as organizações tomem atitudes pró-ativas. Muitas delas, tendo em mente que o custo de aquisição de um novo cliente é quase que sempre superior ao esforço necessário para fidelizar o cliente já adquirido, vêm investindo em programas relacionamento de forma a recompensar seus clientes pelas suas escolhas de consumo.

É neste momento que surgem as redes de fidelidade, também conhecidas como programas de coalizão, por meio das quais se oferecem aos consumidores possibilidades de recompensas que vão muito além daquelas normalmente oferecidas por empresas cujo foco de atuação está fora do mercado de fidelidade. Dowling e Uncles (1997) afirmam que um dos principais objetivos de um programa de fidelidade está em estabelecer um maior nível de retenção de clientes em segmentos rentáveis, proporcionando maior satisfação e valor para um seletivo grupo de consumidores fiéis.

O caso em estudo neste trabalho é a organização Loyalty S.A., player do mercado de fidelização do Brasil cujo *core business* consiste na integração de programas de fidelidade de empresas parceiras. A empresa concentra seus esforços em duas frentes principais, as quais seguem descritas:

- Na frente que trata formação e solidificação da rede, atuam as equipes “comercial” e “estratégica”, visando ampliar cada vez mais o número de parceiros, desde que estes mostrem-se aderentes ao contexto da organização. Mais parceiros, em última instância

quer dizer maiores possibilidades de realização de interação dos clientes com a rede e maior movimentação de *credits* (moeda complementar criada pela organização).

- Já na frente que responsável pelos clientes, detecta-se uma forte preocupação em compreender suas necessidades, de modo a aperfeiçoar esforços para entender cada vez o cliente e assim avançar no objetivo de fidelizá-lo. Como exemplo deste aperfeiçoamento, detalha-se o esforço em descobrir quais campanhas de *E-mail Marketing*, o principal canal de vendas da organização, são mais atrativas para quais tipos de clientes. Além disso, esta frente, ainda ocupa-se atuar na experiência dos clientes no contato com a rede, pois se acredita que clientes que tiverem uma má experiência dificilmente retornam a interagir.

Segundo analistas internos da Loyalty S.A., o canal *E-mail Marketing* é o responsável por cerca de 50% do trânsito de clientes no site da organização. Apesar desta importância nota-se, ainda, a ausência de uma metodologia claramente estabelecida em prol da construção de hipóteses com base teórica e realização de testes neste canal. Procura-se utilizar conceitos baseados no ciclo PDCA (Planejar, Fazer, Medir, Atuar, do inglês), porém, necessita-se de incrementos em prol do estabelecimento de um ciclo de aprendizagem estruturado.

1.2. Problemas

Sendo o *E-mail Marketing* o principal canal de vendas da organização, conforme comentando na seção 1.1., a empresa procura permanentemente alternativas que venham a melhorar seu desempenho, que é medido pelas taxas de entrega (Taxa de entrega no provedor), de interação (Taxas de Abertura e de Cliques) e taxa de conversão.

Segundo analistas internos do time de campanhas da Loyalty S.A., a taxa de entrega no provedor refere-se ao volume de *E-mails* retornados pelo sistema de disparo com o *status* “Entregue” em relação ao total disparado. Essa discrepância ocorre pois nem sempre tudo que é disparado é recebido pelos servidores dos provedores de *E-mail*. As taxas de interação, como o próprio nome já diz, visam medir como foi a interação dos clientes impactados com o *E-mail*. A taxa de abertura é calculada dividindo-se o total de usuários que abriram o *E-mail* pelo total de usuário que o receberam e a taxa de cliques pelo total de usuários que clicaram em pelo menos um *link* da peça em relação ao total que abriu. Já a taxa de conversão mostra o ganho em vendas gerado diretamente pelo *E-mail Marketing*. Ela pode ser calculada

dividindo-se o total de clientes que compraram via site advindos do *E-mail* pelo total de clientes impactados.

Como sugestão do gerente da área responsável pelo *E-mail*, o time de campanhas da Loyalty S.A., que é quem está no dia-a-dia na gestão deste canal de vendas, procura realizar cada vez mais análises e testes de modo a aprimorá-lo. Contudo, segundo os próprios analistas da organização, muitas destas análises acabam esbarrando na ausência de uma modelagem estatística aplicada de forma estruturada e que seja capaz de promover a aprendizagem da organização na busca por melhores indicadores de desempenho associados às campanhas de *E-mail Marketing*.

Logo, uma oportunidade de melhoria levantada pelo time interno da organização está na ausência de um processo claramente estabelecido para geração das hipóteses, realização dos testes e validação dos prognósticos.

Em segundo lugar, destaca-se a necessidade da empresa em obter conhecimento acerca dos fatores mais relevantes na interação cliente – *E-mail Marketing*, de modo, que ela possa potencializar os indicadores deste canal, aumentando ainda mais sua relevância.

1.3. Objetivo

Sintetizando-se os problemas citados anteriormente, destaca-se que o objetivo principal deste trabalho está em atacar a inexistência de um processo estabelecido em prol da construção de hipóteses de melhoria do canal *E-mail Marketing*, via modelagem estatística, e que permita a realização de testes práticos com os clientes para que se possa validar ou descartar tais hipóteses. A construção deste processo baseia-se na metodologia *Lean Startup*, idealizada por Steve Blank (2006) e Eric Ries (2013) e, também, com grande contribuição de Alistar Croll e Ben Yoskovitz (2013), descrita em detalhes na seção destinada a revisão de Literatura. Destaca-se, por hora, que o processo aqui planejado para a Loyalty S.A. baseia-se na utilização dos conceitos do ciclo “Construir–medir-aprender” e da “Aprendizagem Validada”, ambos descritos por Ries (2013).

Com relação a necessidade da organização em conhecer os fatores mais relevantes na interação cliente- *E-mail Marketing*, adotou-se, como fonte de pesquisa, a taxa de abertura das ações de *E-mail* da Loyalty S.A.. Deste modo, toda construção do processo baseado nos conceitos *Lean* aplicados ao contexto da Loyalty S.A. terá como foco de atuação a taxa de

abertura. Enfatiza-se, contudo, que o mesmo processo pode ser replicado para os outros indicadores de desempenho de *E-mail Marketing* comentados na seção anterior.

1.4. Justificativa

Segundo Ries (2013), “o objetivo de uma *Startup* é descobrir a coisa certa a criar – a coisa que os clientes querem e pela qual pagarão- o mais rápido possível. Extrapolando-se esta citação para o caso em estudo, observa-se que o sucesso da organização Loyalty S.A. está atrelado diretamente ao grau de fidelização que esta consegue despertar em seus clientes. Notadamente há uma linha muito tênue entre descobrir a coisa certa, a coisa que os clientes querem e pela qual pagarão e a necessidade de fidelização dos clientes. Quanto mais rápido e mais assertivo for o esforço da organização em detectar e oferecer aquilo que seus clientes realmente esperam, maior será o grau de retenção e fidelização destes com a rede Loyalty S.A.

Sabendo-se que o canal *E-mail Marketing* é responsável direto por mais de 50% dos acessos registrados no site da empresa e, também, que dias com ações de *E-mail Marketing* elevam o volume de “*credits*” acumulados nos parceiros, aumentando o valor da rede que conecta estas empresas, torna-se muito bem vista a existência de esforços e iniciativas que venham a colaborar para o aprimoramento deste canal no contexto da Loyalty S.A., objetivo central deste trabalho.

Ressalta-se ainda, a correlação deste trabalho com a Engenharia de Produção, uma vez que este visa desenvolver um processo para otimização dos resultados do canal de vendas da organização em estudo, a partir de uma metodologia completamente original – a Metodologia *Lean*. Para a criação deste processo baseia-se em conceitos muito estudados pelo autor durante o curso de Engenharia de Produção, tais quais, estatística aplicada, inovação e processos.

Outro ponto de destaque que está de acordo com o que se espera do Engenheiro de Produção é a necessidade de visão sistêmica. Para a criação do processo de otimização do canal *E-mail*, o autor teve de analisar o sistema de *E-mail Marketing* da organização como um todo e também cada uma de suas partes individualmente.

1.5. Estrutura do Trabalho

A estrutura deste trabalho baseia-se na ordem com que as atividades foram executadas.

Em primeiro lugar definiu-se a literatura que iria embasar todo o projeto. Dentro da revisão literária, a primeira parte consistiu em realizar um estudo a cerca do que os autores vêm chamando de Metodologia Lean, procurando levantar seus principais fundamentos e formas de aplicação. A seguir, estudou-se o *E-mail Marketing*. Em terceiro lugar, procurou-se analisar e fundamentar os conceitos do Modelo Estatístico utilizado na análise – a regressão Linear Multivariada.

A seguir desenvolveu-se a Metodologia de Pesquisa deste trabalho, procurando elaborar as principais macro-etapas e as etapas para a sua realização.

O próximo passo foi colocar em prática os conceitos da Literatura, seguindo as etapas determinadas na metodologia de pesquisa para o estudo do caso da Loyalty S.A. Analisou-se, num primeiro momento, o mercado brasileiro de Fidelização, o modelo de negócio da organização em estudo e o ciclo de vida de seus clientes, de forma que o autor pudesse familiarizar-se com o contexto abordado. Após a realização desta análise, definiu, baseando-se na literatura, a melhor forma de abordagem para o problema analisado e, também, em qual momento do ciclo de vida dos clientes os esforços estariam concentrados.

Uma vez definidos estes aspectos, trabalhou-se na construção do Modelo Estatístico de modo a gerar as hipóteses a cerca do comportamento do fenômeno em estudo – a taxa de abertura das ações de *E-mail Marketing* da Loyalty S.A.

A etapa seguinte foi a realização dos testes em ambiente prático de interação com os clientes em prol da validação das hipóteses geradas pelo modelo estatístico. Caso alguma das hipóteses não fosse validada, o modelo seria considerado reprovado e retornaria a etapa anterior para refinamento.

Por fim, após a validação de todas as etapas do modelo, desenhou-se um processo visando internalizar esta metodologia de trabalho na Loyalty S.A.

A última parte deste trabalho apresenta todos os detalhes teóricos e todas as passagens executadas durante a geração e refinamento do modelo. Estas passagens de cunho puramente teórico foram destacados para a seção anexos visando tornar facilitar a leitura deste documento.

Capítulo 2 - Revisão de Literatura

2. Revisão de Literatura

Neste capítulo são apresentados os referenciais teóricos que embasaram a Metodologia de Pesquisa deste trabalho.

Em primeiro lugar descreve-se a Metodologia *Lean* baseando-se nas obras *Lean Startup* (Ries, 2013), *The Four Steps to the Epyphany* (Blank, 2006) e *Lean Analytics* (Croll e Yoskovitz, 2013).

Em segundo lugar, procurou-se buscar na literatura o que os autores vêm descrevendo a cerca do principal canal de vendas da Loyalty S.A. – o *E-mail Marketing*. Utilizou-se então, de uma série de artigos, os quais se encontram referenciados em detalhe na seção destinado a bibliografia.

Por fim, procurou-se analisar o modelo teórico estatístico utilizado na abordagem do problema deste trabalho. Neste caso, as principais obras referenciadas são “*Regression Analysis and its Applications*” (Gunst e Mason, 1980) e Construção de um Modelo de Regressão para Previsão da Inflação (Mori, 2007).

2.1. Metodologia *Lean*

Gerir uma *Startup* e trabalhar em um ambiente de extrema incerteza não é uma tarefa fácil. Por vezes os empreendedores encontram-se frente a frente com situações que requerem uma forma de abordagem distinta do que propõe a administração tradicional que fala em elaboração de planos de negócios, definição de metas, entre outros.

Steve Blank, autor de “*The Four Steps to the Epyphany*” (2006), e Eric Ries, autor de “*The Lean Startup*” (2013) tratam logo no início das suas obras exemplos de empreendedores que viram seus sonhos de sucesso tornarem-se pesadelos em pouco tempo após o lançamento dos produtos de suas empresas. Blank (2006) apresenta o dado alarmante de que nove em cada dez produtos lançados falham logo nos primeiros dias da interação com seus potenciais clientes.

Ries (2013), por sua vez, descreve o caso dos empreendedores Ricardo Sazima e Virgílio Neves, desenvolvedores de um dos primeiros sistemas de educação a distância via web do Brasil e também, sócios e fundadores de uma *Startup* do setor.

Após um início de sucesso, tendo desenvolvido um sistema de *e-learning* para a EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) que chegou a ser utilizado em escala nacional para capacitação dos trabalhadores rurais, Sazima e Neves decidiram fundar sua *Startup*. Uma vez fundada a empresa, os dois sócios passaram quase um ano desenvolvendo seu produto, empregando além de tempo, recursos financeiros e materiais visando atingir um sistema completo e estável.

Após todo este emprego de tempo, esforço e recursos materiais e financeiros, quando, finalmente, os sócios chegaram ao ponto em que consideravam sua solução “pronta” o suficiente para ser lançada no mercado, tiveram uma grande surpresa: ninguém se interessava em adquirir seu produto.

Esta breve história introdutória chama atenção para algumas questões: Será que de fato, o que se imagina dentro da mente dos analistas, coordenadores, gerentes, diretores e presidentes como sucesso garantido em termos de rentabilidade e retorno de seus produtos realmente vai de encontro ao que esperam seus clientes? Os produtos imaginados estão de acordo com as necessidades dos clientes? Incrementos e desenvolvimentos realmente podem sempre serem considerados avanços? Isto é, o cliente sempre tende a se interessar mais por seu produto a partir deles?

Blank (2006) afirma que um dos principais erros dos empreendedores que os leva ao fracasso está em pensar primeiro no ciclo de desenvolvimento do produto antes de sequer ter desenvolvido seus clientes. O ciclo tradicional de desenvolvimento do produto encontra-se detalhado na figura 1.

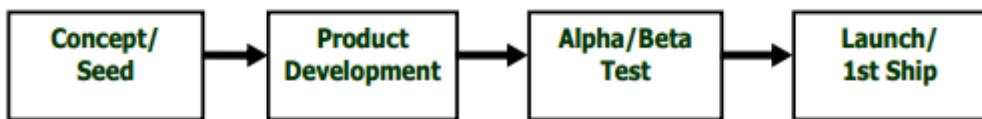


Figura 1: Ciclo tradicional de desenvolvimento do produto (Blank, 2006).

De acordo com Blank (2006) as fases descritas no ciclo da figura anterior são as seguintes:

1. Fase de concepção: nesta etapa os empreendedores desdobram sua visão e princípios para o negócio em ideias, as quais, rapidamente tornam-se um plano de negócio. Em seguida define-se o produto ou serviço-chave da organização, procurando verificar se

ele é possível técnica e financeiramente de ser concebido e quais os recursos necessários para sua concepção.

2. Desenvolvimento do Produto: Nesta fase, iniciam-se os trabalhos de construção do produto propriamente dito. Blank (2006) afirma que todos param de discutir e começam a trabalhar. Cada um dos respectivos departamentos torna-se responsável por uma parte do produto e vai para seu departamento concebê-lo.
 3. Testes A/B: Neste estágio, segundo Blank (2006) os engenheiros trabalham com uma pequena amostra de possíveis usuários de modo a verificar que o produto funciona como planejado (em termos técnicos) e atuam para corrigir seus erros (*bugs*). *Marketing*, por sua vez encarrega-se de desenvolver um completo plano de comunicações.
 4. Lançamento dos Produtos e Primeira Interação com o Cliente: Blank (2006) afirma que é para este momento que toda a empresa se prepara. Com o produto em plena capacidade de funcionamento a empresa vai ao mercado e lança seu produto. E, é a partir daqui que as coisas saem diferente do que foi planejado.

É importante notar que, em nenhum momento no ciclo do Produto houve foco na análise e verificação sobre se as ideias de concepção realmente vão de encontro ao que os clientes esperam, ou o que é ainda pior, não se tem a menor ideia de quem é este cliente e nem por que ele pagaria por este serviço/produto. É justamente neste ponto que a abordagem *Lean* começa a se diferenciar da abordagem tradicional. Blank (2006) enfatiza que primeiro o empreendedor deve se preocupar em conhecer seus clientes e suas respectivas necessidades para só depois procurar conceber seu produto em parceira com estes clientes. Para isso, propõe que a empresa olhe atentamente para o que chama de ciclo de desenvolvimento do cliente, ilustrado a seguir.

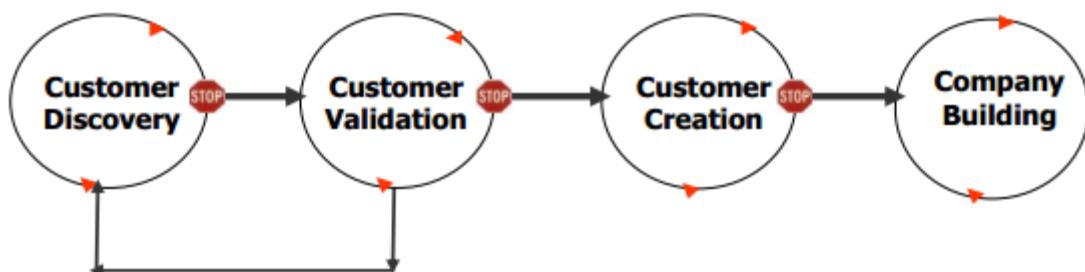


Figura 2: Ciclo de desenvolvimento de clientes de Blank (2006).

Neste ciclo, o empreendedor procura entender primeiro quem são seus verdadeiros clientes para aí sim, e só então, voltar-se para o desenvolvimento do produto, sempre em parceria com estes clientes. De acordo com as definições de Blank (2006), a “Descoberta do Cliente” refere-se a fase em que o empreendedor preocupa-se em entender o consumidor, seus problemas e necessidades; na fase batizada de “Validação do Cliente” o empreendedor preocupa-se em desenvolver um canal de vendas que o permita promoverem seu produto e estar em contato com seu cliente. Na fase de “Criação do Cliente” são desenvolvidos mecanismos para atração de novos clientes. E, por fim em “Construção da Companhia” a *Startup* transita da fase de descobrimento e aprendizagem a despeito de seus clientes para uma fase mais madura em que se divide em departamentos tradicionais e pode iniciar sua preocupação com o desenvolvimento tradicional de seus produtos.

Croll e Yoskovitz (2013) em sua obra “*Lean Analytics*” descrevem um modelo de explicação baseado nas cinco etapas do ciclo de vida dos clientes de modo que se possa extrair todo o valor deles. Valor este, nas palavras dos autores, que provém não só do caráter transacional (vendas), mas também de sua fidelização e promoção. Tal modelo está ilustrado na figura 3.

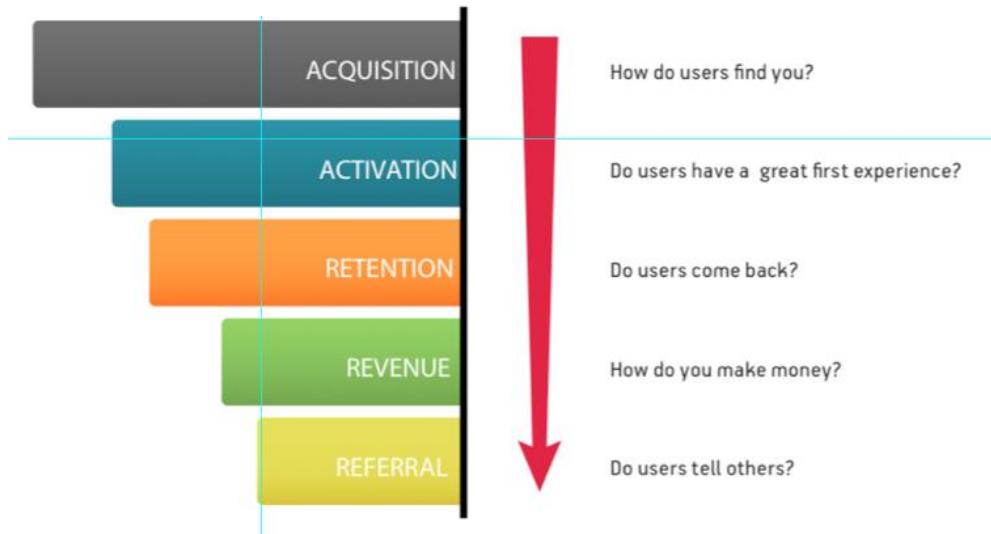


Figura 3: Ciclo de vida dos clientes extraído de autoria de Croll e Yoskovitz (2013).

O detalhe de cada uma das fases está contido a seguir:

- **Aquisição:** Esta fase preocupa-se em criar mecanismos para atração de novos clientes.

- **Ativação:** após a aquisição, o desafio seguinte está em fazer com que os clientes realizem sua primeira transação com a empresa. Estratégias em prol de relacionamento com o objetivo de explicar para o cliente os produtos da organização e como ele pode se beneficiar com isso e, também, realizações de promoções diferenciadas são bem-vindas nesta etapa.
- **Retenção:** A fase seguinte consiste em fazer com o que cliente, após a realização da primeira transação, volte a interagir com a empresa. As ações aqui são mais no sentido de fidelizar este cliente.
- **Revendas:** Esta fase tem o objetivo de potencializar o ganho da empresa a partir das ações de seus clientes. Para se chegar a este etapa é muito importante que a organização tenha uma base de clientes fidelizada ou mesmo um volume de trânsito em seu site que permita explorá-lo por meio de propagandas de terceiros, por exemplo.
- **Referenciamento:** Esta fase é muito interessante e marca o sucesso da organização em termos de sua solidificação no mercado. Aqui, incentivam-se os clientes mais engajados a convidar novos clientes e criam-se mecanismos de recompensa para eles.

Croll e Yoskovitz (2013) apresentam também, o Lean Canvas, o qual trata-se, segundo estes autores, de um processo claro para definir e ajustar o modelo de negócio baseado em desenvolvimento do cliente. Segundo eles, o Lean canvas é um modelo de negócio “vivo” contido em apenas uma página. “Consiste em nove blocos organizados em um pedaço de papel, concebido para fazer com que se reflita acerca de importantes aspectos do negócio”. A figura 4 ilustra este modelo de plano de negócios:

PROBLEM List your top 1-3 problems	SOLUTION Outline a possible solution for each problem	UNIQUE VALUE PROPOSITION Single, clear, compelling message that turns an unaware visitor into an interested prospect	UNFAIR ADVANTAGE Something that can't be easily copied or bought	CUSTOMER SEGMENTS List your target customers and users
EXISTING ALTERNATIVES List how these problems are solved today	KEY METRICS List the key numbers that tell you how your business is doing	HIGH-LEVEL CONCEPT List your X for Y analogy (e.g. YouTube = Flickr for videos)	CHANNELS List your path to customers	EARLY ADOPTERS List the characteristics of your ideal customers
COST STRUCTURE List your fixed and variable costs	REVENUE STREAMS List your sources of revenue			

Figura 4: Lean Canvas de Croll e Yoskovitz (2013)

O Lean Canvas é, segundo os autores, fantástico para que se identifiquem as áreas de maior risco. Sempre que for preciso verificar se está frente a frente com uma real oportunidade de negócios, os autores sugerem que se considere o seguinte:

- **Problema:** Identificaram-se de maneira correta as necessidades das pessoas?
- **Segmentos de Clientes:** Há uma certeza sobre o público alvo? Há a noção clara de como impactá-los de maneira segmentada?
- **Proposição de Valor Única:** Há uma forma clara, distinta e memoriável de explicar o porquê de o negócio ser melhor ou diferente em relação ao mercado?
- **Soluções:** Os problemas estão sendo solucionados da maneira correta?
- **Canais:** Como é feito o contato com os clientes? De que forma podem consumir os produtos/serviços e pagar por isso?
- **Fontes de Receita:** De onde vem o dinheiro? Esta fonte de receitas é recorrente ou pontual?
- **Estrutura de custos:** Quais são os custos associados a prestação do serviço ou fabricação do produto?
- **Métricas:** Tem-se em mente os números corretos para verificar o progresso do negócio?

- **Vantagens:** Qual é a “Força Multiplicadora” que tende a fazer com que os esforços dos empreendedores tenham maior impacto que os dos concorrentes?

Deste modo, utilizou-se o Lean Canvas para o entendimento, por parte do autor, do contexto e do momento do negócio da Loyaty S.A.

Estudando-se com maior profundidade a metodologia Lean, encontram-se seus princípios, apresentados por Ries (2013), que buscam fazer com que emprego de tempo e dinheiro realmente traga retornos. Assim como propôs Blank (2006) e detalharam Croll e Yoskovitz (2013), as ideias de Ries (2013) também colocam o cliente e suas necessidades no centro da discussão de novos produtos, novas empresas, novas soluções ou mesmo produção de incrementos.

A metodologia *Lean* tal qual descrita por Ries (2013) possui cinco princípios, os quais estão descritos a seguir:

- A. **Empreendedores estão por toda parte.** Ries (2013) destaca que qualquer pessoa que trabalha dentro de *Startups* (“instituições humanas projetadas para criar novos produtos e serviços sob condições de extrema incerteza”, Ries (2013)) pode ser considerada um empreendedor.
- B. **Empreender é administrar.** Ries (2013) destaca que uma *Startup* é uma instituição que requer um novo tipo de gestão, especificamente constituída para seu contexto de extrema incerteza.
- C. **Aprendizagem Validada.** Os empreendedores devem estar constantemente testando cada elemento de sua visão, de modo a validar científicamente cada uma de suas hipóteses (Ries, 2013).
- D. **Construir-medir-aprender.** Ries (2013) destaca que a principal atividade de uma *Startup* é transformar ideias em produtos, medir como seus clientes reagem e, então, aprender se é o caso de pivotar (alterar a estratégia e suas ações) ou perseverar. Este conceito trabalha diretamente com o anterior e, ambos, podem ser vistos como o cerne deste trabalho.
- E. **Contabilidade para inovação.** Necessita-se de uma forma de medir o progresso, definir marcos e priorizar o trabalho. Segundo Ries (2013), tudo isso requer um novo conceito de contabilidade focado em *Startups*.

Ries (2013) destaca, ainda, que o objetivo de qualquer *Startup* é “descobrir a coisa certa a criar, a coisa que os clientes querem e pela qual pagarão o mais rápido possível”. Neste sentido, a metodologia *Lean* pode ser vista como uma maneira inovadora de se considerar o desenvolvimento e aperfeiçoamento de produtos com ênfase na interação rápida e constante com os clientes.

Esta definição abre espaço direto para a primeira correlação entre a metodologia *Lean* e o caso em estudo – a Loyalty S.A. Conforme explicado em detalhes nas seções seguintes, esta organização gerencia uma rede de parceiros em que seus clientes podem consumir com “*credits*” (espécie de moeda complementar da Loyalty S.A.) com o objetivo de fidelizá-los. Em outras palavras, como deseja fidelizar seus clientes, esta empresa tem a necessidade de estar em contato com seu cliente e receber feedbacks de maneiras constante, pois, apenas consegue avançar em seu objetivo de fidelização se gerar soluções que vão de encontro a suas necessidades.

Outro fator que aproxima a metodologia *Lean* do contexto em estudo está no elevado grau de incerteza para o qual a primeira foi concebida e que a segunda enfrenta constantemente. Segundo Ries (2013) “qualquer pessoa que está criando um novo produto ou negócio sob condições de extrema incerteza é um empreendedor”. O contexto de *E-mail Marketing* da Loyalty S.A., a medida como está concebido hoje, torna-se um ambiente de extrema incerteza, já que cada uma de suas comunicações de *E-mail Marketing* é quase que concebida a partir do zero.

A metodologia *Lean Startup* apesar de seu nome referenciar diretamente empresas em seu início do processo de desenvolvimento, também se mostra aplicável em um contexto de empresas já consolidadas, com ciclo de vida de seus clientes definido e com produtos desenvolvidos. O parágrafo a seguir ilustra esta proposição.

Como exemplo da aplicação da abordagem da *Lean Startup* no contexto de uma grande empresa Ries (2013) apresenta em seu livro a história da Intuite S.A. (empresa criada com foco em automatizar a contabilidade empresarial, concebida há mais de 20 anos e que conta com cerca de sete mil colaboradores).

No modelo antigo, a organização levava 5,5 anos, em média, para que um novo produto bem sucedido começasse a gerar receita de cerca de 50 milhões de dólares. Hoje em

dia, pós adoção da metodologia *Lean*, segundo Brad Smith (CEO da Intuit), a empresa gerou 50 milhões de dólares em produtos que não existiam há doze meses.

Ries (2013) conclui este trecho deixando a lição de que “o tempo que uma empresa tem de liderança de mercado para explorar suas inovações pioneiras está encolhendo, e isso cria um imperativo para que até mesmo as empresas mais bem sucedidas invistam em inovação”. Ressalta, ainda, que, sob seu ponto de vista, “o único caminho sustentável de uma empresa para o crescimento econômico de longo prazo é desenvolver uma “fábrica de inovação” utilizando as técnicas da *Startup* enxuta para criar inovações radicais continuamente”.

Este exemplo permite que o autor faça um paralelo muito interessante em relação a Loyalty S.A. Em primeiro lugar, o caso da Intuit, mostra como a metodologia *Lean* pode ser aplicada na organização em estudo, com a forte possibilidade de trazer ganhos reais em receita. Ambas são organizações já estabelecidas, porém, com espaço para crescimento baseado em revisão de seus processos e de sua metodologia.

Assim como o Turbo Tax (produto da Intuit) representava uma forte fonte de receita para a Intuit S.A., as comunicações de *E-mail* enviadas aos clientes da Loyalty S.A. deslocam sua curva de receita de maneira considerável. Segundo analistas, o *E-mail* representa hoje cerca de 50% do trânsito de pessoas no site da empresa. Ou seja, assim como se mostrou viável para a melhoria do Turbo Tax, a metodologia *Lean* mostra-se interessante ao processo de *E-mail* da Loyalty S.A.

Vale destacar que a “fábrica de inovação” a que Ries (2013) refere-se tem um paralelo muito próximo ao que se denomina “Laboratório de campanhas” dentro da Loyalty S.A.. Ambos são espaços destinados a produção de conhecimentos via teste de possíveis soluções baseando-se nos conceitos de aprendizagem validada e do ciclo de Realimentação “Construir-medir-aprender”.

Ou seja, esta metodologia, apesar de seu nome “*Lean Startup*” também é muito factível para empresas já consolidadas e tende a potencializar seus resultados.

Pensando-se um pouco mais a fundo num dos conceitos fundamentais da metodologia *Lean* focados neste trabalho, encontra-se a definição de Ries (2013, página 34) para a aprendizagem validada. Segundo o autor “a aprendizagem validada é o processo de demonstrar empiricamente que uma equipe descobriu verdades valiosas acerca da perspectiva

do negócio presentes e futuras [...].” Trabalhar com este conceito-chave quer dizer que todas as propostas de melhorias só podem ser consideradas avanços uma vez que seus protótipos forem testados e tenham aceitação por parte de seus clientes.

Pensando-se no caso do *E-mail Marketing* da Loyalty S.A., todas as hipóteses que venham a melhorar a interação dos clientes com o *E-mail* da empresa e com isso sua experiência com a rede, só podem ser consideradas válidas e prontas para serem implementadas se, efetivamente, provocarem um aumento do interesse dos clientes pelos *E-mails* da organização, medido através da taxa de abertura das ações de *E-mail*. Imagine, por exemplo, o fracasso que seria se supusesse, via modelo teórico ou via análise interna do canal, sem a efetiva interação direta com o cliente, que segmentar as ações de *E-mail* irá melhorar a experiência do cliente com o canal. Imagine agora que fossem empregados tempo e recursos materiais e financeiros visando desenvolver e implementar um modelo de segmentação e, que pós construção, o retorno deste modelo fosse pífio, ou seja, a taxa de abertura não aumentasse. Ou seja, seguindo-se as ideias de Ries (2013) tal modelo de segmentação só seria construído se o fato de segmentar realmente trouxesse ganhos diretos a este canal, o que previne a organização de esforços que não venham a trazer resultados.

O modo de experimentação deve, segundo Ries (2013), ser iniciado através de uma hipótese clara, que prognostica o que pode acontecer. Em seguida, testam-se tais prognósticos (hipóteses) de forma empírica. O objetivo deste experimento *Lean* é descobrir como desenvolver uma melhoria sustentável no produto/ empresa em questão, verificando se os prognósticos devem fazer parte desta melhoria.

Outro tópico muito importante da metodologia *Lean* para este trabalho é o ciclo “Construir-medir-aprender”, exemplificado na figura 5:

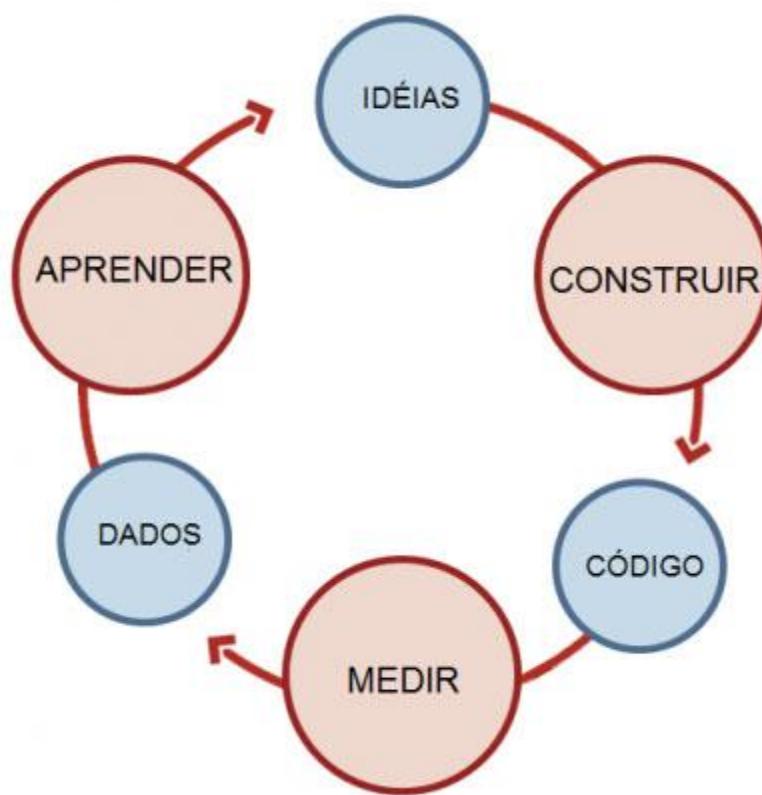


Figura 5: Ciclo “Construir – Medir- Aprender”. Fonte Ries (2013).

O primeiro passo deste ciclo está em construir um MPV (Mínimo Produto Viável) a partir das ideias que pode vir a despertar interesse dos clientes. Um ponto fundamental do MPV é que ele caracteriza-se por ser a versão mínima do produto que permite dar uma iteração completa do ciclo, com o mínimo de esforço e o menor tempo de desenvolvimento possível.

Uma vez prototipado este MPV deve ser colocado no mercado, em contato com os clientes. Deve-se juntamente a essa entrada no mercado criar os indicadores de desempenho utilizados para medir o sucesso deste MPV em contato com o cliente. A despeito das métricas, Croll e Yoskovitz (2013) apresentam algumas regras básicas para o que consideram ser boas métricas (aqueles que tendem a provocar alterações desempenho do negócio):

- **Boas métricas são comparáveis.** Croll e Yoskovitz (2013) afirmam que comparar indicadores atuais com aqueles obtidos em outros períodos de tempo, outro grupo de clientes, ou mesmo com os dos competidores, ajuda os empreendedores a entender qual caminho está sendo seguido e abre a possibilidade de pivotamento de forma mais ágil.

- **Boas métricas devem ser de fácil compreensão:** Segundo Croll e Yoskovitz (2013), quando as pessoas não conseguem lembrar e nem discutir as métricas, torna-se muito difícil transformá-las em motor de mudanças da cultura organizacional.
- **Boas métricas devem ser taxas,** já que taxas são mais fáceis de ser comparadas e de agir-se sobre elas.
- **Boas métricas devem mudar o comportamento da organização.** Esta informação é uma das mais importantes a despeito das métricas. Métricas devem ser concebidas de modo a incentivar mudanças comportamentais a partir de mudanças em seus valores. Caso contrário, a organização está perdendo tempo com métricas que, de fato, não proporcionam nada.

Deve-se colher e analisar os dados relativos a interação do MPV com os clientes chegando a aprendizados (aprendizagem validada) os quais realimentam o ciclo para o aperfeiçoamento do produto.

Traçando uma linearidade com o caso do *E-mail Marketing* da Loyalty S.A. e o projeto em questão, a fase de construção associa-se à modelagem, através da qual se constroem hipóteses (prognósticos) a cerca dos fatores que influenciam a interação com os clientes com o *E-mail*. O próximo passo está em disparar ações seguindo o que determina as hipóteses, de modo a colocar os prognósticos em testes práticos. A medição destes testes é feita pela própria taxa de interação destas campanhas.

Por fim, e mais importante, está o pivô. Uma vez testadas, deve-se analisar os resultados destas campanhas, visando determinar se deve perseverar ou pivotar. Pivotar quer dizer, retornar a fase de reestruturação, redesenhar o modelo e reenviá-lo a fase de testes, dando mais uma volta no ciclo.

2.2. *E-mail Marketing*

O *E-mail Marketing*, classificados por analistas do time interno da Loyalty S.A. como seu principal canal de vendas, é composto por ofertas enviadas ao clientes através de seus *E-mails*. O benefício do *E-mail Marketing* dentro do núcleo de canais digitais (*E-mail*, Redes Sociais, SMS) é que estes podem ser personalizados, no limite, a cada cliente da organização.

Tassabehji e Vakola (2005) afirmam, baseando-se em estudo realizado em 2004 com as organizações do Reino Unido, que o *E-mail* têm tornado-se uma parte fundamental no

trabalho das pessoas, uma vez que praticamente todos os respondentes de sua pesquisa disseram usar o *E-mail* diariamente. A exemplo do que foi citado pelo time de analistas da Loyalty, a respeito da importância do canal *E-mail* para seu negócio, DuFrene et al. (2006) destacam que o *E-mail* têm tornado-se cada vez mais uma importante frente do *Marketing* estratégico de muitas empresas, uma vez que elas utilizam-se deste canal para “notificar novas promoções e serviços; adquirir novos clientes; incrementar vendas; e, o que é mais importante, desenvolver e nutrir um diálogo e relacionamento contínuos com seus clientes”.

Ansari et al. (2001) afirmam que a customização pode formar a base de relacionamentos duradouros em *Marketing*. Ainda segundo estes autores, a comunicação personalizada pode reduzir a sobrecarga de informações e ajudar nas decisões dos clientes, sem contar o fato de que produtos altamente relevantes tendem a melhorar a satisfação dos clientes. A fidelização dos clientes resultante da customização pode gerar aumentos no fluxo de caixa e, consequentemente, lucros a organização.

O problema, segundo afirma O’connor (2008) é que a eficiência do *E-mail Marketing* vêm sendo reduzida ao longo anos devido ao crescimento de mensagens do tipo SPAM, as quais são *E-mails* promocionais não solicitados pelo cliente. Tais comunicações, ainda segundo este autor, têm tornado-se um problema pois, em contraste com métodos tradicionais de comunicação, enviar um *E-mail* adicional tem praticamente um custo marginal zerado, desde que o sistema já encontre-se sedimentado. Este processo de viralização do *E-mail* devido a seu baixíssimo custo foi comentado por Krishnamurthy (2001). Segundo este autor, “do ponto de vista econômico, o custo marginal de enviar *E-mails* adicionais é baixo, custando praticamente o mesmo para um organização enviar 1 mil, 100 mil ou 500 mil *E-mails*.

Como resultado disto, Montague (2006) comenta que os níveis de SPAM circulando pela rede continuam a subir. Ainda segundo este autor, a empresa de segurança da Internet – Postini estimou que, em Novembro de 2006, cerca de 91% de todo o tráfego de *E-mails* pela web correspondia a SPAM. Isso quer dizer cerca de 7 bilhões de mensagens enviadas sem o consentimento dos clientes.

Este cenário torna cada vez mais difícil para que as empresas legítimas utilizem-se do canal *E-mail* como ferramenta de negócio, uma vez que grande parte dos clientes, segundo Rainie e Fallws (2004) encontram-se frustrados pelo elevado volume de mensagens

irrelevantes que vêm recebendo. Baseando-se neste cenário e na importância que o canal *E-mail* tem para a Loyalty S.A., justifica-se a necessidade levantada inicialmente pelo time de campanhas da organização em focar esforços neste canal, procurando inovar e otimizar a relação de seus clientes com seu *E-mail*.

2.3. Análise Estatística

No desenvolvimento deste trabalho, conforme destacado em seções seguintes utilize-se do Modelo Estatístico conhecido como Regressão Linear, como ferramenta de geração das hipóteses a ser testadas em prol da verificação dos principais fatores que influenciam na taxa de abertura dos clientes da Loyalty S.A. Esta seção, descreve em detalhes esta técnica.

Segundo Gunst e Mason (1980), a regressão linear é uma técnica estatística focada no estudo das relações entre uma variável, a variável dependente, e uma ou mais variáveis, as chamadas preditoras ou independentes. Ainda segundo estes autores, está implícita na regressão linear o fato de que o modelo necessariamente endereça como as variáveis preditoras influenciam, descrevem ou controlam a variável resposta. Logo, a relação aplicada é unidirecional, ou seja, despreza-se o fato de a variável dependente exercer influência nas independentes.

Pindyck e Rubinfeld (1976) afirmam que para o estudo da relação entre duas variáveis X e Y, com o objetivo de descrever esta relação estatisticamente, é necessário que se obtenha uma série de observações de cada variável, denominada amostra, e uma hipótese a respeito da relação entre estas variáveis. No modelo de regressão linear, tal relação, conforme o próprio nome já diz é representada por uma reta.

No início do projeto, verificou-se a existência de uma variável que não apresentava-se de maneira clara para que pudesse ser trabalhada num modelo estatístico. Tratava-se do “Assunto inserido as comunicações”. Tal variável nada mais é do que uma sentença, compilada de maneira livre pelo time de campanhas da Loyalty S.A., com o objetivo de incentivar os clientes a interagir com o *E-mail Marketing*. Portanto, antes de que se realizasse quaisquer alterações e adaptações nas variáveis necessitava-se olhar com atenção para este caso em específico.

No que tange a abordagem da situação descrita acima, Minzuno (1993) apresenta um ferramenta denominada Diagrama de Afinidade que, segundo o autor, “esclarece problemas ou situações importantes, cujo estado inicial é confuso, desordenado ou inexplorado.” KArsak

et al. comentam que em coletas de dados verbais sobre determinados problemas, o Diagrama de Afinidade permite o agrupamento dos mesmos em diversos conjuntos seguindo suas afinidades e relações naturais.

Almeida et al (2006) afirmam que o Diagrama de Afinidades trata-se de um processo de agrupamento lento em que é necessário captar a essência de cada uma das ideias, em cada um dos dados. Ainda segundo estes autores, um mesmo evento pode ser definido por diversos nomes a_i ($i = 1$ até n) e que compete a equipe definir o evento e nomeá-lo por um nome padrão.

No trabalho realizado para o tratamento da variável “assunto”, conforme comentado na seção de Modelagem, as várias possibilidades de assuntos convergiram para oito categorias definidas por característica recorrentes nos assuntos, tal qual a categoria Promocional, a qual caracterizou-se pela presença de palavras como Desconto ou Bônus nos assuntos.

Com relação às variáveis categóricas, Gunst e Mason (1980) comentam que frequentemente pesquisadores costumam deparar-se com variáveis não-numéricas, as quais podem ser nominais ou ordinais. Variáveis nominais são aquelas cujas categorias não expressam uma ordem específica. Sexo, estado civil e ocupação são os exemplos citados pelos autores deste tipo de variável. Já as variáveis ordinais são aquelas em que há uma ordem claramente definida entre as categorias. São elas: Classes sociais, escalas de atitude (discorda fortemente, discorda, neutrom, concorda, concorda plenamente), nível de escolaridade, entre outras.

Segundo os autores, tais variáveis categóricas, para que possam ser analisadas segundo o modelo de regressão linear, devem ser transformadas em números. No caso de variáveis do tipo *dummie*, as quais assumem apenas dois valores possíveis, esta transformação é bem simples, bastando associar a cada categoria um número específico. Gunst e Mason (1980) demonstram em sua obra “Regression Analysis and its Application” que o número associado a cada uma das duas categorias não interfere no resultado final da variável dependente.

Porém, ainda segundo Gunst e Mason (1980), “quando uma variável categórica apresenta três ou mais possíveis valores, as dificuldades em definir corretamente a influência desta variável na variável dependente tornam-se mais agudas”. Neste caso, segundo os autores deve ser feita uma transformação binária, de modo a dividir as categorias desta variável em outras variáveis. O exemplo a seguir ilustra esta transformação:

Imagine que uma das variáveis dependentes consideradas num modelo de regressão seja o nível de satisfação dos clientes com a empresa XPTO. Esta variável pode assumir os valores: “Plenamente Satisfeito, Satisfeito, Não Satisfeito, Indiferente”. Ou seja, possui 4 categorias. Visando medir sua influência na variável dependente estudada, deve-se realizar a seguinte transformação binária:

$$n = \log_2 c , \text{ onde:}$$

- n = número de variáveis em que decompõe a variável em questão
- c = número de categorias assumidas pelos dados desta variável.

Logo, a variável em questão será decomposta em duas, da seguinte forma:

Tabela 1: Conversão da variável “Satisfação dos Clientes” em variáveis binárias

Categoria	X1_satisfação	X2_satisfação
Plenamente satisfeito	1	0
Satisfeito	0	1
Não satisfeito	0	0
Indiferente	1	1

Desenvolvido pelo próprio autor

Ou seja, aplicando esta transformação, recai-se no caso em que as variáveis apresentam apenas dois valores possíveis, podendo-se, portanto prosseguir-se com a análise.

Prosseguindo-se com a preparação dos dados para a realização da regressão linear multivariada, o próximo passo, segundo Gunst e Mason (1980) está em eliminar os efeitos de escala. Este passo é importante principalmente quando trabalha-se com variáveis cujos dados encontram-se em escalas muito distintas, uma vez que este fato, segundo os autores, acaba influenciado nos valores dos coeficientes da equação de regressão e conduzindo os pesquisadores a conclusões equivocadas. A transformação executada consiste na aplicação da seguinte função:

$$X_{transf.} = \frac{X - \bar{X}}{\sigma_x},$$

Onde \bar{X} representa a média dos valores assumidos pela variável em questão e σ_x representa o desvio-padrão desta amostra.

De acordo com Mori (2007), o modelo de regressão linear multivariada pressupõe a seguinte equação:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \dots + \beta_n X_n + \varepsilon, \quad (3-1)$$

Em que Y é variável dependente, $X_1, X_2, X_3 \dots X_n$ as variáveis preditoras, α equivale o coeficiente linear, β_n os coeficientes angulares e ε a soma das diferenças entre cada um dos valores de Y observado e os previstos.

Trabalhando-se no nível da série de n observações independentes em que Y associa-se a X_n , o modelo completo fica da seguinte forma:

$$Y_1 = \alpha + \beta_1 X_{11} + \beta_2 X_{21} + \beta_3 X_{31} + \beta_4 X_{41} + \dots + \beta_n X_{n1} + \varepsilon_1, \quad (3-2)$$

$$Y_2 = \alpha + \beta_1 X_{12} + \beta_2 X_{22} + \beta_3 X_{32} + \beta_4 X_{42} + \dots + \beta_n X_{n2} + \varepsilon_2,$$

•

•

•

$$Y_n = \alpha + \beta_1 X_{1n} + \beta_2 X_{2n} + \beta_3 X_{3n} + \beta_4 X_{4n} + \dots + \beta_n X_{nn} + \varepsilon_n,$$

Mori (2007) destaca que o objetivo do problema, está em determinar α e β_n de modo a minimizar o erro ε . A respeito disso Gunst e Mason (1980) destacam que, usualmente, a solução deste problema passa pelo uso do método dos mínimos quadrados, o qual possui como critério de determinação da reta que minimiza o erro (ε), a minimização da soma dos quadrados dos erros. Para uma amostra de n observações, a fórmula que representa este critério esta descrita a seguir:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \alpha - \beta_i X_i)^2 \quad (3-3)$$

A demonstração feita a seguir baseia-se no que descreve Mori (2007) a respeito da Regressão Linear.

Visando facilitar a demonstração do cálculo dos coeficientes linear (α) e angulares (β_i), trabalha-se com a equação no seguinte formato:

$$Y_i = \alpha + \beta X_i + \varepsilon, \quad (3-4)$$

Denomina-se a e b os parâmetros utilizados como estimadores dos coeficientes linear (α) e angular (β), respectivamente, que tendem a minimizar a soma S .

Visando encontrar os valores de a e b deve-se calcular, em primeiro lugar, as derivadas parciais da equação (3-4) em função de α e β , como se segue:

$$\frac{\partial S}{\partial \alpha} = -2 \sum_{i=1}^n (Y_i - \alpha - \beta X_i) \quad (3-5)$$

$$\frac{\partial S}{\partial \beta} = -2 \sum_{i=1}^n X_i (Y_i - \alpha - \beta X_i) \quad (3-6)$$

Já a segunda derivada parcial da equação (3-4) em função, novamente de α e β , é sempre positiva, conforme demonstrado a seguir:

$$\frac{\partial^2 S}{\partial \alpha^2} = \sum_{i=1}^n 2 \quad (3-7)$$

$$\frac{\partial^2 S}{\partial \beta^2} = \sum_{i=1}^n 2 X_i^2 \quad (3-8)$$

Logo, sendo as segundas derivadas parciais sempre positivas, pode-se afirmar que as derivadas parciais de primeira ordem são sempre crescentes. Igualando-se as equações (3-5) e (3-6) a zero, encontra-se o ponto de mínimo global de (3-4), solucionando-se o sistema de equações, tal qual:

$$a = \bar{Y} - b \bar{X} \quad (3-9)$$

$$b = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (3-10)$$

Generalizando-se, agora, a regressão para o caso citado nas equações (3-1) e (3-2) em que há mais de uma variável explicativa e empregando a linguagem matricial, temos:

$$Y = X\beta + \varepsilon, \quad (3-11)$$

Onde:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1n} \\ 1 & X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nn} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix},$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Via cálculo matricial, pode-se determinar o vetor $\hat{\boldsymbol{\beta}}$, o qual contém os estimadores a, b_1, b_2, \dots, b_n , referentes aos parâmetros $\alpha, \beta_1, \beta_2, \dots, \beta_n$. Tal cálculo é feito através da seguinte equação:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (3-12).$$

Uma vez demonstrado o princípio da regressão linear, Mori (2007) afirma que o próximo passo consiste em verificar se o modelo se aproxima de maneira considerável a realidade. Para isso utiliza-se de uma ferramenta conhecida ANOVA (Análise de Variância). Para realizar esta análise define-se os seguintes parâmetros:

- Variância total ou soma dos quadrados total (S_{Total}) é dada pela soma dos quadrados das diferenças entre as observações Y_i e a média \bar{Y} :

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (3-13).$$

- Variância da regressão ou soma dos quadrados da regressão (S_{Reg}) é calculada pela soma dos quadrados das diferenças entre as previsões \hat{Y}_i e a média \bar{Y} :

$$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (3-14).$$

- Variância do erro ou soma de quadrados residual ($S_{residual}$) é dada pela soma dos quadrados do erro:

$$S_{residual} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (3-15).$$

- Quadrado médio da regressão (QM_{reg}) é obtido dividindo-se a soma de quadrados da regressão por seu número de graus de liberdade, dado por k :

$$QM_{reg} = \frac{S_{reg}}{k}$$

- Quadrado médio residual ($QM_{residual}$) é obtido dividindo-se a soma de quadrados residual por seu número de graus de liberdade, dado por $n - k - 1$:

$$QM_{residual} = \frac{S_{residual}}{n-k-1}$$

Por fim, calcula-se o fator F por meio da razão entre QM_{reg} e $QM_{residual}$. A estatística F avalia a hipótese de todos estimadores (exceto α) serem iguais a zero:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_n = 0$$

$$H_1: \exists \text{ pelo menos um } \beta_i \neq 0, \forall 1 < i < k.$$

Se o valor de F calculado for maior que o da distribuição F-snecedor tabelado, rejeita-se H_0 . No software estatístico utilizado para análise,
rejeita – se H_0 sempre que a significância for menor do que 5%.

Usualmente, os resultados do teste de regressão são mostrados na tabela de Análise de Variância, ilustrada a seguir:

Tabela 2: Análise de Variância com mais de uma Variável independente

Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio	F
Regressão	k	$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	$\frac{S_{reg}}{k}$	
Residual	$n - k - 1$	$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	$\frac{S_{residual}}{n - k - 1}$	$\frac{QM_{reg}}{QM_{residual}}$
Total	$n - 1$	$\sum_{i=1}^n (Y_i - \bar{Y})^2$		

Retirado de Mori (2007), página 35

Gunst e Mason (1980) apresentam um coeficiente, denominado R^2 , o qual mede a proporção a variância total explicada pelo modelo de regressão. Mori (2007) afirma que “se a variância da regressão explicar boa parte da variância total do modelo, tendo a variação residual pouca influência, pode-se dizer que o modelo possui um bom ajuste”. Tal coeficiente varia entre zero e um e é dado pela seguinte expressão:

$$R^2 = \frac{S_{reg}}{S_{Total}} = \frac{\text{variação da variável dependente explicada pelo modelo}}{\text{variação total da variável dependente}} \quad (3.16)$$

Mori (2007) comenta ainda que no caso da regressão múltipla deve-se estar atento ao fato de haver uma contribuição sempre positiva ao coeficiente de determinação à medida que se adicionam novas variáveis preditoras ao modelo mesmo no caso de estas não agregarem ao poder explicativo. Neste caso, segundo autor, deve-se trabalhar com o coeficiente $R^2_{ajustado}$, o qual é dado pela seguinte expressão:

$$R^2_{ajustado} = 1 - \frac{n-1}{n-k-1} (1 - R^2) \quad (3.17)$$

O próximo passo, pós-realização da regressão linear consiste em realizar a análise de resíduos. Gunst e Mason (1980) destacam que este passo é importante, pois é através dele que se valida, em caráter teórico, as suposições do modelo construídas pelo modelo de regressão linear, verificando a confiabilidade de seus resultados.

Calcula-se o resíduo (ε_i) pela diferença seguinte expressão:

$$\varepsilon_i = |Y_i - \hat{Y}_i| = |Y_i - \alpha - \beta_i X_i| \quad (3.18)$$

Ainda segundo Gunst e Mason (1980), para que o modelo possa ser considerado apropriado, os resíduos devem apresentar as propriedades homocedasticidade, normalidade e independência. Para verificar se tais princípios estão sendo obedecidos, utiliza-se das seguintes ferramentas:

- A. **Gráfico dos resíduos versus valores ajustados:** Comprova a homocedasticidade do modelo
- B. **Teste de Kolmogorov-Smirnova e Teste de Shapiro-Wilk:** Comprova a normalidade dos resíduos
- C. **Teste de Durbin Watson:** Verifica a independência dos resíduos.

O gráfico dos resíduos versus valores trata-se de um gráfico de dispersão em que em um dos eixos estão alocados os resíduos e no outro os valores ajustados. A verificação da homocedasticidade do modelo é feita por meio de inspeção visual. Para que possa ser considerado homocedástico, os pontos deste gráfico devem apresentar distribuição aleatória, conforme ilustrado na figura 6:

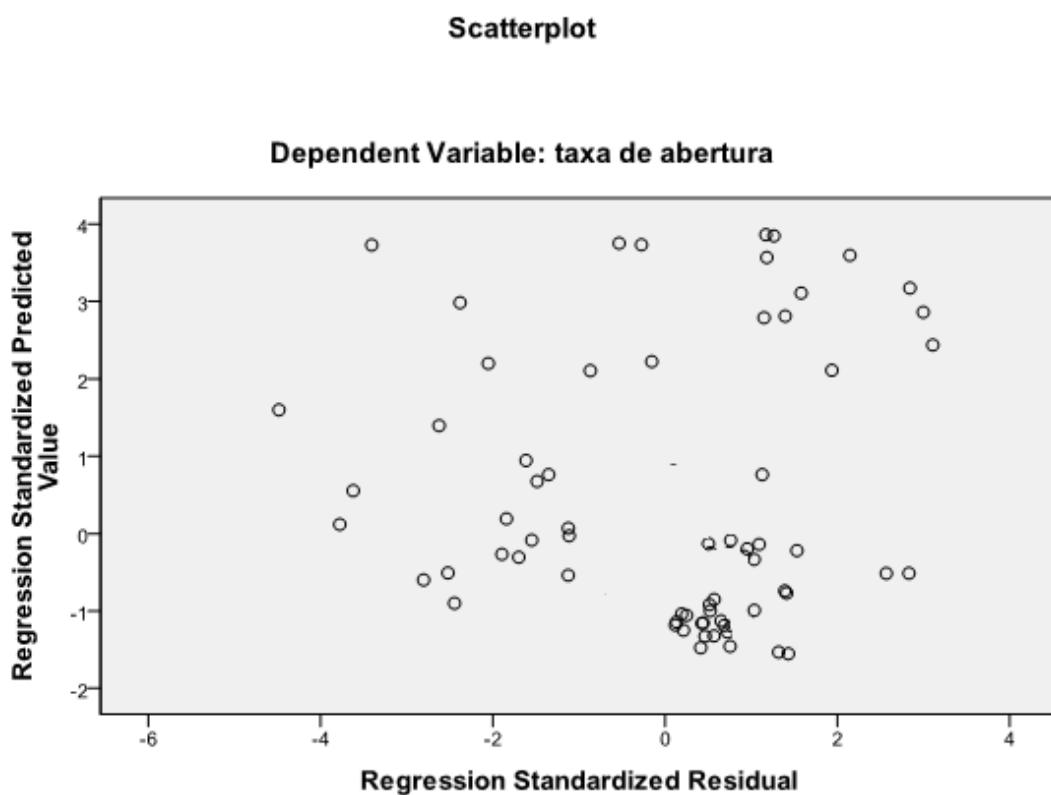


Figura 6: Exemplo de *Scatterplot* para modelo homocedástico.

Os testes de Kolmogorov-Smirnova e de Shapiro-Wilk tem por objetivo determinar a normalidade dos resíduos. O manual de utilização do software estatístico utilizado (PASW Statistics) cita que, para que possam ser considerados normais, os dados submetidos a análise devem apresentar significância superior a 5%. Caso contrário, verifica-se uma diferença significativamente válida da curva normal.

O teste de Durbin-Watson, por sua vez, é utilizado para verificar a independência dos resíduos. Com relação a este teste específico, o software estatístico utilizado não apresenta a significância. Neste caso, ele apresenta sempre um valor entre 1 e 3 para o teste de Durbin Watson. Quando o valor está próximo de 2, significa que o teste não é significativo e que e que os erros são independentes, podendo-se prosseguir com a análise.

Caso alguma das hipóteses acima não seja verificada Matos (1995), sugere que os pesquisados realizem transformações nos dados, procurando inserir variáveis não consideradas na primeira etapa de modelagem ou, mesmo, trabalhando-se com o quadrado ou raiz dos dados dispostos.

Outra técnica utilizada junto a regressão, conforme descrita nas seções seguintes foi a análise de correlações. Gunst e Mason (1980) definem a correlação como sendo uma técnica que leva em conta relações bidirecionais. Enquanto na regressão despreza-se a interferência da variável dependente das independentes na análise de correlações, estas relações são levadas em conta e pode-se analisar a relação de todas as variáveis entre si.

O coeficiente utilizado para determinar a correlação denomina-se coeficiente de correlação Pearson e é calculado da seguinte forma:

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X).Var(Y)}} \quad (3.19)$$

Onde, $Cov(X,Y)$ representa a covariância de X e Y e $Var(x)$ e $Var(Y)$ as variâncias de X e Y respectivamente. Lembrando que, neste caso, X e Y podem ser quaisquer variáveis numéricas.

O coeficiente de Pearson (ρ) assume valores entre -1 e 1, cuja interpretação é a seguinte:

- $\rho = -1$, significa correlação perfeita negativa entre as variáveis.
- $\rho = 0$ indica que não há correlação entre as variáveis.
- $\rho = 1$ revela correlação positiva perfeita entre as variáveis.

Com isso encerra-se este capítulo dedicado a revisão de literatura e inicia-se na próxima seção a descrição da Metodologia de pesquisa deste trabalho.

O software estatístico utilizado para análise foi o PASW *Statistics* 17.0. Segundo seu manual de usuário, a ferramenta trata-se de um sistema completo para a análise de dados. O PASW *Statistics* pode ler dados a partir de praticamente qualquer tipo de arquivo e usá-los para gerar relatórios tabulados, gráficos e parcelas de distribuições e tendências. Além, é claro, de realizar o cálculo de estatísticas descritivas (média, mediana, desvio-padrão e variância) e análises estatísticas mais complexas (modelagem estatística).

No que tange a regressão o software trabalha com o modelo linear (bivariada ou multivariada) e também com modelos mais sofisticados tais quais regressão logística e regressão não-linear.

Capítulo 3: Metodologia de Pesquisa

3. Metodologia de Pesquisa

Este projeto parte do uso da metodologia *Lean*, de Ries (2013) e Blank (2006) e visa institucionalizar o conceito de aprendizagem validada na organização em estudo. Para isto, foi estruturada a metodologia de pesquisa apresentada na figura 7:

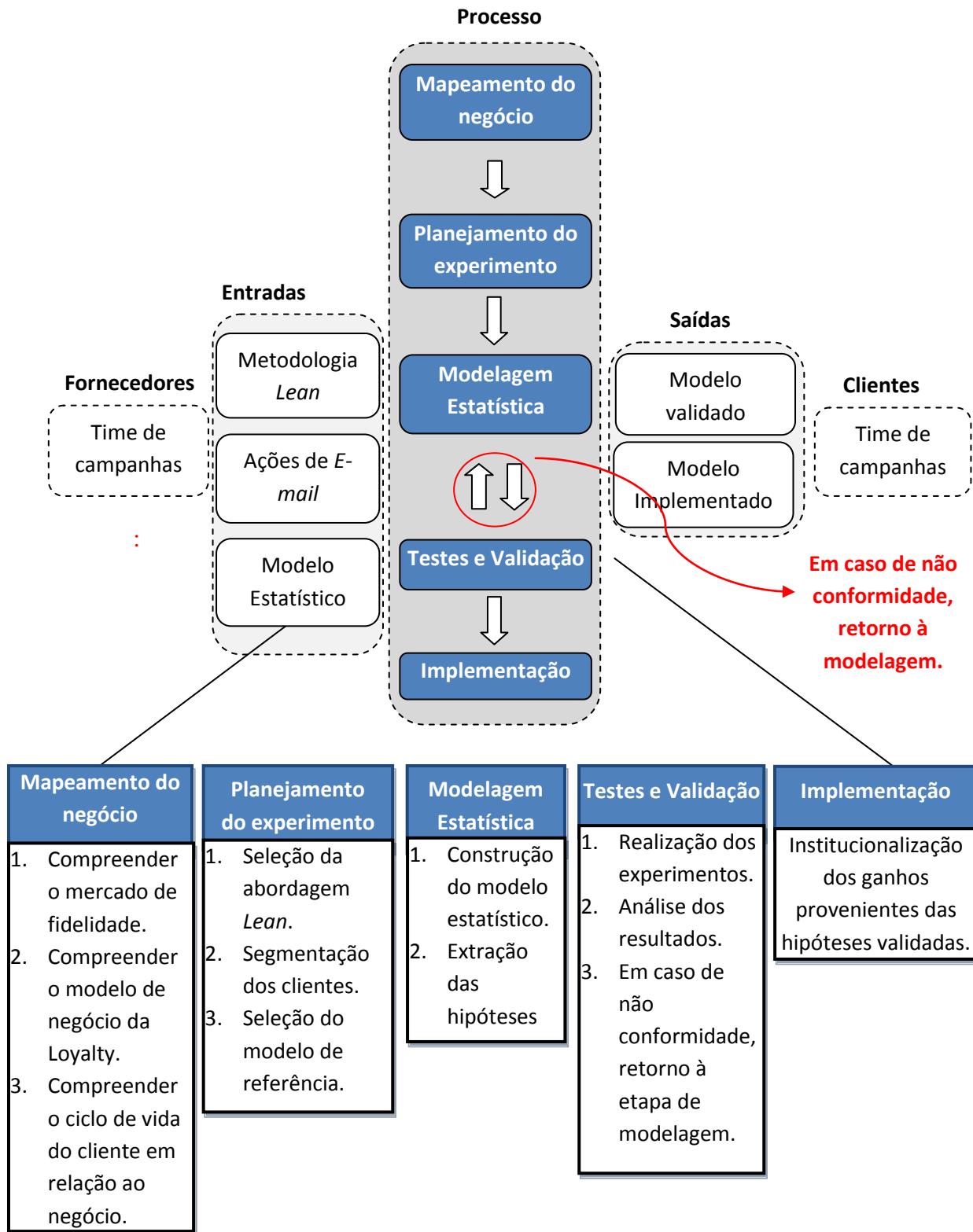


Figura 7: Interligações das macro-étapas da pesquisa

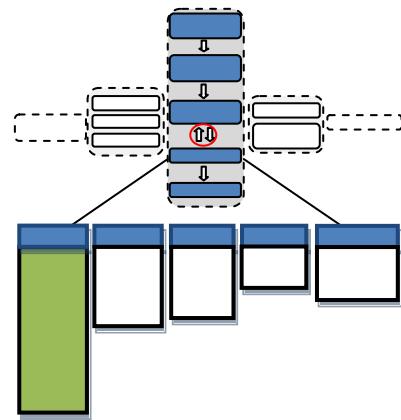
Cada uma das cinco macro-étapas (Mapeamento, Planejamento, Modelagem, Testes e Implementação) teve associada a si etapas de trabalho, as quais deveriam ser completamente concluídas e validadas antes do avanço a macro-etapa seguinte. As descrições de cada uma delas estão a seguir:

3.1. Mapeamento do negócio

A macro-etapa conhecida como mapeamento tratou do início processo. Nela foram realizadas atividades que visaram a compreensão do caso em estudo. São elas:

3.1.1. Compreender o mercado de fidelidade

Esta etapa teve como foco o entendimento do contexto considerado o segmento de mercado que a organização atua. As principais informações que buscadas foram a origem do conceito do mercado de fidelidade, os players e o crescimento da indústria no Brasil.



Ela é importante pois, parafraseando as ideias de Blank (2006) para que se desenvolva uma solução aceitável é necessário que se trabalhe “fora do prédio”, indo “às ruas” para compreender o que de fato os clientes necessitam. Contudo, antes que se possa “dialogar” com os clientes, é fundamental que se tenha conhecimento acerca do mercado de atuação da empresa.

3.1.2. O modelo de negócio da Loyalty S.A.

A etapa “Modelo de negócio da Loyalty S.A.” objetivou o entendimento do modelo de negócio desta organização específica. Para isso, recorreu-se a uma ferramenta que destaca de maneira clara todas as frentes do negócio – o *Lean Canvas*. Além disso, explorou-se a questão da mecânica do negócio sob o ponto de vista dos clientes da Loyalty S.A.

Novamente, seguindo-se as ideias de Blank (2006) decidiu-se compreender o plano de negócio da organização em estudo antes de ir trabalhar em contato com os clientes “fora do prédio”.

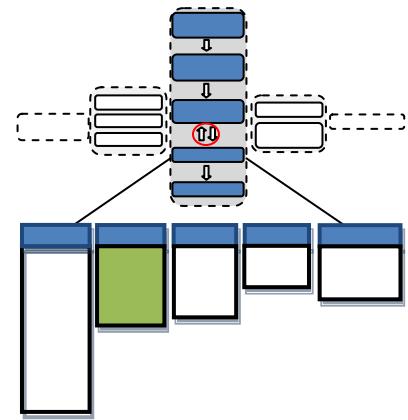
3.1.3. Ciclo de vida dos Clientes da Loyalty S.A.

Esta etapa visou explorar o modelo do ciclo de vida dos clientes dentro da organização baseando-se no conceito de Ciclo de Vida descrito por Croll e Yoskovitz (2013) em seu livro *Lean Analytics*. Ao final desta etapa objetivou-se ter-se em mãos toda a estratégia de ciclo de vida dos clientes, possibilitando, em etapas posteriores, definir, juntamente ao time de analistas internos da organização em estudo em qual etapa o projeto irá concentrar-se.

Seguindo os conceitos de Ries (2013) e Blank (2006) é muito importante que se conheça profundamente o ciclo de vida dos clientes, uma vez que, para aplicar o conceito de aprendizagem validada, necessita-se testar em ambiente prático as hipóteses geradas, o que, só é possível, desde que se tenha em mente o conjunto de clientes abordado.

3.2. Planejamento do experimento

Esta macro-etapa englobou tarefas relacionadas com a seleção da metodologia utilizada para a abordagem do problema e, também em qual momento do ciclo de vida dos clientes que o projeto concentraria seus esforços. Ela foi fundamental, uma vez que as escolhas aqui feitas embasaram todo o método de análise e, consequentemente, os resultados obtidos.



3.2.1. Seleção da abordagem *Lean*

Esta etapa visou selecionar a forma de abordagem escolhida para atacar o problema definido no capítulo 1 deste trabalho. Esta definição foi de suma importância uma vez que, a forma de abordagem pode ser encarada como envoltório dentro do qual toda a metodologia foi desenvolvida.

3.2.2. Segmentação dos clientes

Uma vez definida a forma de abordagem, o passo seguinte foi a escolha de em qual momento do ciclo de vida dos clientes da organização em que o projeto estaria focado.

Esta etapa foi muito importante pois cada um dos momentos do ciclo de vida requer uma abordagem específica, uma vez que os clientes detém objetivos e conhecimentos distintos sobre a organização. Clientes recém cadastrados na organização requerem, por exemplo, uma estratégia de impacto distinta daqueles que estão em pleno engajamento.

3.2.3. Seleção do modelo de referência

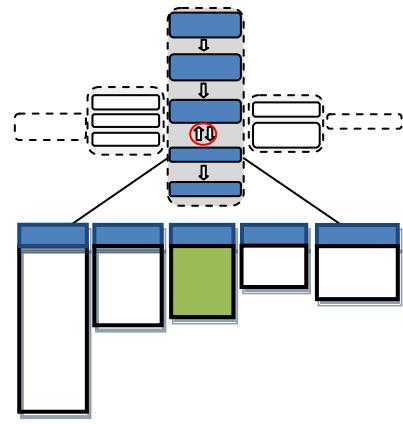
O principal foco desta etapa esteve na definição sobre como o problema iria ser atacado, ou seja, qual o método utilizado para construção das hipóteses a serem testadas em etapa subsequente. Tal seleção levou em conta os seguintes fatores:

- Problema em estudo
- Expectativas referentes a solução
- Revisão literária a cerca do poder do método.

Esta etapa foi fundamental pois é o modelo que produziu os *inputs* das hipóteses que foram testadas em ambiente prático. Ou seja, a escolha de um modelo inadequado tenderia a comprometer todo o processo de geração de ganhos relacionados ao *E-mail Marketing*.

3.3. Modelagem Estatística

Na macro-etapa conhecida como Modelagem estiveram alocadas todas as etapas relacionadas a construção do modelo, desde o tratamento das variáveis, passando pela construção do modelo estatístico propriamente dito e culminando com o levantamento das hipóteses. É importante destacar que, seguindo a metodologia *Lean* de Ries (2013) e Blank (2006), esta etapa trabalhou em conjunto com a etapa subsequente destinada aos testes, num processo de validação ou pivotamento.



3.3.1. Construção do Modelo Estatístico

A primeira atividade desta etapa foi o mapeamento das variáveis a serem utilizadas na explicação da variável em estudo, a taxa de abertura das ações de *E-mail Marketing*. Após este levantamento, levantou-se os dados das campanhas, tendo sido considerado para análise o período Janeiro a Maio/2014.

A tarefa subsequente consistiu em tratar estes dados adequando-os ao que a literatura pressupõe para construção do modelo selecionado, conforme detalhado no capítulo 2 (Transformação binária e padronização).

A seguir realizou-se as atividades relacionadas a construção do modelo propriamente dito. Para isso, utilizou-se da ferramenta estatística “*PASW Statistics*”. Questões relacionadas a validação teórica do modelo também estiveram alocadas a esta etapa.

A importância desta etapa está no fato de que foi nela que trabalhou-se com o modelo propriamente dito, iniciando-se com os dados e chegando aos principais fatores, possíveis de serem testados em etapas subsequentes, em termos de importância para a explicação da variação da taxa de abertura.

3.3.2. Extração das Hipóteses

Nesta etapa objetivou-se, a partir dos resultados obtidos na etapa de modelagem, construir as hipóteses referentes a solução do problema, que foram objeto de teste e validação na etapa subsequente.

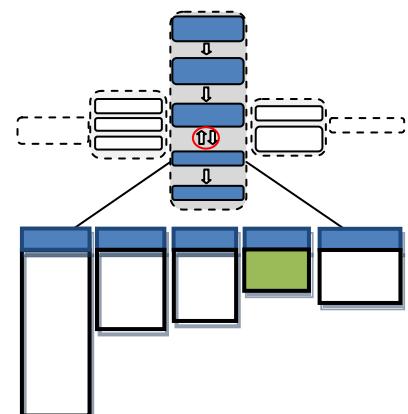
Destaca-se que o objetivo central deste trabalho está em gerar aprendizagem possível de ser testada e confirmada em ambiente prático no laboratório de campanhas, a “Aprendizagem Validada” de Ries (2013). Deste modo, para que se possa construir as hipóteses necessita-se que se detenha, a partir do modelo desenvolvido na macro-etapa de modelagem, os fatores relevantes e também os valores assumidos que, de maneira significante, tendem a potencializar a taxa de abertura das ações de *E-mail Marketing* da Loyalty S.A.

Sendo assim, na extração das hipóteses trabalhou-se com os coeficientes da equação de regressão que mostraram ser significativos estatisticamente e cuja correlação com a variável dependente também mostrou-se significativa. Fatores que não apresentaram correlação significativa que permitiu determinar quais valores assumidos por eles que tendiam a maximizar o fenômeno em estudo foram excluídos da etapa de geração das hipóteses.

3.4. Testes e Validação

A macro-etapa de testes objetivou a verificação do comportamento em ambiente prático (interação com clientes) das hipóteses levantadas na macro-etapa anterior, procurando comprovar se efetivamente geram resultados superiores na taxa de abertura. Ela trabalhou sempre em conjunto com a etapa anterior, pois, em caso de as hipóteses não serem validadas nos testes, pivotou-se (retornou-se a etapa anterior gerando novas hipóteses)

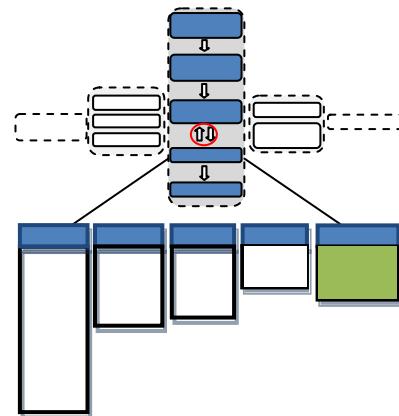
Pode-se ser dizer que nestas duas etapas está o principal da metodologia *Lean* aplicada a este trabalho, pois apenas serão consideradas válidas aquelas hipóteses que



trouxerem incrementos mensuráveis através dos indicadores desenhados na relação cliente - *E-mail Marketing*.

3.5. Implementação

Uma vez validadas em ambiente de testes as propostas de soluções, estas efetivamente transformaram-se em soluções e devem ser documentadas, implementadas e divulgadas. Em termos de documentação, procurou-se registrar a metodologia utilizada e a solução definida, para que estivesse de acordo e pudesse ser incluída no book de conhecimentos da Loyalty S.A.



Pensando-se em implementação, sugeriu-se que o time de campanhas da organização em estudo passasse a construir as campanhas de *E-mail Marketing* levando em conta, quando relevante, a solução proposta por este Trabalho de Formatura.

Capítulo 4: Resultados

4. Resultados

Esta seção apresenta os resultados obtidos com a aplicação da metodologia descrita no capítulo 3, baseada na literatura apresentada no capítulo 2. É nesta seção que efetivamente encontram-se os resultados de todo o trabalho feito.

4.1. Mapeamento do negócio

Esta seção visa realizar um mapeamento do negócio da Loyalty S.A.em termo de seu segmento de atuação, modelo de negócios e, também, ciclo de de seus clientes.

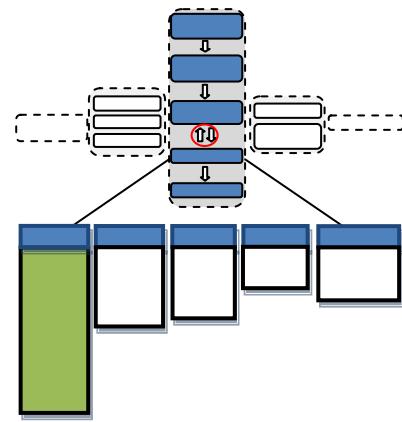
4.1.1. O Mercado de Fidelidade

O conceito de mercado de fidelidade tem suas origens no *Marketing* de relacionamento, o qual pode ser visto como uma frente de *Marketing* que visa não somente um aumento temporário das vendas, mas, segundo Pescador (2004), busca a criação do envolvimento e lealdade de forma a construir um laço permanente com o cliente.

Gronroos (1994) afirma que o *Marketing* de relacionamento tem por função estabelecer, manter e aumentar os relacionamentos com clientes e outros parceiros, de forma lucrativa e de maneira com que os objetivos das partes envolvidas sejam atingidos.

Mckenna (1992) afirma em sua obra “*Marketing* de relacionamento” que o *Marketing* de relacionamento originou-se nas décadas de 1970 e 1980, com o amadurecimento das empresas de serviços. Tal amadurecimento decorre, também, dos avanços tecnológicos na área da informática possibilitando as empresas armazenar e processar informações acerca de seus clientes, vitais para que possam considerá-los de maneira quase que individual, sempre visando sua fidelização.

No Brasil, o mercado de Fidelidade tem uma característica muito peculiar. Segundo, Papadatos (2013), especialista em Programas de Fidelidade do instituto de pesquisa Colloquy, o mercado brasileiro originou-se, em grande parte, dos programas de milhagens das companhia aéreas e, diferentemente do que ocorre em outros países em que cada programa está associado a um banco específico, no Brasil, os principais players Multiplus e Smiles estão associados a quase todos os bancos atuantes no país.



A figura 8 mostra a penetração de alguns programas de fidelidade do mundo em relação a população de cada um destes países. Nota-se a distância entre os principais players brasileiros Multiplus e Smiles , com 6,1% e 4,8%, respectivamente, e o principal player do Reino Unido, a Nectar, com 53,9% de penetração. Ou seja, verifica-se no Brasil a existência de um grande patamar de penetração e crescimento deste mercado, uma vez que, segundo dados divulgados pelo instituto americano Technology Advice, 82,4% dos consumidores preferem fazer compras em lojas com programas de Fidelidade. Esta pesquisa foi feita com base no público americano.

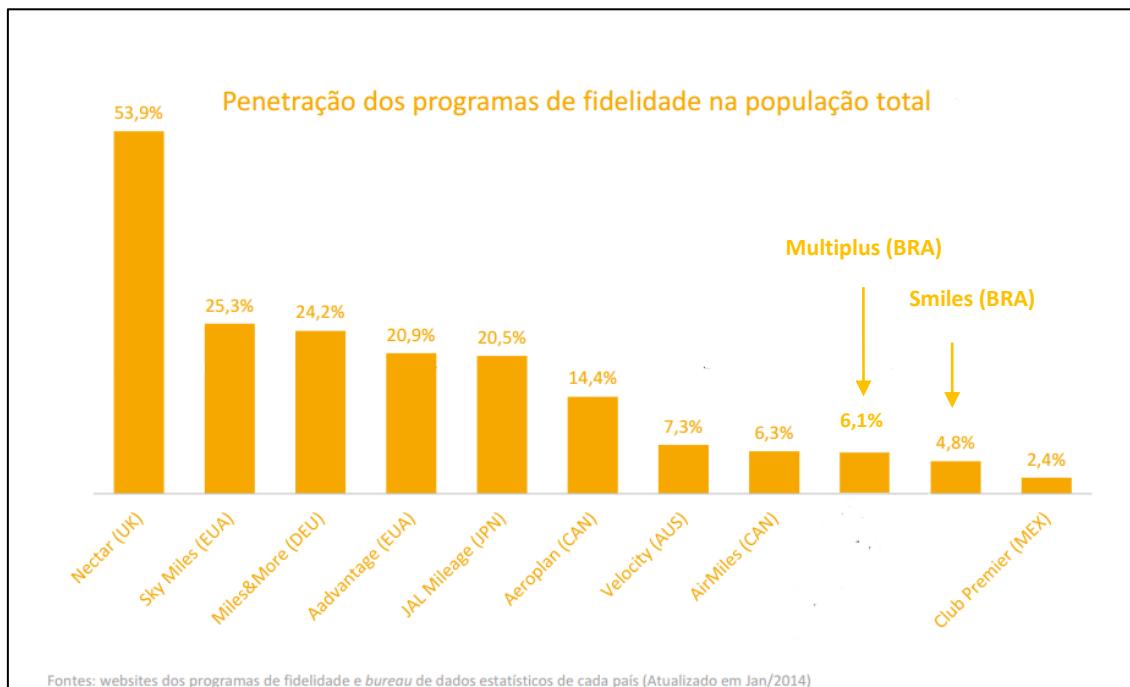


Figura 8: Penetração dos programas de fidelidade na população total (%). Destaque para as brasileiras Multiplus e Smiles. Fonte: Divulgação de Resultados da Multiplus (Ago/2014)

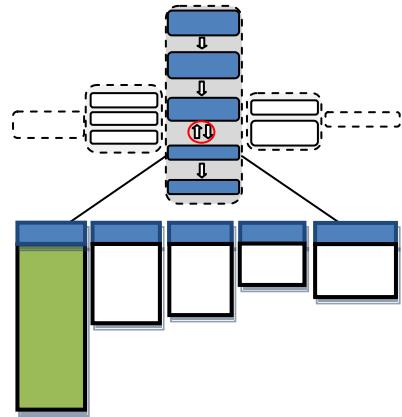
O principal player do Mercado, a Multiplus Fidelidade, originou-se do programa de Fidelidade da TAM Linhas aéreas, tendo sua abertura de capital (IPO, do inglês) sido realizada em 2010. A Multiplus conta hoje, segundo informações divulgadas no último balanço da companhia com cerca de 12,9 milhões de clientes e 477 empresas parceiras, com um faturamento anual em R\$ 2,0 Bilhões em 2013.

A Smiles, segundo maior player do Brasil, é atualmente pertencente ao grupo da Gol Linhas Aéreas. Seu IPO ocorreu em Abril de 2013, influenciado, em grande pelo sucesso obtido pela Multiplus. Segundo dados de Agosto/2014, a companhia conta hoje com cerca de

10 milhões de usuários cadastrados, registrando um faturamento de R\$ 197,4 milhões no segundo trimestre de 2014.

4.1.2. O modelo de negócio da Loyalty S.A.

Conforme apresentado no capítulo 2, o mercado de fidelidade nasce com o objetivo de suprir a necessidade das empresas, principalmente daquelas que adotam um modelo B2C (*business to customer*), de identificar e reter seus melhores clientes. Verifica-se, no modelo de fidelidade, uma forma de gerar vantagens competitivas premiando seus clientes por escolherem determinadas empresas para suprir suas necessidades de consumo.



Neste contexto que surgiu a Loyalty S.A., empresa atuante no mercado de fidelidade no Brasil. Nesta organização a unidade básica, tida como moeda de troca em sua rede de coalizão foi batizada de “*credit*”. O valor em espécie monetária de cada unidade de *credit* está sujeito a diversas variáveis definidas em contratos estabelecidos entre a Loyalty S.A. e seus diversos parceiros.

Do ponto de vista dos clientes desta rede (pessoa física), a mecânica resume-se a três etapas:

- Ao consumir ou adquirir produtos em cada uma das empresas que compõe a rede, o cliente tem a possibilidade de acumular *credits*. Um fator-chave do modelo aplicado é que o cliente não paga nada, além do valor estabelecido na compra do produto ou do serviço desejado; para acumular *credits*, basta informar ao atendente sobre este desejo ao final de cada compra.
- Ao acumular determinado número de *credits* (o qual varia segundo as regras de cada uma das empresas parceiras), o cliente tem a opção de transferi-los para sua conta na Loyalt S.A.
- Uma vez dentro da Loyalt S.A., o cliente pode usar estes “*credits*” como moeda de troca em outros programas de parceiros da rede de coalizão, podendo adquirir prêmios e ser recompensado por sua escolha inicial de consumo.

A fim de explorar e compreender melhor o funcionamento desta organização decidiu-se construir o modelo de negócio *Lean Canvas*, o qual está descrito na figura 4.2. Destaca-se

o uso do *Lean Canvas* ao invés do *Canvas Business Model* convencional, uma vez que o primeiro apresenta-se, na visão do autor, mais robusto em termos de informações levantadas sobre a empresa do que o segundo.

No contexto específico da Loyalty S.A., o *Lean Canvas* apresentou as seguintes diferenças em relação ao *Business Model Canvas* tradicional:

Proposição de valor única: Acredita-se que a proposição de valor da empresa, aquilo que desperta realmente valor e vai de encontro às necessidades de seus clientes pensada de modo único é mais aderente ao modelo de análise do caso da Loyalty S.A, uma vez que ela apresenta uma proposta de valor clara: Reconhecer os clientes por suas escolhas de consumo integrando empresas de diferentes setores.

Adotantes Iniciais: Os adotantes iniciais descritos por Ries (2013) são aqueles clientes que tendem a experimentar e analisar o produto da *Startup* antes de todos os outros nichos de clientes. Pensando-se em termos da Loyalty S.A., segundo analistas internos, os adotantes iniciais tornaram-se os principais clientes da organização, também conhecidos como “*heavy users*”) e representam uma parte importante de seu grupo de clientes, estando, em sua maioria, concentrados hoje, nas estratégias de retenção da organização. Não se pode, portanto, realizar qualquer análise da empresa sem pensar neste grupo.

Conceito de Alto Nível: Permite refletir a respeito do principal conceito de valor que a organização entrega a seus clientes. No contexto da Loyalty S.A. pode-se dizer a organização trata-se de uma gerenciadora do consumo com pontos.

Alternativas existentes: Pensando-se em termos de *Startups*, este fator do *Lean Canvas* objetiva analisar as alternativas de solução já existentes e que tendem a concorrer com o projeto que está sendo concebido. No caso de uma empresa já estruturada, pode-se realizar aqui, uma análise de concorrência e de produtos substitutos. No contexto da Loyalty S.A. verifica-se a possibilidade de trocas dos pontos acumulados em viagens pela Brazilian Airlines S.A. diretamente por novas viagens, sem o intermédio destes pontos na Loyalty S.A., as parcerias bilaterais existentes entre algumas empresas, o oferecimento aos clientes de cupons de vale-desconto e oferta, ainda que restrita, de bens para troca dos pontos como alternativas existentes a rede de fidelização da organização.

Problemas: O *Lean Canvas* permite-se analisar quais os problemas existentes em termos de necessidades dos clientes que os produtos e serviços da empresa visam solucionar.

No contexto da Loyalty, os problemas que a organização visa solucionar passam pela dificuldades das empresas em encontrar uma forma de recompensar seus clientes por terem escolhido consumir seus produtos e, por parte dos clientes, dificuldade em encontrar opções realmente interessante para troca dos pontos que possuem em vários nos programas de fidelidade.

Dentro de todos os blocos de análise do *Lean Canvas* da Loyalty S.A. destacam-se as soluções que esta organização visa entregar a seus clientes e parceiros, as quais são:

- Fornecer às empresas parceiras uma forma de recompensar seus clientes por suas escolhas de consumo;
- Estabelecer uma rede de parceiros de forma a ampliar o poder de compra dos “*credits*”, aumentando assim o poder de compra dos clientes;
- Apresentar aos clientes novas opções em consumo a partir dos “*credits*”, oferecendo as empresas parceiras um novo nicho de mercado a ser explorado.

Para que seu modelo de negócio possa ser considerado viável sob o ponto de vista financeiro é necessário, em primeira instância que a receita associada aos “*credits*” vendidos seja superior ao custo dos “*credits*” utilizados pelos clientes. Esta relação custo benefício está em domínio da Loyalty S.A., uma vez que é ela quem firma por contrato o fator de conversão de “*credits*” em unidade monetária. Tal receita ainda é potencializada pelas aplicações feitas pela organização e também pela “ausência” de custos quando o cliente deixa de usar seus *credits*.

Para monitorar e verificar o comportamento da organização, a empresa desenvolveu indicadores ligados a performance de clientes, detalhados a seguir:

- **percentual de participantes ativos em relação ao tamanho da base:** Este indicador trata do percentual de clientes que tiveram alguma interação transacional com a Loyalty S.A. nos últimos 12 meses em relação a todos os usuários que possuem cadastro na rede. Trata-se de um indicador de performance do poder retenção dos clientes.
- ***credits* em estoque em relação a base total:** Este indicador é construído somando-se o saldo em *credits* de todos os clientes e dividindo-se este número pelo total de usuários cadastrados. Visa verificar o percentual de moeda complementar em estoque da base.

- **volume de *credits* utilizados mensalmente em relação ao total em estoque:** Este indicador revela uma ideia sobre como está a queima de *credits* por parte dos clientes em relação a todos os *credits* disponíveis. Fornece um indicador a respeito da movimentação média dos clientes.
- **volume de *credits* acumulados no mês em relação ao total em estoque:** Este indicador econômico-financeiro permite verificar como se dá a entrada de “*credits*” em relação ao total em estoque. Mostra como está a renovação desta moeda.
- **volume de *credits* expirados mensalmente em relação ao volume em estoque:** Este indicador revela o volume de “*credits*” não utilizados pelos clientes e que expiram-se a cada mês. Lembrando que os *credits* não utilizados expiram-se em três anos.

Para que se possa promover e, também, entrar em contato direto com seus clientes, sejam eles os parceiros ou os clientes da rede, a Loyalty S.A. desenvolveu uma série de canais, dentre os quais destaca-se o *E-mail Marketing*, o qual, segundo dados internos da organização, é responsável por um elevado percentual de movimentação de *credits* (acúmulo e resgate) e que é o tema central deste estudo.

Pensando-se em termos dos segmentos de clientes da organização em estudo, destaca-se, em primeiro lugar, que as próprias empresas parceiras que compõe a rede Loyalty S.A. não deixam de representar um nicho específico de clientes, para os quais, existe a possibilidade de gerar vantagens competitivas frente a seus concorrentes, independentemente da indústria a qual pertencerem, por meio da recompensa (doação de *credits*) a seus clientes por escolherem-nas para realizar seu consumo.

Pensando-se em termos dos clientes, pessoas físicas, pode-se dizer que apresentam-se segmentados em termos de sua experiência em interação com a Loyalty S.A. Não se pode, por exemplo, adotar as mesmas estratégias para os clientes recém-cadastrados e para aqueles que já tiveram um número elevado de interações transacionais com a rede. A seção a seguir descreve o ciclo de vida dos clientes da Loyat.

Uma vez feitas todas estas considerações segue apresentado na figura 9 o *Lean Canvas* para a organização Loyalty S.A:

Loyalty S.A. - Lean Canvas

PROBLEM	SOLUTION	UNIQUE VALUE PROPOSITION	UNFAIR ADVANTAGE	CUSTOMER SEGMENTS
Dificuldade de entendimento de como pode obter vantagens por sempre voar na companhia Brazilian Airlines. Ser reconhecido por sua "fidelidade" ao e-commerce escolhido Dificuldade em converter os pontos que ganha em seu cartão de crédito em prêmios de seu real interesse. Não existência de uma empresa única que gerencia os pontos e permita aos clientes utilizarem em um grande número de locais. Dificuldade na fidelização de clientes e divulgação da marca e geração de novos clientes	Apresentar aos clientes uma rede de parceiros em que os "credits" obtidos podem ser trocados produtos e serviços de diferentes segmentos Oferecer "credits" que podem ser utilizados como moeda de troca em outras empresas Oferecer a possibilidade de converter os pontos dos cartões em "credits" que são universalmente aceitos em todos os parceiros que compõe a rede. Criação e potencialização constante de uma rede que integre empresas de diversos setores.	Aqui na Loyalty você tem a possibilidade de ser reconhecido por suas escolhas de consumo, tendo a disposição um amplo leque de opções em produtos e serviços, sem ter de gastar nada para isso. HIGH-LEVEL CONCEPT Loyalty S.A. = gerenciadora do consumo com pontos	Catálogo de produtos e serviços oferecidos pela empresa Conhecimento sobre os padrões de consumo dos clientes da base	Pessoas que voam pelo menos 1 vez a cada 4 meses na empresa aérea Brazilian Airlines Pessoas com frequência de compra (mais de 3 vezes) nos e-commerces parceiros (artigos esportivos, eletrônicos e eletrodomésticos) Pessoas que tem por hábito a utilização do cartão de crédito dos bancos parceiros
EXISTING ALTERNATIVES	KEY METRICS		CHANNELS	EARLY ADOPTERS
Possibilidade de trocar os pontos de passagem aérea em novas passagens aéreas. Adoção de programas de incentivo como vale-descontos. Oferta restrita de bens para troca dos pontos. Parcerias bilaterais entre algumas empresas.	Clientes ativos em relação a base total Pontos em estoque em relação a base total "credits" utilizados mensalmente em relação ao volume em estoque "Credits" acumulados mensalmente em relação ao volume em estoque Volume de "Credits" expirados a cada mês em relação ao volume de credits em estoque		Aplicativo Mobile Email Marketing Portal Loyalty na WEB Banners em Apps de geolocalização Materiais nos Pontos de Venda dos partners	Pessoas com frequencia mensal de voo na companhia aérea parceira Portadores do cartões black, infinity, platinum dos bancos parceiros.
COST STRUCTURE		REVENUE STREAMS		
Preço pago a empresa parceira associada a compra de produtos e serviços para entrega aos clientes quando estes optarem pelo uso de seus "credits" Despesas associadas ao funcionamento da organização (Administrativas)		Receita atrelada aos pontos vendidos as empresas parceiras, para que estas doem ou convertam os pontos de seus programas de fidelidade em "Credits". Aplicação em fundos do dinheiro obtido por esta venda até a geração de seu custo associado (utilização dos credits) Não utilização dos "credits" pelos clientes da rede.		

Figura 9: *Lean canvas* da organização Loyalty S.A.

Seguindo a metodologia proposta no capítulo 3, o próximo passo, após a definição do modelo de negócio da Loyalty S.A. está na apresentação de seu modelo de ciclo de vida dos clientes. Tal modelo está detalhado na seção a seguir.

4.1.3. Ciclo de vida dos clientes

Por tratar-se de uma empresa em contexto de mercado ainda pouco explorado no Brasil – mercado de empresas cujo objetivo principal é a fidelização de clientes, não existe na literatura um modelo de ciclo de vida completamente desenvolvido considerando o caso específico deste estudo.

Baseando-se na obra “*Lean Analytics*”, de autoria de Croll e Yoskovitz (2013) decidiu-se utilizar o modelo de ciclo de vida, ilustrado na figura 10 e detalhado em seguida. Nesta etapa, contou-se com a colaboração dos analistas relacionados a análises focadas nos clientes da Loyalty S.A.

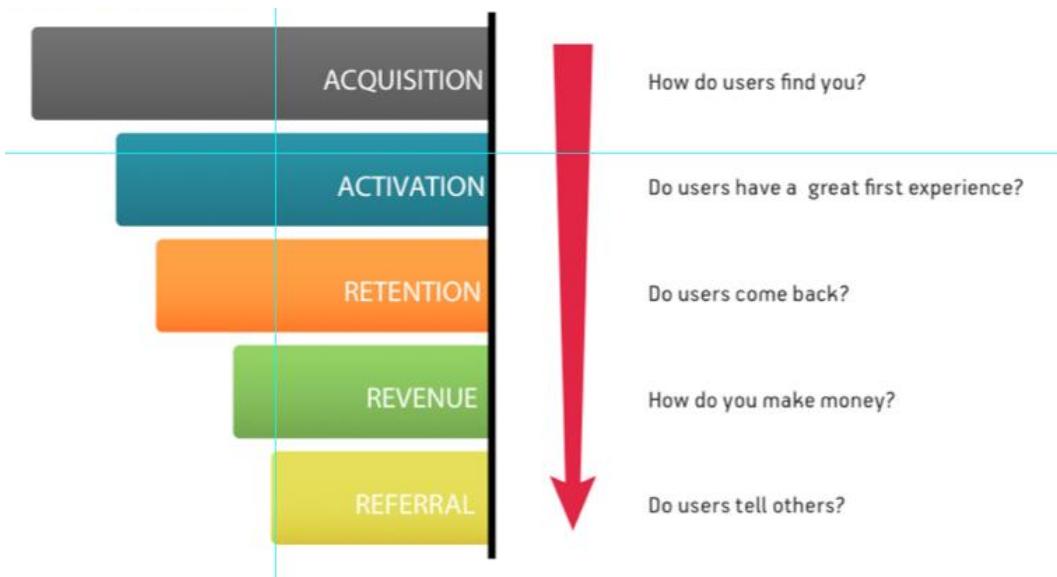
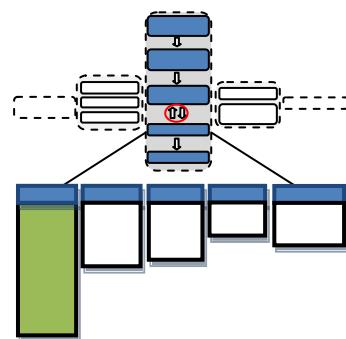


Figura 10: Ciclo de vida dos clientes das Loyalty S.A. baseado no diagrama do ciclo de vida de Croll e Yoskovitz (2013).

A figura anterior apresenta o modelo de ciclo de vida idealizado por Croll e Yoskovitz (2013) e que é adotado no contexto deste trabalho para descrever o ciclo de vida dos clientes da Loyalty S.A.. Conforme comentado na seção destinada a revisão de literatura, cada uma destas etapas requer esforços diferentes por parte da organização em termos de fidelizar e recompensar seus clientes. Nos itens a seguir estão descritos cada um destes momentos:

4.1.3.1. Aquisição

Esta etapa engloba o primeiro momento em que os clientes tornam-se parte da rede. Ao contrário do que parece intuitivo, neste momento, ainda não é certo que os clientes apresentam real convicção sobre os benefícios de estabelecerem relacionamentos com a empresa. A experiência prática mostra que, na maioria dos casos, ao se cadastrarem ou terem

o cadastro efetivado na empresa, os clientes não sabem ao certo a real dimensão ou, o que é ainda mais contraditório, os benefícios que podem obter com a parceria com a empresa.

Há basicamente dois motores responsáveis pela condução do cliente até a empresa em questão:

- O primeiro deles, e que soa mais natural, é o **cadastramento direto**, ou seja, o cliente se interessa pela empresa e ativamente realiza seu cadastro através de seu portal eletrônico.
- O segundo é mediante a realização de alguma compra em alguns dos parceiros que permitem o acúmulo direto de *credits* da rede Loyalty S.A., como é o caso de algumas empresas varejistas que compõe a rede.

Portanto, cada um destes dois segmentos de clientes entra na rede em caráter distinto (os primeiros fizeram seu cadastro diretamente na rede, o que leva a crer que possuem interesse genuíno em conhecer e participar da Loyalty S.A.; já o segundo segmento optou muitas vezes por receber os primeiros *credits* sem deter conhecimento sobre a rede Loyalty S.A.)

4.1.3.2. Ativação

No contexto em estudo, com a absorção de participantes ocorrendo via motores completamente distintos, entende-se que a ativação deve ir além da simples realização de alguma transação, seja ela de acúmulo ou resgate na rede, pois caso contrário, poderia-se chegar a uma situação em que a empresa apresentasse um elevado número de participantes ativos (tendo realizado alguma transação) sendo que grande parte destes efetivamente realizou apenas uma transação de acúmulo via motor de condução dois, sem que, sequer, detenha conhecimento sobre a rede.

Desta forma, decidiu-se, considerar como participante ativo aquele que tenha feito pelo menos 1 transação na rede no último ano. Estabelecido de tal forma, pode-se, portanto, encontrar o cenário em que um cliente ativado na rede, caso fique um ano sem interação com a rede, deixar de ser ativo. Existem aqui uma série de possibilidades e planos de como prever ou interagir com clientes que deixam de interagir com a Loyalty S.A., porém, sua descrição está além do escopo deste trabalho.

4.1.3.3. Retenção

Seguindo-se o modelo do ciclo de vida de Croll e Yoskovitz (2013), a próxima fase denomina-se retenção. Na situação em estudo, acredita-se que pode-se considerar como retido o cliente que realizar mais de 1 interação transacional com a rede no período de 1 ano.

Deve-se, entretanto, levar em conta que o intervalo entre a realização de duas transações consecutivas varia entre os parceiros. É natural que leve um tempo característico, não necessariamente idêntico, para que os clientes acumulem e levem seus “*credits*” a Loyalty S.A. entre os diversos parceiros de acúmulo. O intervalo de transferência de bancos, por exemplo, varia com os gastos que os clientes tem com seus cartões de crédito. Já nos parceiros do varejo, a cada compra pontual há a possibilidade de ganhar-se *credits* na rede.

4.1.3.4. Revenda

A próxima etapa do modelo de ciclo de vida de Croll e Yoskovitz (2013) diz respeito ao potencial de geração de receita dos clientes para a organização.

Vale ressaltar que, como o lastro dos “*credits*” é definido pela Loyalty S.A. mediante acordos bilaterais com cada um de seus parceiros, nem todas as transações de acúmulo de “*credits*” geram a mesma receita. Isso ocorre, também, quando os clientes convertem seus “*credits*” em produtos. Por questões de confidencialidade, não de pode divulgar aqui o ranking em termos de parceiros. Pensando-se nisso, procurou-se dividir os clientes em três categorias descritas a seguir.

- A. Alto Valor: Realizam mais de 1 transação em parceiros estratégicos, aqueles que maximizam o ganho na conta “Preço_venda – preço_utilização” no período de 1 ano.
- B. Médio Valor: Realizam 1 transação em parceiros estratégicos, aqueles que maximizam o ganho na conta “Preço_venda – preço_utilização” no período de 1 ano.
- C. Baixo Valor: Realizam pelo menos 1 transação de acúmulo ou de resgate em parceiro fora do grupo dos estratégicos no período de 1 ano.

Está em desenvolvimento interno na organização uma forma de reconhecimento dos clientes do grupo alto valor e, também, de incentivo aos de baixo valor para que migrem a níveis superiores.

4.1.3.5. Referenciamento

A próxima fase do ciclo de vida denomina-se “*Referenciamento*” e tem como foco incentivar os próprios clientes a divulgarem a marca da empresa e, também, convidarem pessoas de seu ciclo de amizades a tornarem-se clientes.

A Loyalty S.A. utiliza-se, hoje, de uma pesquisa com os grupos de clientes, via canais E-mail Marketing e redes sociais, visando determinar quais já promoveram a marca. Além disso, está em fase de desenvolvimento, um novo produto que vise bonificar os clientes com *credits* por amigos indicados e cadastrados na rede.

4.1.3.6. Perda de Clientes

Apesar do modelo considerado não explorar a questão da perda de clientes (conhecida no mercado como *churn*), entende-se que ela é fundamental a Loyalty S.A., uma vez que esta lida diariamente com grupos de clientes que fazem apenas 1 transação e depois não retornam mais a empresa. Com relação a isso, adota-se duas formas de *churn* dentro da organização:

- *Churn* Transacional: Clientes deste tipo são aqueles que estão há mais de 2 anos sem realizar qualquer tipo de transação com a rede Loyalty S.A..
- *Churn* Completo: Se enquadram neste grupo aqueles participantes que estão há mais de 2 anos sem realizar qualquer tipo de transação com a rede Loyalty S.A. e há mais de 6 meses sem interação em qualquer um dos canais da empresa (site, *E-mail*, redes sociais).

Sendo assim, encerra-se esta macro-etapa inicial de Mapeamento do negócio. Nela estudou-se os principais fatores que compõe a indústria de fidelidade no Brasil, analisou-se os modelo de negócio específico da Loyalty S.A. através do *Lean Canvas* e, também, o ciclo de clientes desta organização. Tais informações foram muito importantes para a continuidade do trabalho.

Ressalta-se que contou-se, para a realização desta macro-etapa com grande participação do time interno da Loyalty S.A. e, também, do professor especialista em aplicação da metodologia *Lean*.

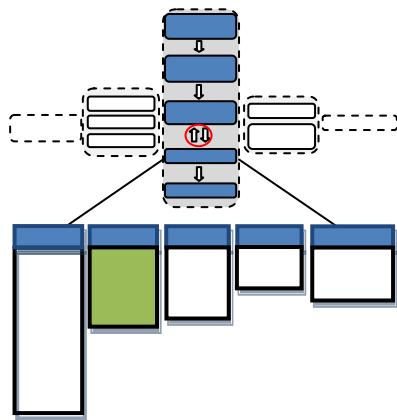
4.2. Planejamento do experimento

Esta etapa visou, baseando-se nas informações levantadas em etapa anterior selecionar qual a abordagem metodológica mais apropriada para o caso em questão e, também, em qual momento do ciclo de vida dos clientes os esforços de melhoria do canal E-mail Marketing devem estar concentrados; além disto, também é selecionado o modelo de análise estatística mais apropriado para a descoberta dos fatores significantes na explicação da taxa de abertura destas ações.

4.2.1. Seleção da abordagem

Lean

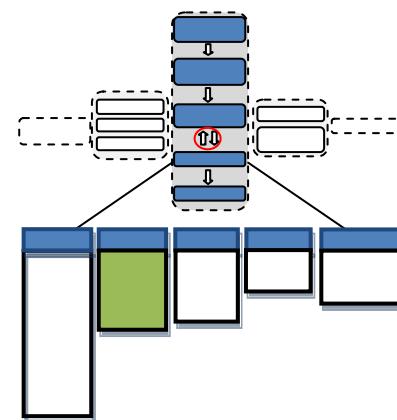
Ries (2013) e Blank (2006) deixam claro que o sucesso no desenvolvimento e melhoria contínua de produtos e serviços decorre da interação correta entre empresa e cliente. Ries (2013) destaca que apenas podem ser consideradas melhorias aquelas modificações que vão de encontro às necessidades dos clientes; fora deste contexto, podem ser vistas como desperdício de tempo e recursos.



A rede Loyalty S.A., com seu foco em fidelização de clientes através do método de recompensas por suas escolhas de consumo, tem, no centro de seu negócio, a necessidade de satisfazer os anseios de seus clientes e dos clientes de seus parceiros. Deste modo, encontrou-se na abordagem *Lean*, de Ries (2013) e Blank (2006), uma metodologia que permite lidar diretamente com clientes da rede Loyalty S.A., idealizando e testando produtos que venham a aumentar o grau de fidelização de sua base de clientes.

4.2.2. Segmentação dos clientes

O primeiro passo na realização deste projeto consiste em determinar em qual momento do ciclo de vida dos clientes da organização a pesquisa estará focada. Deve-se tomar tal cuidado pois, conforme comentado em mais detalhes na seção destinada a revisão de literatura, cada um dos 5 (cinco) elementos categorizados por Croll e Yoskovitz (2013) requerem esforços e ações específicas.



Com o auxílio dos analistas do time de campanha da Loyalty S.A. e partindo-se do princípio que o principal canal da organização está focado, em sua maioria, em promover ações de forma a incentivar os clientes a estarem sempre em interação transacional com a rede gerenciada pela empresa, concluiu-se que os maiores ganhos gerados neste canal serão originados quando o foco estiver nos clientes durante a fase de retenção. A figura a seguir ilustra este direcionamento:

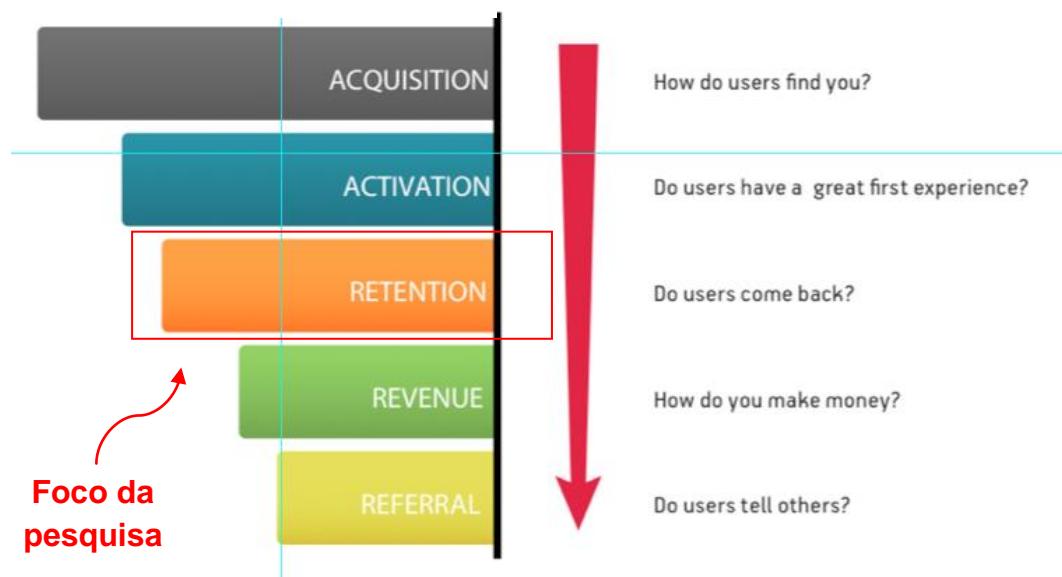
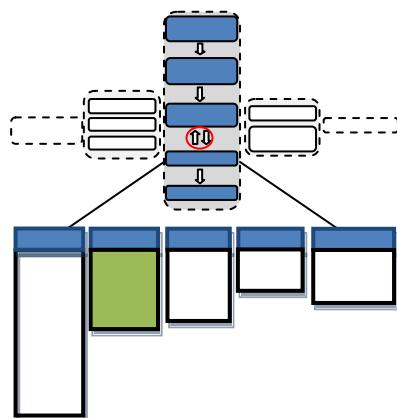


Figura 11: Determinação do foco do estudo baseado no diagrama do ciclo de vida de Croll e Yoskovitz (2013).

Conforme ilustrado na figura anterior, foca-se em ações de E-mail Marketing que visam promover a constante interação transacional entre os clientes e a rede. Para que o objetivo dedicado a este canal seja atingido é necessário, em primeiro lugar, que os clientes tenham uma interação direta com o canal e-mail, medida através da taxa de abertura das campanhas, foco deste estudo.

4.2.3. Seleção do modelo de referência

Uma vez escolhido a metodologia *Lean* e as ações de E-mail Marketing com foco na retenção dos clientes, a fim de atingir o objetivo de gerar hipóteses de melhoria a ser testadas e implementadas, necessita-se escolher um formato



de geração destas hipóteses.

Decidiu-se, a partir de uma análise referente aos indicadores das campanhas, dos dados disponíveis na base da organização, e à necessidade do time de campanhas em inserir modelos estatísticos em suas análises, trabalhar-se com um modelo estatístico para a geração de tais hipóteses.

Procurando-se determinar qual na vasta literatura de modelos estatísticos seria o mais interessante para o caso em questão um professor especialista em estatística aplicada foi consultado. Em reunião com este professor, foi sugerido que o autor tivesse em mente para a escolha do modelo, o tipo característico da variável dependente (assume valores categóricos ou contínuos) e o que deseja-se, de fato, descobrir a respeito desta variável. O especialista sugeriu ainda que o autor considerasse utilizar o modelo de regressão linear.

Após conversa com este especialista em modelos estatísticos, decidiu-se procurar na literatura a respeito do modelo de regressão linear multivariada. Conforme comentando na seção destinada a revisão de literatura, utilizou-se os conceitos propostos por Gunst & Mason (1980), publicadas em sua obra “*Regression Analysis and its Application*”. Segundo estes autores a regressão linear pode ser usada para descrever as relações entre uma variável, a variável dependente, e uma ou mais variáveis, as variáveis preditoras. Ou seja, está de acordo com o primeiro objetivo do caso em estudo que é verificar as relações significativas entre a taxa de abertura e as variáveis independentes, descritas em detalhes nas seções seguintes.

Ainda segundo Gunst & Mason (1980), em algumas situações, como em experimentos de laboratório, os pesquisadores podem selecionar alguns valores das variáveis independentes e então observar a magnitude da resposta da variável dependente para estes valores. Isto está completamente de acordo com a necessidade de integração do modelo com a metodologia *Lean*, utilizando-se dos conceitos de “aprendizagem validada” e ciclo “Construir-medir-aprender”.

Filho e Junior (2009) em seu artigo “Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r)*” afirmam que a correlação é uma ferramenta utilizada para verificar se duas variáveis X e Y possuem algum tipo de relacionamento entre si e que, através dos valores assumidos pelo coeficiente, pode-se verificar se esta correlação é positiva ou negativa. Deste modo, pode-se identificar quais valores assumidos pelas variáveis

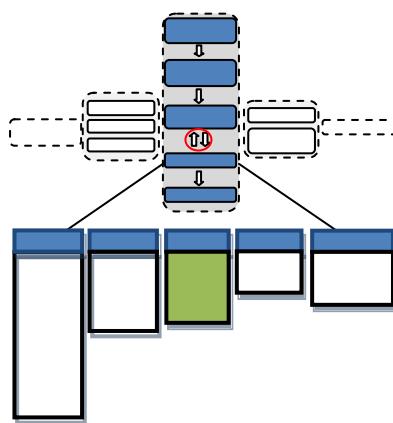
independentes tendem a maximizar a taxa de abertura e, assim, construir as hipóteses a ser testadas em etapas subsequentes.

Logo, ao final desta macro-etapa, escolheu-se a metodologia de análise e, também, o modelo de referência a ser utilizado na geração das hipóteses. Sendo assim, o próximo passo está em construir este modelo, tem abordado nas seções seguintes.

4.3. Modelagem estatística

4.3.1. Construção do modelo estatístico

Nesta seção forma desenvolvidas as atividades referentes à concepção do Modelo Estatístico propriamente dito.



4.3.1.1. Mapeamento das variáveis do modelo

O objetivo específico desta etapa está em mapear dentro do ambiente em estudo quais são as variáveis em que há margem de atuação e que são candidatas a relacionarem-se e interferir diretamente no desempenho da taxa de interação com uma campanha de E-mail Marketing do caso em estudo.

No mapeamento de tais variáveis levou-se em consideração especificamente o conhecimento adquirido pelo time de analistas de campanhas da Loyalty S.A. junto com as sugestões de variáveis destacadas por Croll e Yokovitz na obra Lean Analytics (2013), as quais foram apresentadas em detalhe na seção destinada a revisão de literatura.

Com isso chegou-se as seguintes variáveis em estudo:

- Dia da semana

A determinação do dia da semana de envio das campanhas está em completo controle do time de campanhas da Loyalty S.A. e, também, vêm de encontro ao princípio destacado por Croll e Yokovitz (2013) de que uma variável adequada tende a mudar o comportamento da organização. Caso conclua-se, pós fase de testes que determinados dias tendem a ser mais

assertivos do que outros, abre-se a possibilidade para que a empresa redesenhe sua estratégia de impactos.

Os valores assumidos por esta variável são: “segunda-feira”, “terça-feira”, “quarta-feira”, “quinta-feira”, “sexta-feira”, “sábado” e “domingo”.

- Tipo de comunicação

No plano de comunicações de E-mail Marketing da organização em estudo, há dois tipos básicos de movimentação de “*credits*” e, consequentemente, de movimentação financeira da organização. São elas:

- Junte *Credits*: Estão atribuídas neste grupo aquelas que visam incentivar os participantes da Loyalty S.A. a obterem “*credits*” a partir de suas compras e/ou transferências dos programas de fidelidade das empresas parceiras.
- Use seus *credits*: São comunicações que visam com que o participante use os *credits* acumulados.

- Assunto inserido no título da comunicação

De acordo com Stephanie Miller (2007) da organização especializada em campanhas de *E-mail Marketing* Return Path S.A., um fator preponderante na decisão do cliente em abrir ou não determinado E-mail é o assunto inserido no título da comunicação. Seguindo novamente os princípios de Croll e Yokovitz (2013), o assunto é um fator que está sob completo domínio da organização em estudo, uma vez que é sua equipe interna de Marketing que define qual assunto considera-se o mais adequado e atrativo para cada comunicação.

Utilizando-se de uma ferramenta conhecida como Diagrama de Afinidade descrito em detalhes na seção destinada a revisão de literatura e trabalhando-se em parceira com dois analistas do time de campanhas da Loyalty S.A., chegou-se a oito categorias, descritas a seguir:

- Com informação da peça: Estão presentes nesta categoria, assuntos que contêm informações claras sobre o que está inserido na peça do E-mail Marketing analisado. Considere, por exemplo, uma peça de E-mail focada na venda de um produto específico, por exemplo, um celular. Seu assunto

estaria enquadrado nesta categoria, se nele estivesse contido algo do seguinte modo: “Celular XPTO por X “*credits*””. Ou seja, são assuntos que deixam claro por si só o que o cliente irá ver ao abrir o E-mail.

- Detalhado: Estão presentes nesta categoria, assuntos que passam ao cliente a ideia completa do que se espera que façam. São mais comuns em peças de acúmulo, relatando que o participante pode acumular até X “*credits*” participando das ofertas descritas na mensagem. Diferem dos “Com informação da peça”, pois estes dão a ideia completa do que o cliente deve fazer para ganhar aqueles *credits*, enquanto os “Com informação da peça” apenas trazem a informação do que está contido no Email. Um exemplo deste tipo de assunto é: “Compre no parceiro X e ganhe até X *credits*”
- Com parceiros: Foram inseridos nesta categoria assuntos que fazem alusão ao(s) parceiro(s) ofertado(s) naquele E-mail. Tomando-se o mesmo exemplo destacado acima do E-mail com foco em promover o celular, seu assunto estaria enquadrado nesta categoria caso fosse construído do seguinte modo: “Produtos do Parceiro XPTO destacados para você”.
- Genéricos: Pertencem a esta categoria assuntos genéricos os assuntos que não deixam claro o que o cliente irá encontrar ao abrir a peça e nem fazem quaisquer alusões às características do cliente ou do parceiro em destaque. Ainda considerando o exemplo do celular, seu assunto estaria nesta categoria caso fosse algo parecido com “Ofertas selecionadas para você.”
- Personalizados: Estão nesta categoria, assuntos que trazem consigo alguma característica do cliente, proporcionando a sensação de personalização do E-mail. No exemplo do celular, seu assunto estaria alocado nesta categoria caso fosse do seguinte modo: “Fulano, produtos especiais para você” ou “Oferta dedicada a você que adora eletrônicos”.
- Misto: Foram alocados a esta categoria, assuntos que trazem uma característica do cliente associados ao parceiro ofertado ou que contém informações sobre o(s) produto(s) contido(s) na peça do E-mail. Ainda

tomando-se como exemplo o celular, seu assunto estaria nessa categoria caso fosse do seguinte modo: “Fulano, celular exclusivo para você por X “*credits*” ou “Fulano, confira as ofertas do parceiro X”.

- Promocionais: Estão enquadrados nesta categoria, assuntos que trazem ou fazem alguma alusão a alguma condição exclusiva e/ou por tempo determinado. Tais assuntos são mais comumente utilizados em comunicações de promoções que oferecem bônus ao cliente associados ao acúmulo de “*credits*”. Um exemplo deste tipo de assunto é o seguinte: “OFERTA EXCLUSIVA: Ganhe X% de bônus ao transferir seus “*credits*” do parceiro X”.
 - Promocionais – Urgência: Foram incluídos nesta categoria, assuntos promocionais que procuram destacar o caráter de últimos dias de uma promoção em destaque, procurando incentivar o cliente a realizar sua transação antes que a promoção acabe. Um exemplo deste tipo de assunto é o seguinte: “ÚLTIMOS DIAS: Ganhe % de bônus ao transferir seus “*credits*””.
- Saldo médio da base

Há uma percepção interna na organização de que o saldo médio da base impacta diretamente na taxa de abertura das mensagens. Acredita-se, a priori, que ações focadas no segmento da base com maior saldo tendem a gerar mais interação do que aquelas dedicadas à base com menor saldo. Deseja-se, portanto, verificar se esta percepção é estatisticamente válida, se realmente impacta e qual o potencial deste impacto nas taxas de abertura das ações de E-mail marketing da organização em estudo.

- Segmentação utilizada

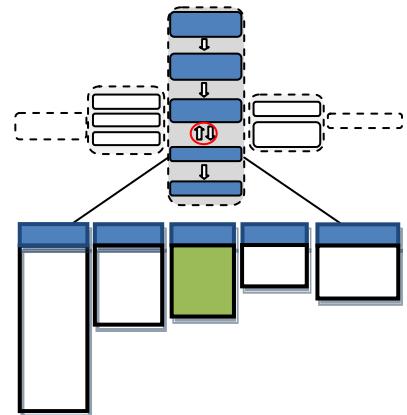
Deseja-se verificar até que ponto o esforço em segmentar as comunicações realmente traz como resultado um incremento na taxa de abertura que seja estatisticamente significativo. A variável aqui utilizada denomina-se percentual impactado, calculado dividindo-se o tamanho da base impactada pelo tamanho total da base de clientes da empresa.

- Comportamento transacional no parceiro do Email Marketing

Por fim, o último fator mapeado para análise estatística procura medir o quanto relevante é o fato de haver uma experiência transacional passada no parceiro destacado no E-mail para a abertura de uma determinada ação. A variável aqui utilizada é o volume de clientes com experiência passada no parceiro em relação ao total da base impactada.

4.3.1.2. Os dados

Os dados coletados combinam informações de “retorno” (variável dependente) e de “planejamento” (variáveis independentes) das ações de E-mail Marketing da empresa em estudo durante os meses de Janeiro a Maio de 2014.



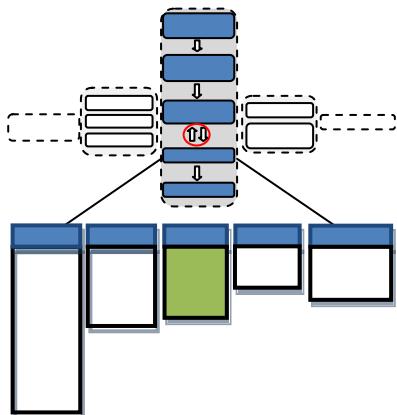
Dispõe-se, no total, de dados referentes a 255 campanhas. É importante destacar que todas possuem valores para todas as 6 variáveis destacadamente em todas estas 255 ações de E-mail Marketing.

Realizou-se, também, nestes dados uma busca por *Outliers* (dados considerados “estranhos” em relação aos demais), a qual não trouxe resultados significativos.

4.3.1.3. Transformação das variáveis

Conforme citado as variáveis “Dia da Semana”, “Tipo de comunicação” e “Tipo de Assunto” são categóricas. Para realizar-se a regressão linear multivariada, necessita-se, portanto, realizar a transformação binária descrita em detalhes na seção de revisão de literatura. Detalha-se abaixo a transformação feita para a variável “dia da Semana”. As variáveis “Tipo de comunicação” e “Tipo de Assunto” tem suas transformações detalhadas na seção anexos deste trabalho.

Dia da semana é uma variável categórica que pode assumir sete valores: “Segunda, Terça, Quarta, Quinta, Sexta, Sábado e Domingo”. Aplicando-se a expressão matemática logarítmica ($\log(7;2)$) obtém o valor: 2,81. Realizando-se o arredondamento para o inteiro imediatamente superior, descobre-se que a variável “Dia da Semana” deve ser desdobrada em três outras variáveis, aqui



denominadas: “X1_dia_semana”, “X2_dia_semana”, “X3_dia_semana”.

A transformação dos valores assumidos pela variável “Dia da Semana” nas variáveis auxiliares está descrita na seguinte tabela:

Tabela 3: Transformação dos valores da variável “Dia da Semana”

	X1_dia_semana	X2_dia_semana	X3_dia_semana
Domingo	1	0	0
Segunda-feira	0	1	0
Terça-feira	0	0	1
Quarta-feira	1	1	0
Quinta-feira	1	0	1
Sexta-feira	0	1	1
Sábado	1	1	1

Desenvolvido pelo próprio autor

Com isso, tomando-se como exemplo a categoria “Quarta-feira”, pode-se dizer que esta é representada unicamente pela trinca (1;1;0). É importante ressaltar que a cada uma das categorias está associada uma única trinca e vice-versa.

4.3.1.4. Padronização dos valores

Após a transformação binária das variáveis categóricas, chegou-se a um cenário em que se trabalha com uma variável dependente (taxa_abertura) e 11 variáveis numéricas procurando explicar a variação da variável dependente. Na tabela a seguir estão contidas os possíveis valores assumidos por todas estas 12 (doze) variáveis:

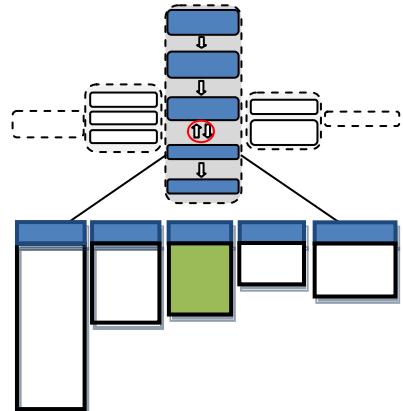


Tabela 4: Tipos das variáveis após a transformação binária das categóricas

Variável	Tipo	Valores assumidos
X1_dia_semana	Independente	Dummie (0 ou 1)
X2_dia_semana	Independente	Dummie (0 ou 1)
X3_dia_semana	Independente	Dummie (0 ou 1)
X1_tipo_comunicação	Independente	Dummie (0 ou 1)
X2_tipo_comunicação	Independente	Dummie (0 ou 1)
X1_assunto	Independente	Dummie (0 ou 1)
X2_assunto	Independente	Dummie (0 ou 1)
X3_assunto	Independente	Dummie (0 ou 1)
Saldo médio	Independente	Valores naturais
Segmentação utilizada	Independente	Percentual
Percentual transacional	Independente	Percentual
Taxa de abertura (y)	Dependente (Y)	Percentual

Desenvolvido pelo próprio autor

Seguindo o que foi destacado na seção destinada a revisão de literatura, para que se obtenha um resultado em que os coeficientes de cada uma das variáveis independentes obtidos na equação da reta após aplicação do modelo de regressão linear multivariada realmente possam ser traduzidos como a importância (peso) daquela variável no comportamento da variação da variável dependente deve-se eliminar os efeitos de escala, associados as ordens de grandeza de cada uma das variáveis.

Visando, portanto, eliminar tais efeitos, aplica-se o modelo de normalização a cada uma destas variáveis. A normalização consistiu na aplicação da seguinte função a cada um dos valores assumidos pelas variáveis:

$$X_{transf.} = \frac{X - \bar{X}}{\sigma_x},$$

Em que, conforme descrito na seção destinada a revisão de literatura, X refere-se ao valor a ser transformado, \bar{X} a média de todos os valores assumidos pela variável em transformação e σ_x ao desvio-padrão destes valores.

A seguir estão destacados a padronização das variáveis “X1_dia_semana”, “X2_dia_semana” e “X3_dia_semana”. As demais padronizações estão descritas em detalhe na seção anexos.

a. Padronização dos valores da variável X1_dia_semana

Conforme comentado na seção de revisão de literatura, a padronização adotada é a normalização. Para o caso específico da variável X1_dia_semana, consiste em aplicar a cada um dos 255 valores assumidos por esta variável a seguinte formula:

$$X_{1_{dia_semana_tranc.}} = \frac{X_{1_dia_semana} - 0,5400}{0,4988},$$

O quadro a seguir nos traz um exemplo com 10 valores associados a X1_dia_semana que sofreram tal transformação:



$X_{1_dia_semana}$	$X_{1_dia_semana_tranc.}$
1,0000	0,9117
0,0000	-1,0925
0,0000	-1,0925
0,0000	-1,0925
1,0000	0,9117
1,0000	0,9117
1,0000	0,9117
1,0000	0,9117
0,0000	-1,0925

Quadro 1: Transformação da variável X1_dia_semana

b. Padronização dos valores da variável X2_dia_semana

O mesmo processo também foi adotado na transformação da variável X2_dia_semana. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:

$$X_{2_{dia_semana_tranc.}} = \frac{X_{2_dia_semana} - 0,7411}{0,4388},$$

O quadro a seguir nos traz um exemplo com 10 valores associados a X2_dia_semana que sofreram tal transformação:

$X_{2_dia_semana}$	$X_{2_dia_semana_tranf.}$
0,0000	-1,6889
1,0000	0,5898
0,0000	-1,6889
0,0000	-1,6889
1,0000	0,5898
1,0000	0,5898
1,0000	0,5898
1,0000	0,5898
1,0000	0,5898

Quadro 2: Transformação da variável X2_dia_semana

c. Padronização dos valores da variável X3_dia_semana

Repete-se, mais uma vez, este processo para a variável X3_dia_semana. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:

$$X_{3_{dia_semana_tranf.}} = \frac{X_{3_dia_semana} - 0,4784}{0,5005},$$

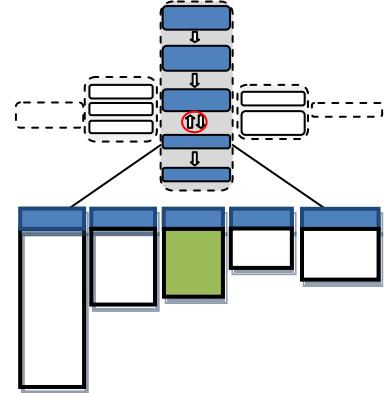
O quadro tabela a seguir nos traz um exemplo com 10 valores associados a X3_dia_semana que sofreram tal transformação:

$X_{3_dia_semana}$	$X_{3_dia_semana_tranf.}$
1,0000	1,0421
0,0000	-0,9559
1,0000	1,0421
1,0000	1,0421
0,0000	-0,9559
0,0000	-0,9559
0,0000	-0,9559
0,0000	-0,9559
0,0000	-0,9559
1,0000	1,0421

Quadro 3: Transformação da variável X3_dia_semana

4.3.2. A regressão linear

Uma vez transformados e padronizados, os dados encontram-se prontos para serem analisados estatisticamente com aplicação da técnica de regressão linear multivariada. Aqui estão descritos diretamente os resultados obtidos após a aplicação da técnica. O passo a passo em detalhes pode ser encontrado na seção anexos.



4.3.2.1. Poder de explicação do modelo

Segundo Johnson e Wichern (2002), o poder de explicação do modelo está associado aos valores assumidos pelos parâmetros R-quadrado (*R square*) e R quadrado ajustado (*Adjusted R square*). A tabela 5 apresenta os valores assumidos por estes parâmetros nesta análise:

Tabela 5: Valores de R e R^2 no modelo em estudo.

R quadrado (R^2)	R quadrado ajustado (R^2 ajustado)
0,680	0,667

Desenvolvido pelo próprio autor

Logo, o modelo desenvolvido é capaz de explicar cerca de 68% da variância da variável em estudo – taxa de abertura, o que pode ser considerado satisfatório.

4.3.2.2. Coeficientes da equação da regressão

Os coeficientes da regressão (peso das variáveis na explicação da variação da taxa de abertura) e sua respectiva significância para o modelo de regressão estão expostos na tabela 6 a seguir. Chegou-se a tais coeficientes mediante realização da análise de regressão linear realizada no software PASW *Statistics*. Todos os passos realizados no software para se chegar a estes valores estão descritos na seção anexos.

Tabela 6: Coeficiente e sua respectiva significância para o modelo 1.

Modelo	Variável	Coeficiente	Significância
1	(Constante)	3,846 E -9	1,000
	X1_dia_semana	-0,033	0,386
	X2_dia_semana	0,039	0,403
	X3_dia_semana	0,016	0,724
	X2_tipo_comunicação	0,063	0,157
	X1_assunto	0,065	0,094
	X2_assunto	0,100	0,023
	X3_assunto	0,154	0,001
	Saldo médio	0,037	0,429
	Percentual impactado	-0,393	0,000
	Grau transacional	0,632	0,000

Desenvolvido pelo próprio autor

Conforme comentado na seção destinada a revisão de literatura, deve-se verificar a significância estatística para determinar se a variável exerce influência considerável na variação da taxa de abertura. Para aquelas que mostrarem-se significantes, deve-se medir seu poder de influencia pelo módulo de seu coeficiente B.

Deste modo, as variáveis em questão são, nesta ordem: “Grau transacional, Percentual impactado, X3_assunto e X2_assunto”.

Retomando-se a seção de revisão de literatura, temos a citação de Moore (2007) que afirma que “a correlação mensura a direção e o grau de relação linear entre duas variáveis quantitativas”. Mediante análise realizada no software PASW *Statistics*, contida em detalhe na seção anexos, chegou-se aos seguintes valores para as variáveis com poder de explicação da taxa de abertura significativo:

Tabela 7: Correlações de Pearson entre a taxa de Abertura e as variáveis independentes com coeficiente significante

Variável	Correlação de Pearson	Sig. (1-tailed)
X2_assunto	0,131	0,019
X3_assunto	0,108	0,042
Percentual impactado	-0,480	0,000
Grau Transacional	0,701	0,000

Desenvolvido pelo próprio autor

Pode-se extrair da tabela 7 que as variáveis que possuem correlação estatisticamente válida (significância de 5%), isto é, aquelas, que possuem um relacionamento estatisticamente válido para que se possa extrair os valores que tendem a maximizar a taxa de abertura, visando os testes em etapas subsequentes são:

- Grau transacional: Correlação positiva (+)
- Percentual impactado: Correlação negativa (-)
- X3_assunto: Correlação positiva (+)
- X2_assunto: Correlação positiva (+)

4.3.2.3. Análise de resíduos

Conforme descrito na seção destinada a revisão de literatura, segundo Gunst e Mason (1980), é necessário realizar a análise de resíduo de modo a verificar se os resultados da regressão são confiáveis. Os principais fatores a serem analisados são:

- Normalidade dos resíduos.
- Homoscedasticidade do modelo, isto é, se a variância dos dados é constante.
- Independência dos resíduos.

Ressalta-se que todas estas condições devem ser atendidas para que o modelo seja válido de acordo com os preceitos teóricos.

Em primeiro lugar, realizou-se os testes de Kolmogorov-Smirnov e Shapiro-Wilk, visando testar se os resíduos apresentam distribuição normal. Para que sejam considerados normais, a significância destes testes deve ser maior do que 5%, uma vez que, para valores inferiores a 5%, pode-se interpretar que os dados apresentam uma diferença significativamente considerável em relação a curva normal.

A tabela 8 apresenta os resultados destes testes:

Tabela 8: Resultados dos testes de normalidade para o modelo 1

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
<i>Unstandardized Residual</i>	0,123	255	0,000	0,911	255	0,000

Desenvolvido pelo próprio autor

De acordo com este teste de normalidade, pode-se concluir que os resíduos não apresentam distribuição normal, o que, segundo a literatura, fere um dos princípios teóricos da regressão linear citados anteriormente.

Neste ponto, procurou- se novamente o professor especialista em modelos estatísticos para análise do caso em questão. De acordo com o especialista verifica-se situação semelhante em muitos casos práticos de aplicação de modelos estatísticos teóricos. Ainda segundo o especialista, nem sempre, todas as considerações teóricas podem ser verificadas a risca em ambiente prático.

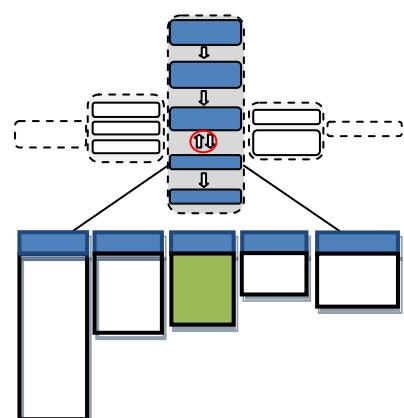
Deste modo, o especialista sugeriu que o autor prosseguisse com os testes da regressão, construindo o modelo e testando os resultados obtidos. Segundo ele, como o autor tomou o cuidado de testar as hipóteses do modelo antes de considerá-las como verdade no contexto em estudo, fatalmente se o modelo tiver produzido resultados não condizentes com a realidade, estes não serão aprovados na etapa dos testes

A análise dos outros parâmetros, bem como o passo a passo da análise de resíduos, estão descritos em detalhes na seção anexos.

4.3.3. Extração das hipóteses

Juntando-se as informações relativas aos coeficientes de correlação de Pearson e dos coeficientes significativos do modelo de regressão, extraí-se, as seguintes Hipóteses:

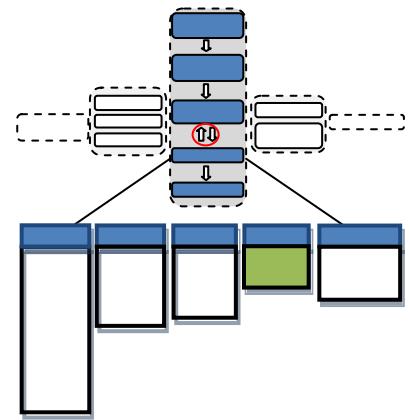
- I. Hipótese I - Clientes tendem a interagir mais com ações de parceiros que já tiveram alguma interação transacional prévia. Este fator influência fortemente na decisão de abertura ou não do *E-mail*



- II. Hipótese II - Quanto mais segmentadas forem as ações, mais interesses terão os clientes em abri-las. Este fator também exerce forte influência na escala de decisão da abertura.
- III. Hipótese III - Assuntos dos tipos “Detalhado, Misto, Personalizado, Promocional” e “Genérico e com parceiros” tendem a influenciar positivamente na decisão dos clientes em interagir com as ações de *E-mail Marketing*, sendo que os primeiros exercem maior poder influenciador do que os segundos.

4.4. Testes e validação das hipóteses

Os testes aqui realizados para a validação das hipóteses construídas em etapa anterior foram do tipo A/B nos moldes sugeridos por Croll e Yoskovitz (2013), em que divide-se a base de disparo ao meio e 50% da base recebe com o fator que pretende-se testar (A) e os outros 50% com o fator de referência (B). Objetiva-se, com isso, verificar como é a performance de A analisando-a comparativamente com B.



Infelizmente, por conta de adequações do calendário de disparos da ações de *E-mail Marketing* visando realizar os testes, foi possível a realização de apenas 1 teste prático para cada uma das hipóteses. Isso impede a realização da análise estatística conhecida como Teste de Hipóteses que tem por objetivo verificar se a igualdade das médias apresenta significância estatística. Deste modo, utilizou-se a comparação numérica para verificar se as hipóteses produziam de fato incrementos nas taxas de aberturas das ações de *E-mail Marketing*.

Por questões de confidencialidade das informações não foi possível divulgar na integra todas as informações referentes aos testes realizados. O assunto inserido as comunicações e a taxa de abertura de cada um dos grupos do teste A/B não puderam ser publicado. Deste modo, apresenta-se a seguir o resultado em termos incrementais, ou seja, o quanto, em percentual, a taxa de abertura de determinado grupo foi superior a do outro.

a. Teste da sensibilidade da taxa de abertura à Hipótese I.

Com o auxílio do analista de planejamento de campanhas e do time responsável por disparo e geração dos relatórios de retorno, realizou-se teste A/B com uma campanha de *E-*

mail no mês de Setembro/2014. Neste teste, “A” refere-se a base com experiência de compra (com *credits*) no parceiro do *E-mail* e “B” ao restante da base elegível.

Tomou-se o cuidado de incluir o nome do parceiro no assunto da ação e, manter-se o mesmo assunto as duas comunicações. Como o fator estudado, taxa de abertura, não depende diretamente da peça do *E-mail*, já que o cliente só vê a peça após tê-lo aberto, os detalhes da peça não são considerados neste teste. Além disso, enviou-se ambas as peças no mesmo dia, e verificou-se que o saldo médio encontrava-se distribuído nas duas bases.

Como resultado, percebeu-se que o grupo A apresentou taxa de abertura 6% superior a B, o que comprova a Hipótese I.

b. Teste da sensibilidade da taxa de abertura à Hipótese II.

O teste II visa determinar o quanto a segmentação traz ganhos diretos à taxa de abertura. A segmentação passa por diversos fatores, tais quais saldo, tempo de cadastro, idade, sexo, ou alguma característica relevante de algum grupo específico da base.

Para este teste, decidiu-se adotar uma característica peculiar de determinado grupo versus o restante da base. Por questões de confidencialidade e estratégia interna, não se pode descrever aqui a regra utilizada nesta segmentação. Do mesmo modo que foi feito em teste anterior, neste, assegurou-se que os disparos realizados aos grupos A e B fossem feitos no mesmo dia e com assuntos semelhantes.

Como resultado verificou-se que um incremento de 45% na taxa de abertura.

c. Teste da sensibilidade da taxa de abertura à Hipótese III.

A hipótese III consiste basicamente no teste de assunto. Para que esta hipótese pudesse ser testada, necessitou-se que as outras variáveis relacionadas a E-mail (público e dia do disparo) fossem controladas. Os tipos de assunto que estão fora dos mapeados pelo modelo que tendem a elevar a taxa de abertura são “Com informação da peça” e “Promocional – urgência”. Como o promocional – urgência está associado a promoções pontuais realizadas pela empresa, seu teste é mais complexo.

Logo, com auxílio dos analistas do time de campanhas da Loyalty S.A., decidiu-se testar os *assuntos* classificados pelo modelo *versus* o tipo “Com informação da peça”.

Para isso escolheu-se um disparo característico da empresa e testou-se os assuntos “Com parceiros” vs. “Com informação da peça”. Os outros assuntos não puderam ser testados uma vez que a realização dos testes depende de espaço no calendário de disparos da organização. Deste modo, caso os dois representantes mostrem-se válidos, considera-se que a hipótese é verdadeira.

Para a realização deste teste A/B selecionou-se um disparo em que parte da base recebeu o E-mail com o assunto que continha os produtos ofertados na peça (Com informação da peça) e a outra parte um assunto que contenha apenas o nome do parceiro (com parceiros). Como resultado, a base que recebeu o assunto B apresentou taxa de abertura 18% superior ao A. Portanto, neste teste, verificou-se que o assunto com “Com parceiros” trouxe mais retorno em termos de abertura de Email, validando a Hipótese III.

A tabela a seguir resume as hipóteses geradas pelo modelo e validadas em ambiente prático:

Tabela 9: Resumo das hipóteses validadas em ambiente prático

Número	Hipótese	Validada?
I	Clientes tendem a interagir mais com ações de parceiros que já tiveram alguma interação transacional prévia	Sim
II	Quanto mais segmentadas forem as ações, mais interesses terão os clientes em abri-las.	Sim
III	Assuntos dos tipos “Detalhado, Misto, Personalizado, Promocional” e “Genérico e com parceiros” tendem a influenciar positivamente na decisão dos clientes em interagir com as ações de <i>E-mail Marketing</i>	Sim

Desenvolvido pelo próprio autor

Todavia, antes de considerar o modelo aceito e considerando a experiência prática do time de campanhas, que observou com dúvidas o fato do modelo não ter relacionado o saldo médio como um fator relevante, decidiu-se verificar se este fator realmente não provoca diferenças significativas na taxa de abertura. Os analistas levantaram a hipótese, percebida por eles em seu dia-a-dia, de que as campanhas com segmentação focada em clientes com maior saldo normalmente possuem maior taxa de abertura do que as focadas nos clientes de menor saldo.

Como a empresa utiliza-se bastante esta estratégia de segmentação, pôde-se realizar testes A/B em três disparos e trabalhar com a média de abertura de cada um destes, procurando fixar as outras variáveis. Como resultado verificou-se que as ações focadas no maior saldo tiveram uma taxa de abertura 16% superior às ações com saldo menor.

Sendo assim, encontrou-se por solicitação do time de campanhas, um fator relevante, testado por solicitação do time de campanhas, o qual possui expertise e sensibilidade, por contato do dia-a-dia com as ações de Email da Loyalty S.A., sobre os fatores candidatos a relevantes na interação dos clientes com o Email. Logo, por precaução decidiu-se reenviar o modelo para a fase de modelagem para que fosse refinado, aprimorado e testado novamente.

Todavia, vale destacar que o modelo inicial já nos trouxe importantes informações como “Clientes tendem a interagir mais com ações de parceiros que já tiveram alguma interação transacional prévia” e “Quanto mais segmentadas forem as ações, mais interesses terão os clientes em abri-las” que foram validadas na prática.

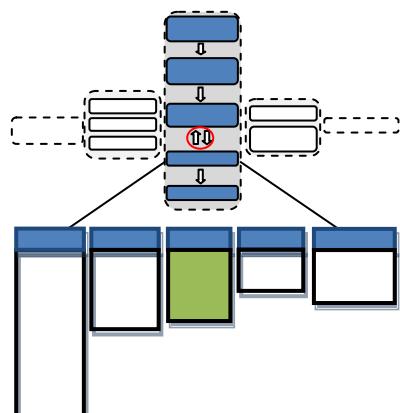
4.5. Refinamento do modelo

Visando redesenhar e refinar o modelo estatístico utilizado e seguindo-se as indicações da teoria a respeito de regressão linear multivariada, a qual, conforme comentada na seção destinada a revisão de literatura, sugere realizar-se novamente a regressão trabalhando com dados quadráticos ou com a raiz dos dados (GUNST; MASON, 1980). Além da literatura, o especialista em modelos estatísticos também comentou ser válida as transformações quadráticas e aplicação da raiz cúbica no reprocesso do modelo.

A literatura descrita trata especificamente da revisão do modelo por conta de ter sido ferido o princípio de normalidade dos resíduos. Todavia, aproveitou-se deste gancho literário para reprocessar o modelo pelo fato de ele não ter previsto uma hipótese importante para a explicação da taxa de abertura.

4.5.1. Transformação A

Seguindo-se os preceitos da literatura descritos por Gunst e Mason (1980), a transformação executada foi a aplicação da função quadrática ($Y = X^2$), trabalhando-se com o quadrado dos valores assumidos pelas variáveis mapeadas



para execução da regressão linear e construção de um modelo explicativo do comportamento da variável taxa de abertura.

Nos quadros a seguir constam exemplos das transformações executadas e seus resultados para as variáveis “X1_dia_semana” e “X1_tipo_comunicacao”. As transformações referentes as outras variáveis constam integralmente na seção anexos.

“(X1_dia_semana)”	“(X1_dia_semana) ² ”
0,9117	0,8313
-1,0925	1,1936
-1,0925	1,1936
-1,0925	1,1936
0,9117	0,8313
0,9117	0,8313
0,9117	0,8313
0,9117	0,8313
0,9117	0,8313
0,9117	0,8313



Quadro 4: Exemplo de transformação da variável X1_dia_semana

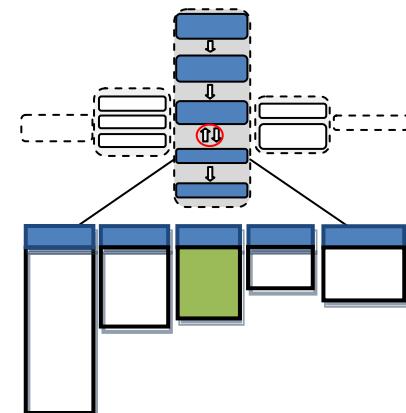
“(X1_tipo_comunicação)”	“(X1_tipo_comunicação) ² ”
-0,7182	0,5158
-0,7182	0,5158
-0,7182	0,5158
1,3869	1,9235
-0,7182	0,5158
-0,7182	0,5158
-0,7182	0,5158
-0,7182	0,5158
-0,7182	0,5158
-0,7182	0,5158



Quadro 5: Exemplo de transformação da variável X1_tipo_comunicação

4.5.2. A regressão linear com os dados quadráticos

Nesta etapa foi realizada novamente a regressão linear, trabalhando com os dados quadráticos de modo a colher novas hipóteses a serem testadas em ambiente prático, contando com a ajuda dos analistas da Loyalty S.A. dedicados a campanhas de Email. Novamente, destaca-se a seguir os resultados mais importantes para a validação do modelo. A análise de regressão está descrita por completo na seção anexos.



4.5.2.1. Poder de explicação do modelo

Segundo Johnson e Wichern (2002), o poder de explicação do modelo está associado aos valores assumidos pelos parâmetros R-quadrado (*R square*) e R quadrado ajustado (*Adjusted R square*). A tabela 10 exprime os valores assumidos por estes parâmetros nesta análise:

Tabela 10: Valores de R e R² no modelo em estudo.

R quadrado (R ²)	R quadrado ajustado (R ² ajustado)
0,739	0,728

Desenvolvido pelo próprio autor

Logo, o modelo desenvolvido é capaz de explicar cerca de 73% da variância da variável em estudo – taxa de abertura, o que pode ser considerado satisfatório.

4.5.2.2. Coeficientes da equação da regressão

Os coeficientes da regressão (peso das variáveis na explicação da variação da taxa de abertura) e sua respectiva significância para o modelo de regressão são apresentados na tabela 11 a seguir. Chegou-se a tais coeficientes mediante a realização da análise de regressão linear realizada no software PASW *Statistics*. Todos os passos realizados no software para se chegar a estes valores estão descritos na seção anexos.

Tabela 11: Coeficiente e sua respectiva significância para o modelo 2.

Modelo	Variável	Coeficiente	Significância
2	(Constante)	-0,697	0,698
	X1_dia_semana	-0,264	0,637
	X2_dia_semana	-0,033	0,765
	X3_dia_semana	-1,765	0,197
	X2_tipo_comunicação	0,476	0,005
	X1_assunto	2,532	0,035
	X2_assunto	-0,259	0,039
	X3_assunto	0,271	0,007
	Saldo médio	-0,060	0,000
	Percentual impactado	-0,090	0,234
	Grau transacional	0,892	0,000

Desenvolvido pelo próprio autor

Conforme comentado na seção destinada a revisão de literatura, deve-se verificar a significância estatística para determinar se a variável exerce influência considerável na variação da taxa de abertura. Para aquelas que mostrarem-se significantes, deve-se medir seu poder de influencia pelo módulo de seu coeficiente B.

Deste modo, as variáveis em questão são, nesta ordem: “X1_assunto, Grau transacional, X2_tipo_comunicacao, X3_assunto, X2_assunto e saldo médio”.

Retomando-se a seção de revisão de literatura, temos a citação de Moore (2007) que afirma que “a correlação mensura a direção e o grau de relação linear entre duas variáveis quantitativas”. Mediante análise realizada no software PASW *Statistics*, contida em detalhe na seção anexos, chegou-se aos seguintes valores para as variáveis com poder de explicação da taxa de abertura significativo:

Tabela 12: Correlações de Pearson entre a taxa de abertura e as variáveis independentes com coeficiente significante

Variável	Correlação de Pearson	Sig. (1-tailed)
X1_assunto	0,088	0,079
Grau transacional	0,829	0,000
X2_tipo_comunicacao	0,361	0,000
X3_assunto	0,213	0,000
X2_assunto	0,007	0,455
Saldo médio	0,242	0,000

Desenvolvido pelo próprio autor.

Pode-se extrair da tabela 12 que as variáveis que possuem correlação estatisticamente válida (significância de 5%), isto é, aquelas, que possuem um relacionamento estatisticamente válido para que se possa extrair os valores que tendem a maximizar a taxa de abertura, visando os testes em etapas subsequentes são:

- Grau transacional: Correlação positiva (+)
- X2_tipo_comunicação: correlação positiva (+)
- X3_assunto: Correlação positiva (+)
- Saldo médio: Correlação positiva (+)

4.5.2.3. Análise de resíduos

A análise de resíduos foi feita, em primeiro lugar, utilizando-se dos testes de Kolmogorov-Smirnov e Shapiro-Wilk, para testar se os resíduos apresentam distribuição normal, fator necessário à validação da regressão linear desenvolvida. Para que sejam considerados normais, a significância destes testes deve ser maior do que 5%, uma vez que, para valores inferiores a 5%, pode-se interpretar que os dados apresentam uma diferença significativamente considerável em relação a curva normal.

A tabela 13 nos traz os resultados destes testes:

Tabela 13: Resultados dos testes de normalidade para o modelo 2

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
<i>Unstandardized Residual</i>	0,256	255	0,000	0,670	255	0,000

Desenvolvido pelo próprio autor

De acordo com este teste de normalidade, pode-se concluir que os resíduos não apresentam distribuição normal, o que, segundo a literatura, fere novamente um dos princípios teóricos da regressão linear citados anteriormente.

Neste ponto, procurou- se o professor especialista em modelos estatísticos, apresentando para ele o caso em questão.

O professor especialista afirmou que verifica-se situação semelhante em muitos casos práticos de aplicação de modelos estatísticos teóricos. Segundo o professor especialista, nem sempre, todas as considerações teóricas podem ser verificadas à risca em ambiente prático.

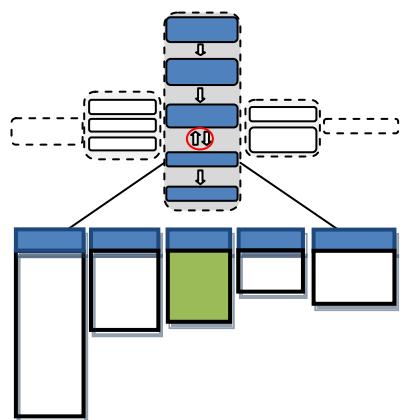
Deste modo, o professor especialista sugeriu novamente que o autor prosseguisse com os resultados da regressão, construindo o modelo e testando os resultados. Segundo ele, como o autor tomou o cuidado de testar as hipóteses do modelo antes de considerá-las como verdade no contexto em estudo; fatalmente se modelo tiver produzido resultados não condizentes com a realidade, estes não serão aprovados pela etapa dos testes.

A análise dos outros parâmetros, bem como o passo a passo da análise de resíduos, estão descritos em detalhes na seção anexos.

4.5.3. Extração das hipóteses pós 1 etapa do refino

Juntando-se as informações relativas aos coeficientes de correlação de Pearson e dos coeficientes significativos do modelo de regressão, extraí-se, as seguintes hipóteses:

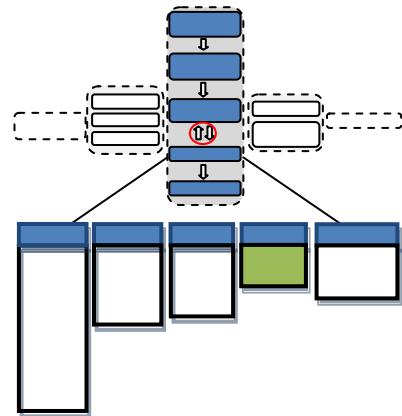
- I. Hipótese I - Clientes tendem a interagir mais com ações de parceiros que já tiveram alguma interação transacional prévia. Este fator influênciaria fortemente na decisão de abertura ou não do *E-mail*



- II. Hipótese II - Comunicações de resgate tendem a possuir maior taxa de abertura do que comunicações de acúmulo.
- III. Hipótese III - Assuntos dos tipos “Detalhado, Misto, Personalizado e Promocional” tendem a influenciar positivamente na decisão dos clientes em interagir com as ações de *E-mail Marketing*.
- IV. Hipótese IV - Clientes com maior saldo tendem interagir mais com ações de *E-mail* do que aqueles que possuem menor saldo.

4.5.4. Testes e validação das hipóteses

Assim como ocorreu na etapa de testes das Hipóteses geradas pelo modelo 1, os testes realizados para a validação das hipóteses do modelo 2 basearam-se em testes do tipo A/B, também nos moldes por Croll e Yoskovitz (2013), em que divide-se a base de disparo ao meio e 50% da base recebe com o fator que pretende-se testar (A) e os outros 50% com o fator de referência (B). Objetiva-se, com isso, verificar como é a performance de A analisando-a comparativamente com B.



Novamente, por conta do calendário de disparos da Loyalty S.A., foi possível a realização de apenas 1 teste prático para cada uma das hipóteses, o que impediu a realização da análise estatística conhecida como Teste de Hipóteses (verifica se a igualdade das médias apresenta significância estatística). Deste modo, utilizou-se, mais uma vez, a comparação numérica para verificar se as hipóteses produziam de fato incrementos nas taxas de aberturas das ações de *E-mail Marketing*.

Por questões de confidencialidade das informações não foi possível divulgar na íntegra todas as informações referentes aos testes realizados. O assunto inserido as comunicações e a taxa de abertura de cada um dos grupos do teste A/B não puderam ser publicado. Deste modo, apresenta-se a seguir o resultado em termos incrementais, ou seja, o quanto, em percentual, a taxa de abertura de determinado grupo foi superior a do outro.

- a. Teste da sensibilidade da taxa de abertura à Hipótese I.

Na etapa de construção das hipóteses do Modelo 1 já estava contida a hipótese I gerada pelo Modelo 2. Deste modo, esta hipótese já havia sido testada e validada nos testes do Modelo 1.

Considera-se, portanto, o resultado replicado para o Modelo 2 e, a hipótese I validada.

b. Teste da sensibilidade da taxa de abertura à Hipótese II.

Realizar um teste que confronte ações de acúmulo com as de resgate e mantenha as outras variáveis (público e assunto) idênticas esbarra na estratégia interna da organização de, salvo em épocas de grandes promoções, enviar ações de acúmulo para a base que não tenha saldo e as de resgate para aqueles que possuem saldo mínimo para realizar um resgate. Encontrou-se, todavia, um espaço na grade em que oferecia-se, primeiro, uma condição especial de acúmulo para clientes que possuíam determinado cartão de crédito de um banco parceiro e, logo em seguida, uma oferta de resgate.

A única ressalva existente neste teste fica por conta do assunto, uma vez que torna-se praticamente impossível enviar o mesmo assunto numa ação de acúmulo e numa com foco em resgate. Visando minimizar tal ponto, utilizou-se, em ambas as ações, assuntos genéricos.

Como resultado deste teste, verificou-se que a ação incentivando o resgate teve uma taxa de abertura 10% superior aquela que incentiva o acúmulo, o que permite-nos concluir que a Hipótese II está validada.

c. Teste da sensibilidade da taxa de abertura à Hipótese III.

A hipótese III do modelo 1 (Assuntos dos tipos “Detalhado, Misto, Personalizado, Promocional” e “Genérico e com parceiros” tendem a influenciar positivamente na decisão dos clientes em interagir com as ações de *E-mail Marketing*) contém a hipótese III do Modelo 2 (Assuntos dos tipos “Detalhado, Misto, Personalizado e Promocional” tendem a influenciar positivamente na decisão dos clientes em interagir com as ações de *E-mail Marketing*).

No modelo 1, testou-se a efetividade do assunto “Com parceiros” vs. “Com informação da peça”, a qual gerou um resultado que permitiu a validação da hipótese dentro daquela análise. Lembrando que, durante os testes do modelo 1, não pôde-se testar todos os assuntos por questões de calendário.

Deste modo, adotando-se uma postura conservadora deseja-se realizar um novo teste, utilizando-se desta vez um tipo de assunto contido na Hipótese III do Modelo 2 e que não tenha sido testado no Modelo 1. Escolheu,-se, com o auxílio dos analistas do time de campanhas da Loyalty, realizar o teste utilizando assunto “Detalhado” *versus* “Com informação da peça”.

Infelizmente assim como ocorrera nos testes do Modelo 1, os outros assuntos trazidos na Hipótese III do Modelo 2 não puderam ser testados uma vez que a realização dos testes depende de espaço no calendário de disparos da organização.

Para a realização deste teste A/B selecionou-se um disparo em que parte da base recebeu o E-mail com o assunto que continha os produtos ofertados na peça (Com informação da peça) e a outra parte um assunto que continha o nome do parceiro e o nome do cliente, categorizando um assunto do tipo Misto. Os outros fatores da base, tal qual Saldo e percentual de participantes com interação prévia no parceiro foram fixados de forma a não prejudicar o resultados dos testes.

Como resultado deste teste, verificou-se que a base que recebeu o assunto do tipo A (Com informação da peça) apresentou uma taxa de abertura 14% superior a base que recebeu a comunicação com o assunto do tipo B (Misto). Portanto, verifica-se um caso prático em que um dos assuntos da Hipótese III teve performance inferior a um outro assunto não contido nesta hipótese. Logo, a hipótese III não pode ser comprovada em ambiente de testes práticos. A tabela a seguir resume as hipóteses geradas pelo modelo e validadas em ambiente de testes práticos.

d. Teste da sensibilidade da taxa de abertura à Hipótese IV.

A Hipótese IV do Modelo 2 foi justamente aquela levantada pelo time de campanhas da Loyalty S.A., cuja validação acabou fazendo com que o Modelo 1 não pudesse ser considerado válido.

Deste modo, considera-se esta hipótese como validada.

A tabela 14 apresenta os resultados das hipóteses levantadas pelo Modelo 2 e testadas esta etapa:

Tabela 14: Resumo das hipóteses validadas em ambiente prático

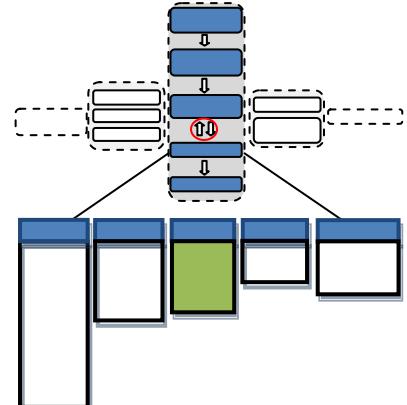
Número	Hipótese	Validada?
I	Clientes tendem a interagir mais com ações de parceiros que já tiveram alguma interação transacional prévia	Sim
II	Comunicações de resgate tendem a possuir maior taxa de abertura do que comunicações de acúmulo.	Sim
III	Assuntos dos tipos “Detalhado, Misto, Personalizado e Promocional” tendem a influenciar positivamente na decisão dos clientes em interagir com as ações de <i>E-mail Marketing</i> .	Não
IV	Clientes com maior saldo tendem interagir mais com ações de <i>E-mail</i> do que aqueles que possuem menor saldo.	Sim

Desenvolvido pelo próprio autor

Deste modo, seguindo-se a abordagem *Lean* deve-se, retornar, mais uma vez, a macro- etapa de modelagem já que a Hipótese III não mostrou-se válida em ambiente de testes práticos.

4.5.5. Transformação B

Seguindo-se os preceitos da literatura descritos por Gunst e Mason (1980), a segunda tentativa executada no sentido de refinar o modelo foi a aplicação da função raiz cúbica ($Y = \text{raiz}(X,3)$), trabalhando-se com a raiz cúbica dos valores assumidos pelas variáveis mapeadas para execução da regressão linear. Não se utilizou da raiz quadrada pelo fato de as variáveis transformadas (processo de padronização) assumirem valores negativos.



A seguir estão listados exemplos destas transformações para a variável “X1_dia_semana” e “X1_tipo_comunicacao”. As transformações das demais variáveis estão contidas na integra na seção anexos.



“(X1_dia_semana)”	Raiz
0,9117	0,9697
-1,0925	-1,0299
-1,0925	-1,0299
-1,0925	-1,0299
0,9117	0,9697
0,9117	0,9697
0,9117	0,9697
0,9117	0,9697
0,9117	0,9697
0,9117	0,9697

Quadro 6: Exemplo de transformação da variável X1_dia_semana



“(X1_tipo_comunicação)”	Raiz
-0,7182	-0,8955
-0,7182	-0,895
-0,7182	-0,8955
1,3869	1,1152
-0,7182	-0,8955
-0,7182	-0,8955
-0,7182	-0,8955
-0,7182	-0,8955
-0,7182	-0,8955
-0,7182	-0,8955

Quadro 7: Exemplo de transformação da variável X1_tipo_comunicação

4.5.6. A regressão linear pós transformação B

Realiza-se, novamente, a regressão linear trabalhando-se, agora, com a raiz cúbica dos dados, visando colher hipóteses a ser testadas no laboratório de campanhas com a ajuda dos analistas da Loyalty S.A. dedicados a campanhas de *E-mail*. Novamente, destaca-se a seguir

os resultados mais importantes para o modelo e, os outros, estão foram alocados na seção anexos.

4.5.6.1. Poder de explicação do modelo

Segundo Johnson e Wichern (2002), o poder de explicação do modelo está associado aos valores assumidos pelos parâmetros R-quadrado (*R square*) e R quadrado ajustado (*Adjusted R square*). A tabela 15 exprime os valores assumidos por estes parâmetros nesta análise:

Tabela 15: Valores de R e R^2 no modelo em estudo.

R quadrado (R^2)	R quadrado ajustado (R^2 ajustado)
0,572	0,554

Desenvolvido pelo autor

Logo, o modelo desenvolvido é capaz de explicar cerca de 55% da variância da variável em estudo – taxa de abertura, o que pode ser considerado satisfatório.

4.5.6.2. Coeficientes da equação da regressão

Os coeficientes da regressão (peso das variáveis na explicação da variação da taxa de abertura) e sua respectiva significância para o modelo de regressão estão expostos na tabela X a seguir. Chegou-se a tais coeficientes mediante à realização de análise de regressão linear utilizando o software PASW *Statistics*. Todos os passos realizados no software para se chegar a estes valores estão descritos na seção anexos.

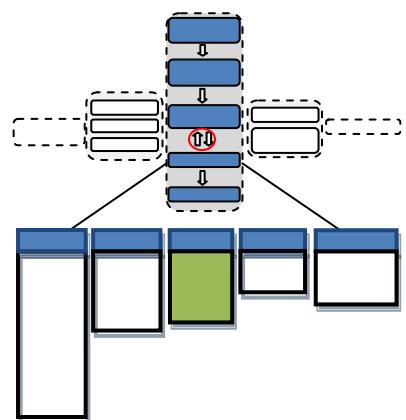


Tabela 16: Coeficiente e sua respectiva significância para o modelo 3.

Modelo	Variável	Coeficiente	Significância
3	(Constante)	0,093	0,064
	X1_dia_semana	0,062	0,078
	X2_dia_semana	-0,015	0,749
	X3_dia_semana	0,013	0,750
	X2_tipo_comunicação	0,131	0,003
	X1_assunto	0,001	0,977
	X2_assunto	0,084	0,055
	X3_assunto	0,134	0,005
	Saldo médio	0,576	0,000
	Percentual impactado	-0,386	0,000
	Grau transacional	0,194	0,002

Desenvolvido pelo próprio autor

Conforme comentado na seção destinada a revisão de literatura, deve-se verificar a significância estatística para determinar se a variável exerce influência considerável na variação da taxa de abertura. Para aquelas que mostrarem-se significantes, deve-se medir seu poder de influencia pelo módulo de seu coeficiente B.

Deste modo, as variáveis em questão são, nesta ordem: “Saldo médio, Percentual impactado, Grau transacional, X3_assunto e X2_tipo_comunicação.”

Retomando-se a seção de revisão de literatura, temos a citação de Moore (2007) que afirma que “a correlação mensura a direção e o grau de relação linear entre duas variáveis quantitativas”. Mediante análise realizada no software PASW *Statistics*, contida em detalhe na seção anexos, chegou-se aos seguintes valores para as variáveis com poder de explicação da taxa de abertura significativo:

Tabela 17: Correlações de Pearson entre a taxa de Abertura e as variáveis independentes com coeficiente significante

Variável	Correlação de Pearson	Sig. (1-tailed)
Saldo médio	0,580	0,000
Percentual impactado	-0,582	0,000
Grau transacional	0,273	0,000
X3_assunto	0,035	0,288
X2_tipo_comunicação	0,185	0,001

Desenvolvido pelo próprio autor

Pode-se extrair da tabela 17 que as variáveis que possuem correlação estatisticamente válida (significância de 5%), isto é, aquelas, que possuem um relacionamento estatisticamente válido para que se possa extrair os valores que tendem a maximizar a taxa de abertura, visando os testes em etapas subsequentes são:

- Saldo médio: Correlação positiva (+)
- Percentual impactado: Correlação negativa (-)
- Grau transacional: Correlação positiva (+)
- X2_tipo_comunicação: Correlação positiva (+)

4.5.6.3. Análise de resíduos

A análise de resíduos foi feita, em primeiro lugar, utilizando-se dos testes de Kolmogorov-Smirnov e Shapiro-Wilk, para testar se os resíduos apresentam distribuição normal, fator necessário a validação da regressão linear desenvolvida. Para que sejam considerados normais, a significância destes testes deve ser maior do que 5%, uma vez que, para valores inferiores a 5%, pode-se interpretar que os dados apresentam uma diferença significativamente considerável em relação a curva normal.

A tabela 18 nos traz os resultados destes testes:

Tabela 18: Resultados dos testes de normalidade para o modelo 3

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
<i>Unstandardized Residual</i>	0,123	255	0,004	0,911	255	0,003

Desenvolvido pelo próprio autor

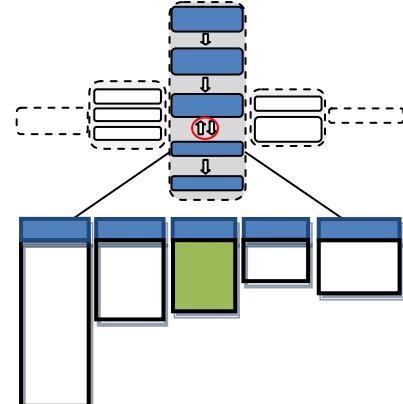
De acordo com este teste de normalidade, pode-se concluir que os resíduos não apresentam distribuição normal, o que, segundo a literatura, fere um dos princípios teóricos da regressão linear citados anteriormente.

Utilizou-se, mais uma vez, da sugestão do especialista em modelos prosseguiu-se com os resultados da regressão construindo o modelo e testando os resultados.

A análise dos outros parâmetros, bem como o passo a passo da análise de resíduos, estão descritos em detalhes na seção anexos.

4.5.7. Extração das hipóteses

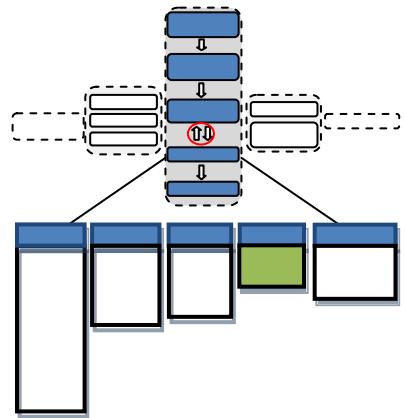
Juntando-se as informações relativas aos coeficientes de correlação de Pearson e dos coeficientes significativos do modelo de regressão, extraí-se, as seguintes hipóteses:



- I. Hipótese I - Clientes com saldo maior tendem a interagir mais com as ações de *E-mail Marketing* do que aqueles que possuem menor saldo.
- II. Hipótese II - Quanto mais segmentadas forem as ações, mais interesses terão os clientes em abri-las. Este fator também exerce forte influência na escala de decisão da abertura.
- III. Hipótese III - Clientes tendem a interagir mais com ações de parceiros que já tiveram alguma interação transacional prévia. Este fator influência fortemente na decisão de abertura ou não do *E-mail Marketing*.
- IV. Hipótese IV - Comunicações de resgate tendem a possuir maior taxa de abertura do que comunicações de acúmulo.

4.5.8. Experimentação e Validação

Os testes aqui realizados para a validação das hipóteses construídas em etapa anterior foram do tipo A/B nos moldes sugeridos por Croll e Yoskovitz (2013)., em que divide-se a base de disparo ao meio e 50% da base recebe com o fator que pretende-se testar (A) e os outros 50% com o fator de referência (B). Objetiva-se, com isso, verificar como é a performance de A analisando-a comparativamente com B.



a. Teste da sensibilidade da taxa de abertura à Hipótese I.

Este teste também havia sido feito nos testes do modelo 1, verificando a suspeita levantada pelo time de analistas da Loyalty S.A. o saldo têm correlação positiva com a taxa de abertura. Foi este teste, inclusive o responsável por invalidar o modelo 1, uma vez que ele não conseguiu este fator como relevante.

Para guiar a leitura, o autor repete a seguir a estrutura e o resultado deste teste:

“Como a empresa utiliza-se bastante desta estratégia de segmentação, pôde-se realizar testes A/B em três disparos e trabalhar-se com a média de abertura de cada um deles, procurando-se fixar as outras variáveis. Como resultado verificou-se que as ações focadas no maior saldo tiveram uma taxa de abertura 16% superior as com saldo menor. “

b. Teste da sensibilidade da taxa de abertura à Hipótese II.

Esta hipótese já havia sido testada no modelo 1 e, como resultado verificou-se um incremento de 45% na taxa de abertura. A seguir, o autor repete a estrutura deste site, facilitando o entendimento do leitor:

“O teste I visa determinar o quanto a segmentação traz ganhos diretos a taxa de abertura. A segmentação passa por diversos fatores, tais quais saldo, tempo de cadastro, idade, sexo, ou alguma característica relevante de algum grupo específico da base.

Para este teste, decidiu-se adotar uma característica peculiar de determinado grupo vs. o restante da base. Por questões de confidencialidade e estratégia interna, não pode-se descrever aqui a regra utilizada nesta segmentação. Do mesmo modo que foi feito em teste

anterior, neste, assegurou-se que os disparos realizados aos grupos A e B fossem feitos no mesmo dia e com assuntos semelhantes.”

c. Teste da sensibilidade da taxa de abertura à Hipótese III.

Este teste também fora feito na etapa de testes do modelo 1, e seu resultado validou a hipótese, uma vez que o grupo A (experiência anterior no parceiro) apresentou taxa de abertura 6% superior ao B (restante da base). Para ajudar na leitura, o autor repete a seguir a estrutura deste teste:

“Com o auxílio do analista de planejamento de campanhas e do time responsável por disparo e geração dos relatórios de retorno, realizou-se teste A/B com uma campanha de *E-mail* no mês de Setembro/2014. Neste teste, A refere-se a base com experiência de compra (com *credits*) no parceiro do *E-mail* e B ao restante da base elegível.

Tomou-se o cuidado de incluir o nome do parceiro no assunto da ação e, manter-se o mesmo assunto as duas comunicações. Como o fator estudado, taxa de abertura, não depende diretamente da peça do *E-mail*, já que o cliente só vê a peça após tê-lo aberto, os detalhes da peça não são considerados neste teste. Além disso, enviou-se ambas as peças no mesmo dia, e verificou-se que o saldo médio encontrava-se distribuído nas duas bases.”

d. Teste da sensibilidade da taxa de abertura à Hipótese IV.

O teste da Hipótese IV do Modelo 3, “Comunicações de resgate tendem a possuir maior taxa de abertura do que comunicações de acúmulo”, fora feito quando da validação do Modelo 2, uma vez que esta hipótese também foi gerada pro aquele modelo.

Como resultado deste teste, verificou-se que a ação incentivando o resgate teve uma taxa de abertura 10% superior aquela que incentiva o acúmulo, o que permite-nos concluir que a Hipótese II está validada.

A tabela 19 apresenta os resultados das hipóteses levantadas pelo Modelo 3 e testadas nesta etapa:

Tabela 19: Resumo das hipóteses validadas em ambiente prático

Número	Hipótese	Validada?
I	Clientes com saldo maior tendem a interagir mais com as ações de <i>E-mail Marketing</i> do que aqueles que possuem menor saldo	Sim
II	Quanto mais segmentadas forem as ações, mais interesses terão os clientes em abri-las.	Sim
III	Clientes tendem a interagir mais com ações de parceiros que já tiveram alguma interação transacional prévia	Sim
IV	Comunicações de resgate tendem a possuir maior taxa de abertura do que comunicações de acúmulo.	Sim

Desenvolvido pelo próprio autor

Portanto, todas as Hipóteses geradas pelo Modelo 3 puderam ser validadas em ambiente de testes práticos, o que valida os resultados deste modelo, baseando na metodologia Lean e no conceito de Aprendizagem Validada. A seção seguinte traz as considerações do autor a cerca deste resultado.

4.5.9. Analise dos resultados

O modelo validado na etapa anterior mostrou-nos que as variáveis relacionadas ao público (Saldo Médio, percentual impactado e grau transacional) e ao tipo de comunicação são as mais representativas em termos de explicação do comportamento da taxa de abertura das ações da Loyalty S.A.

O fato curioso está no menor peso dado as variáveis próprias da campanha (Assunto e dia do disparo). Ao que parece a decisão de abrir um *E-mail Marketing* está muito mais associada a expectativa do cliente em interagir com a rede da Loyalty S.A. e, também, ao quão assertiva a empresa é em segmentar sua base.

Como possibilidade de melhoria deste resultado, o time de campanhas da Loyalty S.A., sugeriu o teste de um modelo que leve em conta apenas o assunto das ações, de modo que o modelo retorne a ordem de motivar o cliente a abrir a ação associada a este fator. Tal análise fica como um possível oportunidade de continuação deste projeto.

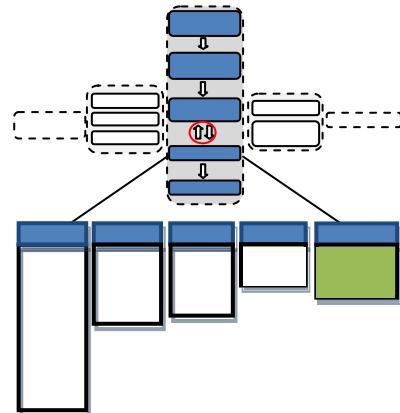
Sendo assim, conclui-se, em princípio, a parte do trabalho que visou a aplicação do conceito de “Aprendizagem Validada”, tal qual descrita por Ries (2013) no ambiente em

estudo. Note que, todos os conceitos verificados de maneira teórica só foram considerados como ganhos de fato quando puderam ser verificados em ambiente prático de teste e, também, mostraram que a hipótese de fato demonstrou ser válida em caráter prático. Este é o cerne do que diz Ries (2013) quando trata da “Aprendizagem Validada”. Deste modo, entrega-se ao time de campanhas da Loyalty S.A. informações muito valiosas e realmente comprovadas acerca de como otimizar o canal E-mail.

A etapa subsequente trata da institucionalização da metodologia abordada até aqui no contexto da Loyalty S.A., dialogando diretamente com o ciclo “Construir medir aprender”.

4.6. Institucionalizando a metodologia

O resultado deste projeto gerou ganhos tanto em termos de produtos quanto de processos para o time de campanhas da Loyalty S.A. Pensando-se em termos de produtos, o principal ganho está nas informações geradas acerca do que, de fato, promove uma maior interação dos clientes com as ações de *E-mail Marketing* da Loyalty S.A..



Toda este estudo teve como pano de fundo a metodologia *Lean* de Blank (2006), Ries (2013) e Croll e Yoskovitz (2013), principalmente no que tange a “Aprendizagem Validada”. Deste modo, quando o time de campanhas da Loyalty S.A. estiver construindo uma peça de *E-mail*, já tem em mãos aspectos advindos de um modelo estatístico e validados em ambiente prático para que possa apoiar-se a fim de decidir a estratégia da ação.

Outro fator de suma importância está na apresentação do processo de análise construído. Desenvolveu-se, neste projeto, uma metodologia em torno do ciclo “Construir – medir- aprender” (Ries, 2013), a qual engloba tanto aspectos teóricos (modelos estatísticos) quanto de testes e validação.

Montando-se um paralelo com o ciclo “Construir-medir-aprender” de Ries (2013), pode dizer que a fase “Construir” englobou a macro-etapa relacionada a Modelagem Estatística, em que construiu-se o modelo e as hipóteses. A fase “Medir” caminhou lado a lado com a macro-etapa “Testes e Validação”, uma vez que foi nesta macro-etapa que mediuse, via taxa de abertura, se as hipóteses de fato convergiam para a realidade. Já a fase

“Aprender”, engloba esta última etapa de implementação da metodologia que é onde, de fato, o conhecimento produzido deve ser instalado na organização.

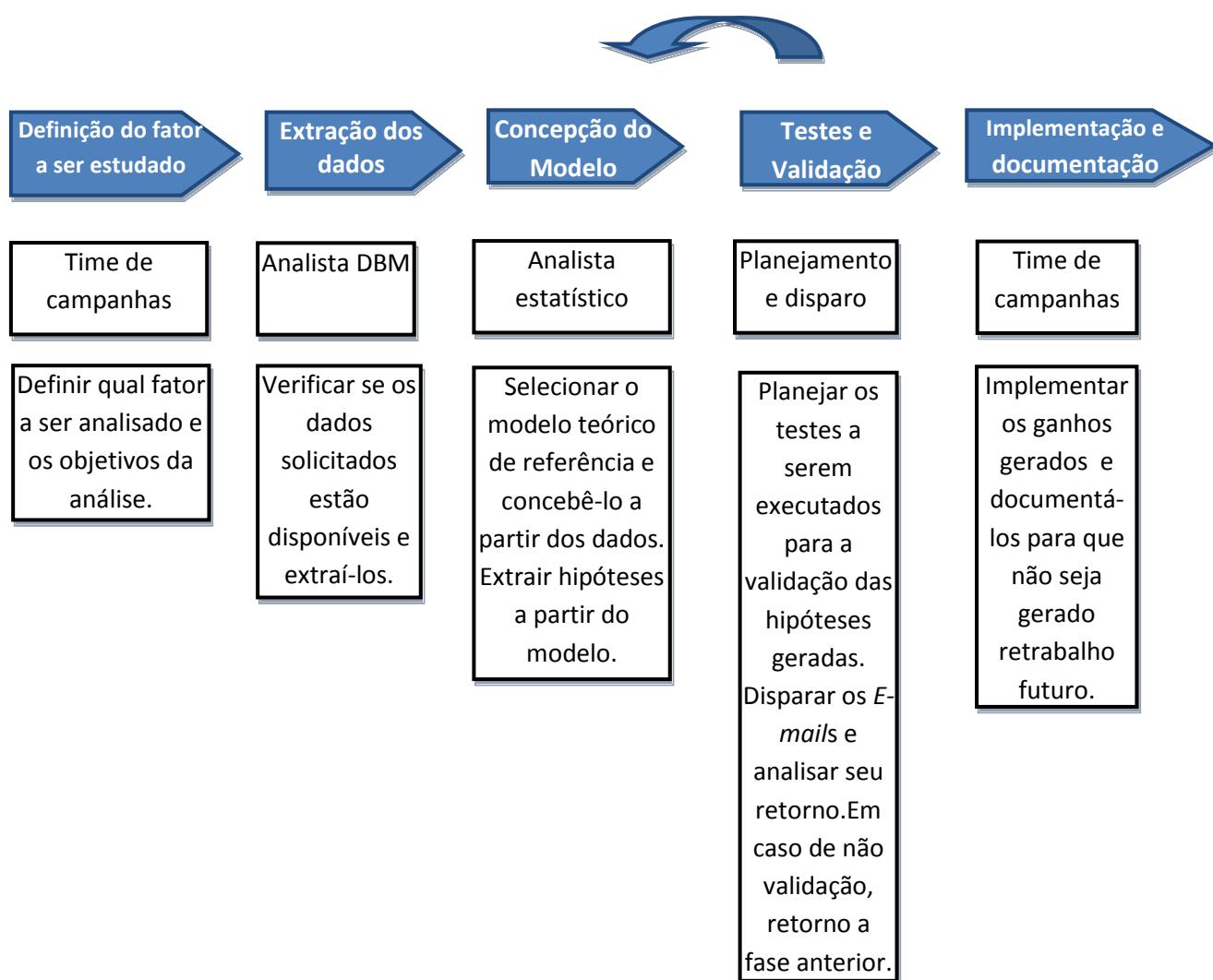
Considera-se este processo estabelecido com base no ciclo “Construir medir aprender” um fator de suma importância, pois é um ganho exclusivo da Loyalty S.A. e difícil de ser copiado por suas concorrentes, o que, tende a gerar vantagens competitivas consideráveis a organização.

Colocados numa escala de importância, considera-se os ganhos do processo superiores aos ganhos do produto, uma vez que enquanto estes são específicos para a taxa de abertura, aqueles podem replicados para outros fatores muito relevantes do *E-mail Marketing*, tal qual taxa de conversão, taxa de cliques, taxa de *Optout*, entre outros.

Pensando-se em termos de implementação da metodologia em ambiente organizacional da Loyalty S.A., sugere-se a criação de um grupo de trabalho composto pelos seguintes profissionais:

- Um analista de banco de dados, responsável por toda extração e tratamento dos dados a serem executados no modelo.
- Um analista com conhecimento estatístico que permita-o julgar o modelo mais adequado para o caso em estudo e, também, verificar os parâmetros de saída do modelo e construir as hipóteses a ser testadas.
- Um analista ligado ao planejamento das campanhas, de modo, que as hipóteses geradas pelos modelos possam ser testadas e validadas em ambiente prático.

No diagrama a seguir estão descritas as principais funções e seus respectivos responsáveis para que a metodologia desenvolvida neste projeto possa ser implementada na organização em estudo e replicada para outros fatores que a Loyalty S.A. julgar apropriado:



Quadro 8: Processo de testes baseado no Ciclo “Construir Medir Aprender” de Ries (2013) aplicado a Loyalty S.A.

Para que todo o resultado deste projeto tome corpo e o processo desenhado anteriormente, com os respectivos responsáveis por cada etapa possam tomar conhecimento, procurou-se estabelecer reunião com o gerente ligado a canais e com o coordenador do *E-mail* Marketig da organização. Nesta reunião, objetiva-se apresentá-los toda esta forma de abordagem baseada em *Lean* e que fora concebida juntamente ao time interno de campanhas de *E-mail* da organização em estudo.

Capítulo 5: Conclusão

5. Conclusão

Este trabalho foi desenvolvido para solucionar o problema levantado pelo time de campanhas da Loyalty S.A. com relação a ausência de metodologia específica para a gestão de seu principal canal de vendas – o *E-mail Marketing*.

Um dos pontos de destaque deste estudo está na metodologia inovadora utilizada para a criação do processo de gestão do canal E-mail, a Metodologia Lean. Conforme pôde-se perceber nas referências feitas a esta metodologia durante todo o texto e enfatizadas na seção de revisão da literatura, trata-se de uma forma de abordagem muito recente. Sua primeira referência é datada do ano de 2006 através dos estudos de Steve Blank. Todavia, as primeiras obras que tratam especificamente desta metodologia, citando pela primeira vez o termo Lean no contexto de *Startup*, têm data de 2013 (*Lean Startup* de Ries e *Lean Analytics* de Croll e Yoskovitz).

O processo desenvolvido neste trabalho, combinando Regressão Linear Multivariada com conceitos inovadores como Aprendizagem Validada e Ciclo “Construir-Medir-Aprender” representou uma forma prática de mesclar conhecimentos de cunho teórico-acadêmico com um modelo baseado em conceitos e demonstrações empíricas realizadas em ambiente empresarial.

Ressalta-se que o objetivo atingido decorrente desta união de conceitos está de acordo com que se espera do profissional graduando em Engenharia de Produção, uma vez que é de sua responsabilidade encontrar alternativas que venham a potencializar os ganhos e gerar vantagens competitivas as organizações.

Por fim, entrega-se ao time de campanhas e a gerência da Loyalty S.A. um processo estruturado em que coloca-se o cliente da organização no centro das discussões e estabelece-se formas de geração, validação e implementação de alternativas que venham a impulsionar os ganhos no relacionamento cliente - Loyalty S.A.

Além disso, como forma de testar o processo, procedeu-se o estudo dos fatores associados a abertura de *E-Mail Marketing* por parte dos clientes da Loyalty, descobrindo-se quais variáveis tendem a provocar elevação desta taxa de interação.

Como próximos passos deste projeto, destaca-se em primeiro lugar a implementação deste processo na organização em estudo e, uma vez implantando, a replicação de seu *modus*

operandi para o estudo dos outros fenômenos associados ao relacionamento cliente – organização.

Já em termos da academia, este trabalho representa uma integração entre conceitos completamente inovadores e ferramentas matemáticas tradicionais e, também, uma colaboração no avanço das pesquisas associadas a esta nova metodologia emergente – A *Lean Startup*.

Capítulo 6: Referências Bibliográficas

6. Referências Bibliográficas

ALMEIDA, Dagoberto A. LEAL, Fabiano. PINHO, Alexandre F. FAGUNDES, Liliane D. Gestão do Conhecimento na análise de falhas: mapeamento de falhas através de sistema de informação, Universidade Federal de Itajubá, 2006.

ANSARI, Assim. DILLARD William T. E-Customization, Columbia and Duke Universities, 2001.

BLANK, Steve. The Four Steps to the Epyphany, Lulu.com, 2006

CROLL, Alistair. YOSKOVITZ, Benjamin. *Lean Analytics: Use Data to Build a Better Startup Faster*. O'reilly, 2013

DOWLING, Grahame R., UNCLES Marking. Do Customer Loyalty Programs Really Work? Sloan Management Review, 1997.

DUFRENE, Debbie D. ENGELLAND, Brian T. LEHMAN, Carol M. PEARSON Rodney A. Changes in customer attitudes resulting from participation in a permission *E-mail Campaign*, Journal of Current Issues and Research in Advertising, 2005.

FILHO, Dalson B. F. e JUNIOR, José A. S. Desvendando os Mistérios do Coeficiente de Pearson (r)*, Revista Política Hoje, Vol. 18, n.1, 2009

GUNST, Richard F. e MASON, Robert L. Analysis and Application: A Data-Oriented Approach, Hardcover, 1980

JONHSON, Richard A. WICHERN, Dean W. Applied multivariate Statistical Analysis. Prentice Hall, 2009

KRISHNAMURTHY, S., A Comprehensive Analysis of permission *Marketing*, Journal of Computer mediated Communication, 2001.

MATOS, Manuel A. Manual Operacional para a Regressão Linear, Faculdade de Engenharia da Universidade do Porto (FEUP), 1995.

MCKENNA, Regis. *Marketing de Relacionamento – Estratégias Bem-Sucedidas para a Era do Cliente*, Elsevier, 1994.

MINGOTI, Sueli A. “Análise de Dados Através de Métodos de Estatística Multivariada – uma Abordagem Aplicada”, Editora UFMG, 2007

MINZUNO, S. Gerência para Melhoria da Qualidade – As Sete novas ferramentas do controle da Qualidade – Editora LTC, 1993.

MONTAGUE , B. Spam, spam, spam, spam... you've got mail. Sunday Times, 2006

MORI, Danilo T. Construção de um modelo de Regressão para Previsão da Inflação, Escola Politécnica da Universidade de São Paulo, 2007

O'CONNOR, Peter. *E-mail Marketing by International Hotel Chains: An Industry-Practices Update*, Sage, 2008.

PAPADATOS, Carolina. Global strategies, Local Tactics – How global Economies are sparking creativity for Loyalty Fundamentals, Colloquy, 2013.

PESCADOR, Paula C.C., *Marketing de Relacionamento: Estudo de Caso da Implantação dos Programas de Fidelidade na AMBEV – CIA Brasileira de Bebidas S/A – Centro de Distribuição Florianópolis*, Universidade Federal de Santa Catarina, 2004.

RIES, Eric, *A Startup Enxuta (The Lean Startup)*, Leya, 2013

SERRA, Fernando A. Ribeiro. TORRES, Alexandre P. TORRES, Maria Cândida S. Administração Estratégica: conceitos, roteiros práticos e casos. Reichmann & Affonso Editores, 2004.

TASSABEHJI, Rana e VAKOLA Maria, Business *E-mail: The Killer Impact*, ACM, 2005

Referências Disponíveis em meio eletrônico:

Return Path. <blog.returnpath.com/blog/stephanie-mill>. Acesso em 19 de Outubro de 2014.

Multiplus S.A. <ri.pontosmultiplus.com.br> Acesso em 20 de Outubro de 2014

Smiles S.A.<ri.smiles.com.br> Acesso em 20 de Outubro de 2014

Capítulo 7: Anexos

7. Anexos

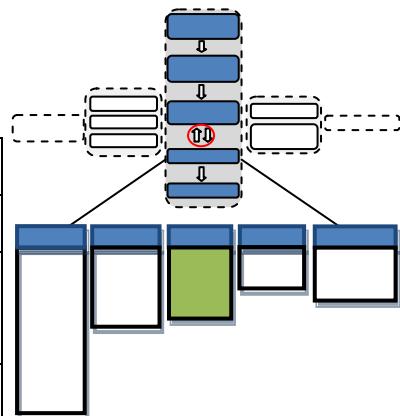
7.1. Anexo I: Tipos das variáveis utilizadas no modelo

A tabela a seguir, descreve o tipo de cada uma das variáveis mapeadas anteriormente:

Tabela 20: Descrição das variáveis

Índice	Variável	Tipo associado	
1	Dia da Semana	Categórica	Discreta
2	Tipo de comunicação	Categórica	Discreta
3	Tipo de Assunto	Categórica	Discreta
4	Percentual Impactado	Percentual	Contínua
5	Saldo Médio	Numérica	Contínua
6	Grau transacional	Percentual	Contínua

Desenvolvido pelo próprio autor

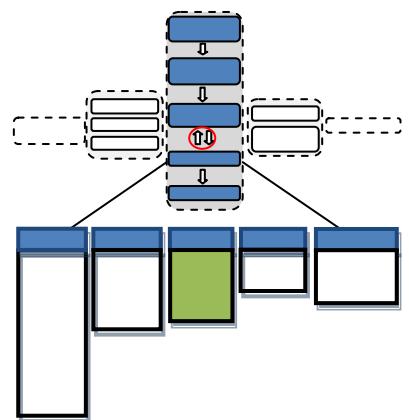


Por variável categórica, entende-se aquelas que cujos valores assumidos são categorias, como por exemplo, cada um dos dias da semana.

As variáveis percentuais e numéricas, diferentemente das categóricas, são variáveis contínuas, cujos valores são números.

7.2. Anexo II: Transformação binária das variáveis “Tipo de comunicação” e “Tipo de Assunto”

A variável categórica “Tipo de Comunicação” assume apenas dois valores “Accural” e “Redemption”. Aplicando-se, também, a expressão matemática função logaritmo ($\log(2;2)$) obtém-se como resultado o número, já inteiro, 2 (dois). Deste modo, deve-se desdobrar a variável “Tipo de Comunicação” em duas variáveis: “X1_tipo_comunicação” e “X2_tipo_comunicação”.



A tabela a seguir mostra a transformação binária para os valores da variável categórica “Tipo de Comunicação”.

Tabela 21: Transformação dos valores assumidos pela variável “Tipo de Comunicação”

	X1_tipo_comunicação	X2_tipo_comunicação
Acúmulo	1	0
Resgate	0	1

Desenvolvido pelo próprio autor

Deste modo, a categoria “Acúmulo”, por exemplo, relaciona-se de modo único ao par (1;0). Assim como ocorre na transformação da variável “Dia da Semana” as variáveis auxiliares de “Tipo de Comunicação” assumem pares de “credits” únicos, assim como a cada par de pontos também corresponde uma única categoria.

Por fim, a variável “Tipo de Assunto”, categorizada utilizando-se o “Diagrama de Afinidade”, pode assumir 8 (oito) valores, que representam os tipos de assunto das comunicações de *E-mail Marketing* do case em estudo: “Com informação da peça”, “Com parceiros”, “Genérico”, “Personalizado”, “Misto”, “Promocional”, “Promocional – Urgência”.

Aplicando-se, mais uma vez, a expressão matemática conhecida como função logarítmica ($\log(8;2)$) encontrou-se o número inteiro 3 (três). Com isso, conclui-se que a variável categórica “Tipo de Assunto” deve ser subdividida nas variáveis auxiliares “X1_assunto”, “X2_assunto” e “X3_assunto”.

A tabela a seguir nos mostra como foi feita a transformação binária para esta variável categórica:

Tabela 22: Transformação binária dos valores assumidos pela variável categórica “Tipo de Assunto”

	X1_assunto	X2_assunto	X3_assunto
Com informação da peça	1	0	0
Com parceiros	0	1	0
Detalhado	0	0	1
Genérico	1	1	0
Misto	1	0	1
Personalizado	0	1	1
Promocional	1	1	1
Promocional – Urgência	0	0	0

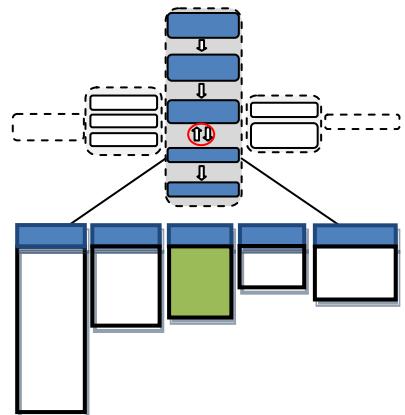
Desenvolvido pelo próprio autor

Tomando-se como exemplo a categoria “Genérico”, pode-se dizer que é representada unicamente pela trinca (1;1;0). Novamente, é importante destacar que a cada categoria corresponde uma única trinca e vice-versa, garantido, portanto, a unicidade da solução da análise multivariada.

7.3. Anexo III: Padronização das variáveis

- a. Padronização dos valores da variável X1_tipo_comunicação

Repete-se, mais uma vez, o mesmo processo para a variável X1_tipo_comunicação. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:



$$(X1_tipo_comunicação - 0,3412) / 0,4750.$$

O quadro a seguir nos traz um exemplo com 10 valores associados a X1_tipo_comunicação que sofreram tal transformação:

X1_tipo_comunicação	X1_tipo_comunicação
0,0000	-0,7182
0,0000	-0,7182
0,0000	-0,7182
1,0000	1,3869
0,0000	-0,7182
0,0000	-0,7182
0,0000	-0,7182
0,0000	-0,7182
0,0000	-0,7182
1,0000	1,3869

Quadro 9: Transformação da variável X1_tipo_comunicação

- b. Padronização dos valores da variável X2_tipo_comunicação

Repete-se, mais uma vez, o mesmo processo para a variável X2_tipo_comunicação. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:

$$(X2_tipo_comunicação - 0,6588) / 0,4750.$$

O quadro a seguir nos traz um exemplo com 10 valores associados a X2_tipo_comunicação que sofreram tal transformação:

X2_tipo_comunicação
1,0000
1,0000
1,0000
0,0000
1,0000
1,0000
1,0000
1,0000
1,0000
0,0000

→

X2_tipo_comunicação
0,7182
0,7182
0,7182
-1,3869
0,7182
0,7182
0,7182
0,7182
0,7182
-1,3869

Quadro 10: Transformação da variável X2_tipo_comunicação

c. Padronização dos valores da variável X1_assunto

Repete-se, mais uma vez, o mesmo processo para a variável X1_assunto. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:

$$(X1_assunto - 0,4704) / 0,5005.$$

Quadro 5.3 a seguir nos traz um exemplo com 10 valores associados a X1_assunto que sofreram tal transformação:

X1_assunto	X1_assunto
1,0000	1,0421
1,0000	1,0421
0,0000	-0,9559
0,0000	-0,9559
0,0000	-0,9559
0,0000	-0,9559
0,0000	-0,9559
0,0000	-0,9559
1,0000	1,0421
0,0000	-0,9559

Quadro 11: Transformação da variável X1_assunto

d. Padronização dos valores da variável X2_assunto

Repete-se, mais uma vez, o mesmo processo para a variável X2_assunto. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:

$$(X2_{assunto} - 0,7098) / 0,4547.$$

O quadro a seguir nos traz um exemplo com 10 valores associados a X2_assunto que sofreram tal transformação:



X2_assunto	X2_assunto
0,0000	-1,5609
1,0000	0,6382
1,0000	0,6382
1,0000	0,6382
1,0000	0,6382
1,0000	0,6382
1,0000	0,6382
1,0000	0,6382
1,0000	0,6382
0,0000	-1,5609

Quadro 12: Transformação da variável X2_assunto

e. Padronização dos valores da variável X3_assunto

Repete-se, mais uma vez, o mesmo processo para a variável X3_assunto. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:

$$(X3_{assunto} - 0,2471) / 0,4321.$$

O quadro a seguir nos traz um exemplo com 10 valores associados a X3_assunto que sofreram tal transformação:



X3_assunto	X3_assunto
0,0000	-0,5717
0,0000	-0,5717
0,0000	-0,5717
0,0000	-0,5717
0,0000	-0,5717
0,0000	-0,5717
0,0000	-0,5717
0,0000	-0,5717
0,0000	-0,5717
1,0000	1,7423

Quadro 13: Transformação da variável X3_assunto

f. Padronização dos valores da variável “saldo médio”

Repete-se, mais uma vez, o mesmo processo para a variável “saldo médio”. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:

$$(\text{“saldo médio”} - 70.891,9791)/3.263,8123.$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “saldo médio” que sofreram tal transformação:



“saldo médio”	“saldo médio”
6.851	-0,2152
30.766	-0,2078
5.044	-0,2157
15.568	-0,2125
972.688	0,0808
974.641	0,0814
171.483	-0,1647
171.517	-0,1647
1.102	-0,2169
4.030	-0,2160

Quadro 14: Transformação da variável “saldo médio”

g. Padronização dos valores da variável “segmentação utilizada”

Repete-se, mais uma vez, o mesmo processo para a variável “segmentação utilizada”. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:

$$(\text{“segmentação utilizada”} - 0,2598) / 0,2432.$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “segmentação utilizada” que sofreram tal transformação:

“percentual impactado”	“percentual impactado”
0,37	0,4729
0,14	-0,5071
0,49	0,9618
0,27	0,0459
0,03	-0,9568
0,03	-0,9568
0,05	-0,8774
0,05	-0,8776
0,67	1,6846
0,81	2,2637

Quadro 15: Transformação da variável “saldo médio”

h. Padronização dos valores da variável “grau transacional”

Repete-se, mais uma vez, o mesmo processo para a variável “percentual transacional”. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:

$$(\text{“grau transacional”} - 0,1361) / 0,2410.$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “percentual transacional” que sofreram tal transformação:



“grau transacional”	“grau transacional”
0,00	0,5517
0,00	-0,5516
0,00	0,5607
0,72	2,4243
0,08	-0,2405
0,08	-0,2518
0,07	-0,2745
0,07	-0,2758
0,08	-0,2507
0,00	-0,5603

Quadro 16: Transformação da variável “grau transacional”

i. Padronização dos valores da variável “taxa de abertura”

Por fim, executa-se o mesmo processo para a variável “taxa de abertura”. Por motivos confidenciais, os valores destas variável encontram-se multiplicados por uma constante K. Para o caso específico desta variável, a normalização consiste em aplicar a cada um de seus 255 valores a seguinte formula:

$$(\text{“taxa de abertura”} - 0,6776) / 0,3248.$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “taxa de abertura” que sofreram tal transformação:

The diagram illustrates the transformation of the "taxa de abertura" variable. On the left, a vertical list of raw values is shown in a table. An arrow points to the right, where another table shows the corresponding standardized values.

"taxa de abertura"
0,55
0,49
0,44
0,31
0,88
0,88
0,76
0,76
0,41
0,40

→

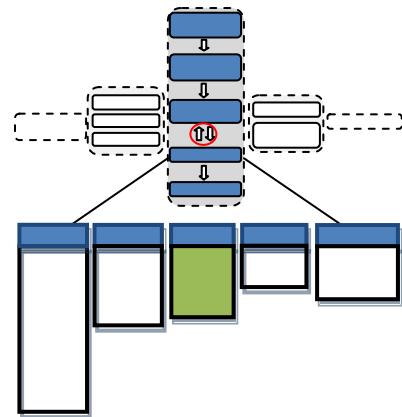
"taxa de abertura"
-0,4051
-0,5876
-0,7192
-1,1207
0,6273
0,6114
0,2662
0,2559
-0,8147
-0,8624

Quadro 17: Transformação da variável “taxa de abertura”

7.4. Anexo IV: Regressão Linear no PASW Statistics

a. Importação dos dados

O primeiro passo para a realização da análise no software estatístico PASW Statistics é a importação de dados do banco de convertido contido no software Microsoft Office Excel 200 para o estatístico.



Isto é feito atualmente, bastando selecionar, na opção de “File -> Open”, o arquivo que contém a planilha de dados que está sendo trabalhada e o formato .xlsx.

A figura 12 nos mostra os dados já importados para o software estatístico:

	X1_dia_semana	X2_dia_semana	X3_dia_semana	x1_TIPO_COMUNICAÇÃO	x2_TIPO_COMUNICAÇÃO	X1_Subject	X2_Subject	X3_Subject	saldomédio	GraudeSegurança	Grautransacional	taxadea
1	.9117	-1.6889	1.0421	-.7182	.7182	1.0421	-1.5609	-,5717	-,2152	,4729	-,5571	
2	-1.0925	.5898	-,9559	-,7182	,7182	1.0421	,6382	-,5717	-,2078	-,5071	-,5516	
3	-1.0925	-1.6889	1.0421	-,7182	,7182	-,9559	,6382	-,5717	-,2157	,9618	-,5607	
4	-1.0925	-1.6889	1.0421	1.3869	-,3869	-,9559	,6382	-,5717	-,2125	,0459	2.4243	
5	,9117	.5898	-,9559	-,7182	,7182	-,9559	,6382	-,5717	,0808	-,9568	-,2405	
6	,9117	.5898	-,9559	-,7182	,7182	-,9559	,6382	-,5717	,0814	-,9568	-,2518	
7	,9117	.5898	-,9559	-,7182	,7182	-,9559	,6382	-,5717	-,1647	,8774	-,2745	
8	,9117	.5898	-,9559	-,7182	,7182	-,9559	,6382	-,5717	-,1647	,8776	-,2758	
9	,9117	.5898	-,9559	-,7182	,7182	1.0421	,6382	-,5717	-,2169	,1.6846	-,2507	
10	-1.0925	.5898	1.0421	1.3869	-,3869	-,9559	-1.5609	1.7423	-,2160	2.2637	-,5623	
11	,9117	.5898	1.0421	1.3869	-,3869	1.0421	,6382	1.7423	-,2160	2.2637	2.1447	
12	-1.0925	.5898	-,9559	-,7182	,7182	-,9559	,6382	-,5717	-,1654	,4856	-,5613	
13	-1.0925	.5898	-,9559	-,7182	,7182	-,9559	,6382	-,5717	-,2137	,2696	-,5615	
14	-1.0925	.5898	-,9559	1.3869	-,3869	1.0421	,6382	-,5717	-,2172	,8724	-,5569	
15	-1.0925	-1.6889	1.0421	-,7182	,7182	-,9559	-1.5609	1.7423	-,2142	,0180	,7550	
16	-1.0925	-1.6889	1.0421	1.3869	-,3869	-,9559	,6382	-,5717	-,2050	,3106	-,4190	
17	,9117	.5898	-,9559	-,7182	,7182	-,9559	,6382	-,5717	-,0637	,8528	-,2231	
18	,9117	.5898	-,9559	-,7182	,7182	-,9559	,6382	-,5717	-,1900	,6899	-,2552	
19	,9117	.5898	-,9559	-,7182	,7182	1.0421	,6382	-,5717	-,2133	,3140	-,2452	
20	,9117	.5898	-,9559	1.3869	-,3869	-,9559	-1.5609	1.7423	-,2172	,8958	,0137	
21	-1.0925	.5898	1.0421	-,7182	,7182	1.0421	,6382	-,5717	-,2161	,2.2697	-,5629	
22	-1.0925	.5898	-,9559	-,7182	,7182	-,9559	,6382	-,5717	-,2141	,1.703	-,5616	
23	-1.0925	.5898	-,9559	-,7182	,7182	-,9559	,6382	-,5717	-,1900	,3957	-,5617	
24	-1.0925	.5898	-,9559	1.3869	-,3869	1.0421	-1.5609	-,5717	-,2172	1.0258	-,5497	

Figura 12: Amostra de dados no software PASW Statistics 17

b. Regressão Linear

Como todos os dados já encontram-se previamente trabalhados para a realização da regressão linear multivariada, pode-se partir, então, direto a sua realização.

A figura a seguir nos mostra os parâmetros definidos na análise estatística:

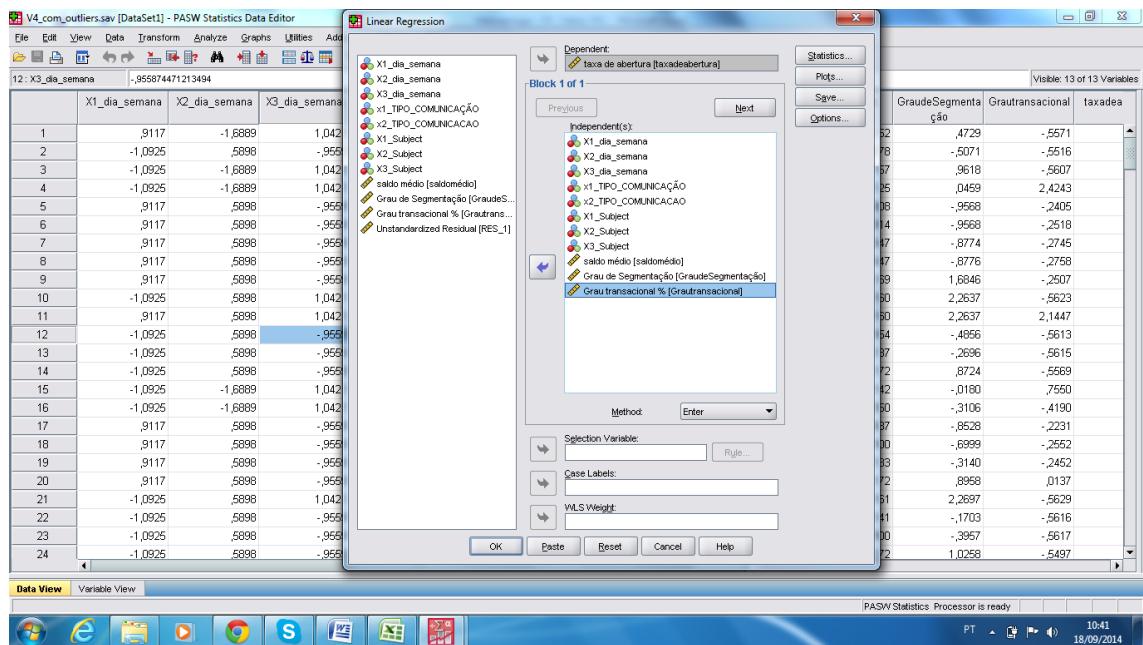


Figura 13: Inserindo os parâmetros da regressão no software estatístico

Conforme ilustrado na figura 13, seleciona-se a variável “Taxa de Abertura” como a variável dependente (aquele que deseja-se estudar o comportamento a partir de outras variáveis) e as demais (X1_dia_semana, X2_dia_semana, X3_dia_semana, X1_tipo_comunicação, X2_tipo_comunicação, X1_assunto, X2_assunto, X3_assunto, “Saldo médio”, “Segmentação utilizada” e “Grau transacional”) como as variáveis independentes (as utilizadas para explicar o comportamento da dependente).

Na aba Statistics, ilustrada na figura 14., seleciona-se os parâmetros de controle da análise a ser mostrados nas telas de saída do software.

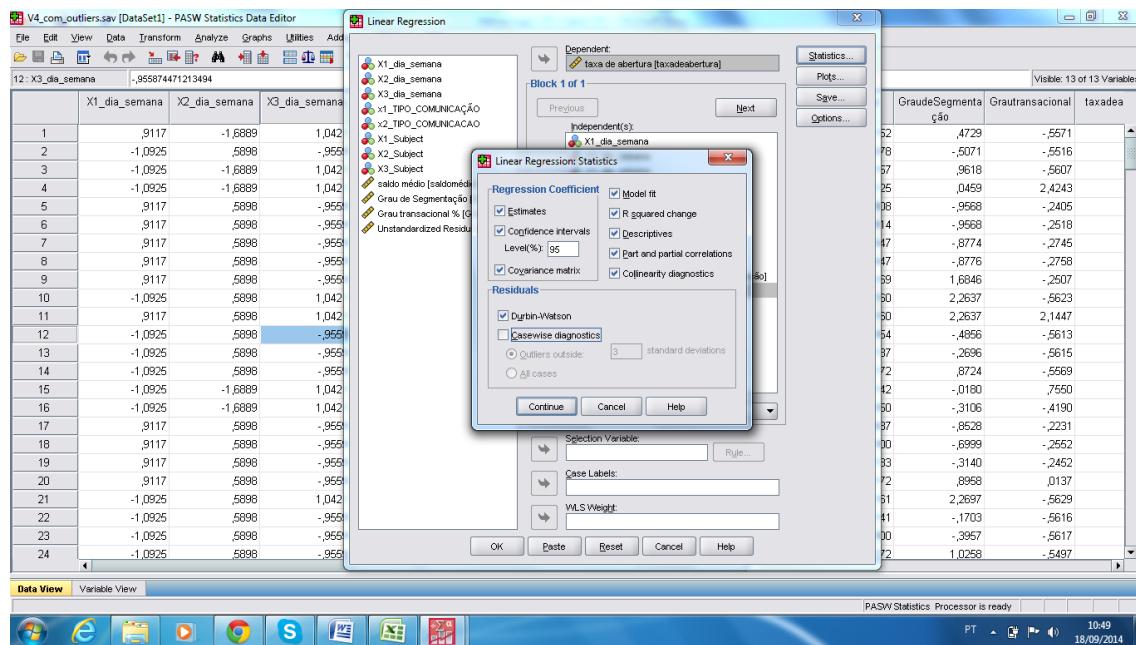


Figura 14: Seleção dos parâmetros de controle a ser mostrados na tela de saída da análise multivariada

Conforme detalhado na seção de revisão literária, um ponto fundamental para validação do modelo de regressão utilizado está na realização da análise de resíduos. Na aba “Plots”, seleciona-se os parâmetros “*ZPRED” E “*ZRESID”, para construção do gráfico que é parte da análise residual a priori, obtida junto a regressão linear multivariada.

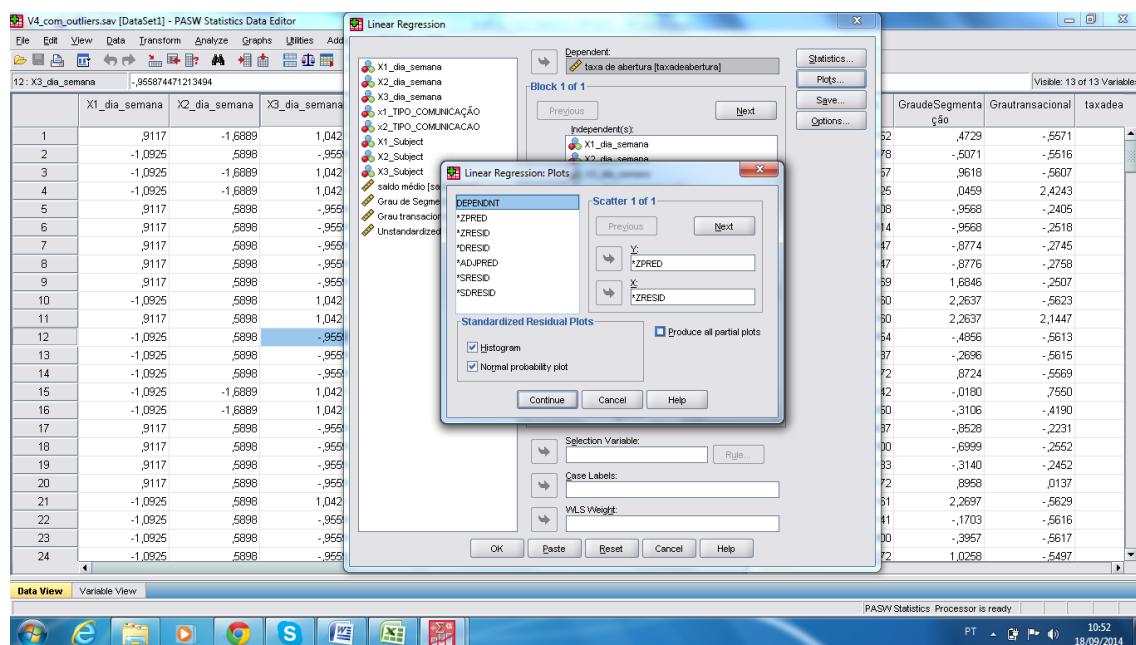


Figura 15: Parâmetros da análise residual a priori.

Visando realizar a análise de resíduos em sua completude, deve-se, na tela save, salvar o parâmetro “Residuals unstandardized”. A figura 16 ilustra este passo.

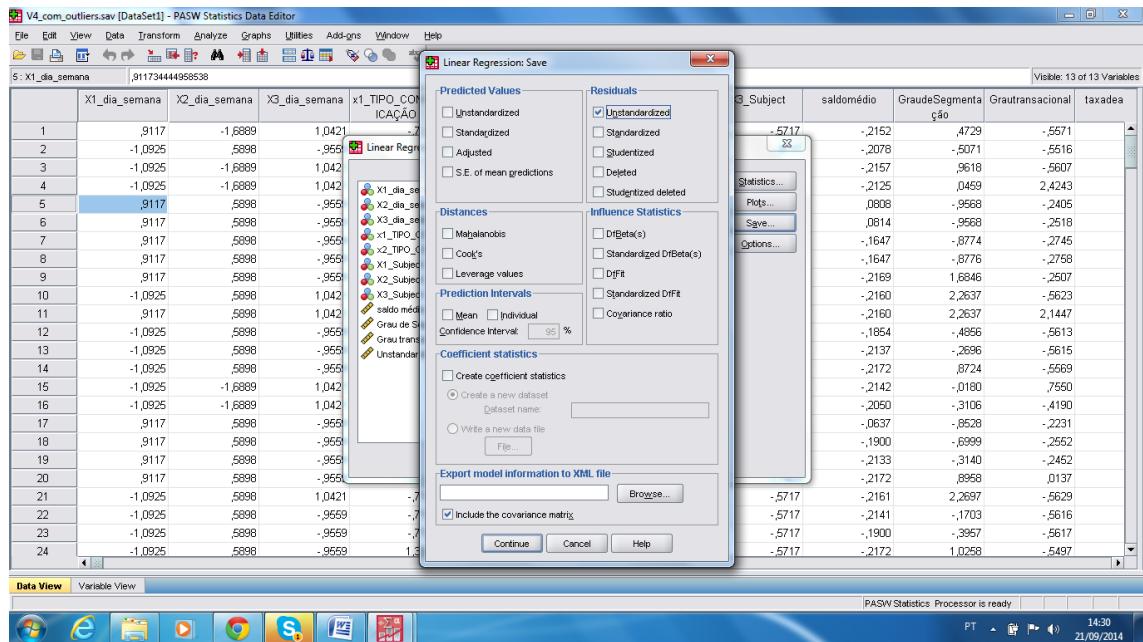


Figura 16: Parâmetros residuais para a análise de resíduos

Na guia “Options”, o PASW Statistics já traz, por default, o critério utilizado na análise de regressão linear, conforme citado na seção de revisão literária. Desta forma, o método utiliza-se da probabilidade F, tendo 0,05 como parâmetro de entrada e 0,10 como parâmetro de saída. Além disso, a opção “Include Constant in equation”, relacionada a inclusão da constante na equação da regressão já vêm selecionada. A figura 17 exibe estas informações:

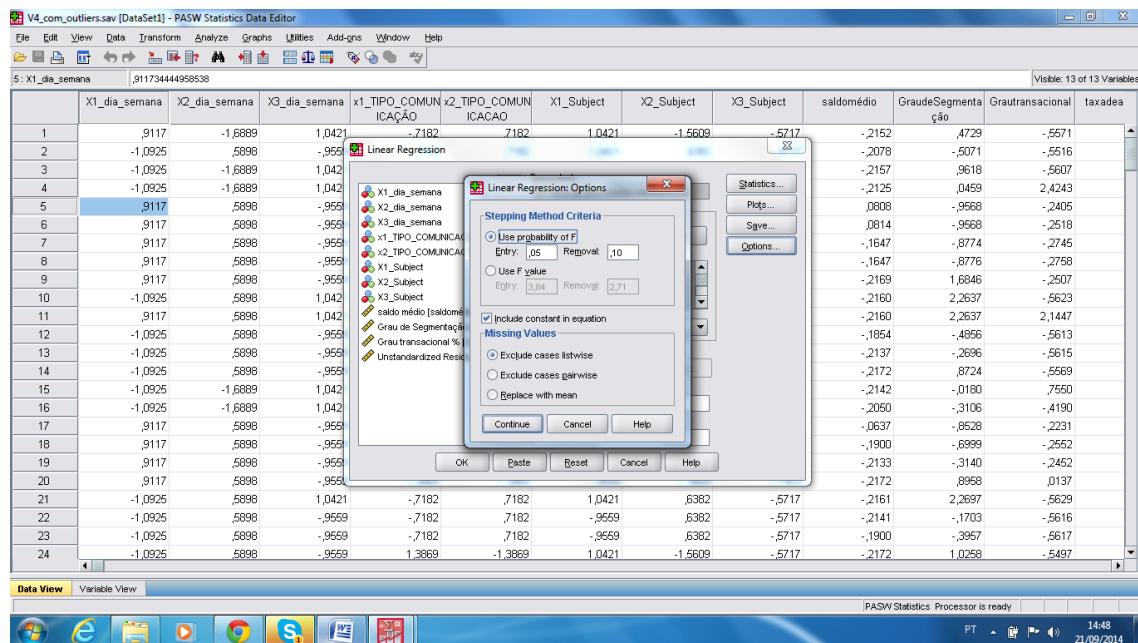


Figura 17: Critérios utilizados na regressão linear multivariada.

Pós execução de todas essas passagens, o próximo passo consiste em realizar a análise propriamente no software estatístico.

c. Análise de resíduos

Pós-execução da análise estatística via técnica de regressão linear multivariada, deve-se, procurando assegurar os pressupostos deste método detalhados na seção destinada a revisão literária, promover a análise de resíduos.

No software estatístico PASW Statistics 17, tal análise pode ser feita selecionando-se a opção “Analyze => Descriptive Statistics => Explore”. Selecionada tal opção, o software nos exibe a tela ilustrada na figura 18:

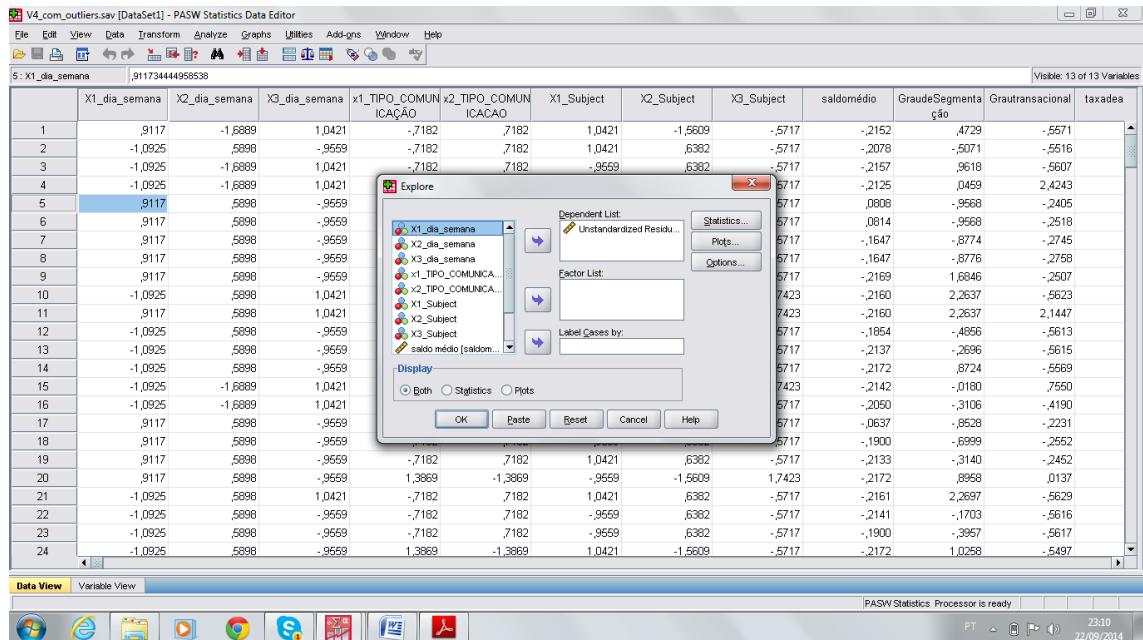


Figura 18: Tela inicial da análise de resíduos

O parâmetro base desta análise é o “Unstandardized Residual”. Tal parâmetro foi gerado e armazenado gravado na opção “Save” durante a realização da regressão linear.

O próximo passo consiste em selecionar as análises e os gráficos que deseja-se que o Software nos exiba a fim de embasar a análise residual.

Seguindo-se a literatura considerada neste estudo, a primeira análise a ser feita é o teste de normalidade (*normality test*) procurando identificar se os resíduos apresentam distribuição normal, condição fundamental na regressão linear. A figura a seguir exibe a tela em que se seleciona tal opção:

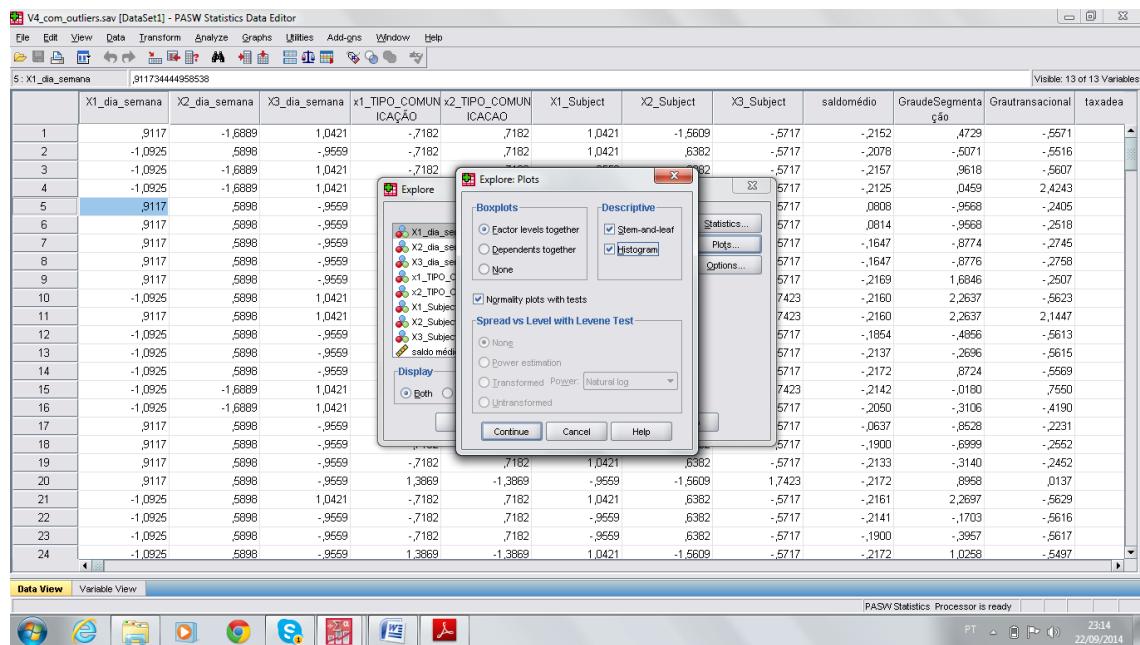


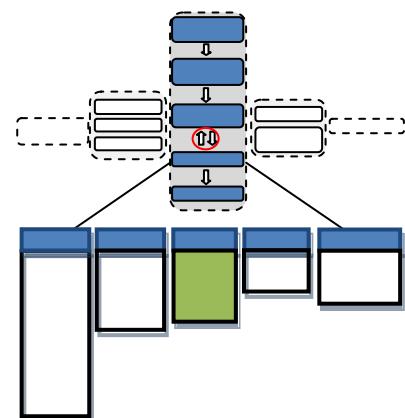
Figura 19: Selecionando-se os parâmetros e gráficos a ser exibidos na análise de resíduos

Pode-se perceber que a figura 19 exibe, além da seleção do teste de normalidade, os gráficos que deseja-se que o software exiba para embasar a análise realizada. São eles:

- *Histogram;*
- *Normal Q-Q Plot of Unstandardized Residual;*
- *Detrended Normal Q-Q Plot of Unstandardized Residual; e*
- *Observed Value.*

7.5. Anexo V: Resultado da regressão linear I

Denomina-se regressão linear I aquela realizada com os dados pós padronização, sem as transformações com aplicação das transformações quadrática ou com uso da raiz cúbica dos dados. Trata-se da primeira tentativa de geração de hipóteses a ser validadas.



a. Estatísticas Descritivas

Tabela 23: Resultados das estatísticas descritivas para regressão com dados padronizados

Variável	Média	Std. Deviation	N
taxa de abertura	,000000	1,0000000	255
X1_dia_semana	,000000	1,0000000	255
X2_dia_semana	,000000	1,0000000	255
X3_dia_semana	,000000	1,0000000	255
x1_TIPO_COMUNICAÇÃO	,000000	1,0000000	255
x2_TIPO_COMUNICACAO	,000000	1,0000000	255
X1_Assunto	,000000	1,0000000	255
X2_Assunto	,000000	1,0000000	255
X3_Assunto	,000000	1,0000000	255
saldo médio	,000000	1,0000000	255
Grau de Segmentação	,000000	1,0000000	255
Grau transacional %	,000000	1,0000000	255

Desenvolvido pelo próprio autor

A tabela 23 nos traz os resultados das estatísticas descritivas aplicadas aos dados pós-padronização. O efeito da normalização nos dados é bastante claro, uma vez que todas as variáveis apresentam média 0 (zero) e desvio-padrão unitário (1). Destaca-se que todas as variáveis assumem 255 valores possíveis, os quais, conforme comentado em seções anteriores remetem a disparos de ações de *E-mail Marketing*, foco deste estudo.

b. Análise de correlações

Um ponto importante da análise realizada está na verificação das correlações entre a variável dependente (taxa de abertura) e as variáveis independentes, de modo a, após determinação dos coeficientes da equação da regressão, possa-se verificar como se dá a relação dentre elas e a taxa de abertura.

Tabela 24: Correlações de Pearson entre a taxa de Abertura e as variáveis independentes

Variável	Correlação de Pearson	Sig. (1-tailed)
Taxa de Abertura	1,000	-
X1_dia_semana	0,000	0,4990
X2_dia_semana	-0,127	0,022
X3_dia_semana	0,067	0,143
X1_tipo_comunicação	0,037	0,279
X2_tipo_comunicação	-0,037	0,279
X1_assunto	0,086	0,085
X2_assunto	0,131	0,019
X3_assunto	0,108	0,042
Saldo Médio	0,523	0,000
Percentual impactado	-0,480	0,000
Grau Transacional	0,701	0,000

Desenvolvido pelo próprio autor

Pode-se extrair da tabela 24 que as variáveis que possuem correlação estatisticamente válida, a significância padrão de 5%, com a variável em estudo são:

- X2_dia_semana: Correlação positiva (+)
- X1_assunto: Correlação positiva (+)
- X2_assunto: Correlação positiva (+)
- X3_assunto: Correlação positiva (+)
- Saldo médio: Correlação positiva (+)
- Percentual impactado: Correlação negativa (-)
- Grau transacional: Correlação positiva (+)

c. Variáveis utilizadas no modelo

Com exceção da variável “X1_tipo_comunicação”, todas as outras foram utilizadas no modelo em análise. Tal variável foi retirada por apresentar uma forte correlação negativa (-1,000) com a variável “X2_tipo_comunicação”. Esta retirada se dá de forma automática na modelagem via software estatístico.

d. ANOVA

A fim de medir-se a robustez do modelo, pode-se realizar a análise de variância (ANOVA) e verificar o valor assumido por sua significância, lembrando-se, que, de acordo com a revisão de literatura, valores inferiores a 5% são considerados aceitáveis.

Na tabela 25 está expressa o resultado da ANOVA realizada:

Tabela 25: ANOVA para modelo 1 – após a padronização das variáveis

Modelo		Soma de Quadrados	Graus de liberdade	Quadrado Médio	F	Significância
1	Regressão	172,698	10	17,270	51,830	0,000
	Residual	81,302	244	0,333		
	Total	254,000	254			

Desenvolvido pelo próprio autor

Logo, a significância neste caso é igual a 0,000, o que permite-se concluir que o modelo é robusto.

e. Resumo do Modelo

A tabela 26 nos mostra algumas informações relacionadas ao modelo aplicado:

Tabela 26: resumo das informações do modelo 1

Modelo	R	R ²	Durbin-Watson
1	0,680	0,667	2,030

Desenvolvido pelo próprio autor

Conforme ilustrado na tabela 26, o modelo apresenta um poder de explicação da variação da taxa de abertura da ordem de 67% (R²).

Já o teste de Durbin-Watson, parte da análise de resíduos, apresenta um valor próximo a 2, o que, conforme comentado na seção destinada a revisão de literatura garante que os erros são independentes.

f. Coeficientes da equação da regressão

Os coeficientes da regressão (peso das variáveis na explicação da variação da taxa de abertura) e sua respectiva significância para o modelo de regressão estão expostos na tabela 27 a seguir:

Tabela 27: Coeficiente e sua respectiva significância para o modelo 1.

Modelo	Variável	Coeficiente	Significância
1	(Constante)	3,846 E -9	1,000
	X1_dia_semana	-0,033	0,386
	X2_dia_semana	0,039	0,403
	X3_dia_semana	0,016	0,724
	X2_tipo_comunicação	0,063	0,157
	X1_assunto	0,065	0,094
	X2_assunto	0,100	0,023
	X3_assunto	0,154	0,001
	Saldo médio	0,037	0,429
	Percentual impactado	-0,393	0,000
	Grau transacional	0,632	0,000

Desenvolvido pelo próprio autor

Conforme comentado na seção destinada a revisão de literatura, deve-se verificar a significância estatística para determinar se a variável exerce influência considerável na variação da taxa de abertura. Para aquelas que mostrarem-se significantes, deve-se medir seu poder de influencia pelo módulo de seu coeficiente B.

Deste modo, as variáveis em questão são, nesta ordem: “Grau transacional, Percentual impactado, X3_assunto e X2_assunto”.

g. Validando os resultados

Conforme comentado na seção destinada a descrição da metodologia de pesquisa e da revisão bibliográfica, utilizou-se o método de análise de resíduos a fim de validar os resultados encontrados.

O primeiro passo para verificar se os resíduos apresentam distribuição normal pode ser dado plotando-se os resíduos no papel de distribuição normal e analisando-se seu comportamento. A figura 20 exibe este teste:

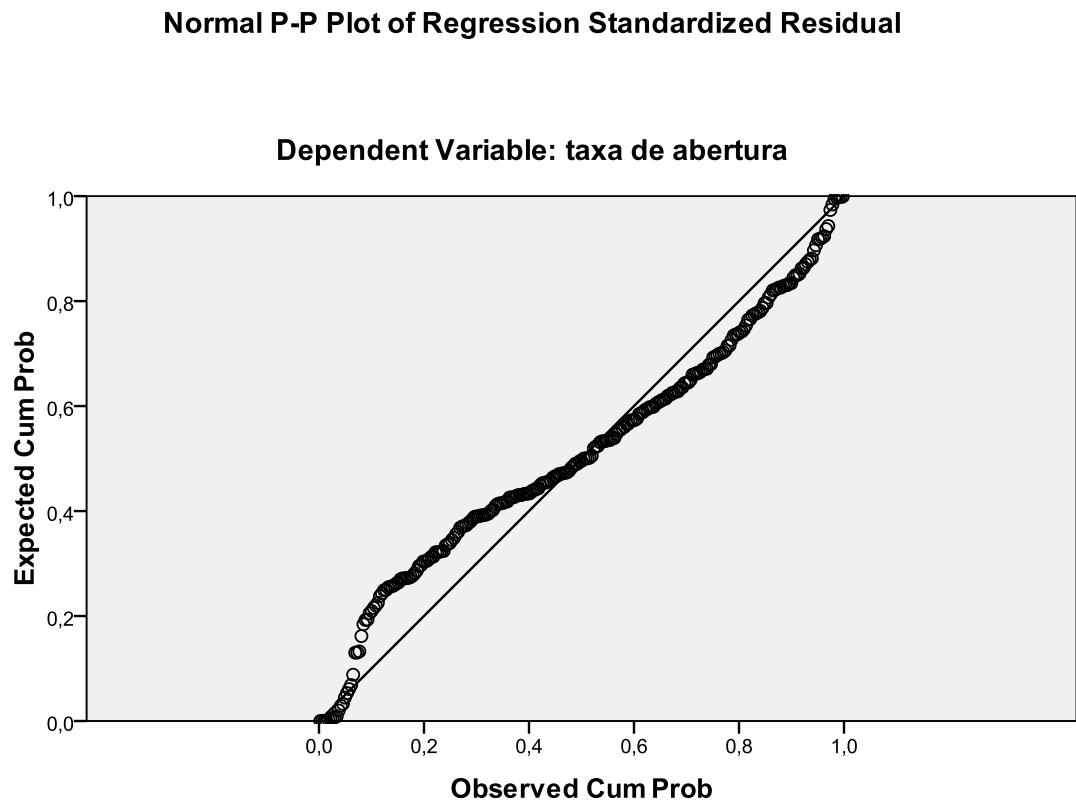


Figura 20: Normal *Plot-test* para análise de resíduos.

De fato, o resultado da primeira análise de resíduos não é nada animador, uma vez que os resultados não se comportam como uma reta no papel de probabilidade normal. Porém, antes de concluir-se algo a respeito dos resíduos, proceder-se-á dois outros testes – o *Scatterplot* e os testes de Kolmogorov-Smirnov e Shapiro-Wilk.

No *Scatterplot*, conforme comentando na seção destinada a revisão de literatura, dados normais distribuem-se aleatoriamente pelo plano. A figura a seguir traz o *Scatterplot* referente aos resíduos do modelo em análise:

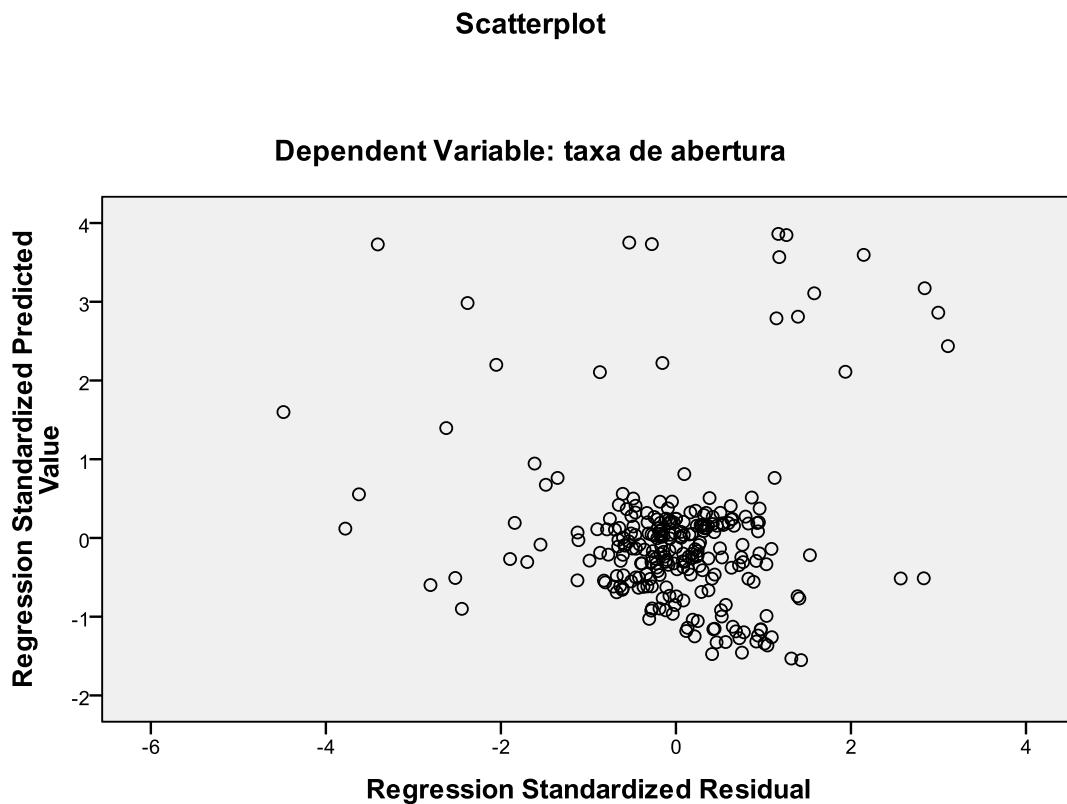


Figura 21: *Scatterplot* dos resíduos do modelo em análise

Novamente, os resíduos apresentaram resultados que os afastam da possibilidade de serem normais, uma vez que no *Scatterplot* fica claro que os resíduos apresentam uma concentração em torno do ponto (0;0).

Por fim, realizou-se o testes de Kolmogorov-Smirnov e Shapiro-Wilk. Para que sejam considerados normais, a significância destes testes deve ser maior do que 5%, uma vez que, para valores inferiores a 5%, pode-se interpretar que os dados apresentam uma diferença significativamente considerável em relação a curva normal.

A tabela 28 traz os resultados destes testes:

Tabela 28: Resultados dos testes de normalidade para o modelo 1

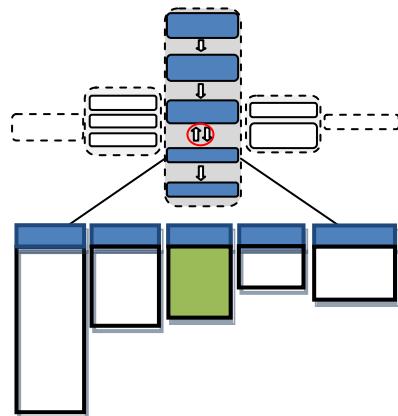
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	DF	Sig.
<i>Unstandardized Residual</i>	0,123	255	0,000	0,911	255	0,000

Desenvolvido pelo próprio autor

Logo, de acordo com os três testes de normalidade feitos, pode-se concluir que os resíduos não apresentam distribuição normal, o que não está de acordo com os princípios da regressão linear, conforme comentado na seção destinada a revisão de literatura.

7.6. Anexo VI: Transformação das variáveis para o refinamento do modelo (quadráticos)

Visando refinar o modelo, seguiu-se as sugestões retiradas da literatura trabalhando-se, num primeiro momento, com o quadrado dos dados transformados. Os quadros a seguir ilustram esta transformação para os primeiros 10 valores de cada um dos fatores:



A. Transformação da variável X1_dia_semana

O procedimento executado para transformação desta variável é o seguinte:

$$X1_dia_semana = (X1_dia_semana)^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X1_dia_semana” que sofreram tal transformação:

“X1_dia_semana”	(“X1_dia_semana”)²
0,91173	0,83126
-1,09251	1,19358
-1,09251	1,19358
-1,09251	1,19358
0,91173	0,83126
0,91173	0,83126
0,91173	0,83126
0,91173	0,83126
0,91173	0,83126
-1,09251	1,19358

Quadro 18: Transformação da variável “X1_dia_semana”

B. Transformação da variável X2_dia_semana

O procedimento executado para transformação desta variável é o seguinte:

$$X2_dia_semana = (X2_dia_semana)^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X2_dia_semana” que sofreram tal transformação:



“X2_dia_semana”	(“X2_dia_semana”)²
-1,68891	2,85241
0,58978	0,34786
-1,68891	2,85241
-1,68891	2,85241
0,58978	0,34786
0,58978	0,34786
0,58978	0,34786
0,58978	0,34786
0,58978	0,34786

Quadro 19: Transformação da variável “X2_dia_semana”

C. Transformação da variável X3_dia_semana

O procedimento executado para transformação desta variável é o seguinte:

$$X3_{dia_semana} = (X2_{dia_semana})^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X3_dia_semana” que sofreram tal transformação:



“X3_dia_semana”	(“X3_dia_semana”)²
1,04206	1,085889
-0,95587	0,91370
1,04206	1,085889
1,04206	1,085889
-0,95587	0,91370
-0,95587	0,91370
-0,95587	0,91370
-0,95587	0,91370
-0,95587	0,91370
1,04206	1,085889

Quadro 20: Transformação da variável “X3_dia_semana”

D. Transformação da variável X1_tipo_comunicação

O procedimento executado para transformação desta variável é o seguinte:

$$X1_tipo_comunicação = (X1_tipo_comunicação)^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X1_tipo_comunicação” que sofreram tal transformação:

“X1_tipo_comunicação”	(X1_tipo_comunicação) ²
-0,71821	0,51583
-0,71821	0,51583
-0,71821	0,51583
1,38689	1,92346
-0,71821	0,51583
-0,71821	0,51583
-0,71821	0,51583
-0,71821	0,51583
-0,71821	0,51583
1,38689	1,92346

Quadro 21: Transformação da variável “X1_tipo_comunicação”

E. Transformação da variável X2_tipo_comunicação

O procedimento executado para transformação desta variável é o seguinte:

$$X2_tipo_comunicação = (X2_tipo_comunicação)^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X2_tipo_comunicação” que sofreram tal transformação:

“X2_tipo_comunicação”	$(X2_tipo_comunicação)^2$
0,71821	0,51583
0,71821	0,51583
0,71821	0,51583
-1,38689	1,92346
0,71821	0,51583
0,71821	0,51583
0,71821	0,51583
0,71821	0,51583
-1,38689	1,92346

Quadro 22: Transformação da variável “X2_tipo_comunicação”

F. Transformação da variável X1_assunto

O procedimento executado para transformação desta variável é o seguinte:

$$X1_assunto = (X1_assunto)^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X1_assunto” que sofreram tal transformação:

“X1_assunto”	(“X1_assunto”) ²
1,04206	1,08589
1,04206	1,08589
-0,95587	0,91370
-0,95587	0,91370
-0,95587	0,91370
-0,95587	0,91370
-0,95587	0,91370
1,04206	1,08589
-0,95587	0,91370

Quadro 23: Transformação da variável “X1_assunto”

G. Transformação da variável X2_assunto

O procedimento executado para transformação desta variável é o seguinte:

$$X2_{assunto} = (X2_{assunto})^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X2_assunto” que sofreram tal transformação:

“X2_assunto”	(“X2_assunto”) ²
-1,56088	2,43635
0,63815	0,40724
0,63815	0,40724
0,63815	0,40724
0,63815	0,40724
0,63815	0,40724
0,63815	0,40724
0,63815	0,40724
0,63815	0,40724
-1,56088	2,43635

Quadro 24: Transformação da variável “X2_assunto”

H. Transformação da variável X3_assunto

O procedimento executado para transformação desta variável é o seguinte:

$$X3_{assunto} = (X3_{assunto})^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X3_assunto” que sofreram tal transformação:

“X3_assunto”	(“X3_assunto”) ²
-0,57170	0,32684
-0,57170	0,32684
-0,57170	0,32684
-0,57170	0,32684
-0,57170	0,32684
-0,57170	0,32684
-0,57170	0,32684
-0,57170	0,32684
-0,57170	0,32684
1,74232	3,03567

Quadro 25: Transformação da variável “X3_assunto”

I. Transformação da variável “Saldo Médio”

O procedimento executado para transformação desta variável é o seguinte:

$$\text{“saldo Médio”} = (\text{“Saldo Médio”})^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “Saldo Médio” que sofreram tal transformação:

“Saldo Médio”	(“Saldo Médio”) ²
-0,21518	0,04630
-0,20785	0,04320
-0,21573	0,04654
-0,21251	0,04516
0,08082	0,00653
0,08142	0,00663
-0,16472	0,02713
-0,16471	0,02713
-0,21694	0,04706
-0,21604	0,04667

Quadro 26: Transformação da variável “Saldo Médio”

J. Transformação da variável “Percentual impactado”

O procedimento executado para transformação desta variável é o seguinte:

$$\text{“Percentual impactado”} = (\text{“Percentual impactado”})^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “Percentual Impactado” que sofreram tal transformação:

“Percentual impactado”	(“Percentual impactado”) ²
-0,21518	0,22363
-0,20785	0,25710
-0,21573	0,92509
-0,21251	0,00210
0,08082	0,91546
0,08142	0,91540
-0,16472	0,76991
-0,16471	0,77009
-0,21694	2,83782
-0,21604	5,12435

Quadro 27: Transformação da variável “Percentual impactado”

K. Transformação da variável “Grau transacional”

O procedimento executado para transformação desta variável é o seguinte:

$$\text{“Grau transacional”} = (\text{“Grau transacional”})^2$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “Grau transacional” que sofreram tal transformação:

“Grau transacional”	(“Grau transacional”) ²
-0,55712	0,31038
-0,55163	0,30430
-0,56074	0,31443
2,42431	5,87729
-0,24049	0,05784
-0,25177	0,06339
-0,27451	0,07536
-0,25577	0,07605
-0,25070	0,06285
-0,56229	0,31617

Quadro 28: Transformação da variável “Grau transacional”

L. Transformação da variável “Taxa de abertura”

O procedimento executado para transformação desta variável é o seguinte:

$$\text{“Taxa de abertura”} = (\text{“Taxa de abertura”})^2$$

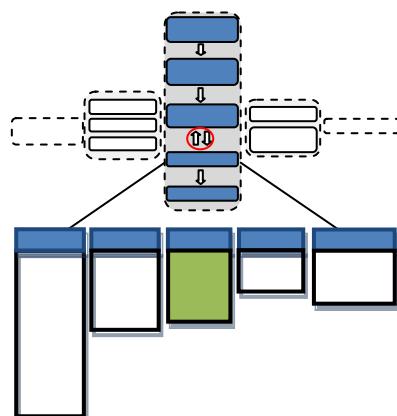
O quadro a seguir nos traz um exemplo com 10 valores associados a “taxa de abertura” que sofreram tal transformação:

“Taxa de abertura”	(“Taxa de abertura”)²
-0,45921	0,21087
-0,66613	0,44373
-0,81525	0,66463
-1,27041	1,61394
0,71107	0,50562
0,69301	0,48026
0,30170	0,09102
0,29012	0,08417
-0,92348	0,85282
-0,97759	0,95568

Quadro 29: Transformação da variável “Taxa de abertura”

7.7. Anexo VII: Resultado da regressão linear II

Denomina-se regressão linear II aquela realizada com os dados quadráticos, pós-realização da primeira transformação com foco em refinar o modelo gerado anteriormente, uma vez que seus resultados não se mostraram válidos em ambiente de testes prático.



a. Estatísticas Descritivas

Tabela 29: Resultados das estatísticas descritivas para regressão com dados padronizados

Variável	Média	Std. Deviation	N
taxa de abertura	,996078	2,9130340	255
X1_dia_semana	,996078	0,1807749	255
X2_dia_semana	,996078	1,0991299	255
X3_dia_semana	,996078	0,0861854	255
x1_TIPO_COMUNICAÇÃO	,996078	0,6686787	255
x2_TIPO_COMUNICACAO	,996078	0,6686787	255
X1_Assunto	,996078	0,0861854	255
X2_Assunto	,996078	0,9227316	255
X3_Assunto	,996078	1,1706191	255
saldo médio	,996078	8,0110115	255
Grau de Segmentação	,996078	1,3274981	255
Grau transacional %	,996078	2,8442704	255

Desenvolvido pelo próprio autor

A tabela 29 nos traz os resultados das estatísticas descritivas aplicadas aos dados quadráticos. Com a transformação realizada, nota-se que, agora, os dados não são mais normais, uma vez que apresentam média e variância diferentes de 1 e 0, respectivamente. Novamente, todas as variáveis assumem 255 valores possíveis, os quais, conforme comentado em seções anteriores remetem a disparos de ações de *E-mail Marketing*, foco deste estudo.

b. Análise de correlações

Um ponto importante da análise realizada está na verificação das correlações entre a variável dependente (taxa de abertura) e as variáveis independentes, de modo a, após determinação dos coeficientes da equação da regressão, possa-se verificar como se dá a relação dentre elas e a taxa de abertura.

Tabela 30: Correlações de Pearson entre a taxa de Abertura e as variáveis independentes

Variável	Correlação de Pearson	Sig. (1-tailed)
Taxa de Abertura	1,000	-
X1_dia_semana	0,043	0,248
X2_dia_semana	0,224	0,000
X3_dia_semana	0,136	0,015
X1_tipo_comunicação	0,361	0,000
X2_tipo_comunicação	0,361	0,000
X1_assunto	0,088	0,079
X2_assunto	0,007	0,455
X3_assunto	0,213	0,000
Saldo Médio	0,242	0,000
Percentual impactado	0,026	0,337
Grau Transacional	0,829	0,000

Desenvolvido pelo próprio autor

Pode-se extrair da tabela 30 que as variáveis que possuem correlação estatisticamente válida, a significância padrão de 5%, com a variável em estudo são:

- X2_dia_semana: Correlação positiva (+)
- X3_dia_semana: Correlação positiva (+)
- X1_tipo_comunicação: Correlação positiva (+)
- X2_tipo_comunicação: Correlação positiva (+)
- X3_assunto: Correlação positiva (+)
- Saldo médio: Correlação positiva (+)
- Grau transacional: Correlação positiva (+)

c. Variáveis utilizadas no modelo

Com exceção da variável “X1_tipo_comunicação”, todas as outras foram utilizadas no modelo em análise. Tal variável foi retirada por apresentar uma forte correlação positiva (1,000) com a variável “X2_tipo_comunicação”. Esta retirada se dá de forma automática na modelagem via software estatístico.

d. ANOVA

A fim de medir-se a robustez do modelo, pode-se realizar a análise de variância (ANOVA) e verificar o valor assumido por sua significância, lembrando-se, que, de acordo com a revisão de literatura, valores inferiores a 5% são considerados aceitáveis.

Na tabela 31 está expressa o resultado da ANOVA realizada:

Tabela 31: ANOVA para modelo 2

Modelo		Soma de Quadrados	Graus de liberdade	Quadrado Médio	F	Significância
2	Regressão	1592,140	10	159,214	68,972	0,000
	Residual	563,245	244	2,308		
	Total	2155,385	254			

Desenvolvido pelo próprio autor

Logo, a significância neste caso é igual a 0,000, o que permite-nos concluir que o modelo é robusto.

e. Resumo do Modelo

A tabela 32 nos mostra algumas informações relacionadas ao modelo aplicado:

Tabela 32: Resumo das informações do modelo 2

Modelo	R	R ²	Durbin-Watson
2	0,739	0,728	1,967

Desenvolvido pelo próprio autor

Conforme ilustrado na tabela 32, o modelo apresenta um poder de explicação da variação da taxa de abertura da ordem de 73% (R²).

Já o teste de Durbin-Watson, parte da análise de resíduos, apresenta um valor próximo a 2, o que, conforme comentado na seção destinada a revisão de literatura garante que os erros são independentes.

f. Coeficientes da equação da regressão

Os coeficientes da regressão (peso das variáveis na explicação da variação da taxa de abertura) e sua respectiva significância para o modelo de regressão estão expostos na tabela 33 a seguir:

Tabela 33: Coeficiente e sua respectiva significância para o modelo 2.

Modelo	Variável	Coeficiente	Significância
2	(Constante)	-0,697	0,698
	X1_dia_semana	-0,264	0,637
	X2_dia_semana	0,033	0,765
	X3_dia_semana	-1,765	0,197
	X2_tipo_comunicação	0,476	0,005
	X1_assunto	2,532	0,035
	X2_assunto	-0,259	0,039
	X3_assunto	0,271	0,007
	Saldo médio	-0,060	0,000
	Percentual impactado	-0,090	0,234
	Grau transacional	0,892	0,000

Desenvolvido pelo próprio autor

Conforme comentado na seção destinada a revisão de literatura, deve-se verificar a significância estatística para determinar se a variável exerce influência considerável na variação da taxa de abertura. Para aquelas que mostrarem-se significantes, deve-se medir seu poder de influencia pelo módulo de seu coeficiente B.

Deste modo, as variáveis em questão são, nesta ordem: “X1_assunto, Grau transacional, X2_tipo_comunicação, X3_assunto, X2_assunto e saldo médio”.

g. Validando os resultados

Conforme comentado na seção destinada a descrição da metodologia de pesquisa e da revisão bibliográfica, utilizou-se o método de análise de resíduos a fim de validar os resultados encontrados.

O primeiro passo para verificar-se se os resíduos apresentam distribuição normal pode ser dado plotando-se os resíduos no papel de distribuição normal e analisando-se seu comportamento. A figura 22 exibe este teste:

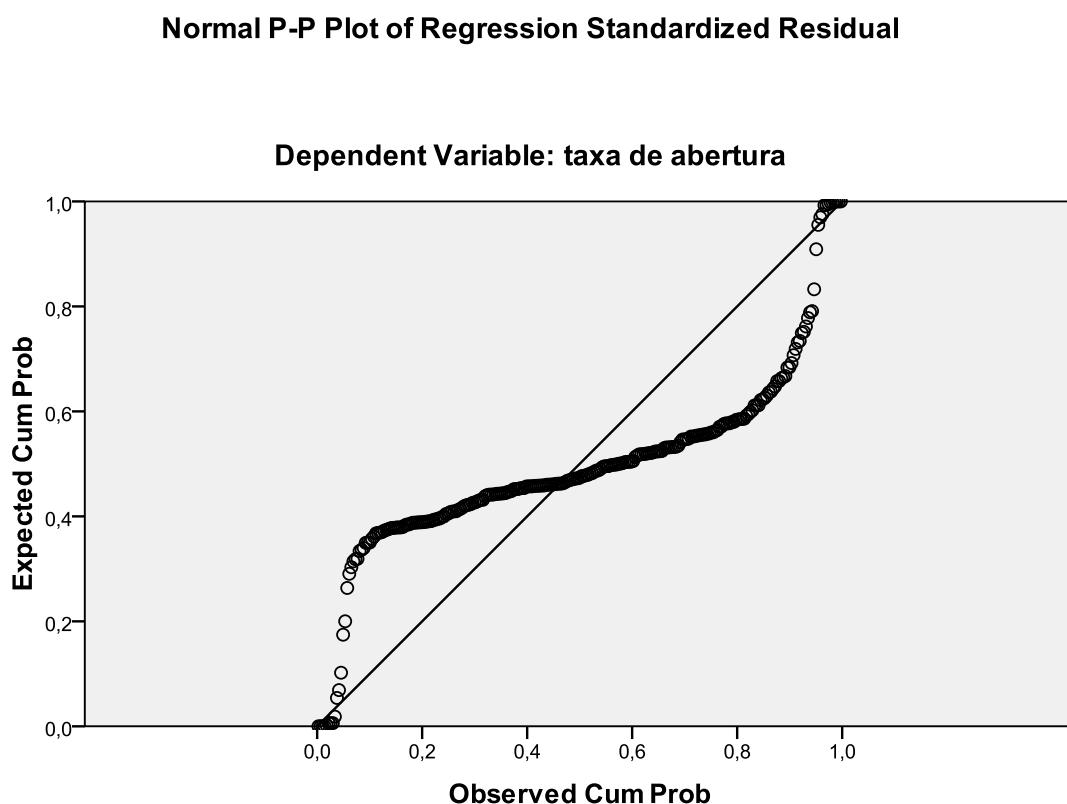


Figura 22: Normal *Plot-test* para análise de resíduos.

De fato, o resultado da primeira análise de resíduos não é nada animador, uma vez, que os resultados não se comportam como uma reta no papel de probabilidade normal. Porém, antes de concluir-se algo a respeito dos resíduos, proceder-se-á dois outros testes – o *Scatterplot* e os testes de Kolmogorov-Smirnov e Shapiro-Wilk.

No *Scatterplot*, conforme comentando na seção destinada a revisão de literatura, dados normais distribuem-se aleatoriamente pelo plano. A figura a seguir traz o *Scatterplot* referente aos resíduos do modelo em análise:

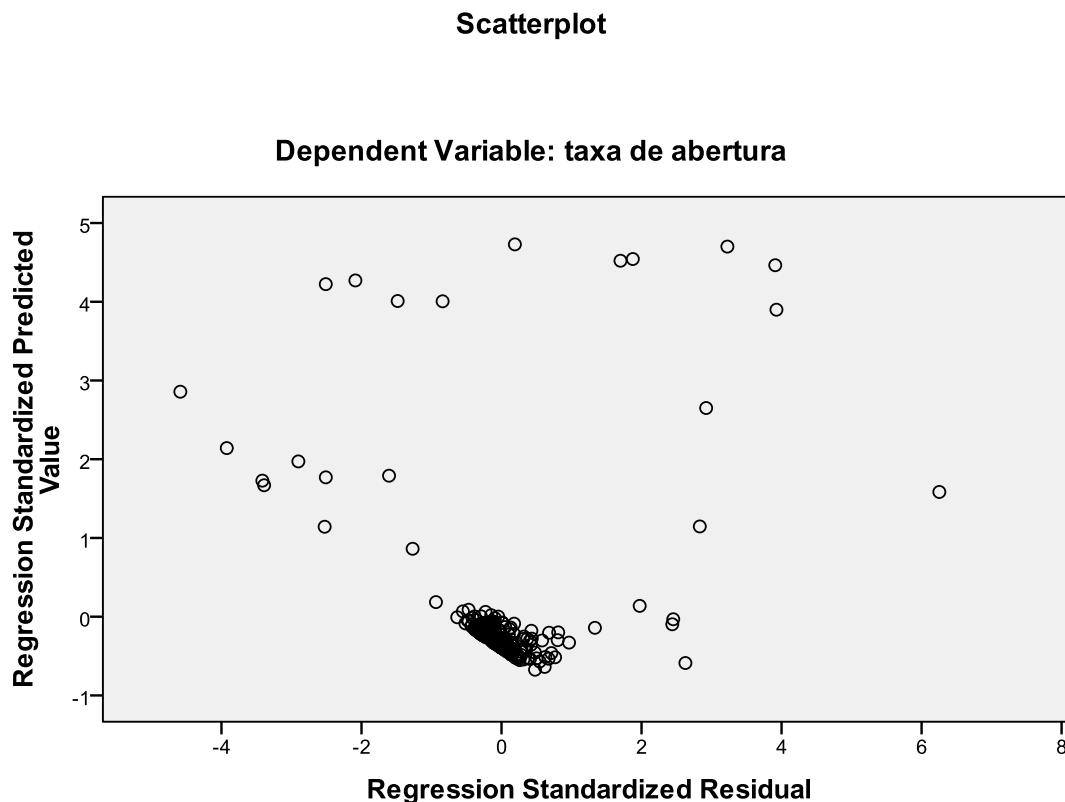


Figura 23: *Scatterplot* dos resíduos do modelo em análise

Novamente, os resíduos apresentaram resultados que os afastam da possibilidade de serem normais, uma vez que no *Scatterplot* fica claro que os resíduos apresentam uma concentração em torno do ponto (0;0).

Por fim, realizou-se o testes de Kolmogorov-Smirnov e Shapiro-Wilk. Para que sejam considerados normais, a significância destes testes deve ser maior do que 5%, uma vez que, para valores inferiores a 5%, pode-se interpretar que os dados apresentam uma diferença significativamente considerável em relação a curva normal.

A tabela 34 nos traz os resultados destes testes:

Tabela 34: Resultados dos testes de normalidade para o modelo 2.

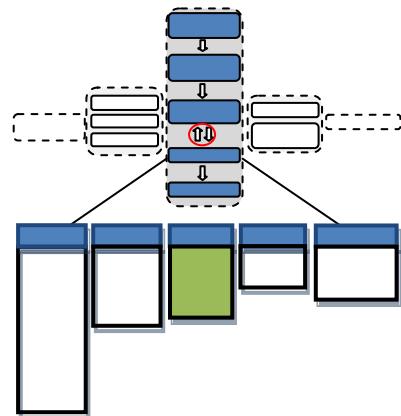
	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
<i>Unstandardized Residual</i>	0,256	255	0,000	0,670	255	0,000

Desenvolvido pelo próprio autor

Logo, de acordo com os três testes de normalidade feitos, pode-se concluir que os resíduos não apresentam distribuição normal, o que não está de acordo com os princípios da regressão linear, conforme comentado na seção destinada a revisão de literatura.

7.8. Anexo VIII: Transformação das variáveis para o refinamento do modelo (raiz cúbica)

Para a segunda etapa de refinamento, trabalhou-se com a raiz cúbica dos dados. (Não utilizou-se a raiz quadrada pois trabalha-se com valores negativos). Os quadros a seguir ilustram esta transformação para os primeiros 10 valores de cada um dos fatores:



A. Transformação da variável X1_dia_semana

O procedimento executado para transformação desta variável é o seguinte:

$$X1_dia_semana = \text{raiz}((X1_dia_semana;3))$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X1_dia_semana” que sofreram tal transformação:

“X1_dia_semana”	Raiz cúbica
0,91173	0,96967
-1,09251	-1,02993
-1,09251	-1,02993
-1,09251	-1,02993
0,91173	0,96967
0,91173	0,96967
0,91173	0,96967
0,91173	0,96967
0,91173	0,96967
-1,09251	-1,02993

Quadro 30: Transformação da variável “X1_dia_semana”

B. Transformação da variável X2_dia_semana

O procedimento executado para transformação desta variável é o seguinte:

$$X2_dia_semana = \text{raiz}((X2_dia_semana;3))$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X2_dia_semana” que sofreram tal transformação:

The diagram illustrates a transformation process. On the left, there is a vertical stack of 10 rectangular boxes, each containing a numerical value. This stack is labeled "X2_dia_semana". To the right of this stack is a large, thin black arrow pointing towards the right. To the right of the arrow is another vertical stack of 10 rectangular boxes, each containing a numerical value. This second stack is labeled "Raiz".

“X2_dia_semana”	→	Raiz
-1,68891		-1,19088
0,58978		0,83861
-1,68891		1,19088
-1,68891		1,19088
0,58978		0,83861
0,58978		0,83861
0,58978		0,83861
0,58978		0,83861
0,58978		0,83861

Quadro 31: Transformação da variável “X2_dia_semana”

C. Transformação da variável X3_dia_semana

O procedimento executado para transformação desta variável é o seguinte:

$$X3_{dia_semana} = \text{raiz}(X3_{dia_semana}; 3)$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X3_dia-semana” que sofreram tal transformação:



“X3_dia_semana”	Raiz
1,04206	1,01383
-0,95587	-0,98507
1,04206	1,01383
1,04206	1,01383
-0,95587	-0,98507
-0,95587	-0,98507
-0,95587	-0,98507
-0,95587	-0,98507
1,04206	1,01383

Quadro 32: Transformação da variável “X3_dia_taxa de abertura”

D. Transformação da variável X1_tipo_comunicação

O procedimento executado para transformação desta variável é o seguinte:

$$X1_tipo_comunicacao = raiz((X1_tipo_comunicacao;3)$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X1_tipo_comunicação” que sofreram tal transformação:



“X1_tipo_comunicação”	Raiz
-0,71821	-0,89554
-0,71821	-0,89554
-0,71821	-0,89554
1,38689	1,11519
-0,71821	-0,89554
-0,71821	-0,89554
-0,71821	-0,89554
-0,71821	-0,89554
-0,71821	-0,89554
1,38689	1,11519

Quadro 33: Transformação da variável “X2_tipo_comunicação”

E. Transformação da variável X2_tipo_comunicação

O procedimento executado para transformação desta variável é o seguinte:

$$\text{X2_tipo_comunicacao} = \text{raiz}((\text{X2_tipo_comunicacao};3)$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X2_tipo_comunicação” que sofreram tal transformação:



“X2_tipo_comunicação”	Raiz
0,71821	0,89554
0,71821	0,89554
0,71821	0,89554
-1,38689	-1,11519
0,71821	0,89554
0,71821	0,89554
0,71821	0,89554
0,71821	0,89554
0,71821	0,89554
-1,38689	-1,11519

Quadro 34: Transformação da variável “X2_tipo_comunicação”

F. Transformação da variável X1_assunto

O procedimento executado para transformação desta variável é o seguinte:

$$X1_assunto = \text{raiz}((X1_assunto;3)$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X1_assunto” que sofreram tal transformação:

“X1_assunto”	Raiz
1,04206	1,01383
1,04206	1,01383
-0,95587	-0,98507
-0,95587	-0,98507
-0,95587	-0,98507
-0,95587	-0,98507
-0,95587	-0,98507
1,04206	1,01383
-0,95587	-0,98507

Quadro 35: Transformação da variável “X1_assunto”

G. Transformação da variável X2_assunto

O procedimento executado para transformação desta variável é o seguinte:

$$X2_assunto = \text{raiz}(X2_assunto;3)$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X2_assunto” que sofreram tal transformação:

“X2_assunto”	Raiz
-1,56088	-1,16000
0,63815	0,86094
0,63815	0,86094
0,63815	0,86094
0,63815	0,86094
0,63815	0,86094
0,63815	0,86094
0,63815	0,86094
0,63815	0,86094
-1,56088	-1,16000

Quadro 36: Transformação da variável “X2_assunto”

H. Transformação da variável X3_assunto

O procedimento executado para transformação desta variável é o seguinte:

$$X3_assunto = \text{raiz}((X3_assunto;3)$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “X3_assunto” que sofreram tal transformação:



“X3_assunto”	Raiz
-0,57170	-0,82996
-0,57170	-0,82996
-0,57170	-0,82996
-0,57170	-0,82996
-0,57170	-0,82996
-0,57170	-0,82996
-0,57170	-0,82996
-0,57170	-0,82996
-0,57170	-0,82996
1,74232	1,20330

Quadro 37: Transformação da variável “X3_assunto”

I. Transformação da variável “Saldo Médio”

O procedimento executado para transformação desta variável é o seguinte:

$$\text{“Saldo Médio”} = \text{raiz}(\text{“Saldo Médio”}; 3)$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “Saldo Médio” que sofreram tal transformação:



“Saldo Médio”	Raiz
-0,21518	-0,59924
-0,20785	-0,59236
-0,21573	-0,59975
-0,21251	-0,59675
0,08082	0,432348
0,08142	0,433413
-0,16472	-0,54817
-0,16471	-0,54816
-0,21694	-0,60087
-0,21604	-0,60004

Quadro 38: Transformação da variável “Saldo Médio”

J. Transformação da variável “Percentual impactado”

O procedimento executado para transformação desta variável é o seguinte:

$$\text{“Percentual impactado”} = \text{raiz}(\text{“Percentual impactado”}; 3)$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “Percentual Impactado” que sofreram tal transformação:

“Percentual impactado”	Raiz
-0,21518	0,77909
-0,20785	-0,79742
-0,21573	0,98711
-0,21251	0,35799
0,08082	-0,98539
0,08142	-0,98538
-0,16472	-0,95735
-0,16471	-0,95739
-0,21694	1,18986
-0,21604	1,31303

Quadro 39: Transformação da variável “Percentual impactado”

K. Transformação da variável “Grau transacional”

O procedimento executado para transformação desta variável é o seguinte:

$$\text{“Grau transacional”} = \text{raiz}(\text{“Grau transacional”}; 3)$$

O quadro a seguir nos traz um exemplo com 10 valores associados a “Grau transacional” que sofreram tal transformação:

“Grau transacional”	Raiz
-0,55712	-0,82284
-0,55163	-0,82013
-0,56074	-0,82462
2,42431	1,34337
-0,24049	-0,62187
-0,25177	-0,63145
-0,27451	-0,64991
-0,25577	-0,65090
-0,25070	-0,63055
-0,56229	-0,82538

Quadro 40: Transformação da variável “Grau transacional”

L. Transformação da variável “Taxa de abertura”

O procedimento executado para transformação desta variável é o seguinte:

$$\text{“Taxa de abertura”} = \text{raiz}(\text{“Taxa de abertura”}; 3)$$

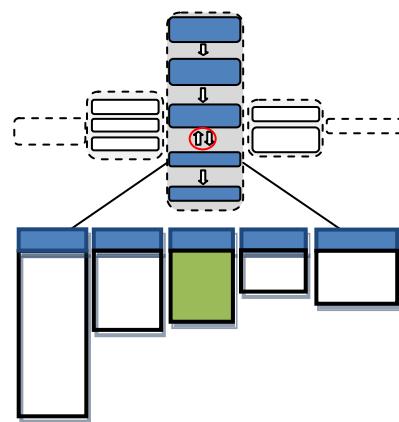
O quadro a seguir nos traz um exemplo com 10 valores associados a “Taxa de abertura” que sofreram tal transformação:

“Taxa de abertura”	Raiz
-0,45921	-0,77150
-0,66613	-0,87335
-0,81525	-0,93418
-1,27041	-1,08305
0,71107	0,89256
0,69301	0,88494
0,30170	0,67069
0,29012	0,66200
-0,92348	-0,97381
-0,97759	-0,99247

Quadro 41: Transformação da variável “Taxa de abertura”

7.9. Anexo IX: Resultado da regressão linear III

Denomina-se regressão linear III aquela realizada com a raiz cúbica dos dados utilizados na regressão linear I, pós-realização da segunda transformação com foco em refinar o modelo gerado anteriormente, uma vez que seus resultados não se mostraram válidos em ambiente de testes prático.



a. Estatísticas Descritivas

Tabela 35: Resultados das estatísticas descritivas para regressão com dados padronizados

Variável	Média	Desvio-padrão	N
taxa de abertura	-0,143890	0,7927802	255
X1_dia_semana	0,060046	0,9976825	255
X2_dia_semana	0,313334	0,8906442	255
X3_dia_semana	-0,028734	1,0004821	255
x1_TIPO_COMUNICAÇÃO	-0,209526	0,9551679	255
x2_TIPO_COMUNICACAO	-0,209526	0,9551679	255
X1_Assunto	-0,028734	1,0004821	255
X2_Assunto	0,274474	0,9190128	255
X3_Assunto	-0,327622	0,8786729	255
saldo médio	-0,430495	0,4781424	255
Grau de Segmentação	-0,184920	0,9084715	255
Grau transacional %	-0,339420	0,7277538	255

Desenvolvido pelo próprio autor

A tabela 35 traz os resultados das estatísticas descritivas aplicadas aos dados modelo de regressão III. Com a transformação realizada, nota-se que, agora, os dados não são mais normais, uma vez que apresentam média e variância diferentes de 1 e 0, respectivamente. Novamente, todas as variáveis assumem 255 valores possíveis, os quais, conforme comentado em seções anteriores remetem a disparos de ações de *E-mail Marketing*, foco deste estudo.

b. Análise de correlações

Um ponto importante da análise realizada está na verificação das correlações entre a variável dependente (taxa de abertura) e as variáveis independentes, de modo a, após determinação dos coeficientes da equação da regressão, possa-se verificar como se dá a relação dentre elas e a taxa de abertura.

Tabela 36: Correlações de Pearson entre a taxa de Abertura e as variáveis independentes

Variável	Correlação de Pearson	Sig. (1-tailed)
Taxa de Abertura	1,000	-
X1_dia_semana	0,112	0,037
X2_dia_semana	-0,044	0,243
X3_dia_semana	-0,028	0,325
X1_tipo_comunicação	-0,185	0,001
X2_tipo_comunicação	0,185	0,001
X1_assunto	0,041	0,258
X2_assunto	0,217	0,000
X3_assunto	0,035	0,288
Saldo Médio	0,580	0,000
Percentual impactado	-0,582	0,000
Grau Transacional	0,273	0,000

Desenvolvido pelo próprio autor

Pode-se extrair da tabela 36 que as variáveis que possuem correlação estatisticamente válida, a significância padrão de 5%, com a variável em estudo são:

- Percentual impactado: Correlação negativa (-)
- Saldo médio: Correlação positiva (+)
- Grau transacional: Correlação positiva (+)
- X2_assunto: Correlação positiva (+)
- X2_tipo_comunicação: Correlação positiva (+)
- X1_tipo_comunicação: Correlação negativa (-)
- X1_dia_semana: Correlação positiva (+)

c. Variáveis utilizadas no modelo

Com exceção da variável “X1_tipo_comunicação”, todas as outras foram utilizadas no modelo em análise. Tal variável foi retirada por apresentar uma forte correlação negativa (-1,000) com a variável “X2_tipo_comunicação”. Esta retirada se dá de forma automática na modelagem via software estatístico.

d. ANOVA

A fim de medir-se a robustez do modelo, pode-se realizar a análise de variância (ANOVA) e verificar o valor assumido por sua significância, lembrando-se, que, de acordo com a revisão de literatura, valores inferiores a 5% são considerados aceitáveis.

Na tabela 37 está expressa o resultado da ANOVA realizada:

Tabela 37: ANOVA para modelo 3

Modelo		Soma de Quadrados	Graus de liberdade	Quadrado Médio	F	Significância
3	Regressão	91,318	10	9,132	32,613	0,000
	Residual	68,321	244	0,280		
	Total	159,639	254			

Desenvolvido pelo próprio autor

Logo, a significância neste caso é igual a 0,000, o que permite-nos concluir que o modelo é robusto.

e. Resumo do Modelo

A tabela 38 mostra algumas informações relacionadas ao modelo aplicado:

Tabela 38: resumo das informações do modelo 3

Modelo	R	R ²	Durbin-Watson
2	0,572	0,554	1,775

Desenvolvido pelo próprio autor

Conforme ilustrado na tabela 38, o modelo apresenta um poder de explicação da variação da taxa de abertura da ordem de 73% (R²).

Já o teste de Durbin-Watson, parte da análise de resíduos, apresenta um valor próximo a 2, o que, conforme comentado na seção destinada a revisão de literatura garante que os erros são independentes.

f. Coeficientes da equação da regressão

Os coeficientes da regressão (peso das variáveis na explicação da variação da taxa de abertura) e sua respectiva significância para o modelo de regressão estão expostos na tabela 39 a seguir:

Tabela 39: Coeficiente e sua respectiva significância para o modelo 3.

Modelo	Variável	Coeficiente	Significância
3	(Constante)	0,093	0,064
	X1_dia_semana	0,062	0,078
	X2_dia_semana	-0,015	0,749
	X3_dia_semana	0,013	0,750
	X2_tipo_comunicação	0,131	0,003
	X1_assunto	0,001	0,977
	X2_assunto	0,084	0,055
	X3_assunto	0,134	0,005
	Saldo médio	0,576	0,000
	Percentual impactado	-0,386	0,000
	Grau transacional	0,194	0,002

Desenvolvido pelo próprio autor

Conforme comentado na seção destinada a revisão de literatura, deve-se verificar a significância estatística para determinar se a variável exerce influência considerável na variação da taxa de abertura. Para aquelas que mostrarem-se significantes, deve-se medir seu poder de influencia pelo módulo de seu coeficiente B.

Deste modo, as variáveis em questão são, nesta ordem: “Saldo médio, Percentual impactado, Grau transacional, X3_assunto e X2_tipo_comunicação”.

g. Validando os resultados

Conforme comentado na seção destinada a descrição da metodologia de pesquisa e da revisão bibliográfica, utilizou-se o método de análise de resíduos a fim de validar os resultados encontrados.

O primeiro passo para verificar-se se os resíduos apresentam distribuição normal pode ser dado plotando-se os resíduos no papel de distribuição normal e analisando-se seu comportamento. A figura 24 exibe este teste:

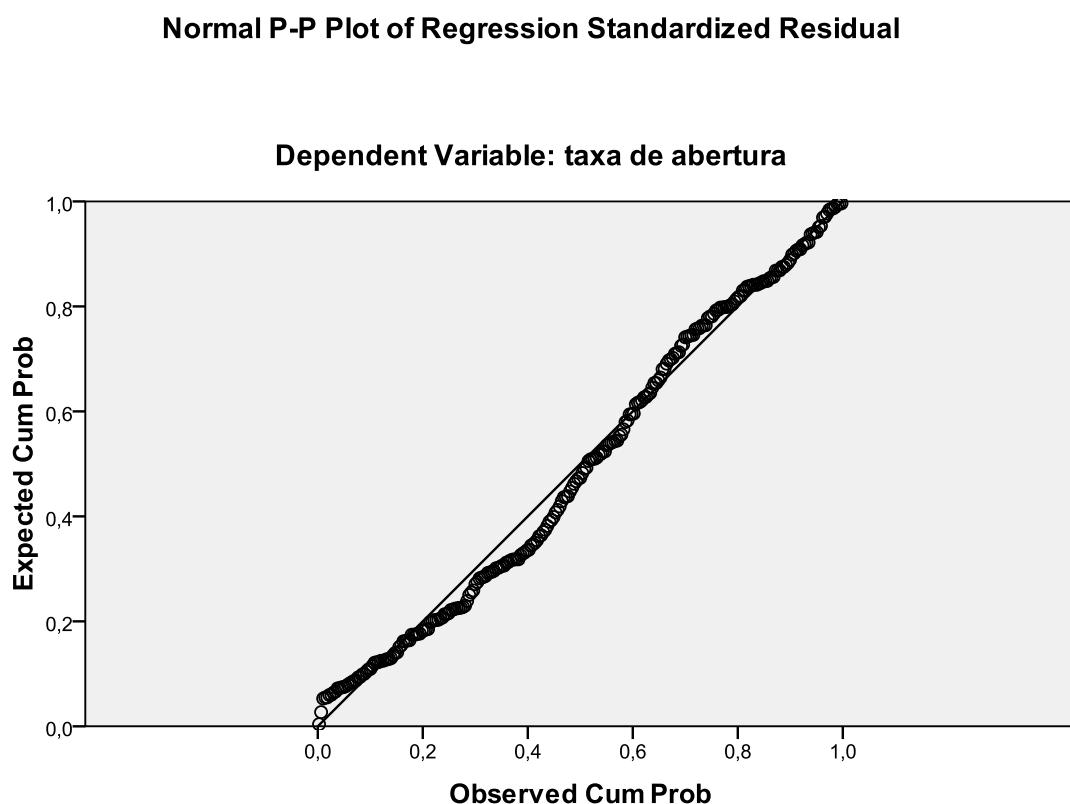


Figura 24: Normal *Plot-test* para análise de resíduos.

De fato, o resultado da primeira análise de resíduos não é nada animador, uma vez que os resultados não se comportam como uma reta no papel de probabilidade normal. Porém, antes de concluir-se algo a respeito dos resíduos, proceder-se-á dois outros testes – o *Scatterplot* e os testes de Kolmogorov-Smirnov e Shapiro-Wilk.

No *Scatterplot*, conforme comentando na seção destinada a revisão de literatura, dados normais distribuem-se aleatoriamente pelo plano. A figura a seguir traz o *Scatterplot* referente aos resíduos do modelo em análise:

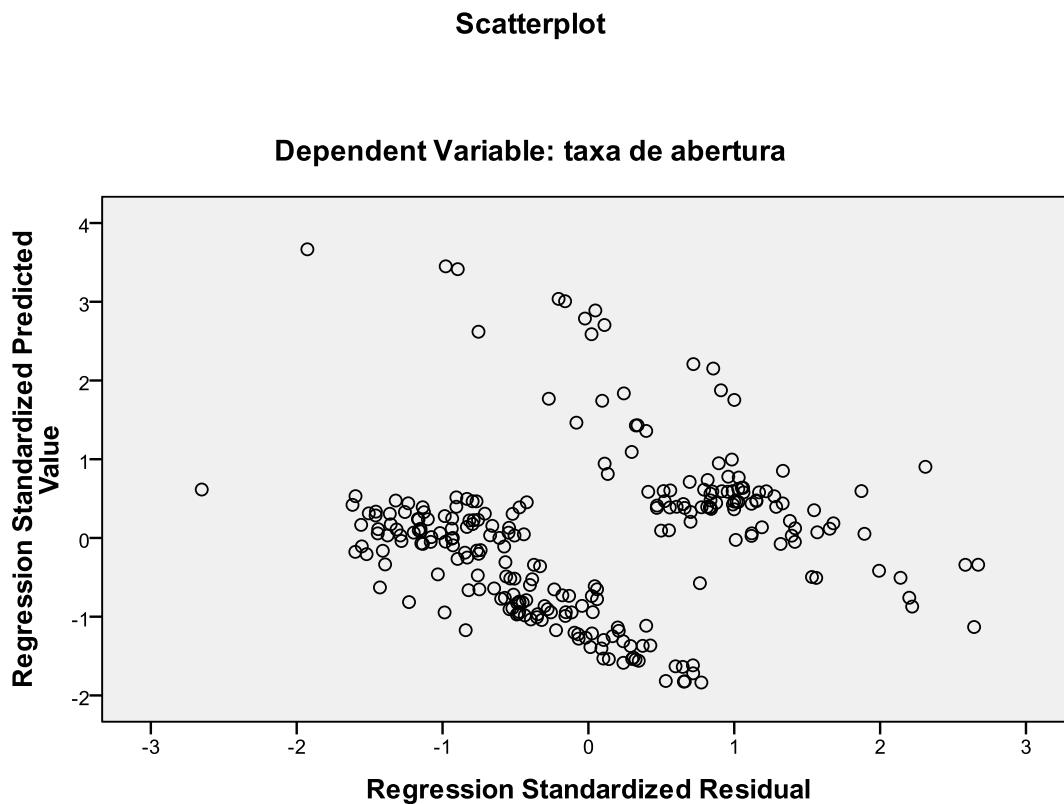


Figura 25: *Scatterplot* dos resíduos do modelo em análise

Novamente, os resíduos apresentaram resultados que os afastam da possibilidade de serem normais, uma vez que no *Scatterplot* fica claro que os resíduos apresentam uma concentração em torno do ponto (0;0).

Por fim, realizou-se o testes de Kolmogorov-Smirnov e Shapiro-Wilk. Para que sejam considerados normais, a significância destes testes deve ser maior do que 5%, uma vez que, para valores inferiores a 5%, pode-se interpretar que os dados apresentam uma diferença significativamente considerável em relação a curva normal.

A tabela 40 nos traz os resultados destes testes:

Tabela 40: Resultados dos testes de normalidade para o modelo 3

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
<i>Unstandardized Residual</i>	0,071	255	0,004	0,982	255	0,003

Desenvolvido pelo próprio autor

Logo, de acordo com os três testes de normalidade feitos, pode-se concluir que os resíduos não apresentam distribuição normal, o que não está de acordo com os princípios da regressão linear, conforme comentado na seção destinada a revisão de literatura.