

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ADMINISTRAÇÃO (EA)  
DEPARTAMENTO DE CIÊNCIAS ADMINISTRATIVAS (DCA)**

**Rodrigo Costa Lima Ferreira**

**Atualização da informação – um modelo matemático de apoio à decisão  
como contribuição à Qualidade da Informação.**

**Porto Alegre**

**2012**

**Rodrigo Costa Lima Ferreira**

**Atualização da informação – um modelo matemático de apoio à decisão  
como contribuição à Qualidade da Informação.**

Trabalho de conclusão de curso de graduação  
apresentado ao Departamento de Ciências Administrativas  
da Universidade Federal do Rio Grande do Sul, como  
requisito parcial para a obtenção do grau de Bacharel em  
Administração.

**Orientador: Prof. Antônio Carlos Gastaud Maçada**

**Porto Alegre**

**2012**

**Rodrigo Costa Lima Ferreira**

**Atualização da informação – um modelo matemático de apoio à decisão  
como contribuição à Qualidade da Informação.**

Trabalho de conclusão de curso de graduação  
apresentado ao Departamento de Ciências Administrativas  
da Universidade Federal do Rio Grande do Sul, como  
requisito parcial para a obtenção do grau de Bacharel em  
Administração.

**Orientador: Prof. Antônio Carlos Gastaud Maçada**

**Conceito final:**

**Aprovado em ..... de .....de.....**

**BANCA EXAMINADORA**

---

**Prof. Dr. Raquel Janissek-Muniz– UFRGS**

---

**Orientador – Prof. Dr. Antônio Carlos Gastaud Maçada – UFRGS**

**Porto Alegre**

**2012**

## **AGRADECIMENTOS**

Registro aqui meus mais sinceros agradecimentos a todos aqueles que, de uma forma ou de outra, contribuíram para a concretização deste trabalho. Agradeço ao meu professor orientador Antônio Carlos Gastaud Maçada, por me apresentar durante a graduação o tema deste estudo, por me encorajar a realizar a pesquisa em um campo pouco desbravado e por me inspirar segurança nos momentos de instabilidade. Agradeço a empresa que não hesitou em abrir as portas para que este trabalho pudesse ser realizado. A todos os entrevistados que encontraram, em suas concorridas agendas, tempo para contribuir com a pesquisa acadêmica. São mentes como as destes profissionais que permitem enriquecer o conhecimento na área de administração.

Deixo um agradecimento especial a minha namorada, Gabriela, pois sem a sua paciência, carinho e apoio este trabalho não teria sido possível. Agradeço ao meu grande pai João, a minha preciosa mãe Marta, a minha doce irmã Rafaela e a toda a minha querida família, o carinho e a compreensão a mim dedicados durante a realização deste trabalho, bem como fizeram em toda minha vida.

## RESUMO

A informação ganhou, nas últimas décadas, posição de destaque entre os recursos consumidos pelas organizações. Especialmente em empresas privadas, a quantidade de informação disponibilizada pelas novas tecnologias trouxe inúmeros ganhos. Com a ascensão da área da qualidade total, tipicamente aplicada à produção industrial, surgiu a preocupação com a qualidade da informação consumida pela empresa. Tema de grande diversidade de pesquisas científicas, questões como a precisão, a relevância, a volatilidade da informação, entre muitas outras, passaram a compor o rol de dimensões nas quais a qualidade da informação (QI) pode ser avaliada e gerenciada, o que permitiu que diversas metodologias que viabilizassem a gestão da QI também fossem desenvolvidas. O presente estudo contribui com a área de QI, com enfoque na gestão da QI, ao apresentar um modelo de apoio à decisão para a dimensão volatilidade no nível operacional. Quirin Görz (2011) propôs um modelo que permite a tomada da decisão de menor custo entre *atualizar* ou *não atualizar* a informação contida em um determinado atributo. Este estudo considera o fato de que o custo de atualizar a informação em um conjunto de múltiplos atributos em uma única ação é, no máximo, igual ao custo de realizar uma ação para cada atributo do conjunto. Sob esta premissa, realizou-se uma ampliação do modelo de Görz (2011) a fim de torná-lo aplicável para um conjunto de múltiplos atributos. Em seguida, realizou-se uma simulação de aplicação do modelo ampliado através do estudo de um caso prático de uso de informação em um processo de venda de apólices de seguros no Banco B. Finalmente, uma simulação de aplicação do modelo de Görz (2011) para o mesmo caso, permitiu realizar uma comparação entre os modelos.

## ABSTRACT

In the last decades, Information has acquired prominent position among the resources consumed by Organizations. Especially in the private Companies, the amount of information provided by new technologies has brought numerous gains. The rising prominence of quality management field, classically applied in the industrial production area, has incited the concern about the quality of information consumed by the companies. Topic of a wide diversity in scientific researches, issues as accuracy, relevance, currency of information, among many others, were included in the list of dimensions in which the quality of information can be assessed and managed and, therefore, several methodologies that provide IQ management were also be developed. This study contributes to the IQ field, with a focus on IQ management by proposing a support decision model for the dimension currency in an operational level. In prior paper, Quirin Görz (2011) proposed a model which supports the best economics-driven decision between *update* or *not to update* the information contained in an attribute. This study considers that the cost of updating information of multiple attributes in a single action is at most the same as the cost of performing an update action for each attribute. Under this assumption, this study extended the model proposed by Görz (2011) in order to make it applicable to a set of multiple attributes. Then, we conducted a simulation of the extended model by studying a case of information use in a process of selling insurance renewals at Bank B. Finally, a simulated application of the model of Görz (2011) for the same case, allowed a brief comparison between the both models.

## LISTA DE FIGURAS

Figura 1- Avaliação Subjetiva e Objetiva da QI. Adaptado de PIPINO, LEE e WANG (2002)....	24
Figura 2 - Esquema entidades e relacionamentos. Elaborado pelo autor. ....	28
Figura 3- Gráfico da distribuição da probabilidade de a informação estar atualizada em cada t. Elaborado pelo autor.....	30
Figura 4 - Gráfico Exemplo da distribuição da função. Elaborado pelo autor. ....	45
Figura 5 - Aplicação esquemática do modelo ampliado. Elaborada pelo autor. ....	50
Figura 6 - Fluxograma do processo de venda de apólices de renovação. Elaborado pelo autor. ....	60
Figura 7 - Gráfico da função da probabilidade de a informação estar desatualizada para Endereço em função de t, com a idade da informação ( $i_x$ ) nula em $t=0$ . Elaborado pelo autor. ....	68
Figura 8 - Gráfico da função da probabilidade de a informação estar desatualizada para Telefone em função de t, com a idade da informação ( $i_x$ ) nula em $t=0$ . Elaborado pelo autor. ....	70
Figura 9 - Gráfico da função da probabilidade de o próximo uso da informação de Endereço ocorrer em cada de t. Elaborado pelo autor. ....	71
Figura 10 – Gráfico da função da probabilidade de o próximo uso da informação de Telefone ocorrer em cada t. Elaborado pelo autor. ....	72

## LISTA DE QUADROS

Quadro 1 – Demonstração da sequência de cálculos que levaram à definição deste parâmetro do modelo. Elaborado pelo autor. ....	38
Quadro 2 - Demonstração do método usado para encontrar o custo em questão. Elaborado pelo autor. ....	39
Quadro 3 - Técnica usada para estimar os custos de telefone e trabalho por minuto. Elaborado pelo autor.....	40
Quadro 4 - Técnica empregada para estimar o custo da ação de atualizar a informação. Elaborado pelo autor. ....	41
Quadro 5 - Demonstração dos Parâmetros Necessários para a Construção do Modelo. Elaborado pelo autor.....	46
Quadro 6 - Demonstração dos <i>Inputs</i> da planilha eletrônica desenvolvida. Elaborado pelo autor. ....	47
Quadro 7 - Demonstração dos <i>Outputs</i> da planilha eletrônica desenvolvida. Elaborado pelo autor. ....	47
Quadro 8 - Demonstração do custo total de aplicação do modelo original. Elaborado pelo autor. ....	48
Quadro 9 - Natureza dos custos no modelo original. Elaborado pelo autor. ....	51
Quadro 10 - Natureza dos custos no modelo ampliado. Elaborado pelo autor. ....	54
Quadro 11 - Demonstração do cálculo da receita esperada pelo uso do atributo Endereço. Elaborado pelo autor.....	61
Quadro 12 - Demonstração do cálculo da receita esperada pelo uso do atributo Telefone. Elaborado pelo autor.....	62
Quadro 13 - Demonstração do cálculo do custo indireto unitário de usar o atributo Telefone. Elaborado pelo autor.....	63
Quadro 14 - Demonstração do cálculo do custo direto unitário de usar o atributo Telefone. Elaborado pelo autor.....	64
Quadro 15 - Demonstração do cálculo do custo unitário de usar o atributo Telefone. Elaborado pelo autor.....	64
Quadro 16 – Demonstração do cálculo dos custos estimados de trabalho e telefone por minuto (elaborado pelo autor).....	65



Quadro 17 - Demonstração do cálculo da estimação do custo da ação de atualizar a informação para o conjunto de atributos. Elaborado pelo autor. ....	66
Quadro 18 - Demonstração do cálculo da estimação do custo da ação de atualizar a informação para o atributo Endereço. Elaborado pelo autor. ....	66
Quadro 19 - Demonstração do cálculo da estimação do custo da ação de atualizar a informação para o atributo Telefone. Elaborado pelo autor. ....	67
Quadro 20 - Dados inseridos no <i>Curve Expert</i> <sup>®</sup> para cálculo do coeficiente de declínio para o atributo Endereço. Elaborado pelo autor. ....	67
Quadro 21 - Saídas do <i>Curve Expert</i> <sup>®</sup> para o atributo Endereço. Elaborado pelo autor. ....	68
Quadro 22 - Dados inseridos no <i>Curve Expert</i> <sup>®</sup> para cálculo do coeficiente de declínio do atributo Telefone. Elaborado pelo autor. ....	69
Quadro 23 - Saídas do <i>Curve Expert</i> <sup>®</sup> para o atributo Telefone. Elaborado pelo autor. ....	69
Quadro 24- Demonstração da quantidade de propostas de renovação enviadas, pagas e cobradas por telefone. Elaborado pelo autor. ....	71
Quadro 25 – Compilação dos parâmetros dos modelos para o caso. Elaborado pelo autor. .	73

## LISTA DE TABELAS

Tabela 1 - As dimensões da QI de diversos autores. Adaptada de LIMA e MAÇADA (2007)...21	21
Tabela 2 - Descrição das Dimensões da QI. Adaptada de KAHN, STRONG e WANG (2002). ...22	22
Tabela 3 - Estratégias e Técnicas para melhoria da QI. Adaptado de Heinrich, Kaiser e Klier (2007).....25	25
Tabela 4 - Definições da dimensão "Volatilidade". Adaptado de Görz (2011). ....26	26
Tabela 5 - Tabulação das correspondências retornadas pelo motivo "mudou-se". Elaborada pelo autor. ....42	42
Tabela 6 - Percentual de registros desatualizados para cada idade. Elaborada pelo autor. ...42	42
Tabela 7 - Exemplo de tabulação utilizada para encontrar o coeficiente para o atributo Telefone. Elaborada pelo autor. ....43	43
Tabela 8 – Informações utilizadas para encontrar o percentual de registros desatualizados para o atributo Telefone. Elaborada pelo autor.....43	43
Tabela 9 - Decisões da simulação do Modelo Ampliado. Elaborada pelo autor.....73	73
Tabela 10 - Resultado da ação com aplicação do modelo ampliado. Elaborada pelo autor. ..74	74
Tabela 11 – Decisões da simulação do Modelo Original - Etapa 1. Elaborada pelo autor.....75	75
Tabela 12 - Decisões da simulação do Modelo Original - Etapa 2. Elaborada pelo autor. ....75	75
Tabela 13 - Resultado da ação com aplicação do modelo original. Elaborada pelo autor. ....75	75
Tabela 14 – Resultado da ação sem a aplicação de modelos. Elaborada pelo autor.....76	76
Tabela 15 - Comparação do resultado dos modelos e da ação sem aplicação de modelo para o ano 1. Elaborada pelo autor. ....78	78
Tabela 16 - Comparação do resultado dos modelos e da ação sem aplicação de modelo para ano 2. Elaborada pelo autor. ....78	78

## SUMÁRIO

1	Introdução .....	13
2	Justificativa .....	16
3	Objetivos de Pesquisa .....	17
3.1.	Objetivo Geral .....	17
3.2.	Objetivos Específicos .....	17
4	Rererencial Teórico .....	18
4.1.	O Conceito de Informação .....	18
4.2.	O Conceito de QI .....	19
4.3.	As Dimensões da QI .....	20
4.4.	A Gestão da Qualidade da Informação .....	22
4.4.1.	Reconstrução da Situação .....	23
4.4.2.	A Avaliação da QI .....	23
4.4.3.	A Melhoria da QI.....	25
4.5.	A Dimensão Volatilidade.....	26
4.6.	O Modelo de Decisão.....	27
4.6.1.	Entidade, relacionamento e atributo. ....	27
4.6.2.	A Métrica para Volatilidade.....	29
4.6.3.	Erros e Custos dos Erros .....	31
5	Método de Pesquisa.....	34
5.1.	Ampliação do Modelo Original .....	34
5.1.1.	Análise do Modelo Original .....	34
5.1.2.	O Cálculo do Custo Total Projetado.....	35
5.1.3.	O Cálculo do Custo para Múltiplos Atributos .....	36
5.2.	A Escolha da Empresa e do Processo.....	36

5.3.	Construção do Modelo .....	37
5.3.1.	Coleta e Tratamento de Dados.....	37
5.3.2.	Construção do Modelo .....	45
5.4.	Simulando uma aplicação do modelo ampliado.....	46
5.5.	Comparação entre os Modelos.....	47
6	O Modelo Ampliado .....	49
6.1.	Considerações Gerais do Modelo Ampliado.....	49
6.2.	Uma Análise do Modelo Original .....	50
6.3.	A Ampliação do Modelo.....	53
6.3.1.	Considerando múltiplos instantes de tempo. ....	55
6.3.2.	Considerando múltiplos atributos. ....	56
6.3.3.	O Limiar de Decisão do Modelo Ampliado .....	57
7	Uma Simulação de Aplicação do Modelo Ampliado .....	58
7.1.	Descrição do Processo .....	58
7.2.	As Receitas Esperadas.....	61
7.2.1.	Cálculo da Receita R <sub>sperada</sub> pelo Uso do Atributo Endereço.....	61
7.2.2.	Cálculo da Receita Esperada pelo Uso do Atributo Telefone.....	62
7.3.	Os Custos de Usar a Informação .....	62
7.3.1.	O Custo de Usar o Atributo Endereço .....	62
7.3.2.	O Custo de Usar o Atributo Telefone .....	63
7.4.	O Custo da Ação de Atualizar.....	64
7.4.1.	O Custo de Atualizar Endereço e Telefone .....	65
7.4.2.	O Custo de Atualizar os Atributos Separadamente.....	66
7.5.	A probabilidade de a informação estar desatualizada .....	67
7.6.	A Probabilidade de Usar a Informação .....	70
7.7.	A Construção dos Modelos .....	72

7.8. A Simulação de Aplicação dos Modelos .....	73
7.8.1. Simulação de Aplicação do Modelo Original de Görz (2011) .....	74
7.8.1. Simulação da ação sem aplicação de modelos.....	76
7.9. Análise e Comparação das Simulações dos Modelos .....	76
8 Considerações Finais .....	80
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>82</b>
<b>APÊNDICE A – Planilha Eletrônica para aplicação do modelo ampliado. ....</b>	<b>86</b>
<b>APÊNDICE B – Planilha Eletrônica para aplicação do modelo original.....</b>	<b>87</b>

## 1 INTRODUÇÃO

A pobre qualidade da informação (QI) normalmente leva a decisões erradas e altos custos incorridos pelas organizações (FISHER, CHENGALUR-SMITH e BALLOU, 2003).

É fato consensual que, tanto indivíduos quanto organizações, necessitam consumir informações. Empresas privadas necessitam de informações sobre o ambiente, mercado, clientes, não clientes, tecnologias aplicadas no seu setor e informações financeiras para estruturarem e levarem a cabo suas estratégias (DRUCKER, 1995). Logo, a falta de qualidade na informação consumida pelas empresas vem causando forte impacto na efetividade geral das mesmas (WANG e STRONG, 1996). Em uma enquete internacional, 75% dos entrevistados admitiram já ter tomado decisões erradas devido à pobre QI (HARRIS INTERACTIVE, 2006). Em um estudo publicado pela PWC (2004) afirmou que apenas 34% dos executivos chefes de informações (CIO) se dizem “muito confiantes” com a qualidade dos dados em suas empresas.

Não se sabe ao certo a quantidade de dados existente atualmente no mundo, entretanto acredita-se que este montante cresça 60% a cada ano (THE ECONOMIST, 2010a). As novas tecnologias vêm possibilitando que uma grande quantidade de informação seja produzida, armazenada e consumida a velocidade outrora inconcebível. Dessa forma, nos últimos anos, a informação passou de escassa a superabundante o que, ao mesmo tempo em que vem trazendo inúmeros benefícios, imprime uma complexidade muito maior ao seu gerenciamento (THE ECONOMIST, 2010b).

Garantir qualidade à informação produzida, processada e consumida em uma organização passa ser um importante desafio e requer uma eficiente gestão da QI. Visando a encontrar procedimentos eficientes para gerenciamento da QI, uma grande diversidade de autores dedicaram seus esforços na criação de metodologias para gestão da QI (GÖRZ, 2011). Dentre o grande número de metodologias propostas, três etapas ganham destaque por serem comuns à maioria delas: a reconstrução da situação, a avaliação da QI e a melhoria QI (BATINI et al, 2009). A nível gerencial, a gestão da QI envolve projetos de QI,

nível de serviço em QI e investimentos gerais em QI (GÖRZ, 2011) enquanto, a nível operacional, costuma referir-se a esforços para a melhoria da QI no nível do dado, ou seja, do valor contido em um atributo informacional (JIANG et. al., 2007).

Visando enriquecer o campo da gestão da QI no nível operacional, Quirin Görz (2011) propõe um modelo matemático de apoio à decisão orientado ao menor custo que converge com a pesquisa de Even e Kaiser (2009) que afirmam que, embora haja a aceitação de que quanto mais elevada for a QI mais eficaz será o seu uso, o custo dos esforços para melhoria da QI deve ser considerado de modo a identificar um limiar economicamente eficiente. A ideia central do modelo proposto é apoiar a melhor decisão entre *atualizar* ou *não atualizar* determinada informação considerando o custo total de cada decisão. A mensuração de tais custos é feita avaliando a idade da informação armazenada em um banco de dados e o nível de atualização da informação requerido pelo processo que a consumirá. Segundo o autor, o nível de atualização da informação é um fator de imensa relevância para importantes processos tais como a gestão de relacionamento com clientes (CRM) e gestão da cadeia de suprimentos (SCM) o que corrobora a afirmação de que a volatilidade é uma das dimensões mais estudadas na área da QI (WAND e WANG, 2002). O modelo de Görz (2011) resulta em uma atraente e eficiente ferramenta para a gestão da QI, porém, nos moldes em que foi concebido, é capaz de analisar apenas um atributo a cada vez que aplicado.

Sob este contexto o presente trabalho se propõe a responder aos seguintes questionamentos: um modelo de decisão que contemplasse mais de um atributo informacional não levaria a decisões mais eficientes? Como seria a nova configuração do modelo para que este pudesse contemplar esta ampliação? Como se configurariam os parâmetros e métricas de um modelo ampliado? De modo a contribuir com a gestão da QI no nível operacional, este trabalho se empenha em propor uma ampliação do modelo de Görz (2011) a fim de possibilitar decisões mais eficientes. Nesta primeira seção, uma breve introdução ao tema é apresentada. Na segunda seção, aspectos que justifiquem a pesquisa realizada são expostos. Na seção seguinte, são expostos os objetivos gerais e específicos os quais o presente estudo procurou atingir. Em seguida, uma revisão de estudos prévios sobre o tema é realizada. Na quinta seção são apresentados os procedimentos metodológicos que levaram a ampliação do modelo matemático que é foco do presente estudo. Na sexta é

apresentada a proposição de ampliação do modelo matemático em foco. A sétima seção se dedica a uma simulação de aplicação do modelo ampliado e do modelo original para uma mesma situação real. Em seguida, uma comparação entre os resultados obtidos nas simulações é realizada. Finalmente, na nona seção as considerações finais do autor, bem como as sugestões de pesquisas futuras são apresentadas.



## 2 JUSTIFICATIVA

Perdas econômicas geradas por problemas com dados e informação estão estimadas em 600 bilhões de dólares anuais (ECKERSON, 2002), o que justifica, em termos econômicos, a motivação pela pesquisa acadêmica acerca do tema da QI. Dentre as dimensões nas quais a QI pode ser gerenciada, as relacionadas ao tempo como a volatilidade da informação, por exemplo, figuram entre as mais citadas nos estudos publicados na área (WAND e WANG, 1996). Dimensões relacionadas com o tempo são consideradas das mais importantes na gestão da QI (KLEIN et al. 2007, YU et al. 2007 apud HEINRICH e KLIER, 2009).

Somente nos Estados Unidos, cerca de 23% das correspondências são enviadas a endereços errados (HELLMAN, 2004) e uma das causas desta grande penetração de erros na informação “endereço” é a má gestão da QI na dimensão volatilidade. O que acontece é que a informação é dinâmica no mundo real e estática nos bancos de dados das organizações (KAISER, KLIER e HEINRICH, 2009). Isso significa dizer que a informação está em constante mudança no mundo real – as pessoas mudam de endereço, telefone, nome, profissão entre outros atributos – e, a menos que a organização gerencie a volatilidade e realize esforços de atualização, a informação tende a tornar-se obsoleta ao longo do tempo. O modelo referido de Görz (2011), que será detalhado na seção 3, auxilia a decisão de *atualizar* a informação considerando a idade do dado e a sua volatilidade no mundo real. Entretanto o modelo matemático proposto peca ao ser aplicável somente a um atributo informacional, ignorando a hipótese de que, eventualmente, o custo de atualizar numa só ação dois atributos cuja fonte é a mesma, pode ser menor que o custo de atualizar ambos os atributos desencadeando ações independentes. Por exemplo, é razoável pensar que o custo da ação de atualizar endereço, telefone, email, estado civil entre outros atributos de um cliente num único contato é menor que a soma dos custos de realizar um contato e atualizar cada um destes atributos. A razoabilidade da hipótese de que o custo consolidado das decisões de *atualizar* ou *não atualizar* um conjunto de atributos para um grande volume de registros é menor que o custo das decisões de atualizá-los com ações independentes justifica a elaboração do presente estudo, a fim de oferecer uma ferramenta ainda mais eficiente que o modelo matemático de Görz à gestão da QI.

### 3 OBJETIVOS DE PESQUISA

Esta seção visa à descrição dos objetivos gerais e específicos que nortearão o presente estudo. O objetivo geral trata do que se pretende conhecer, enquanto os específicos visam descrever a ordem lógica do processo de construção de conhecimento através do qual este estudo pretende guiar-se a fim de alcançar o objetivo geral.

#### 3.1. OBJETIVO GERAL

O presente estudo tem como objetivo geral propor e validar uma ampliação do modelo matemático proposto por Görz (2011) a fim de contribuir com área da gestão da qualidade informação.

#### 3.2. OBJETIVOS ESPECÍFICOS

Especificamente, este estudo se dedica a:

- Propor ampliação do modelo matemático proposto por Görz (2011) enquanto ferramenta para gestão da QI na dimensão volatilidade que contemple mais de um atributo de informação, bem como considere o custo futuro projetado das decisões.
- Conduzir simulação de aplicação do modelo ampliado através de um caso prático e corriqueiro de uso da informação.
- Realizar comparação, em termos da eficiência, entre os resultados obtidos na aplicação do modelo original com os obtidos na aplicação do modelo ampliado.

## 4 REFERENCIAL TEÓRICO

Nesta seção, primeiramente, os conceitos de informação e qualidade da informação são apresentados. Em seguida apresentam-se as dimensões da QI através de uma compilação de vários autores sobre o tema. Em seguida, a gestão da QI é apresentada através das etapas reconstrução da situação, avaliação e melhoria. Na sequência a dimensão volatilidade, foco do presente estudo, recebe um maior detalhamento na sua conceituação. Finalmente o modelo matemático proposto por Görz (2011) usado como ferramenta para a gestão da QI é apresentado. Uma breve revisão sobre os conceitos de entidade, relacionamento e atributos é realizada. Finalmente apresenta-se a métrica usada pelo modelo, os erros das decisões e seus custos.

### 4.1. O CONCEITO DE INFORMAÇÃO

Vivemos hoje a era da informação, nas últimas décadas o mundo experimentou a migração de uma economia industrial para uma economia da informação na qual ela é a principal força motriz na geração de valor e riqueza (MCGEE e PRUSAK, 1994). A informação pode ser considerada, hoje, o segundo mais importante recurso das organizações, atrás apenas dos recursos humanos (ENGLISH, 2011).

A informação exerce papel importantíssimo na definição das estratégias organizacionais, na execução das mesmas e no processo final de avaliação e aprendizado (MCGEE e PRUSAK, 1994). Peter Drucker (1995) corrobora esta visão afirmando a necessidade de informações sobre clientes, mercado, ambiente, tecnologias, finanças e economia por parte das empresas.

A literatura acadêmica na língua portuguesa, frequentemente, dedica tempo à diferenciação entre os termos *dado* e *informação*. Por outro lado, o mesmo não ocorre na literatura na língua inglesa ao focar o tema QI onde os termos *data quality* e *information quality* são na maioria das vezes apresentados como sinônimos perfeitos (PIPINO, LEE e WANG, 2002). Entretanto, é indispensável uma definição do termo. Para Drucker (1992) a informação pode ser definida como um conjunto de “dados dotados de relevância e

propósito”. McGee e Prusak (1994) afirmam que informação é um conjunto de dados reunidos, organizados, ordenados, aos quais são atribuídos significados e contexto. Afirmam ainda que não se pode pensar a informação de maneira dissociada ao usuário, já que informação representa, precisamente, os dados em uso.

#### 4.2. O CONCEITO DE QI

Podendo ser considerada uma recente área de estudo (LIMA e MAÇADA, 2008), a QI vem sendo contemplada por um amplo número de pesquisas científicas desde os seus primórdios nos anos 90, quando começou a ser pesquisada e os primeiros conceitos foram surgindo a respeito do tema. Um dos estudos pioneiros afirma que a informação de boa qualidade deve refletir a realidade e ser fácil de usar e entender por parte do usuário (WANG, KON e MADNICK, 1993).

Pesquisas que desenvolvem o tema vêm apresentando visões mais amplas do conceito de QI, como Strong, Lee e Wang (1997) que afirmam que uma elevada QI pode ser atribuída à informação que se adéqua a necessidade de quem a consome. Uma visão semelhante é fornecida por Kahn, Strong e Wang (2002), que definem qualidade como sendo a capacidade de um produto atender com sucesso ao propósito de quem o consome e transportam este conceito para o campo da informação.

Thomas Redman (2005) também apresenta uma visão semelhante, entretanto reduz o abarcamento semântico do conceito afirmando que a informação pode ser considerada de alta qualidade quando é adequada ao seu uso, pelos clientes, nas operações, na tomada de decisão e no planejamento.

McGee e Prusak (1994) afirmam que a QI é alcançada através de um “cuidado detalhado com a integridade, precisão, atualidade, interpretabilidade e valor geral da informação, julgado pelos seus clientes”. Percebe-se nesta definição a intenção organizar a QI em diferentes dimensões, tópico que vem sendo estudado por muitos autores e ao qual este estudo oferece atenção especial na subseção seguinte.

Uma visão distinta do conceito de QI é apresentada por English (LIEWELYN, 2008), pesquisador e consultor de projeção internacional cujas pesquisas foram iniciadas no

princípio dos anos 90 e nas quais buscou trazer o conceito de gestão da qualidade total (TQM) para a informação fazendo ampla pesquisa sobre autores clássicos na área de TQM tais como Joseph Juran e, sobretudo, Edwards Deming. English apresenta uma visão distinta sobre o tema, pois afirma não haver se guiado pelas pesquisas que já vinham sendo publicadas naquele tempo (LIEWELYN, 2008). English conceitua a QI, de maneira distinta da dos demais pesquisadores, como a “eliminação de desperdício de informação suja e retrabalho, processos desnecessários e incremento da efetividade do negócio por aumentar a satisfação do consumidor de produtos e serviços de informação” (LIMA e MAÇADA, 2007).

#### 4.3. AS DIMENSÕES DA QI

Nos primórdios dos estudos que a contemplaram, a QI era pensada somente sob seu ponto de vista intrínseco (STRONG, LEE e WANG, 1997). De certo modo, QI poderia ser atribuída a uma informação que fosse livre de erros. As pesquisas de Wang e Strong (1996) possibilitaram ampliar a visão sobre a QI. Tais pesquisadores partiram da ideia de que para que se pudesse melhorar a QI seria importante compreendê-la mais profundamente e de forma mais ampla. Muitos outros autores também vêm desenvolvendo estudos no intuito de definir as dimensões nas quais a QI pode ser considerada. Isso significa dizer que se vem buscando encontrar quesitos através dos quais a QI possa ser testada, melhorada e, conseqüentemente, gerenciada. Na subseção 4.4 a gestão da QI será explorada mais detalhadamente.

Atualmente há muitas proposições distintas sobre as dimensões da QI. Cada autor que dispõe sobre o tema tem a sua particular maneira de arranjar-las. Lima e Maçada (2007) em seu estudo que visa aperfeiçoar o modelo genérico de dimensões da QI para a indústria bancária no Brasil, fazem um resumo dos principais autores e suas visões sobre o tema. Na Figura 5 percebe-se o fato de que alguns autores usam nomenclaturas distintas para designar uma mesma dimensão, bem como nota-se que alguns atingem uma maior profundidade em sua proposta, no sentido de que assumem um maior número de dimensões.

PIPINO; LEE; WANG (2002)	McGEE; PRUSAK (1994)	STRONG; LEE; WANG (1997)	REDMAN (1998)	O'BRIEN, (1999)	OZ (2000)	WIXOM; TODD (2005)
Acessibilidade		x	(Privacidade)			
Quantidade	(Integridade)	x				
Credibilidade		x				
Completeza		x	x	X	x	x
Concisão		x		(Detalhamento)		
Consistência		x	x	X		(de Formato)
Entendimento		(Facilidade de)		(Clareza)		(Acurácia)
Facilidade de Uso						
Livre de Erros	(Precisão)	(Acurácia)	(Acurácia)	(Acurácia)	(Acurácia)	
Interpretabilidade	x	x	(Facilidade de)			
Objetividade		x		(Escopo)		
Relevância		x	x	X	x	
Reputação		x				
Segurança		(de Acesso)	x			
Volatilidade	(Atualidade)	(Temporalidade)	(Atualização)	X	(Atualidade)	(Atualidade)
OUTROS	Valor Geral	Valor	Nível de Detalhe; Posse. Granularidade;	Atualidade; Frequência; Amplit. de Tempo; Performance; Ordem; Apresentação; Mídia.		

Tabela 1 - As dimensões da QI de diversos autores. Adaptada de LIMA e MAÇADA (2007).

Strong, Lee e Wang (1997), através de uma sequência de experimentos estatísticos que visavam excluir a ambiguidade semântica que poderia haver entre duas ou mais dimensões e reduzir o número das mesmas a fim de tornar o modelo mais inteligível chegou-se a quinze dimensões organizadas em quatro categorias: QI intrínseca; QI de acessibilidade; QI contextual; e QI representacional.

Entram na categoria intrínseca as dimensões que dizem respeito às características do dado em si, já a categoria contextual leva em consideração o contexto no qual a tarefa para a qual a informação será usada está inserida, enquanto que as categorias de acessibilidade e representacional enfatizam a importância do sistema como um todo.

A Tabela 2 apresenta uma descrição da definição de cada uma das dimensões da QI. O bom entendimento das definições se mostra importante já que, embora os seus pesquisadores tenham se esforçado em isolá-las semanticamente, ainda há uma proximidade entre alguns termos, o que pode facilmente confundir quem estuda o tema.

<b>Dimensão de QI</b>	<b>Descrição</b>
<b>Acessibilidade</b>	A informação é disponibilizada ou pode ser facilmente e rapidamente extraída.
<b>Completeza</b>	Não há informação faltando e é suficiente para a tarefa.
<b>Concisão</b>	A informação é apresentada de maneira compacta.
<b>Consistência</b>	A informação é apresentada no mesmo formato.
<b>Credibilidade</b>	A informação é considerada confiável e verdadeira.
<b>Facilidade de Compreensão</b>	A informação é de fácil compreensão.
<b>Interpretabilidade</b>	A informação é disponibilizada no idioma apropriado, apresenta símbolos e unidades conhecidos e os termos usados são claros.
<b>Livre de Erros</b>	A informação é correta.
<b>Objetividade</b>	A informação não sofre viés e é imparcial.
<b>Quantidade</b>	O volume de informação disponível não é excessivo nem escasso para a tarefa
<b>Relevância</b>	A informação é aplicável e auxilia na tarefa.
<b>Reputação</b>	A informação tem um histórico positivo em termos de qualidade.
<b>Segurança ao acesso</b>	A informação é restringida apropriadamente para manter a sua segurança.
<b>Valor agregado</b>	A informação beneficia e provê vantagens para quem a usa.
<b>Volatilidade</b>	A informação é suficientemente atualizada para a tarefa.

Tabela 2 - Descrição das Dimensões da QI. Adaptada de KAHN, STRONG e WANG (2002).

#### 4.4. A GESTÃO DA QUALIDADE DA INFORMAÇÃO

Batini et al. (2007) evidenciam as perdas geradas pela QI em Organizações públicas e privadas. Batini et al. (2009) destacam que, para minimizar tais consequências, são utilizadas diversas técnicas, como relacionamento entre bases de dados, regras de negócio e medidas de similaridade. Entretanto, com a complexidade e diversidade das técnicas existentes, algumas pesquisas têm se voltado para a definição de metodologias que buscam selecionar, customizar e aplicar a avaliação da QI e a melhoria de tais técnicas.

Batini et al. (2009) comparam e consolidam as metodologias existentes e evidenciam as três etapas comuns à maioria delas: reconstrução da situação, a avaliação da QI e a melhoria na QI. Tais etapas serão detalhadas a seguir.

#### 4.4.1. RECONSTRUÇÃO DA SITUAÇÃO

Esta etapa consiste em coletar dados, regras de negócio, procedimentos gerenciais, indicadores, custos gerenciais, entre outras variáveis, com o intuito de embasar a análise da informação. Segundo os autores, esta fase pode ser desconsiderada caso a avaliação da QI possa ser baseada em documentação existente e esta não seja considerada obsoleta.

#### 4.4.2. A AVALIAÇÃO DA QI

A segunda etapa da gestão da QI é avaliá-la nas dimensões e critérios estabelecidos na etapa anterior. Para que se possa avaliar a QI de uma organização, necessita-se invariavelmente de métricas. A qualidade, embora possa ser avaliada através de palavras, para que seja possível melhorá-la deve-se encontrar um método que transforme tais palavras em números, possibilitando assim a sua medição gestão (JURAN, 1992 apud. BRODBECK, 1995). Isso se alinha com as afirmações de Scoot Sink e Thomas Tuttle (1993)<sup>1</sup> (apud. BRODBECK, 1995), que acreditam que a medição é indispensável ao controle, gerenciamento e conseqüente aprimoramento de processos organizacionais.

Pipino, Lee e Wang (2002) afirmam que não há um método genérico de avaliação da QI que possa ser considerado ideal para todos os casos. Entretanto há princípios que devem ser seguidos, e sobre os quais se devem desenvolver métricas subjetivas e objetivas de avaliação da QI. A QI é uma disciplina multidimensional (WANG e STRONG, 1996). No que toca o tema da avaliação da QI, Pipino, Lee e Wang (2002) ressaltam a importância de cada dimensão seja submetida a testes objetivos e subjetivos a fim de que se obtenha uma avaliação mais detalhada.

Pipino, Lee e Wang (2002) sugerem a organização dos resultados obtidos nas avaliações subjetivas e objetivas da QI em uma matriz 2x2 (resultado da avaliação subjetiva x

---

<sup>1</sup> Para esclarecimentos consultar SINK, S.; TUTTEL, T.C. Planejamento e Medição para Performance. Ed. Qualitymark. Rio de Janeiro, 1993.



resultado da avaliação objetiva) e a consequente distribuição das dimensões em um dos quadrantes conforme Figura 1.

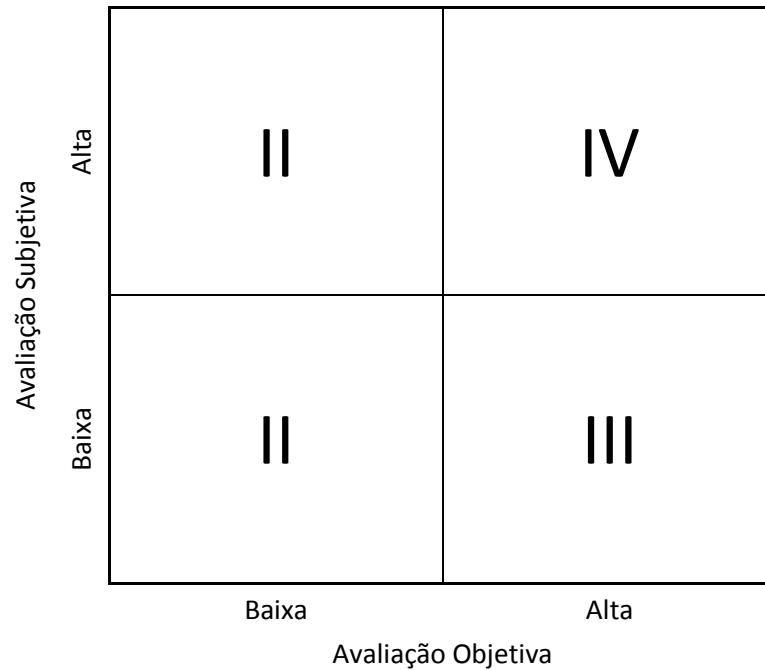


Figura 1- Avaliação Subjetiva e Objetiva da QI. Adaptado de PIPINO, LEE e WANG (2002).

Por outro lado, estudos mais recentes vêm dedicando grande esforço para encontrar métricas aplicáveis que possam ser empregadas de forma metodológica na gestão da QI para sua avaliação em dimensões específicas (HEINRICH, KAISER e KLIER, 2007; HELFERT et al., 2009; BATINI et al., 2007).

Em meio ao processo de avaliação da QI duas importantes definições da QI se tornam pertinentes: a QI como qualidade do design que se refere ao grau de correspondência entre as necessidades dos consumidores e as especificações dos sistemas de informação, e a qualidade de conformidade que representa o grau de correspondência entre as especificações dos sistemas de informação e o que é de fato incorporado pelo sistema. Heinrich, Kaiser e Klier (2007) atentam para a diferença no momento de avaliar a QI para as duas distintas definições. Para avaliar a qualidade de design, recomenda-se uma avaliação subjetiva que perpassa os consumidores da informação através de questionários e outras ferramentas semelhantes, enquanto que para avaliar a qualidade da conformidade a uma análise objetiva que quantifique a correspondência entre o especificado e o valor existente

registrado (KAISER, KLIER e HEINRICH, 2009). Em linhas gerais a qualidade de conformidade está prioritariamente relacionada com o valor registrado no banco de dados, com aspectos intrínsecos do dado (HEINRICH, KAISER e KLIER, 2007). A dimensão volatilidade – foco deste estudo – e todas as dimensões ligadas ao tempo de maneira geral enquadram-se na definição de qualidade de conformidade (HEINRICH e KLIER, 2009).

#### 4.4.3. A MELHORIA DA QI

A Terceira etapa consiste em definir os passos, técnicas e estratégias para melhorar a QI.

As metodologias analisadas por Heinrich, Kaiser e Klier (2007) apresentam, em geral, dois tipos de estratégia para melhorar a QI: A orientada a processo e a orientada a dados. A primeira melhora a QI por meio da alteração dos valores dos dados. A segunda estratégia redesenha os processos que criam ou modificam os dados. A ambas as estratégias se aplicam variadas técnicas e os autores destacam algumas delas, conforme ilustrado na Tabela 3.

ESTRATÉGIA	TÉCNICA
<b>Orientada a dados</b>	<ul style="list-style-type: none"> <li>- Aquisição de novos dados</li> <li>- Padronização (ou normalização)</li> <li>- Relacionamento entre bases de dados (<i>record linkage</i> – RL)</li> <li>- Integração de esquema de bancos (<i>database schema</i>) e dados</li> <li>- Confiabilidade de fontes de informação</li> <li>- Localização e correção de erros</li> <li>- Otimização de custos</li> </ul>
<b>Orientada a processos</b>	<ul style="list-style-type: none"> <li>- Controle do processo</li> <li>- Redesenho do processo</li> </ul>

Tabela 3 - Estratégias e Técnicas para melhoria da QI. Adaptado de Heinrich, Kaiser e Klier (2007).

Redman<sup>2</sup> apud (BATINI et al, 2009) discorre sobre os ganhos que cada uma das técnicas pode proporcionar. O autor faz uma comparação e conclui que, no longo prazo, as técnicas orientadas a processos apresentam resultados superiores às técnicas de orientadas

---

<sup>2</sup> REDMAN, T. *Data Quality for the Information Age*. Artech House. (1996)

a dados. Em uma perspectiva de curto prazo, entretanto, redesenhar o processo pode se tornar economicamente inviável. Por outro lado, no curto prazo, as técnicas orientadas a processos podem ser mais eficientes e econômicas. As técnicas orientadas a dados são apropriadas para aplicações únicas e por isso recomendadas para dados estáticos. O presente estudo propõe uma ampliação em um modelo matemático que é ferramenta que possibilita a melhoria da QI de forma recorrente e sistemática, configurando, assim, a adoção de uma estratégia orientada a processo.

#### 4.5. A DIMENSÃO VOLATILIDADE

Diversos autores aprofundaram suas pesquisas no intuito de definir as distintas dimensões da QI. Conforme visto anteriormente, a organização das dimensões varia de autor para autor, por isso se faz necessária à revisão dos conceitos para a dimensão volatilidade. Uma revisão compilada dos conceitos da dimensão em questão é contemplada pela Tabela 4.

Ballou and Pazer [1995, p. 153] Termo original: <i>Timeliness</i>	Refere-se a se o valor registrado não está desatualizado [...]. Um valor armazenado que se torna desatualizado caso encontre-se diferente do valor vigente no mundo real.
[Wang and Strong 1996, p. 32] Termo original: <i>Timeliness</i>	É a grandeza em que a idade vigente de um dado é apropriada para a tarefa a ser executada.
[Redman 1996, p. 258] Termo original: <i>Currency</i>	Refere-se ao grau em que um dado em questão está atualizado. O dado é considerado atualizado se está correto apesar de possíveis discrepâncias causadas por mudanças relacionadas ao tempo.
[Pipino et al. 2002, p. 212] Termo original: <i>Timeliness</i>	É a grandeza em que o dado está suficientemente atualizado para a tarefa a ser executada.
[Batini and Scannapieco 2006, p. 29] Termo original: <i>Timeliness</i>	Expressa o quão atual do dado está para a tarefa a ser executada.

Tabela 4 - Definições da dimensão "Volatilidade". Adaptado de Görz (2011).

#### 4.6. O MODELO DE DECISÃO

As metodologias de gestão da QI existentes frequentemente negligenciam métodos para determinar o nível desejado de QI ou costumam delegar atenção ao nível gerencial deixando de lado o nível operacional (Görz, 2011). O modelo matemático proposto por Görz (2011) contempla o enfoque operacional da gestão da QI e tem como ideia principal comparar o grau de atualização da informação com o seu grau de atualização desejado de forma que se permita decidir sobre a ação de atualizá-la ou não, tendo como premissa o menor custo.

O modelo contempla uma métrica específica para a dimensão volatilidade, os custos envolvidos no uso da informação (esteja ela desatualizada ou não) e o custo de atualizá-la. Na prática, o evento que desencadeia sua aplicação é a necessidade de utilizar determinada informação em um banco de dados. Neste instante, cada registro do banco considera-se a data da última atualização como entrada para calcular a idade da informação contida no atributo que está sendo analisado. A métrica para a volatilidade empregada pelo modelo, os custos relacionados ao uso da informação, as receitas esperadas devido ao seu uso e o custo de executar a ação de atualizá-la, compõe o rol de parâmetros do modelo, o qual proporciona como saída a decisão de *atualizar* ou não a informação contida no atributo. Para o melhor entendimento do modelo em questão a definição de alguns conceitos se faz necessária.

##### 4.6.1. ENTIDADE, RELACIONAMENTO E ATRIBUTO.

Para o entendimento do modelo é necessário conhecer conceitos utilizados no processo de organização de dados em um banco de dados, tais como entidade, relacionamento e atributo informacional. Tais conceitos foram criados por Peter Chen (1976) a fim de organizar a informação existente no mundo real em bancos de dados. Entende-se por entidade algo que, no mundo real, pode ser distintamente identificado. Exemplos corriqueiros de entidades em bancos de dados são “pessoa”, “empresa”, “projeto”, “departamento”, entre muitos outros. Num banco de dados, as entidades podem ter relacionamentos entre si, como por exemplo, o relacionamento entre uma pessoa e uma

empresa pode configurar a existência de um vínculo de emprego onde “pessoa” é o empregado de “empresa”.

Cada entidade possui um conjunto de atributos informacionais. Um atributo informacional pode ser formalmente definido como a função que mapeia uma entidade ou relacionamento entre entidades em um conjunto de valores ou produto cartesiano de um conjunto de valores. Por exemplo, a entidade “pessoa” tem uma série de atributos tais como: nome, idade, estado civil, endereço residencial, telefone celular, entre muitos outros. Cada entidade tem uma infinita quantidade de atributos, porém, nos bancos de dados, apenas os atributos relevantes para o sistema de informação em questão costumam ser mapeados. Alguns atributos podem ainda ser desmembrados em conjuntos de valores, por exemplo, o atributo nome que pode ser desmembrado entre: primeiro nome, nome do meio e sobre nome. A Figura 2 auxilia o entendimento do modelo de organização de dados entidade relacionamento.

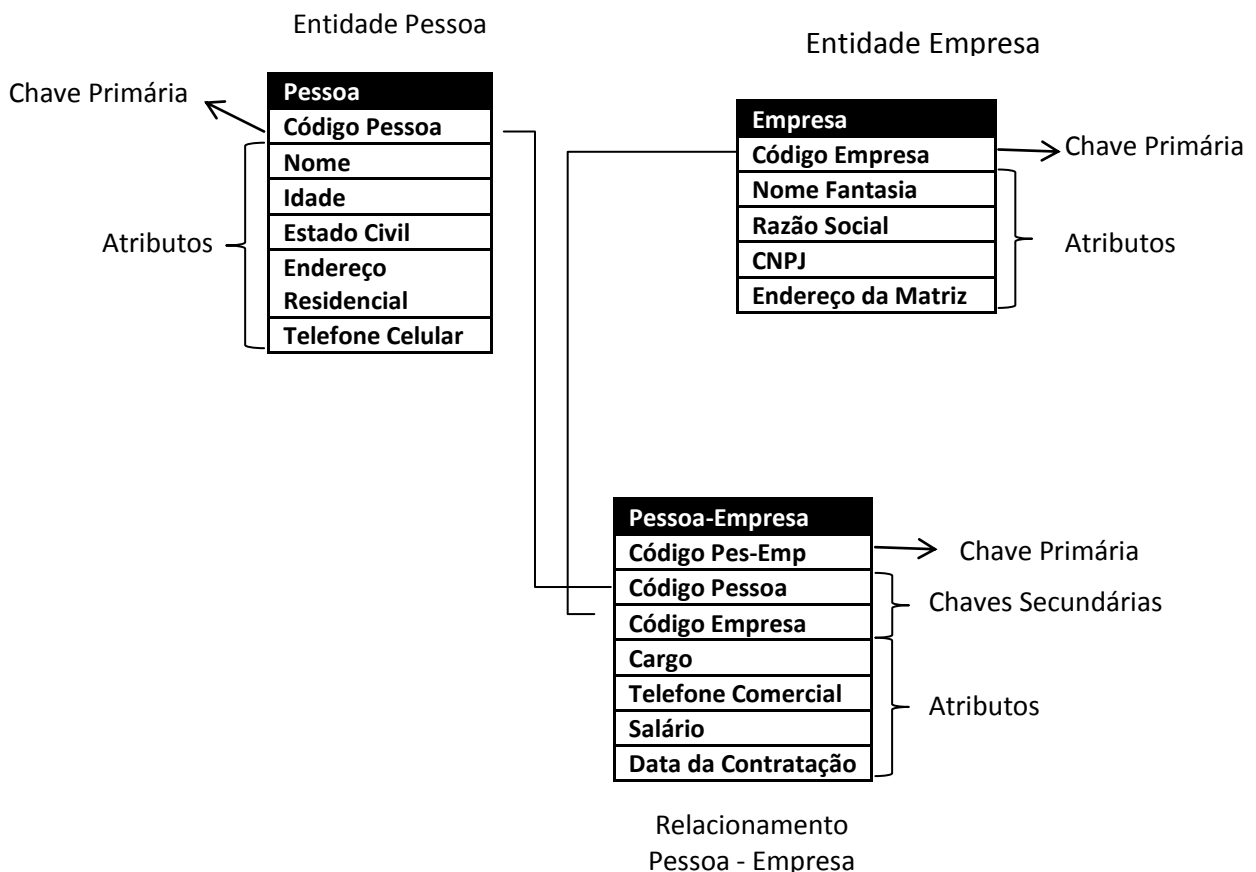


Figura 2 - Esquema entidades e relacionamentos. Elaborado pelo autor.

#### 4.6.2. A MÉTRICA PARA VOLATILIDADE

A métrica utilizada no modelo foi proposta por Heinrich e Klier (2009), a qual se resume em identificar a probabilidade de determinado atributo informacional de um determinado registro estar atualizada em um determinado tempo. O emprego de uma métrica cuja saída é uma probabilidade se justifica, pois, como será apresentado na sequência, o modelo considerará o custo total da decisão de *não atualizar* a informação.

A métrica definida por Heinrich e Klier (2009) é representada por. (1)

$$Q_{\omega curr}(t, w_1, \dots, w_n) = P_{\omega}(T \geq t | W_1 = w_1, \dots, W_n = w_n) \quad (1)$$

Logo a  $Q_{\omega curr}(t, w_1, \dots, w_n)$  é a probabilidade condicional de que o prazo de validade de T seja maior ou igual à idade t da informação do atributo  $\omega$ , dada a informação suplementar  $w_n$ , ou seja de a informação estar atualizada.

Uma distribuição tipicamente empregada para o tempo de vida e comprovadamente utilizável na gestão da qualidade é a distribuição exponencial (HEINRICH e KLIER, 2009). Heinrich, Kaiser e Klier (2007) representam a métrica utilizando a distribuição exponencial conforme (2).

$$Q_{\omega curr}(t) = \exp(-declinio(A) \times t) \quad (2)$$

Onde  $declinio(A)$  é o coeficiente de declínio que indica o percentual de registros cuja informação para o atributo  $\omega$  se torna desatualizada a cada período de tempo  $t$ . Por exemplo, se para o atributo  $\omega$ , a cada período de tempo  $t$ , 20% dos registros em um banco de dados a informação registrada muda no mundo real, o coeficiente de declínio seria 0,2 e em  $t = 5$  aproximadamente 37% dos registros estariam com o valor desatualizado no banco de dados.

A Figura 3 mostra graficamente a distribuição da probabilidade e a sua evolução ao longo de vinte períodos de tempo.

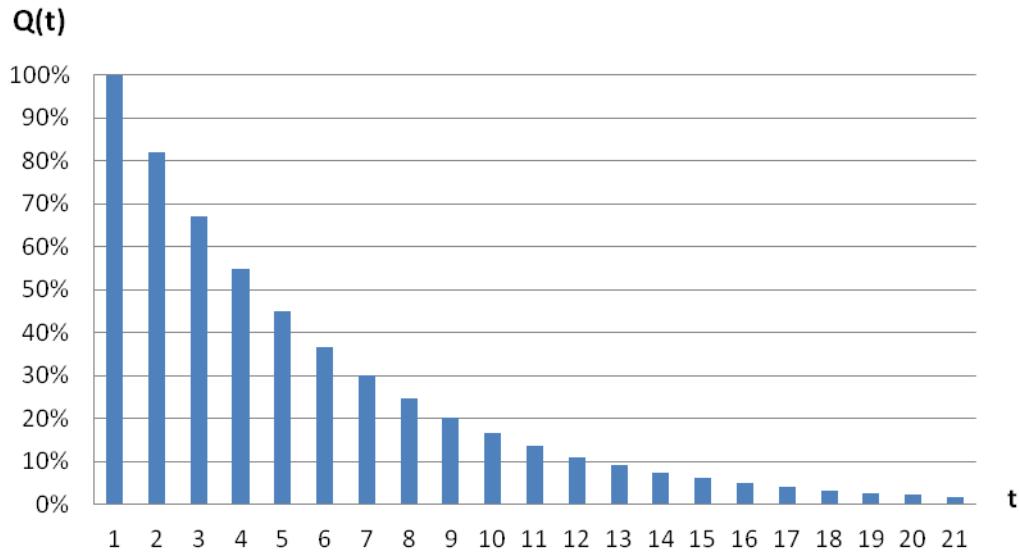


Figura 3- Gráfico da distribuição da probabilidade de a informação estar atualizada em cada t. Elaborado pelo autor.

A informação suplementar  $w_n$  fora incluída na métrica a fim de validar a hipótese de que a volatilidade (ou seja, a frequência em que determinada informação muda no mundo real) de um determinado atributo informacional depende do registro. Logo, a métrica se tornaria mais assertiva se considerasse o registro para o qual o atributo está sendo medido. Na pesquisa, através de um estudo específico, se identificou a existência de agrupamento de registros de uma mesma entidade com volatilidades bastante distinta para um mesmo atributo. E comprovou-se que a taxa de métrica deve considerar a informação contida em outros atributos do mesmo registro para ser mais assertiva. No estudo, entretanto, não vamos nos ater a este nível de detalhamento. As peculiaridades da base e com a qual exemplificaremos o uso da métrica, portanto, deixam de ser relevantes.

Consideraremos então a métrica em sua forma simplificada conforme (4).

$$Q_{\omega,n}(t) = P_{\omega,n}(T_{\omega} \geq t_{\omega,n}) \quad (3)$$

$$Q_{\omega,n}(t) = \exp(-\text{declínio}_{\omega} \times t_{\omega,n}) \quad (4)$$

Onde:

- $Q_{\omega,n}(t)$  é a probabilidade de a informação contida no atributo  $\omega$  do registro  $n$  ser a mesma que a informação no mundo real;
- $declínio_{\omega}$  é o coeficiente de declínio do atributo  $\omega$ ;
- $t_{\omega,n}$  é a idade da informação contida no atributo  $\omega$  do registro  $n$ .

#### 4.6.3. ERROS E CUSTOS DOS ERROS

O gatilho que desencadeia a aplicação do modelo é a necessidade de utilização da informação contida no atributo informacional  $\omega$ . O modelo assume a premissa que o uso da informação contida no atributo  $\omega$ , por uma organização, acarreta a assimilação de uma receita esperada  $R_{\omega}$  e da absorção de um custo esperado  $U_{\omega}$ . O uso do atributo informacional  $\omega$  sugere, portanto, um lucro esperado dado por  $R_{\omega} - U_{\omega}$ . Na circunstância de a informação ser utilizada estando desatualizada a receita  $R_{\omega}$  não é assimilada, já que a informação levará a uma ação ineficaz, o custo de sua utilização, no entanto, é determinístico, ou seja, ocorrerá estando a atualização errada ou não.

Assim, com a incorporação da métrica anteriormente explicada temos que o custo total da decisão de *não atualizar* é o custo de utilizar a informação  $U_{\omega}$  mais a receita esperada pelo uso da informação  $R_{\omega}$  multiplicada pela probabilidade de a informação estar desatualizada no momento  $t$   $P_{\omega}(t)$ , que é, justamente, probabilidade complementar da métrica anteriormente detalhada. A não assimilação da receita esperada  $R_{\omega}$  passa a configurar um custo. Sendo assim temos a equação do custo total da decisão de *não atualizar* em (5).

$$CT(\text{não atualizar}) = R_{\omega} \times P_{\omega}(t) + U_{\omega} \quad (5)$$

A decisão de *atualizar* a informação incorpora o custo  $A_{\omega}$ , que unifica todos os eventuais custos relacionados à ação de *atualizar* a informação contida no atributo. Para análise da decisão de *atualizar* a informação deve-se assumir a premissa de que a tomada desta ação de custo  $A_{\omega}$  acarreta invariavelmente em alimentar o atributo  $\omega$  com a



informação correta e atual, ou seja, deixar a informação nele contida em plena conformidade com o mundo real. Isso significa dizer que, para a validade do modelo deve-se assumir que a ação de atualizar a informação é sempre eficaz e que na decisão de *atualizar*, a probabilidade de a informação estar desatualizada é de 0%. Logo, a receita esperada  $R_\omega$  será assimilada invariavelmente e deixa de configurar custo no modelo. O custo  $U_\omega$  gerado pelo uso da informação, no entanto, é igualmente determinístico nesta decisão, e será somado ao custo da ação de atualizar a informação  $A_\omega$  para compor o custo total da decisão de *atualizar* a informação. Temos, portanto, o custo total da decisão de *atualizar* em (6).

$$CT(\textit{atualizar}) = A_\omega + U_\omega \quad (6)$$

Finalmente, já que o modelo busca encontrar a melhor decisão em termos do menor custo, resta comparar o custo total das duas decisões para que o modelo forneça a resposta. Sempre que o custo total da decisão de *atualizar* a informação for menor que o custo total de *não atualizar*, a decisão de *atualizar* é recomendada. O que leva ao limiar de decisão (7).

$$CT(\textit{atualizar}) < CT(\textit{não atualizar}) \quad (7)$$

Substituindo pelos valor de (5) e (6) chega-se a (8).

$$A_\omega + U_\omega < R_\omega \times (1 - Q_\omega(t)) + U_\omega \quad (8)$$

Derivando (8) algebricamente encontra-se (9).

$$Q_\omega(t) < 1 - \frac{A_\omega}{R_\omega} \quad (9)$$

Logo temos que a decisão de *atualizar* deve ser tomada quanto o lado esquerdo da inequação for menor que o lado direito.

Observando os parâmetros de ambos os lados, percebe-se que no lado direito da inequação todos os valores são constantes para o atributo  $\omega$ , portanto a aplicação do modelo depende apenas do conhecimento dos valores  $A_\omega$  e  $R_\omega$ , e do cálculo da métrica

$Q_{\omega}(t)$  onde  $t$  é a idade do valor contido no atributo explicitado na unidade de tempo utilizada na métrica.

## 5 MÉTODO DE PESQUISA

Nesta seção expõem-se os procedimentos metodológicos e ferramental utilizados no presente estudo. O estudo baseia-se na ampliação do modelo original proposto por Görz (2011) aliado a uma simulação de aplicação dos modelos ampliado e original em um caso prático de uso da informação. Para isso, inicialmente será apresentado o método e ferramentas utilizados na ampliação do modelo original. Em seguida, são explicitados os procedimentos metodológicos que possibilitaram a simulação de aplicação dos modelos: a escolha da empresa e do processo sobre o qual o modelo foi aplicado; as técnicas de coleta e tratamento de dados necessários para a construção do modelo para o caso em questão; métodos e ferramentas usadas para realizar as simulações. Finalmente, nesta seção, se apresenta como se deu a comparação dos resultados obtidos na simulação do modelo ampliado com os obtidos para o mesmo caso com o modelo original.

### 5.1. AMPLIAÇÃO DO MODELO ORIGINAL

O alicerce central do modelo previamente detalhado – a comparação entre os custos totais de duas possíveis decisões – se mantém inalterado para o modelo ampliado. A ampliação do modelo proporciona a sua aplicação para  $n$  atributos e não somente para um, portanto uma série de ajustes se faz necessária. A ampliação do modelo se deu através de uma sequência de etapas iniciada por uma análise dos custos totais das decisões envolvidas no modelo. Na continuação realizou-se a ampliação no sentido de considerar o custo total projetado de cada decisão, ou seja, não somente o custo no instante de tempo da aplicação do modelo mas todos os instantes de tempo subsequentes. Finalmente realizou-se a ampliação no sentido de contemplar múltiplos atributos de informação ao invés de apenas um.

#### 5.1.1. ANÁLISE DO MODELO ORIGINAL

Inicialmente realizou-se uma análise dos custos totais de cada decisão no modelo original, onde a natureza determinística/probabilística de cada custo foi questionada. Nesta

etapa, algumas propriedades matemáticas da teoria das probabilidades tiveram de ser aplicada.

Como afirma Jaynes (1995), a probabilidade de que eventos independentes ocorram simultaneamente é dada pelo produto entre as probabilidades de que os eventos ocorram independentemente. Por exemplo, para os eventos independentes  $A$  e  $B$ , cujas probabilidades são dadas por  $P(A)$  e  $P(B)$ , a probabilidade de que ambos os eventos ocorram simultaneamente (que na notação mais utilizada é representada por  $P(A \cap B)$ ) é representada em (10):

$$P(A \cap B) = P(A) \times P(B) \quad (10)$$

No intuito de calcular a probabilidade de os eventos de a informação estar desatualizada e ser utilizado num determinado período de tempo, esta propriedade é utilizada.

### 5.1.2. O CÁLCULO DO CUSTO TOTAL PROJETADO

A ampliação propriamente dita foi conduzida em duas dimensões. Inicialmente realizaram-se os ajustes necessários para que os custos totais de cada decisão contemplassem não somente o instante de aplicação mas, além deste, todos os instantes de tempo subsequentes. Uma vez que as funções de probabilidade envolvidas no modelo têm como argumento um instante de tempo discreto – ou seja, que pertence ao conjunto de números inteiros – lançou-se mão do conceito matemático de limite, mais especificamente do conceito de limite de uma sequência.

O conceito matemático de limite é usado para identificar o comportamento de uma função contínua  $f(x)$  ou uma de uma sequência  $x_1, x_2, \dots, x_n$  quando o seu argumento tende a se aproximar de um determinado valor tanto quanto se queria, podendo este valor ser infinito. Para uma função contínua  $f(x)$  o cálculo do limite é válido se, e somente se, o valor do argumento  $x$  é um número real é representado por (11).

$$\lim_{x \rightarrow a} f(x) = L : x \in \mathbb{R} \quad (11)$$

O limite de sequências podem ser calculados se, e somente se, o argumento  $x$  é um número natural. O limite de uma sequência é representado por (12)

$$\lim_{n \rightarrow a} x_n = L' : n \in \mathbb{N} \quad (12)$$

O modelo ampliado, ao contrário do modelo original, considera a probabilidade de determinada informação ser utilizada estando desatualizada durante um lapso de tempo que vai do instante da aplicação do modelo ao infinito positivo. O cálculo do custo total de uma decisão é o somatório dos custos desta decisão multiplicado pelo limite do somatório da probabilidade do custo em questão.

### 5.1.3. O CÁLCULO DO CUSTO PARA MÚLTIPLOS ATRIBUTOS

Para viabilizar a ampliação do modelo para múltiplos atributos, lançou-se mão do operador matemático *somatório*, que na notação usual é representado pela letra grega sigma ( $\Sigma$ ) e definido conforme (13).

$$\sum_{i=m}^n x_i = x_m + x_{m+1} + \dots + x_n \quad (13)$$

## 5.2. A ESCOLHA DA EMPRESA E DO PROCESSO

Para simular uma aplicação do modelo ampliado enquanto ferramenta para o gerenciamento da QI e ainda exercer uma comparação desta com a do modelo original, selecionou-se intencionalmente uma organização que atendesse aos seguintes critérios:

- ser uma empresa privada;
- estar geograficamente próxima com fins de facilidade de acesso;
- apresentar predisposição para receber a pesquisa;
- desempenhar algum processo com intenso uso da informação;
- permitir o acesso a pessoas envolvidas com os processos de uso da informação;
- autorizar o uso de informações referentes do processo escolhido;

- ter conhecimento da idade da informação contida nos atributos escolhidos, possibilitando a aplicação do modelo;

Após a escolha preliminar de uma empresa que atendesse os critérios referidos, realizou-se uma reunião de apresentação ao diretor operacional da mesma a fim de receber o apoio da organização e a autorização para o desenvolvimento da aplicação do modelo. Após a autorização, foram conduzidas entrevistas e reuniões com especialistas e pessoas da área operacional até que se definisse qual processo seria utilizado para exemplificação modelo. Optou-se pelo processo de venda de renovações de apólices de seguro, que envolve o envio de proposta pelo correio (atributo Endereço deve estar atualizado) seguido de uma cobrança ativa por telefone (atributo Telefone deve estar atualizado). O processo foi escolhido por ter uma recorrência mensal e atender aos seguintes critérios:

- ser viável identificar os custos da ação de usar da informação;
- ser viável identificar os custos da ação de atualizar a informação;
- ser possível identificar a receita esperada pelo uso da informação;
- a volatilidade dos atributos nele utilizados pode ser estimada através do histórico do seu uso.

### 5.3. CONSTRUÇÃO DO MODELO

Com a escolha da empresa processo a ser estudada e os atributos a serem analisados definidos, partiu-se para a identificação dos parâmetros necessários para a construção do modelo.

#### 5.3.1. COLETA E TRATAMENTO DE DADOS

Para a identificação de cada parâmetro do modelo, uma diferente fonte de dados teve de ser consultada, bem como múltiplas técnicas de tratamento tiveram de ser aplicadas. A coleta e tratamento para cada um dos parâmetros do modelo são apresentados nas seguintes subseções.

### 5.3.1.1. A RECEITA ESPERADA PELO USO DA INFORMAÇÃO

A receita esperada pelo uso da informação para cada atributo foi definida através de um apanhado de informações históricas fornecidas pela empresa a cerca do uso dos atributos Telefone e Endereço, no processo de venda de renovações de apólices de seguro. Colheram-se dados de 12 meses (março de 2011 a fevereiro de 2012) referentes ao uso da informação contida nos atributos Endereço e Telefone para o processo em questão. Através do uso do software Microsoft Excel® somou-se (A) a quantidade de registros do banco de dados usados no período e (B) a quantidade de registros usados com informação errada. Subtraindo A-B, encontrou-se quantidade de registros usados com informação correta. Em seguida contabilizou-se (C) a receita gerada pela venda de renovações de apólices de seguro de todo o período, para cada atributo usado. Finalmente, dividindo-se  $C/(A-B)$  encontrou-se a receita esperada pelo uso do atributo desde que este esteja atualizado. O Quadro 1 auxilia o entendimento da sequência de cálculos que levaram a definição deste parâmetro do modelo.

A	Quantidade de registros do banco de dados usados no período
B	Quantidade de registros usados com informação errada
A-B	Quantidade de registros usados com informação correta
C	Receita gerada pelo uso do atributo
$C/(A-B)$	Receita gerada por atributo com informação correta

Quadro 1 – Demonstração da sequência de cálculos que levaram à definição deste parâmetro do modelo. Elaborado pelo autor.

Cada atributo teve a receita esperada pelo seu uso determinada.

### 5.3.1.2. O CUSTO DE USO DA INFORMAÇÃO

Na simulação apresentada da aplicação do modelo, o custo originado pelo uso da informação contida em um registro se dá, genericamente, contabilizado (A) o custo indireto total relacionado com o custo, dividido pela (B) quantidade de usos executadas, chegando-se ao quociente  $A/B$ . Somando este valor com (C) o custo direto unitário têm-se o custo unitário de uso da informação contida em um determinado atributo. O Quadro 2 possibilita um melhor entendimento do método usado para encontrar o custo em questão.

A	Custo indireto total
B	Quantidade de usos
A/B	Custo indireto unitário
C	Custo direto unitário
A/B + C	Custo unitário do uso da informação

Quadro 2 - Demonstração do método usado para encontrar o custo em questão. Elaborado pelo autor.

Para o atributo Endereço, o custo direto foi facilmente identificado, uma vez que a atividade de envio das correspondências é realizada por uma empresa terceira que cobra um valor fixo unitário para cada correspondência enviada. Nos custos indiretos poderiam ser considerados os relacionados à manipulação da informação antes do seu envio a empresa terceira, porém, por se tratar de uma atividade praticamente toda automatizada pelo sistema de informações da empresa, o tempo dedicado pelo profissional que realiza a atividade foi desprezado por influenciar muito pouco no custo total.

Para calcular o custo do uso do atributo Telefone, foi recomendado pela entrevistada, especialista no processo, estratificar o custo direto do indireto. Contabilizou-se como custo indireto, a despesa média mensal do pessoal ligado à área da empresa que realiza a atividade de uso da informação do atributo Telefone (tendo como base o período entre Março de 2011 e Fevereiro de 2012), multiplicando-se pelo percentual de tempo dedicado a esta atividade (segundo estimativa da entrevistada) e finalmente dividido pelo número de registros (clientes) utilizados (contatados) em média num mês. Como custo direto, considerou-se a despesa de telefone média mensal (tendo como base o período entre Março de 2011 e Fevereiro de 2012), multiplicando-se pelo mesmo percentual de tempo dedicado a atividade e igualmente dividido pelo número médio de registros utilizados.

#### 5.3.1.3. O CUSTO DE ATUALIZAÇÃO DA INFORMAÇÃO

Para a construção do modelo ampliado, a identificação do custo de atualização da informação contida no conjunto de atributos para cada registro é mandatória.

Para chegar-se ao custo da ação de atualizar a informação, se fez necessário compreender como se dá o processo de atualização na empresa usada no estudo. Para que



isso fosse possível, uma entrevista com o profissional responsável pela célula de cadastro da empresa foi realizada.

Na entrevista com o profissional da célula notificou-se que não havia um controle claro dos tempos e recursos consumidos pelas atividades desempenhadas pela célula. Nela os profissionais realizam uma grande variedade de atividades o que torna complexa a identificação dos custos de cada atividade. Para não comprometer a concretização da simulação do modelo, os custos relacionados à ação de atualizar a informação foram estimados baseando-se na experiência da entrevistada com o apoio do autor.

A ação de atualizar é composta por atividades como telefonar (seja para o cliente seja para a concessionária que o atende), acessar o sistema de cadastro, procurar o registro do cliente que deve ser atualizado, registrar as novas informações, entre outras. Identificou-se também que os custos envolvidos nestas atividades são as despesas de telefone e o custo de trabalho, originado pelas despesas de pessoal da área. Das atividades elencadas, apenas telefonar origina a despesa com telefone, porém todas geram custo de trabalho.

Assim, o cálculo do custo da ação de atualizar considerou tomou como base dois custos relevantes, o custo médio mensal de um funcionário e o custo médio de telefone por minuto. Baseado em uma lista de aproximadamente 600 ligações telefônicas realizadas em fevereiro de 2012, identificou-se o custo médio de telefone por minuto. Para estimar o (C) custo por minuto de trabalho, partiu-se do (A) custo médio mensal de um colaborador da área, considerando todas as despesas que incidem sobre o mesmo (salários, encargos, benefícios etc.). O Quadro 3 apresenta um resumo da técnica empregada.

Custo mensal do funcionário	A
Custo diário do funcionário	$A/22$
Custo do minuto do funcionário	$C = A/(400 \times 22)$
Custo médio de telefone por minuto	D

**Quadro 3 - Técnica usada para estimar os custos de telefone e trabalho por minuto. Elaborado pelo autor.**

Em seguida, para cada atributo e bem como para o conjunto de atributos, estimou-se (E) o tempo dedicado a atividades de telefonar que foi multiplicado pela soma do (D) custo de telefone por minuto com o (C) custo de trabalho por minuto para encontrar o (G) custo

total da atividade de *telefonar*. O (F) tempo dedicado às demais atividades foi multiplicado pelo (C) custo de trabalho por minuto, o que permitiu encontrar o (H) custo total estimado das *demais atividades*. A soma de G e H permitiu encontrar o custo estimado da ação de atualizar a informação para cada atributo e para o conjunto de atributos. O

Atividade	Tempo (min.)	Custo de Telefone	Custo de Trabalho	Custo Total
Telefonar	E	E x D	E x C	G = E x (C+D)
Demais	F		F x C	H = F x C
Total				G+H

Quadro 4 - Técnica empregada para estimar o custo da ação de atualizar a informação. Elaborado pelo autor.

#### 5.3.1.4. A PROBABILIDADE DE A INFORMAÇÃO ESTAR DESATUALIZADA

Conforme modelo original de Görz (2011) a probabilidade de a informação utilizada estar desatualizada  $P(i)$  quando tiver a idade  $i$  é dada pela probabilidade complementar (14).

$$Q_{\omega curr}(i) = \exp(-declinio(A) \times i) \quad (14)$$

Logo  $P(i)$  é representada por (15).

$$P(i) = 1 - \exp(-declinio(A) \times i) \quad (15)$$

Assim, para a construção da probabilidade empregada o modelo *declinio* deve ser encontrado.

Para encontrar tal coeficiente de declínio, inicialmente, em uma consulta ao banco de dados cadastrais da empresa identificou-se a data de cadastramento de cada cliente. Assim, pôde-se chegar à data do primeiro registro da informação dos atributos Telefone e Endereço para todos os clientes da empresa. Outra consulta ao banco de dados possibilitou identificar todas as atualizações realizadas sobre a base de clientes para os mesmos atributos a partir de Fevereiro de 2009 (mês de início do registro de *logs* de atualização de cadastro). Logo se cruzou a base com a data de cadastramento e com a data de atualização e identificou-se a

última data de registro do dado, informação que permite identificar a idade da informação de cada cliente no banco de dados.

Para encontrar este coeficiente para o atributo Endereço, uma série de consultas a bancos de dados e tratamentos dos dados através do *software* Microsoft Excel® foi conduzida. Coletou-se uma amostra de cerca de 111.000 registros (clientes) cujo atributo Endereço foi utilizado no envio de correspondências pela empresa estudada entre Março de 2011 e Maio de 2012. Em seguida, uma consulta ao banco de dados cadastrais de clientes permitiu identificar, das correspondências, quais foram retornadas pelo serviço e correios, bem como o motivo do retorno. Filtraram-se apenas as correspondências retornadas pelo motivo “mudou-se” e assim tabularam-se as informações conforme Tabela 5.

Código do Cliente	Idade da informação Endereço	Retornada devido a “mudou-se”
XXX	I	Sim/Não
...	...	...

Tabela 5 - Tabulação das correspondências retornadas pelo motivo "mudou-se". Elaborada pelo autor.

Após isso, lançando mão do ferramental oferecido pelo *software* Microsoft Excel®, contabilizaram-se os percentuais de retorno pelo motivo “mudou-se”, ou seja, percentual de registros desatualizados, para cada idade, conforme a Tabela 6.

Idade	Número de registros aos quais foram enviadas correspondências	Números de correspondências retornadas por motivo “mudou-se”	Percentual de registros desatualizados
0	A	B	$B/A \times 100\%$
1			
2			
3			
4			
...			

Tabela 6 - Percentual de registros desatualizados para cada idade. Elaborada pelo autor.

Para encontrar o coeficiente para o atributo Telefone um processo semelhante foi realizado. Uma lista de clientes contatados (registros utilizados) no mesmo período foi fornecida pela empresa, juntamente com a informação de se o contato teve sucesso ou se o telefone registrado estava errado, conforme demonstrado na Tabela 7.

Código do Cliente	Idade da informação Telefone	Telefone errado
XXX	I	Sim/Não
...	...	...

Tabela 7 - Exemplo de tabulação utilizada para encontrar o coeficiente para o atributo Telefone. Elaborada pelo autor.

Logo, através da mesma ferramenta computacional, a contabilização do número de registros com os quais o telefone estava errado dividido pelo número total de registros para os quais se tentou fazer contato, para cada idade, possibilitaram encontrar o percentual de registros desatualizados para o atributo Telefone, conforme exemplificado na Tabela 8.

Idade	Número de Registros com os quais se tentou estabelecer contato telefônico	Número de registros com telefone errado	Percentual de registros errados
0	A	B	$B/A \times 100\%$
1			
2			
3			
4			
...			

Tabela 8 – Informações utilizadas para encontrar o percentual de registros desatualizados para o atributo Telefone. Elaborada pelo autor.

Após o cálculo dos percentuais de registros errados para cada idade e a idade propriamente dita, os valores foram inseridos no *software* Curve Expert® considerando a função de  $P(i)$  apresentada anteriormente a fim de encontrar por método de regressão o coeficiente *declínio* para cada atributo.

### 5.3.1.5. A PROBABILIDADE DE A INFORMAÇÃO SER UTILIZADA

Considerando o instante de aplicação do modelo como o primeiro instante de tempo ( $t=0$ ), a probabilidade de o próximo uso da informação contida em um atributo neste instante bem como em todos os instantes subsequentes ( $t = 1, t = 2...$ ), passa a compor parte essencial na construção do mesmo. A probabilidade de a informação ser utilizada em um instante  $t$  qualquer deve ser uma constante identificada durante construção do modelo. Aqui chamaremos esta constante de  $C$ . Logo, se considerarmos o instante de aplicação do modelo  $t = 0$ , a probabilidade de o próximo uso da informação ocorrer neste instante será  $C$ . Aqui, chamaremos a probabilidade de o próximo uso da informação contida no atributo  $at$  ocorrer no instante de tempo  $t$  de  $Pu_{at}(t)$ . Esta probabilidade será dada pelo produto da probabilidade de a informação ainda não ter sido utilizada até este instante ( $1 - \sum_{i=0}^{t-1} Pu_{at}(i)$ ) por  $C$ . Logo, essa probabilidade pode ser expressa matematicamente, para todo  $t \in \mathbb{N}$  conforme (16).

$$Pu_{at}(t) = \begin{cases} C, & t = 0 \\ \left(1 - \sum_{i=0}^{t-1} Pu_{at}(i)\right) \times C, & t > 0 \end{cases} \quad (16)$$

Onde:

- $t$  é o instante de tempo em questão;
- $C$  é a probabilidade de a informação ser usada em qualquer instante  $t$ .

Logo, a probabilidade  $Pu_{at}(t)$  é uma função recursiva cujo limite será a 0% se  $t$  tender ao infinito positivo, o qual se encontra representado por (17).

$$\lim_{t \rightarrow \infty} (Pu_{at}(t)) = 0\% \quad (17)$$

O limite do somatório de  $Pu_{at}(t)$  para  $t$  tendendo ao infinito positivo será 100% e é representado por (18).

$$\lim_{t \rightarrow \infty} \left( \sum_{i=0}^t Pu_{at}(i) \right) = 100\% \quad (18)$$

O gráfico apresentado na Figura 4 ilustra um exemplo da distribuição da função para  $C = 25\%$ .

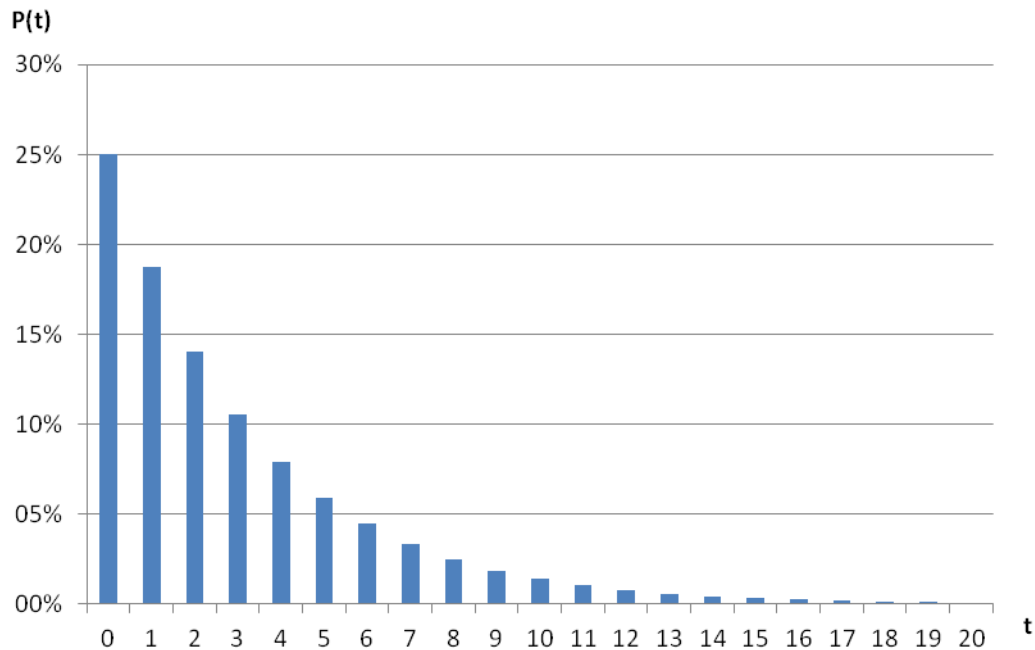


Figura 4 - Gráfico Exemplo da distribuição da função. Elaborado pelo autor.

Logo, para encontrarmos a probabilidade de o próximo uso da informação ocorrer em função do instante de tempo, somente se faz necessário saber a probabilidade de a informação ser utilizada em qualquer instante de tempo e assumir que essa probabilidade é constante. Permitiu-se identificar esta informação através da coleta de informações referentes ao volume registros (clientes) utilizados para cada atributo durante o período entre Março de 2011 e Fevereiro de 2012, fornecidas pela empresa.

### 5.3.2. CONSTRUÇÃO DO MODELO

Finalmente, organizaram-se todos os parâmetros necessários para a construção do modelo conforme Quadro 5 e o limiar de decisão final do modelo foi criado.

Parâmetro	Símbolo	Valor
$R_{\text{endereço}}$	$R_e$	
$R_{\text{telefone}}$	$R_t$	
$U_{\text{endereço}}$	$U_e$	
$U_{\text{telefone}}$	$U_t$	
$A_{\text{endereço}}$	$A_e$	
$A_{\text{telefone}}$	$A_t$	
$A_{\text{endereço,telefone}}$	$A_{e,t}$	
$Pr_{\text{endereço},x}(t)$	$Pr_{e,x}(t)$	
$Pr_{\text{telefone},x}(t)$	$Pr_{e,x}(t)$	
$Pu_{\text{endereço}}(t)$	$Pu_e(t)$	
$Pu_{\text{telefone}}(t)$	$Pu_t(t)$	

Quadro 5 - Demonstração dos Parâmetros Necessários para a Construção do Modelo. Elaborado pelo autor.

#### 5.4. SIMULANDO UMA APLICAÇÃO DO MODELO AMPLIADO.

Para aplicação do modelo escolheu-se aleatoriamente uma amostra de 200 registros da base de clientes usada na ação de venda de renovações de apólices de seguros em Março de 2012. Para tornar mais simples o cálculo dos custos totais das decisões para um grande número de registros, foi elaborada uma planilha eletrônica através do software Microsoft Excel® capaz de informar, de maneira instantânea, a melhor decisão para cada registro da amostra, bem como o somatório dos custos totais das melhores decisões. A referida planilha eletrônica pode ser consultada no apêndice A. Os Quadro 6 e Quadro 7 apresentam as entradas (*inputs*) e saídas (*outputs*) da planilha eletrônica desenvolvida.

Parâmetro	Valor
$R_{\text{endereço}}$	Input
$R_{\text{telefone}}$	Input
$U_{\text{endereço}}$	Input
$U_{\text{telefone}}$	Input
$A_{\text{endereço,telefone}}$	Input

$Pr_{endereço,x}(t)$	Input
$Pr_{telefone,x}(t)$	Input
$Pu_{endereço}(t)$	Input
$Pu_{telefone}(t)$	Input

Quadro 6 - Demonstração dos *Inputs* da planilha eletrônica desenvolvida. Elaborado pelo autor.

Registro	Data da Última Atualização	Custo da Decisão de Atualizar	Custo da Decisão de Não Atualizar	Melhor Decisão	Custo da melhor decisão
1	Input	Output	Output	Output	Output
...	Input	Output	Output	Output	Output
200	Input	Output	Output	Output	Output
Total	N/A	N/A	N/A	N/A	Output

Quadro 7 - Demonstração dos *Outputs* da planilha eletrônica desenvolvida. Elaborado pelo autor.

Para cada registro da base de dados, o modelo foi aplicado, considerando a idade da informação contida nos registros da amostra.

## 5.5. COMPARAÇÃO ENTRE OS MODELOS

Para viabilizar a comparação entre o modelo original e o modelo ampliado, uma simulação de aplicação do modelo original foi executada. A construção do modelo original é feita com os mesmos parâmetros usados para a construção do modelo ampliado a exceção do custo da ação de atualizar a informação, que ao invés de ser somente um para o conjunto de atributos, é um para cada atributo.

Para executar a simulação, uma planilha eletrônica semelhante, porém que atendesse ao modelo original foi desenvolvida. Esta planilha pode ser consultada no Apêndice B. A amostra de 200 registros usados na simulação de aplicação do modelo ampliado foi igualmente usada como base para a aplicação do modelo original. A simulação de aplicação do modelo envolveu duas etapas: a aplicação para o atributo Endereço (etapa 1) e a aplicação para o atributo Telefone (etapa 2).

Para a realização da etapa 1, percorreu-se a amostra de 200 registros, registro por registro aplicando modelo original, considerando a idade da informação contida no atributo



Endereço. O modelo indicou o custo de cada decisão para cada registro, possibilitando que apenas o menor seja contabilizado.

Após o uso de Endereço, uma determinada quantidade de registros já terá sido utilizada com sucesso, ou seja, um número  $n$  de clientes gerou receita na etapa 1, logo não foi utilizado na etapa 2. Assim, montou-se uma nova amostra com os registros cujo uso não reverteu em receita para a empresa. Na etapa 2, de posse da nova amostra de  $200 - n$  registros, o modelo original foi aplicado novamente. Percorrendo a amostra, registro por registro, a melhor decisão é identificada e seu custo contabilizado. Dessa forma o custo total das decisões pode ser calculado.

A soma do custo total das duas etapas leva ao custo total de aplicação do modelo original e foi apresentado conforme o Quadro 8.

<b>Aplicação para atributo</b>	<b>Custo total das decisões</b>
<b>Endereço (etapa 1)</b>	
<b>Telefone (etapa 2)</b>	
<b>Total</b>	

Quadro 8 - Demonstração do custo total de aplicação do modelo original. Elaborado pelo autor.

Finalmente, após o cálculo dos custos totais das decisões das duas simulações, exerce-se uma comparação entre os modelos.

## 6 O MODELO AMPLIADO

A fim de atender aos objetivos levantados, o presente trabalho oferece uma contribuição à gestão da QI através da proposição de uma ampliação do modelo proposto por Görz (2011), de maneira a torná-lo aplicável para um conjunto de atributos informacionais ao invés de um, isoladamente. A justificativa da ampliação proposta se dá através da suposição de que determinados atributos informacionais possuem uma relação entre si no mundo real de forma que o custo de atualizá-los numa mesma ação seja muito inferior à soma dos custos de ações independentes para atualizá-los. Num banco de dados de clientes de uma empresa de serviços, por exemplo, o custo de executar uma ação para atualizar um conjunto de dados cadastrais cuja fonte de atualização é o próprio cliente é menor do que executar uma ação para atualizar cada atributo separadamente. Esta suposição pode ser expressa matematicamente conforme (19), para um conjunto de  $n$  atributos, sendo  $A$  o custo da ação de atualizar a informação.

$$A_{0,\dots,n} \leq \sum_{i=0}^n A_i \quad (19)$$

É importante ressaltar que o presente estudo não atém esforços na validação desta premissa, entretanto irá assumi-la como verdadeira a fim de justificar a aplicação do modelo ampliado.

### 6.1. CONSIDERAÇÕES GERAIS DO MODELO AMPLIADO

Por contemplar um conjunto de  $n$  atributos, o modelo ampliado apoia a melhor decisão entre *atualizar* e *não atualizar* a informação contida nos  $n$  atributos. A circunstância sob a qual se sugere a aplicação do modelo é a necessidade iminente de usar a informação contida em um dos  $n$  atributos do conjunto. Com o modelo previamente construído, cada registro a ser utilizado tem a idade da informação de cada atributo como entrada para o modelo ampliado. A saída do modelo é, entre as decisões de *atualizar* ou *não atualizar*, a que tenha o menor custo total. Dessa forma, aplicando o modelo para uma base de dados o

modelo irá recomendar quais registros devem ser atualizados. A Figura 5 apresenta o funcionamento geral da aplicação do modelo.

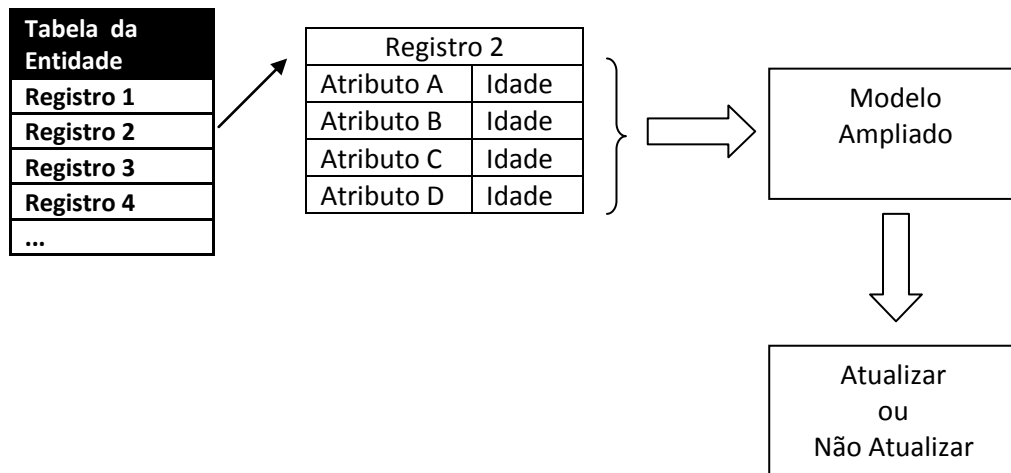


Figura 5 - Aplicação esquemática do modelo ampliado. Elaborada pelo autor.

## 6.2. UMA ANÁLISE DO MODELO ORIGINAL

Uma vez que as decisões envolvidas no modelo ampliado são as mesmas que as do modelo original de Görz (2011), a ampliação se inicia a partir de uma análise sobre a natureza dos custos das decisões. Temos os três custos envolvidos no modelo:

- o custo da ação de atualizar a informação ( $A$ );
- o custo de usar a informação ( $U$ );
- o custo originado pela não assimilação da receita esperada pelo uso da informação ( $R$ ).

Considerando que existe uma probabilidade para que esses custos ocorram são:

- a probabilidade de a ação de atualizar seja executada no instante  $t$  ( $Pa(t)$ );
- a probabilidade de a informação no atributo seja utilizada no instante  $t$  ( $Pu(t)$ );
- a probabilidade de que a informação esteja errada no instante  $t$  ( $Pr(t)$ ).

Define-se uma formulação genérica do custo total de uma decisão em (20).

$$CTDecisão_x = Pa(t) \times A + Pu(t) \times U + Pu(t) \times Pr(t + i_x) \times R \quad (20)$$

Onde:

- $x$  é o registro para o qual se está sendo aplicado o modelo;
- $i_x$  é a idade da informação para o registro  $x$ .

O modelo original considera as probabilidades relacionadas com cada custo para o instante  $t = 0$ , instante da aplicação do modelo. Para a decisão de *atualizar* todos os custos serão determinísticos, pois, mediante a circunstância de aplicação sabe-se que a informação será atualizada ( $Pa(0) = 100\%$ ), que a informação será usada ( $Pu(0) = 100\%$ ) e que a receita esperada será assimilada ( $Pr(0 + 0) = 100\%$ ) pois a informação estará certamente correta devido à atualização que foi executada ( $i_x := 0$ ). Para a decisão de *não atualizar*, os custos  $A$  e  $U$  são determinísticos enquanto o custo  $R$  é probabilístico, pois é certo que a informação não será atualizada ( $Pa(0) = 0\%$ ) e que será usada ( $Pu(0) = 100\%$ ), já a receita só será assimilada na condição de a informação estar correta ( $Pr(0 + i_x)$ ). O Quadro 9 apresenta a natureza de cada custo bem como a sua definição para cada decisão.

		Decisão	
Custo originado pelo(a)	Natureza	Atualizar	Não Atualizar
Ação de atualizar a informação	Determinístico para ambas as decisões	$Pa(0) \times A = 1 \times A$	$Pa(0) \times A = 0 \times A$
Uso a informação	Determinístico para ambas as decisões	$Pu(0) \times U = 1 \times U$	$Pu(0) \times U = 1 \times U$
Não assimilação da receita	Determinístico para <i>atualizar</i> , probabilístico para <i>não atualizar</i>	$Pr(0 + 0) \times R = 0 \times R$	$Pr(0 + i_x) \times R = Pr(i_x) \times R$

Quadro 9 - Natureza dos custos no modelo original. Elaborado pelo autor.

Assim, podemos chegar ao limiar de decisão proposto pelo modelo original através da formulação genérica apresentada em (21).

$$CTDecisão_x = Pa(t) \times A + Pu(t) \times U + Pu(t) \times Pr(t + i_x) \times R \quad (21)$$

Aplicando a formulação genérica para a decisão de *atualizar* encontra-se (22), que pode ser derivada algebricamente até (24).

$$CT_{Atualizar_x} = Pa(0) \times A + Pu(0) \times U + Pu(0) \times Pr(0 + i_x) \times R \quad (22)$$

$$CT_{Atualizar} = 1 \times A + 1 \times U + 1 \times 0 \times R \quad (23)$$

$$CT_{Atualizar} = A + U \quad (24)$$

Aplicando a formulação genérica para a decisão de *não atualizar* encontra-se (25), que pode ser derivada algebricamente até (27).

$$CT_{NãoAtualizar_x} = Pa(0) \times A + Pu(0) \times U + Pu(0) \times Pr(0 + i_x) \times R \quad (25)$$

$$CT_{NãoAtualizar_x} = 0 \times A + 1 \times U + 1 \times Pr(0 + i_x) \times R \quad (26)$$

$$CT_{NãoAtualizar_x} = U + Pr(i_x) \times R \quad (27)$$

Substituindo no limiar de decisão (28) os valores de (24) e (27), pode se chegar à inequação (31) mediando derivação algébrica. Esta equação, caso seja verdadeira, traz como saída do modelo a decisão de *atualizar* a informação.

$$CT_{Atualizar_x} < CT_{NãoAtualizar_x} \quad (28)$$

$$A + U < U + Pr(i_x) \times R \quad (29)$$

$$A < Pr(i_x) \times R \quad (30)$$

$$Pr(i_x) > A/R \quad (31)$$

Assim, se valida a formulação genérica apresentada, que será base para a ampliação proposta, apresentada a partir da subseção seguinte deste estudo.

### 6.3. A AMPLIAÇÃO DO MODELO

O modelo ampliado considera uma série de atributos, portanto a natureza probabilística/determinística dos custos de cada decisão deve ser revisada. No modelo ampliado, para a decisão de *atualizar*, o custo  $A$  é determinístico, pois se sabe que a informação será atualizada no  $t = 0$  e somente em  $t = 0$ , logo  $Pa(0) = 100\%$ . O custo de usar a informação  $U$  se torna probabilístico, pois não se pode afirmar que será usado no instante de aplicação do modelo ( $t = 0$ ). Logo,  $U$ , dependerá da probabilidade  $Pu(t)$ . O custo  $R$  depende da probabilidade de dois eventos acontecerem simultaneamente: a informação contida no atributo estar desatualizada no instante  $t$ , representada por  $Pr(t)$ <sup>3</sup>, e o próximo uso da informação ocorrer em  $t$ , representada por  $Pu(t)$ . Portanto,  $R$  depende da probabilidade combinada de dois eventos dada por  $Pu(t) \times Pr(t)$ , o que faz com que o custo  $R$  seja probabilístico.

Para a decisão de *não atualizar* o custo  $A$  é determinístico, pois é certo que não irá ocorrer, logo  $Pa(t) = 0\%$  para qualquer  $t$ . O custo  $U$ , por outro lado, é probabilístico, pois depende da probabilidade de o próximo uso da informação ser dar em  $t$ , representada por  $Pu(t)$ . O custo  $R$  depende da probabilidade de a informação contida no atributo estar desatualizada em  $t$ , representada por  $Pr(t + i_x)$ <sup>4</sup> e de o próximo uso da informação ocorrer no instante  $t$ , representada por  $Pu(t)$ . Logo o  $R$  também é um custo probabilístico e depende de a probabilidade combinada de dois eventos dada por  $Pu(t) \times Pr(t + i_x)$ . O Quadro 10 apresenta um resumo da natureza dos custos em cada decisão.

Custo originado pelo(a)	Natureza	Decisão	
		Atualizar	Não Atualizar
Ação de atualizar	Determinístico	$1 \times A$	$0 \times A$

<sup>3</sup> Nesta situação  $Pr(t + i_x) = Pr(t)$  pois  $i_x = 0$ , uma vez que informação será atualizada em  $t = 0$  para esta decisão

<sup>4</sup> Neste caso  $i_x$  é a idade da informação para o registro  $x$  quando  $t = 0$ .

a informação	para ambas as decisões		
Uso a informação	Probabilístico para ambas as decisões	$Pu(t) \times U$	$Pu(t) \times U$
Não assimilação da receita	Probabilístico para ambas as decisões	$Pu(t) \times Pr(t) \times R$	$Pu(t) \times Pr(t + i_x) \times R$

Quadro 10 - Natureza dos custos no modelo ampliado. Elaborado pelo autor.

Dessa forma, contabilizando os custos que compõe o custo total da decisão de atualizar a informação encontra-se (32).

$$CTDecisão_x = Pa(t) \times A + Pu(t) \times U + Pu(t) \times Pr(t + i_x) \times R \quad (32)$$

Para considerar que a decisão é referente à atualização ou não da informação contida no atributo  $at$  para o registro  $x$ . Para representar esta alteração, a formulação passa a dispor de do índice de atributo conforme (33).

$$CTDecisão_{at,x} = Pa_{at}(t) \times A_{at} + Pu_{at}(t) \times U_{at} + Pu_{at}(t) \times Pr_{at}(t + i_x) \times R_{at} \quad (33)$$

Baseado na nova notação da formulação genérica do custo total para a decisão de atualizar<sup>5</sup> será expresso por (34), que pode ser simplificada conforme (35).

$$CTAualizar_{at}(t) = 1 \times A_{at} + Pu_{at}(t) \times U_{at} + Pu_{at}(t) \times Pr_{at}(t + 0) \times R_{at} \quad (34)$$

$$CTAtualizar_{at}(t) = A_{at} + Pu_{at}(t) \times U_{at} + Pu_{at}(t) \times Pr_{at}(t) \times R_{at} \quad (35)$$

Para a decisão de *não atualizar* será expresso por (36), que pode ser simplificada conforme (37).

---

<sup>5</sup> Repare que o custo da decisão perde o índice de registro ( $x$ ), já que este deixa de ser relevante pois a informação será atualizada e a idade da informação será 0.

$$CTN\tilde{a}oAtualizar_{at,x} = 0 \times A_{at} + Pu_{at}(t) \times U_{at} + Pu_{at}(t) \times Pr_{at}(t + i_x) \times R_{at} \quad (36)$$

$$CTN\tilde{a}oAtualizar_{at,x} = Pu_{at}(t) \times U_{at} + Pu_{at}(t) \times Pr_{at}(t + i_x) \times R_{at} \quad (37)$$

### 6.3.1. CONSIDERANDO MÚLTIPLOS INSTANTES DE TEMPO.

Conforme explicado na seção 5, o modelo ampliado considera para a tomada da melhor decisão não só o custo total no instante de aplicação do modelo, mas todos os instantes subsequentes até o infinito positivo. Conforme explicado, lança-se mão da teoria de limites para viabilizar este cálculo.

Dessa forma, o somatório dos custos totais da decisão de *atualizar* a informação do atributo *at* para cada instante de tempo entre o instante da aplicação do modelo ( $t = 0$ ) até o infinito positivo é representado por (38).

$$\sum_{t=0}^{\infty} CTAtualizar_{at}(t) = A_{at} + \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t Pu_{at}(k) \right) \times U_{at} + \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t (Pu_{at}(k) \times Pr(k)) \right) \times R_{at} \quad (38)$$

O somatório dos custos totais da decisão de *não atualizar* a informação do atributo *at*, para o registro  $x^6$ , para cada instante de tempo desde o instante da aplicação do modelo ( $t = 0$ ) até o infinito positivo é dado por (39).

$$\sum_{t=0}^{\infty} CTN\tilde{a}oAtualizar_{at,x}(t) = \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t Pu_{at}(k) \right) \times U_{at} + \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t (Pu_{at}(k) \times Pr_{at}(k)) \right) \times R_{at} \quad (39)$$

---

<sup>6</sup> O registro para o qual está sendo calculado o custo desta decisão é relevante pois a idade da informação não é necessariamente 0, logo compõe o argumento de  $Pr(t + i_x)$ .



### 6.3.2. CONSIDERANDO MÚLTIPLOS ATRIBUTOS.

O custo total definitivo do modelo ampliado é dado pelo somatório do custo total de cada atributo em um conjunto de  $n$  atributos. Assim, para a decisão de *atualizar* a informação dos  $n$  atributos, considera-se o somatório para  $at = 0$  até  $n$ , dos somatórios dos custos totais da decisão de *não atualizar* a informação de  $t = 0$  até o infinito positivo. Tal custo é expresso por (40)

$$\sum_{at=0}^n \sum_{t=0}^{\infty} CT_{Atualizar_{at}}(t) = \sum_{at=0}^n A_{at} + \sum_{at=0}^n \left( \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t Pu_{at}(k) \right) \times U_{at} \right) + \sum_{at=a}^n \left( \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t (Pu_{at}(k) \times Pr(k)) \right) \times R_{at} \right) \quad (40)$$

Uma vez que assumido como premissa que determinados atributos possuem uma relação entre si no mundo real de forma que o custo de atualizá-los numa mesma ação seja menor ou igual à soma dos custos de ações independentes para atualizá-lo, o que é expresso matematicamente em (41).

$$A_{0,\dots,n} \leq \sum_{i=0}^n A_i \quad (41)$$

No modelo ampliado se usará o custo possivelmente menor. Logo, substitui-se na formula chegando-se (42).

$$\sum_{at=0}^n \sum_{t=0}^{\infty} CT_{Atualizar_{at}}(t) = A_{0,\dots,n} + \sum_{at=0}^n \left( \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t Pu_{at}(k) \right) \times U_{at} \right) + \sum_{at=a}^n \left( \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t (Pu_{at}(k) \times Pr(k)) \right) \times R_{at} \right) \quad (42)$$

Para a decisão de *não atualizar* a informação dos  $n$  atributos considera-se o somatório para  $at = 0$  até  $n$  dos somatórios dos custos totais da decisão de *não atualizar* a informação de  $t = 0$  até o infinito positivo, para o registro  $x$ . Tal custo é expresso por (43).

$$\sum_{at=0}^n \sum_{t=0}^{\infty} CT_{NãoAtualizar_{at,x}}(t) = \sum_{at=0}^n \left( \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t Pu_{at}(k) \right) \times U_{at} \right) + \sum_{at=0}^n \left( \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t (Pu_{at}(k + i_x) \times Pr_{at}(k + i_x)) \right) \times R_{at} \right) \quad (43)$$

### 6.3.3. O LIMIAR DE DECISÃO DO MODELO AMPLIADO

Com a definição dos custos totais de ambas as decisões pode-se compor o limiar de decisão modelo ampliado que, igualmente ao modelo original, é dado pela inequação (45) a qual, caso seja verdadeira, recomenda-se a decisão de *atualizar* a informação, caso contrário, recomenda-se *não atualizar*.

$$CT_{Atualizar_{at}} < CT_{N\tilde{a}oAtualizar_{at,x}} \quad (44)$$

$$\sum_{at=0}^n \sum_{t=0}^{\infty} CT_{Atualizar_{at}}(t) < \sum_{at=0}^n \sum_{t=0}^{\infty} CT_{N\tilde{a}oAtualizar_{at,x}}(t) \quad (45)$$

Uma vez que o custo de usar a informação, assim como no modelo original, apresenta-se em ambos os lados da inequação, esta pode ser simplificada conforme (46).

$$A_{a,\dots,n} + \sum_{at=0}^n \left( \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t (Pu_{at}(k) \times Pr(k)) \right) \times R_{at} \right) < \sum_{at=0}^n \left( \lim_{t \rightarrow \infty} \left( \sum_{k=0}^t (Pu_{at}(k+i_x) \times Pr_{at}(k+i_x)) \right) \times R_{at} \right) \quad (46)$$

## 7 UMA SIMULAÇÃO DE APLICAÇÃO DO MODELO AMPLIADO

No intuito de tornar mais inteligível o modelo ampliado proposto esta seção se dedica à exposição de uma simulação de aplicação do modelo, através de dados verossímeis de um caso de uso da informação. Cabe reforçar que o modelo proposto é genérico e pode ser aplicado a qualquer caso desde que suas premissas sejam respeitadas.

A seção inicia com uma breve descrição do processo com a qual é exemplificada a aplicação do modelo. Em seguida são apresentados os métodos utilizados para que fosse possível encontrar os parâmetros do modelo. Dando continuidade, é apresentada a métrica utilizada no modelo ampliado, bem como o método utilizado para fosse possível chegar aos parâmetros para cada um dos atributos. Em seguida a simulação de aplicação é apresentada. Finalmente, uma análise e uma comparação entre o modelo ampliado proposto e o modelo original de Görz (2011) são realizadas.

### 7.1. DESCRIÇÃO DO PROCESSO

A simulação de aplicação do modelo ampliado proposto será realizada sob um processo de venda de apólices de seguro de renovação praticada pelo Banco B, em parceria com a corretora de seguros C e companhia seguradora S. Os nomes reais das empresas envolvidas no processo foram preservados por solicitação das mesmas.

O Banco B tem como principal produto operações de financiamento de máquinas e equipamentos. Os valores financiados tem grande variabilidade, podendo ir de R\$ 5.000 até R\$ 4.000.000. Como é de costume neste setor, para mitigação do risco assumido pela instituição financeira, durante o período de financiamento o bem financiado é oferecido como a principal garantia do financiamento. No caso de inadimplência do cliente o bem financiado pode ser retomado pelo banco. Na política de crédito do Banco B, consta a exigência de o cliente proteger o bem financiado com uma apólice de seguro. Devido ao grande número de exigências que o seguro contratado deve apresentar e a fim de acelerar a liberação do financiamento, o Banco oferece uma opção de seguro com características pré-estabelecidas com a Seguradora S, parceira da operação. O produto oferecido tem grande

aceitação por parte dos clientes na primeira apólice de seguro contratada, chegando a uma média de 97% de adesão. Aproximadamente 60 dias antes do vencimento da apólice vigente o Banco B, a Seguradora S e Corretora de seguros C, também parceira na operação, iniciam o processo para oferecer renovação da apólice, já que o bem deve ficar segurado durante todo o período de financiamento. A exemplificação da aplicação do modelo ampliado proposto se dará justamente na ação de renovação de apólices de seguros vendidas através da parceria entre as companhias mencionadas.

O processo de envio de propostas de renovação aos clientes tem início no próprio banco. Próximo ao dia 15 de cada mês o Banco extrai de seu sistema de informações uma relação de todas as apólices com vigência a expirar no dia 15 do segundo mês subsequente. Nessa relação constam, além das informações sobre a apólice que está próxima de expirar, dados cadastrais do cliente tais como nome e dados de contato. A relação é então enviada a corretora, que a recebe e com ela alimenta o seu sistema. A corretora realiza alguns ajustes nas informações relacionadas às apólices através de dados de seu sistema e envia a relação a seguradora. A seguradora então emite as propostas de renovação, e as envia por correio físico para o endereço de cada cliente. Caso o endereço do cliente esteja correto, este receberá a proposta e, de posse dela, poderá optar por pagar e contratar a renovação do seguro ou não. A corretora aguarda o pagamento do cliente até a data de vencimento do prazo do boleto enviado com a proposta. Caso o cliente não efetue o pagamento, a corretora exercerá uma ação ativa de cobrança por telefone. Caso o telefone do cliente esteja correto, esse receberá uma ligação da equipe de negociação da corretora e lhe serão apresentados novos argumentos para que a contratação do seguro seja feita.

Nota-se na operação a relevância da qualidade da informação em dois atributos informacionais da entidade *pessoa* (cliente) – o Endereço e o Telefone. O sucesso do processo de renovação de apólices de seguro depende de que a informação contida nestes atributos, registrados no banco de dados de cadastro de clientes, esteja atualizada. O Banco B registra as informações cadastrais de um cliente durante o processo de formalização da sua primeira operação de financiamento junto ao banco, momento no qual um formulário de dados cadastrais é preenchido e documentos que comprovem veracidade das informações são apresentados, o que torna a informação livre de erros. Sabe-se que erros humanos praticados ao registrar a informação no banco de dados podem ser cometidos, o

que implicaria num erro corriqueiro e gerador de grande perda na QI (LIWELYN, 2008). Esta simulação, no entanto, vai ignorar essa possibilidade e assumirá, a fim de simplificação, que a informação é totalmente livre de erros quando registrada pela primeira vez no banco de dados. Atualmente, a atualização dos atributos Endereço e Telefone somente ocorre por iniciativa dos próprios clientes, que comunicam ao acaso a troca de endereço de correspondência e/ou de telefone. Apesar de boa parte das correspondências enviadas pelo banco, não somente na ação renovação de seguros, serem devolvidas pelo serviço de correios por conterem endereço de correspondência errado, nenhuma ação recorrente de atualização das informações cadastrais é realizada.

O modelo ampliado será aplicado considerando o conjunto de atributos composto pelo Endereço e pelo Telefone. A circunstância de aplicação do modelo é a necessidade de uso do atributo Endereço para o envio pelo correio das propostas de renovação de seguro. No caso dos clientes que não pagarem a proposta enviada, a ação complementar de cobrança pelo Telefone será deverá ser tomada, acarretando assim no uso da informação do atributo Telefone.

O fluxograma apresentado a Figura 6, resume e esquematiza o processo de venda de apólices de seguros.

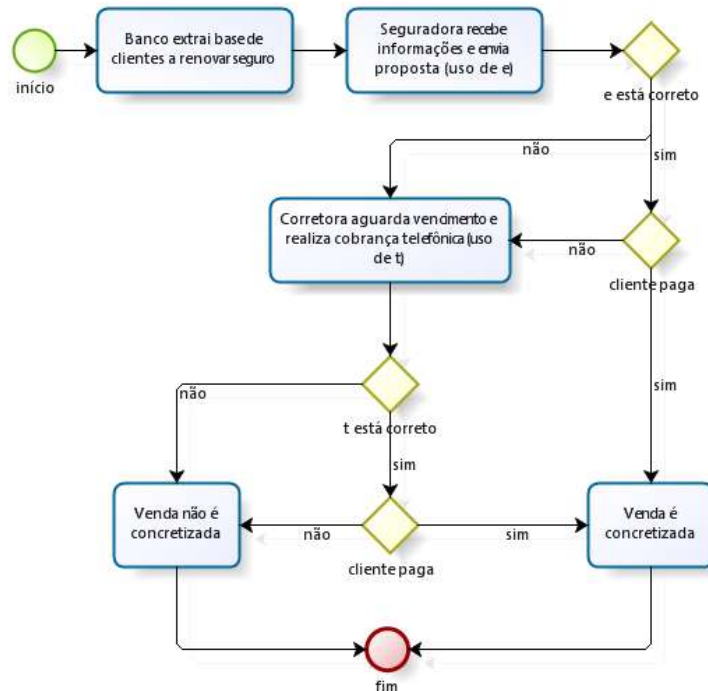


Figura 6 - Fluxograma do processo de venda de apólices de renovação. Elaborado pelo autor.

## 7.2. AS RECEITAS ESPERADAS

Conforme apresentado na seção cinco, a fim de calcular as receitas esperadas do uso de cada atributo coletou-se informações acerca da operação entre os meses de Março de 2011 e Fevereiro de 2012. Para cada atributo, a quantidade de registros usados no período foi subtraída da quantidade de registros usados com a informação errada (ou desatualizada). Em seguida a receita total originada pelo uso de cada atributo foi dividida pela quantidade de atributos usados com informação correta.

### 7.2.1. CÁLCULO DA RECEITA ESPERADA PELO USO DO ATRIBUTO ENDEREÇO.

O cálculo da receita esperada pelo uso do atributo Endereço envolveu a coleta da quantidade de apólices renovação enviadas durante o período (A), subtraída da quantidade de apólices que foram devolvidas pelo serviço de correios por conterem endereço errado (B). A receita total originada no envio das apólices (C) foi dividida pela quantidade de apólices de renovação corretamente recebidas pelos clientes (A-B). O Quadro 11 apresenta os valores mencionados e auxilia o entendimento do cálculo.

		Mar-11	Abr-11	Mai-11	Jun-11	Jul-11	Ago-11	Set-11	Out-11	Nov-11	Dez-11	Jan-12	Fev-12	Total
A	Quantidade de renovações enviadas	3.009	2.998	3.194	3.300	1.812	4.621	4.641	4.809	4.139	3.023	1.708	2.370	39.624
B	Quantidades de Renovações retornadas por problema de endereço	66	54	148	190	137	230	266	196	179	145	61	37	1.709
A-B	Quantidade de renovações recebida	2.943	2.944	3.046	3.110	1.675	4.391	4.375	4.613	3.960	2.878	1.647	2.333	37.915
C	Receita Total envio de apólices (R\$)	826.958	864.366	827.398	958.019	441.221	1.571.200	1.528.076	1.395.592	1.264.141	978.536	456.348	663.725	11.775.580
C/(A-B)	Receita por renovação recebida (R\$)	280,99	293,60	271,63	308,04	263,42	357,82	349,27	302,53	319,23	340,01	277,08	284,49	310,58

Quadro 11 - Demonstração do cálculo da receita esperada pelo uso do atributo Endereço. Elaborado pelo autor.

Portanto temos, para composição do modelo,  $R_{\text{endereço}} = R\$ 310,58$ .

### 7.2.2. CÁLCULO DA RECEITA ESPERADA PELO USO DO ATRIBUTO TELEFONE.

O cálculo da receita esperada pelo uso do atributo Telefone envolveu a coleta da quantidade de clientes chamados (A), subtraída da quantidade de contatos mal sucedidos por conterem o número de telefone errado (B). A receita total originada na cobrança telefônica realizada (C) foi dividida pela quantidade de clientes contatados (A-B). O Quadro 12 apresenta os valores mencionados e auxilia o entendimento do cálculo.

		Mar-11	Abr-11	Mai-11	Jun-11	Jul-11	Ago-11	Set-11	Out-11	Nov-11	Dez-11	Jan-12	Fev-12	Total
A	Quantidade de Clientes chamados	1.264	1.216	1.632	1.818	1.673	2.476	2.425	2.700	2.500	2.415	2.805	2.067	24.991
B	Quantidade de contatos mal sucedidos (telefone errado)	207	231	315	306	318	410	221	518	42	302	593	175	3.638
A-B	Quantidade de clientes contatados	1.057	985	1.317	1.512	1.355	2.066	2.204	2.182	2.458	2.113	2.212	1.892	21.353
C	Receita cobrança telefônica (R\$)	460.812	522.522	477.388	616.848	447.829	831.543	825.747	951.961	787.003	549.143	319.535	563.389	7.353.720
C/(A-B)	Receita por cliente contatado (R\$)	435,96	530,48	362,48	407,97	330,50	402,49	374,66	436,28	320,18	259,89	144,46	297,77	344,39

**Quadro 12 - Demonstração do cálculo da receita esperada pelo uso do atributo Telefone. Elaborado pelo autor.**

Portanto temos, para composição do modelo,  $R_{telefone} = R\$ 344,39$ .

### 7.3. OS CUSTOS DE USAR A INFORMAÇÃO

Conforme explicado na seção cinco, o cálculo do custo de usar a informação de cada atributo baseou-se nas entrevistas realizadas com os especialistas nos processos de uso de ambos os atributos. Dessa forma, identificou-se a melhor maneira de estimar estes custos, necessários para construção do modelo.

#### 7.3.1. O CUSTO DE USAR O ATRIBUTO ENDEREÇO

A atividade de envio de correspondências é totalmente terceirizada pela empresa, o que simplifica muito a identificação do custo de usar o atributo Endereço. A empresa prestadora de serviço de impressão e envio de correspondências cobra o valor fixo de R\$

4,13 por cada apólice enviada, sendo este o custo direto unitário de cada registro usado. O custo indireto de uso do atributo Endereço poderia ser calculado através do custo de manipulação e processamento da informação até que ela seja remetida a empresa prestadora de serviço. Esta atividade, no entanto, tem custo desprezível, pois a extração de um arquivo de dados do sistema do banco B e o seu envio por email para a empresa prestadora de serviços não toma mais de 30 minutos por mês do analista que a realiza e, uma vez realizada, atende a todas as apólices enviadas no mês (cerca de 3.200).

Portanto temos, para composição do modelo,  $U_{\text{endereço}} = R\$ 4,13$ .

### 7.3.2. O CUSTO DE USAR O ATRIBUTO TELEFONE

O cálculo do custo de usar o atributo Telefone se deu através da estimação de custo unitário direto da atividade de usar o atributo (despesas com telefone) somado com a estimação do custo unitário indireto da atividade (despesas com pessoal).

Para estimar o custo unitário indireto (G) contabilizaram-se para o período mencionado os custos de pessoal da célula de *call center* (A), responsável pela atividade de negociação de renovações. Após isso, a profissional entrevistada estimou um percentual do tempo de trabalho dedicado a renovações (B), que representa importante parcela do tempo total. O produto de (A) por (B) é uma estimação do custo de pessoal da atividade de negociação de renovações. Este produto, dividido pelo número de clientes com os quais a célula tentou estabelecer contato telefônico de negociação de renovação (C), o que levou ao valor estimado do custo indireto de uso do atributo Telefone (G). O Quadro 13 apresenta os valores mencionados e auxilia na compreensão do cálculo.

A	Custo indireto da célula de <i>call center</i> (salários + benefícios + encargos) (R\$)	118.300,00
B	Percentual do tempo dedicado à atividade de negociação de renovações	80%
AxB	Custo indireto da atividade de negociação de renovações (salários + benefícios + encargos) (R\$)	94.640,00
C	Número de clientes processados*	24.991
AxB / C = G	Custo indireto unitário (R\$)	3,79

Quadro 13 - Demonstração do cálculo do custo indireto unitário de usar o atributo Telefone. Elaborado pelo autor.



O custo unitário direto (H) foi estimado através da contabilização das despesas com telefone para o período mencionado da célula de *call center* (D). Em seguida, a entrevistada estimou um percentual de ligações para negociação de renovações sobre o total de ligações da célula (E). O produto de (D) por (E) é uma estimativa do custo de telefone da atividade de negociação de renovações. Dividindo este produto, pelo número de clientes com os quais a célula tentou estabelecer contato telefônico de negociação de renovação (C), chegou-se ao valor estimado do custo indireto de uso do atributo telefone (H). O Quadro 14 apresenta os valores mencionados e auxilia na compreensão do cálculo.

D	Custo direto da célula de call center (despesas com telefone) (R\$)	R\$ 35.520,00
E	Percentual de ligações de negociação de renovações sobre o total de ligações	100%
DxE	Custo direto da atividade de negociação de renovações (despesas com telefone) (R\$)	R\$ 35.520,00
C	Número de clientes processados*	24.991
DxE/F = H	Custo direto unitário (R\$)	R\$ 1,42

Quadro 14 - Demonstração do cálculo do custo direto unitário de usar o atributo Telefone. Elaborado pelo autor.

Finalmente, soma dos custos unitários direto (G) e indireto (H) da atividade de negociação de renovações por telefone leva ao custo unitário total, conforme Quadro 15.

G	Custo indireto unitário (R\$)	R\$ 5,04
H	Custo direto unitário (R\$)	R\$ 1,42
G+H	Custo unitário total (R\$)	R\$ 6,46

Quadro 15 - Demonstração do cálculo do custo unitário de usar o atributo Telefone. Elaborado pelo autor.

Portanto temos, para composição do modelo,  $U_{telefone} = R\$ 6,246$ .

#### 7.4. O CUSTO DA AÇÃO DE ATUALIZAR

Conforme apresentado na seção 5, a fim de identificar os custos da ação de atualizar a informação, entrevistou-se a profissional responsável pela célula de cadastro da empresa, área na qual, atividades de cadastramento e atualização de informações cadastrais são realizadas.

Em virtude da inexistência de prévia mensuração do custo das atividades da célula, os valores foram estimados baseando-se na experiência da entrevistada. Estimou-se o custo médio mensal de um funcionário da célula em R\$ 3.600,00, já com todos os benefícios e encargos envolvidos. Esse valor possibilitou a encontrar, descontando-se um tempo de 10 minutos por hora em que o funcionário não está produzido, um custo por minuto do funcionário de R\$ 0,41. A coleta de uma amostra de informações referentes à cerca de 600 chamadas telefônicas realizadas em fevereiro de 2012, considerando o tempo das ligações e seus custos, permitiu estimar um custo de telefone de R\$ 1,23 por minuto. O Quadro 16 apresenta o cálculo realizado.

A	Custo mensal do funcionário	R\$ 3.600,00
A/22	Custo diário do funcionário	R\$ 163,64
A/(22 x 400)	Custo do minuto do funcionário	R\$ 0,41
B	Custo médio de telefone por minuto	R\$ 1,23

**Quadro 16 – Demonstração do cálculo dos custos estimados de trabalho e telefone por minuto (elaborado pelo autor).**

#### 7.4.1. O CUSTO DE ATUALIZAR ENDEREÇO E TELEFONE

Atualmente na empresa estudada, a principal fonte de informação explorada para atualizar dados cadastrais de clientes são os próprios clientes. Normalmente, a célula de cadastro faz um contato telefônico da empresa e se atualizam os dados que necessitam ser atualizados. Porém, quando o atributo Telefone está desatualizado se torna impossível contatá-lo pelo telefone. Portanto, quando o telefone precisa ser atualizado, é feito um contato direto com a concessionária que, por ter um relacionamento muito mais próximo com os clientes, é a usada como fonte de informações paliativa. A solicitação de uma informação junto ao concessionário, normalmente não é atendida instantaneamente, mas em algumas horas ou até dias depois. Isto faz com que o colaborador que está realizando a atualização tenha que, mais de uma vez, acessar o sistema de informações e procurar pelo registro do cliente, o que incrementa o tempo dedicado às demais atividades.

Dessa forma, mediante a necessidade de atualizar o conjunto de atributos Endereço e Telefone, a ação envolve um contato diretamente realizado com a concessionária, nos

quais o tempo dedica para a atividade de *telefonar* e *demais atividades*, bem como seus custos, são apresentadas no Quadro 17.

Atividade	Tempo (min.)	Custo de Telefone	Custo de Trabalho	Custo Total
Telefonar	9	R\$ 11,07	R\$ 3,68	R\$ 14,75
Demais	9		R\$ 3,68	R\$ 3,68
Total	18	R\$ 11,07	R\$ 7,36	R\$ 18,43

Quadro 17 - Demonstração do cálculo da estimação do custo da ação de atualizar a informação para o conjunto de atributos. Elaborado pelo autor.

Portanto temos, para composição do modelo,  $A_{\text{endereço, endereço}} = R\$ 18,43$ .

#### 7.4.2. O CUSTO DE ATUALIZAR OS ATRIBUTOS SEPARADAMENTE

A ação de atualizar a informação somente para o atributo Endereço é mais simples, pois o contato é realizado diretamente com o cliente. O cliente é contatado pelo telefone e a informação atualizada é registrada no sistema da empresa. As atividades envolvidas na ação, bem como seus custos estimados, são apresentadas no Quadro 18.

Atividade	Tempo (min.)	Custo de Telefone	Custo de Trabalho	Custo Total
Telefonar	8	R\$ 9,84	R\$ 3,27	R\$ 13,11
Demais	4		R\$ 1,64	R\$ 1,64
Total	12	R\$ 9,84	R\$ 4,91	R\$ 14,75

Quadro 18 - Demonstração do cálculo da estimação do custo da ação de atualizar a informação para o atributo Endereço. Elaborado pelo autor.

Portanto temos, para composição do modelo,  $A_{\text{endereço}} = R\$ 14,75$ .

Para atualizar a informação somente para o atributo Telefone, como foi mencionado, demanda o contato diretamente com o concessionário, da mesma forma que para atualizar a informação dos dois atributos. As atividades envolvidas na ação, bem como seus custos estimados, são apresentadas no Quadro 19.

Atividade	Tempo (min.)	Custo de Telefone	Custo de Trabalho	Custo Total
Telefonar	8	R\$ 9,84	R\$ 3,27	R\$ 13,11
Demais	8		R\$ 3,27	R\$ 2,27
Total	16	R\$ 9,84	R\$ 6,55	R\$ 16,39

Quadro 19 - Demonstração do cálculo da estimação do custo da ação de atualizar a informação para o atributo Telefone. Elaborado pelo autor.

Portanto temos, para composição do modelo,  $A_{telefone} = R\$ 16,39$ .

### 7.5. A PROBABILIDADE DE A INFORMAÇÃO ESTAR DESATUALIZADA

O cálculo da probabilidade de a informação estar desatualizada para ambos os atributos se deu, conforme explicado na seção cinco, da mesma maneira. Coletou-se uma amostra de registros utilizados no período compreendido entre março de 2011 e fevereiro de 2012, identificou-se a idade da informação de cada registro na data em que foi utilizado e identificaram-se quais estavam desatualizados. A partir daí, contabilizou-se os percentuais de informações desatualizadas para cada idade e usaram-se os dados como entrada para tratamento no *software* Curve Expert®.

Os dados inseridos no *software*, para o atributo Endereço, são apresentados no Quadro 20.

Idade (em anos)	Número de Registros Usados	Números de Registros Desatualizados	Percentual de registros Desatualizados
0	18829	0	0,00000%
1	26389	419	1,58778%
2	18333	297	1,62003%
3	15767	282	1,78855%
4	10635	225	2,11566%
5	3810	132	3,46457%
6	1575	48	3,04762%
7	2221	100	4,50248%
8	1352	87	6,43491%

Quadro 20 - Dados inseridos no *Curve Expert*® para cálculo do coeficiente de declínio para o atributo Endereço. Elaborado pelo autor.

Inserindo os dados de *percentual de registros errados* como variável dependente do dado *idade*, estimando uma relação entre eles conforme a métrica proposta por Heirich e Klier (2009), o *software* apresentou como saídas o Quadro 21.

<b>User-Defined Model: <math>y=1-\exp(-a*x)</math></b>
Coefficient Data:
a = 0,006294243
Standard Error: 0.0075106
Correlation Coefficient: 0.9001817
Comments:
The fit converged to a tolerance of 1e-006 in 14 iterations. No weighting used.

Quadro 21 - Saídas do Curve Expert® para o atributo Endereço. Elaborado pelo autor.

O coeficiente de correlação de aproximadamente 0,90 mostra uma razoável confiabilidade do coeficiente encontrado. Chegou-se ao coeficiente de declínio 0,6294243% que permitiu para composição do modelo definir (47).

$$Pr_{\text{endereço},x}(t) = 1 - \exp(0,00629 \times (t + i_x)) \quad (47)$$

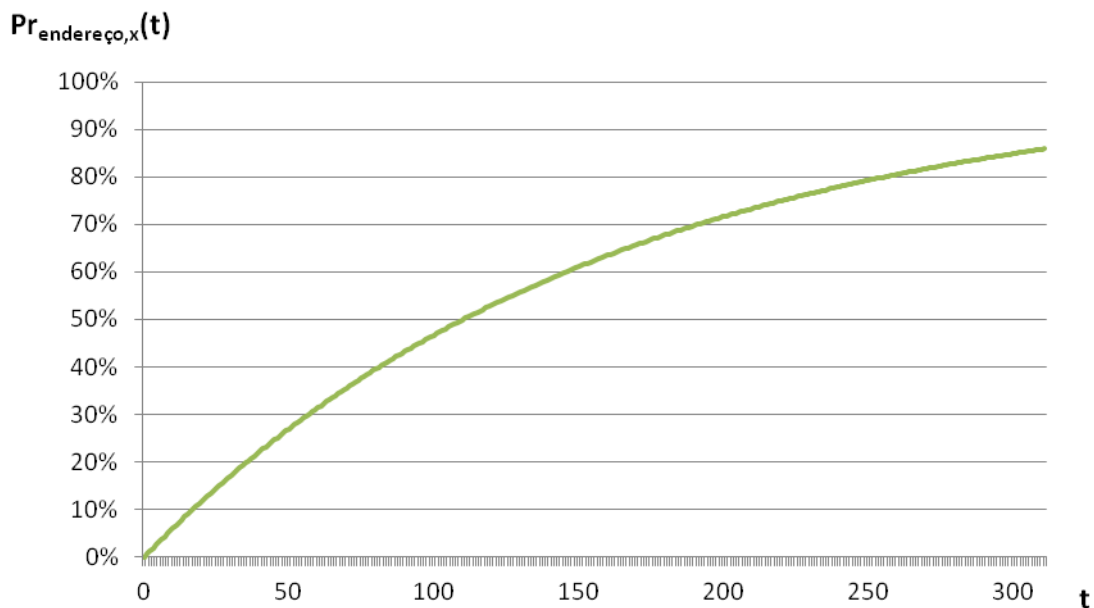


Figura 7 - Gráfico da função da probabilidade de a informação estar desatualizada para Endereço em função de t, com a idade da informação ( $i_x$ ) nula em  $t=0$ . Elaborado pelo autor.

Para o atributo Telefone, os dados inseridos no referido software são apresentados no Quadro 22.

Idade (em anos)	Quantidade de Registros Usados	Quantidade de Registros Desatualizados	Percentual de registros Desatualizados
0	1041	0	0,0000%
1	3850	342	8,8878%
2	6449	512	7,9322%
3	3663	483	13,1843%
4	2984	456	15,2710%
5	957	99	10,3319%
6	700	112	16,0047%
7	957	214	22,3486%
8	625	90	14,3932%

Quadro 22 - Dados inseridos no *Curve Expert*® para cálculo do coeficiente de declínio do atributo Telefone. Elaborado pelo autor.

Para este atributo realizou-se exatamente o mesmo tratamento que para o atributo Endereço. As saídas apresentadas pelo *software* para o atributo Telefone são apresentadas no Quadro 23.

<b>User-Defined Model: <math>y=1-\exp(-a*x)</math></b>
Coefficient Data:
a = 0,0295771228981
Standard Error: 0.0431620
Correlation Coefficient: 0.7247790
Comments:
The fit converged to a tolerance of 1e-006 in 14 iterations. No weighting used.

Quadro 23 - Saídas do *Curve Expert*® para o atributo Telefone. Elaborado pelo autor.

Percebe-se que, para o atributo Telefone a correlação entre as duas variáveis não foi de 0,72, ou seja, não tão alto como para Endereço e, conseqüentemente, menos confiável. Entretanto, para o intuito de simular a aplicação do modelo, o coeficiente de declínio encontrado 2,957712% foi utilizado. Assim, se incorporou ao modelo o seguinte parâmetro conforme (48).

$$Pr_{telefone,x}(t) = 1 - \exp(-0,029577 \times (t + i_x)) \quad (48)$$

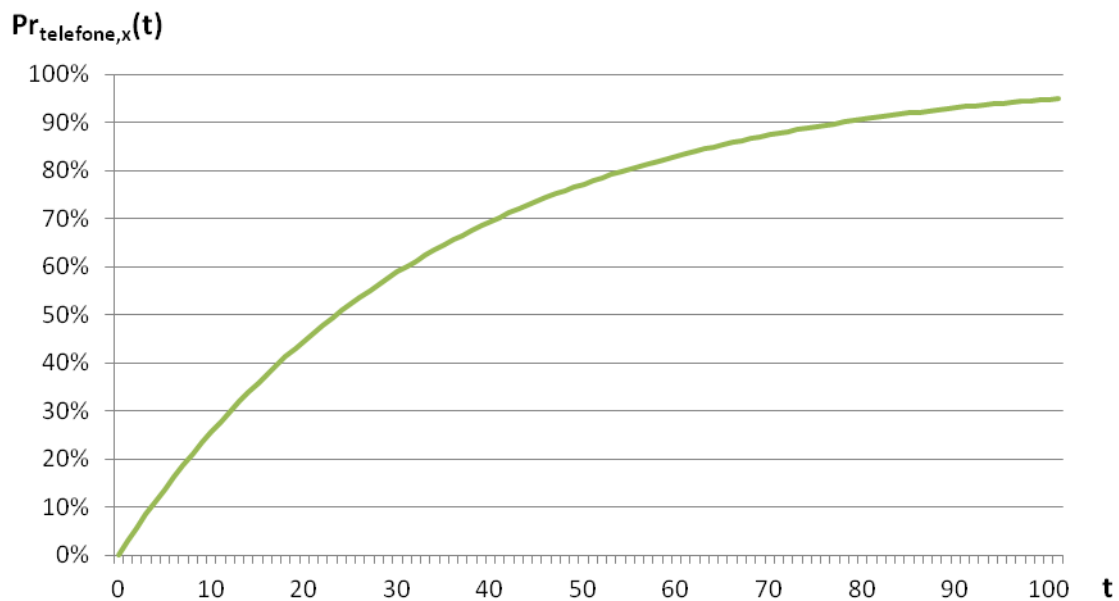


Figura 8 - Gráfico da função da probabilidade de a informação estar desatualizada para Telefone em função de  $t$ , com idade da informação ( $i_x$ ) nula em  $t=0$ . Elaborado pelo autor.

#### 7.6. A PROBABILIDADE DE USAR A INFORMAÇÃO

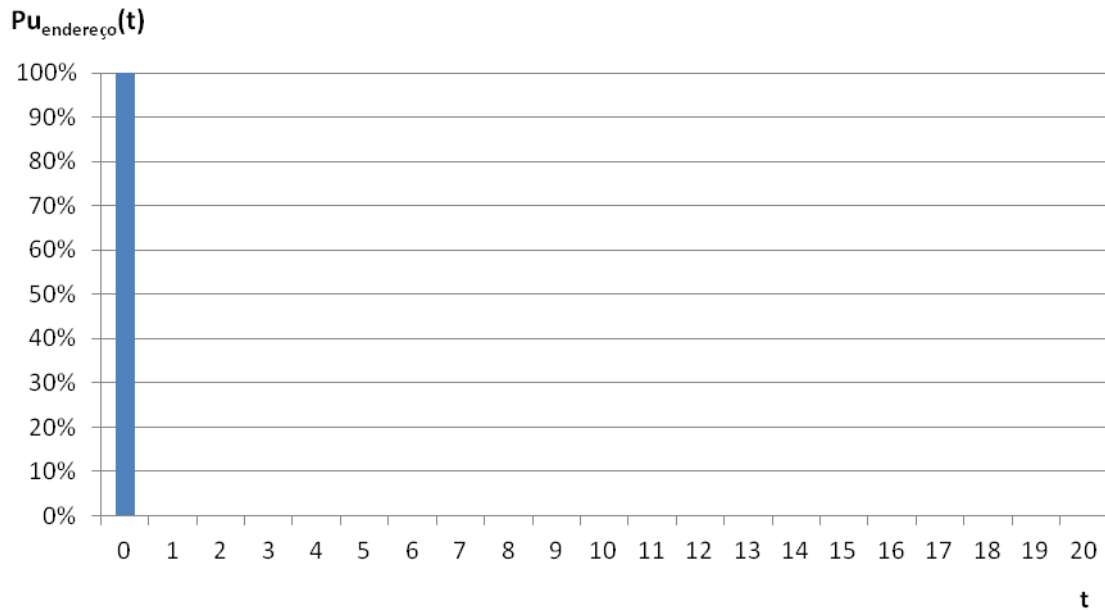
Conforme elucidado na seção cinco, a probabilidade de o próximo uso a informação ocorrer em um determinado instante  $t$  é igual à probabilidade de a informação não ter sido usada desde o instante 0 até o este instante  $t$ , multiplicado pela probabilidade de a informação ser usada em qualquer  $t$ .

A probabilidade de a informação do atributo Endereço ser usada no mês de aplicação do modelo ( $t = 0$ ) é 100%, pois se sabe que a proposta de renovação de seguro será enviada pelo correio para todos os clientes cujo seguro deve ser renovado no mês em questão. Logo, a probabilidade de o próximo uso da informação ocorrer em qualquer instante de tempo depois de 0 será de 0% pois este evento já terá ocorrido em  $t = 0$ . Portanto temos (49), para composição do modelo.

$$Pu_{endereço}(t) = \begin{cases} 100\%, & t = 0 \\ 0\%, & t > 0 \end{cases} \quad (49)$$

O gráfico apresentado na

Figura 9, mostra a distribuição da probabilidade de o próximo uso do atributo ocorrer em cada  $t$ .



**Figura 9 - Gráfico da função da probabilidade de o próximo uso da informação de Endereço ocorrer em cada de t.**  
Elaborado pelo autor.

Dos clientes para os quais é enviada pelo correio a proposta de renovação de seguro, 33,5% fazem o pagamento diretamente pela proposta enviada, sem que haja necessidade de cobrá-los pelo telefone. Logo, o atributo Telefone só é usado para 66,5% dos clientes. O Quadro 24 apresenta os números que levaram a este percentual.

		Mar-11	Abr-11	Mai-11	Jun-11	Jul-11	Ago-11	Set-11	Out-11	Nov-11	Dez-11	Jan-12	Fev-12	Total
A	Quantidade de renovações enviadas	3009	2998	3194	3300	1812	4621	4641	4809	4139	3023	1708	2370	39624
B	Quantidade de Renovações pagas	1085	1055	980	1051	423	1603	1604	1606	1431	1060	579	804	13281
A-B	Quantidade de clientes cobrados pelo telefone	1924	1943	2214	2249	1389	3018	3037	3203	2708	1963	1129	1566	26343
(A-B)/A	Percentual de clientes cobrados pelo telefone	63,9%	64,8%	69,3%	68,2%	76,7%	65,3%	65,4%	66,6%	65,4%	64,9%	66,1%	66,1%	66,5%

**Quadro 24- Demonstração da quantidade de propostas de renovação enviadas, pagas e cobradas por telefone. Elaborado pelo autor.**



Assim, para o atributo Telefone, a probabilidade de a informação ser usada no instante  $t = 0$  é 66,5%, e se torna possível formular a função que expressa a probabilidade de o próximo uso da informação de Telefone acontecer no instante  $t$ . Portanto temos (50), para composição do modelo.

$$Pu_{telefone}(t) = \begin{cases} 66,5\%, & t = 0 \\ \left(1 - \sum_{i=0}^{t-1} Pu_{telefone}(i)\right) \times 66,5\%, & t > 0 \end{cases} \quad (50)$$

A Figura 10 apresenta a distribuição da função apresentada.

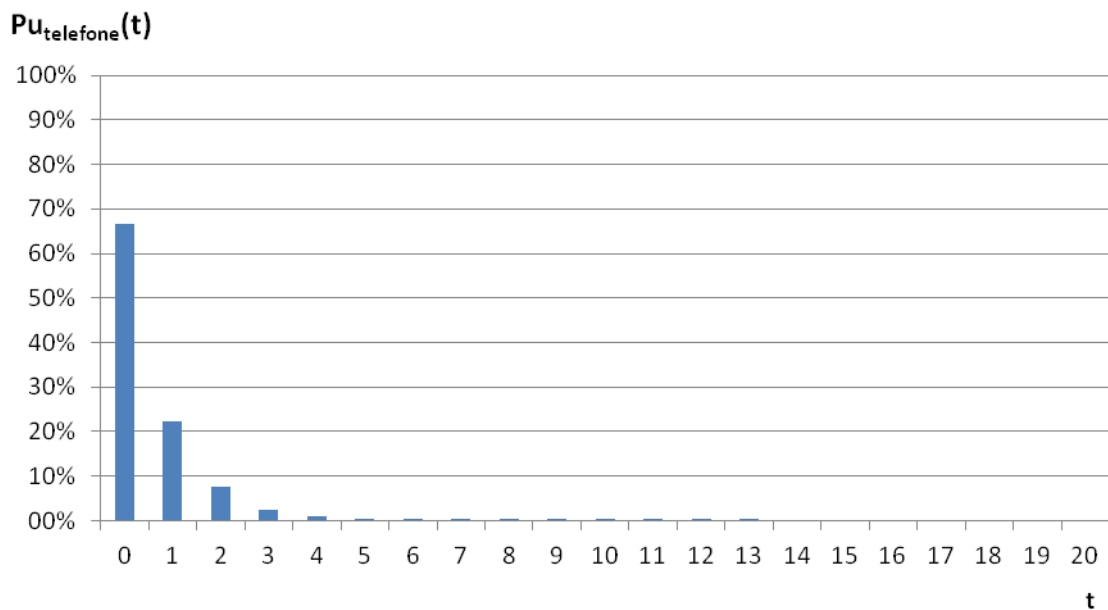


Figura 10 – Gráfico da função da probabilidade de o próximo uso da informação de Telefone ocorrer em cada t. Elaborado pelo autor.

## 7.7. A CONSTRUÇÃO DOS MODELOS

Lançando mão dos custos e receitas apresentadas podem-se estruturar os modelos, conforme Quadro 25:

Parâmetro	Símbolo	Valor
$R_{\text{endereço}}$	$R_e$	R\$ 310,58
$R_{\text{telefone}}$	$R_t$	R\$ 344,39
$U_{\text{endereço}}$	$U_e$	R\$ 4,13
$U_{\text{telefone}}$	$U_t$	R\$ 6,46
$A_{\text{endereço}}$	$A_e$	R\$ 14,75
$A_{\text{telefone}}$	$A_t$	R\$ 16,39
$A_{\text{endereço,telefone}}$	$A_{e,t}$	R\$ 18,43
$Pr_{\text{endereço},x}(t)$	$Pr_{e,x}(t)$	$1 - \exp(-0,00629 \times (t + i_x))$
$Pr_{\text{telefone},x}(t)$	$Pr_{e,x}(t)$	$1 - \exp(-0,02957 \times (t + i_x))$
$Pu_{\text{endereço}}(t)$	$Pu_e(t)$	100%, se $t = 0$ 0%, se $t > 0$
$Pu_{\text{telefone}}(t)$	$Pu_t(t)$	66,5%, se $t = 0$ $\left(1 - \sum_{i=0}^{t-1} Pu_{\text{telefone}}(i)\right) \times 66,5\%$ , se $t > 0$

Quadro 25 – Compilação dos parâmetros dos modelos para o caso. Elaborado pelo autor.

## 7.8. A SIMULAÇÃO DE APLICAÇÃO DOS MODELOS

A aplicação do modelo ampliado foi realizada sobre uma amostra de 200 registros para os atributos estudados, considerando a idade da informação contida em cada registro. O modelo recomendou a atualização das informações para o conjunto Endereço e Telefone para 180 dos 200 clientes testados.

Decisão	Qtd Clientes	%	Custo A (R\$)
Atualizar	180	90%	3.317,40
Não atualizar	20	10%	
Total	200	100%	3.317,40

Tabela 9 - Decisões da simulação do Modelo Ampliado. Elaborada pelo autor.

Com a aplicação do modelo ampliado, o custo total de realizar estas ações de atualização foi de R\$ 3.317,40; o custo de usar as informações foi de R\$ 1.665,00; o custo

originado pela perda de receita devido à informação estar desatualizada foi de 179,43. Uma vez que a receita esperada pelo uso desta informação é de R\$ 106.849,70 o resultado total da ação de venda de renovação de seguros, aplicando o modelo ampliado foi de R\$ 101.724,07. Este resultado é apresentado na Tabela 10.

	Endereço	Telefone	Total
Registros Usados	200	130	
Receita Esperada	62.116,00	44.770,70	106.886,70
R Total (R\$)	(-) 38,95	(-) 140,48	(-) 179,43
U Total (R\$)	(-) 826,00	(-) 839,80	(-) 1.665,00
A Total (R\$)	(-) 3.317,40		(-) 3.317,40
<b>Total (R\$)</b>			<b>101.724,07</b>

Tabela 10 - Resultado da ação com aplicação do modelo ampliado. Elaborada pelo autor.

#### 7.8.1. SIMULAÇÃO DE APLICAÇÃO DO MODELO ORIGINAL DE GÖRZ (2011)

A fim de permitir uma comparação entre o modelo ampliado e o modelo original, conduziu-se uma simulação de aplicação deste. Uma vez que o modelo original contempla a tomada de decisão sobre *atualizar* ou *não atualizar* para apenas um atributo, para que a comparação pudesse ser executada a simulação envolveu duas etapas: uma primeira na qual se aplicou o modelo somente o atributo Endereço sob a mesma amostra usada para simular a ampliação do modelo original; e uma segunda etapa, na qual se aplicou o modelo apenas para os 130 registros, pois se considerou que, mediante média histórica, 35% dos clientes fazem o pagamento diretamente com a proposta enviada pelo correio sem necessidade de cobrança telefônica.

Na primeira etapa, para cada registro uma decisão foi tomada (*atualizar* ou *não atualizar* a informação). O modelo apresentou como saída a decisão de *atualizar* para 18 clientes e a decisão de *não atualizar* para 182 clientes. O custo total das atualizações sugeridas foi de R\$ 250,75. A Tabela 11, permite uma melhor visualização das decisões apresentadas pelo modelo.

Decisão	Qtd Clientes	%	Custo A (R\$)
Atualizar	18	9%	250,75
Não atualizar	182	91%	
Total	200	100%	250,75

Tabela 11 – Decisões da simulação do Modelo Original - Etapa 1. Elaborada pelo autor.

Na segunda etapa, modelo apresentou como saída a decisão de *atualizar* para 116 clientes e a decisão de *não atualizar* para 14 clientes. O custo total das atualizações sugeridas foi de R\$ 1.901,24. A Tabela 12 permite uma melhor visualização das decisões apresentadas pelo modelo.

Decisão	Qtd registros	%	Custo A (R\$)
Atualizar	116	89,2%	1.901,24
Não atualizar	14	10,8%	
Total	130	100%	1.901,24

Tabela 12 - Decisões da simulação do Modelo Original - Etapa 2. Elaborada pelo autor.

Consolidando o resultado das duas etapas, tanto as receitas esperadas quanto os custos de usar a informação seriam os iguais aos incorridos na aplicação do modelo ampliado. Por outro lado, o custo originado pela perda de receita seria de R\$ 1.403,75 e o custo das ações de atualizar a informação totalizou R\$ 2.151,99. O resultado esperado da ação realizada, mediante aplicação do modelo ampliado é apresentado na Tabela 13.

	Etapa 1	Etapa 2	Total
Registros Usados	200	130	
Receita Esperada	62.116,00	44.770,70	106.886,70
R Total (R\$)	(-) 1.273,30	(-) 130,45	(-) 1.403,75
U Total (R\$)	(-) 826,00	(-) 839,80	(-) 1.665,00
A Total (R\$)	(-) 250,75	(-) 1.901,24	(-) 2.151,99
Total (R\$)			<b>101.665,17</b>

Tabela 13 - Resultado da ação com aplicação do modelo original. Elaborada pelo autor.

### 7.8.1. SIMULAÇÃO DA AÇÃO SEM APLICAÇÃO DE MODELOS

Caso a ação houvesse sido realizada sem a aplicação de nenhum dos modelos (i.e sem gestão da QI), a informação seria usada na situação em que encontrava no instante da aplicação o custo. Isto faria com que se incorresse num custo originado pela perda de receita da grandeza de R\$ 6.541,84, por outro lado, o custo da ação de atualizar a informação seria R\$ 0,00. O resultado esperado da ação realizada sem que nenhum modelo seja aplicado é apresentado na Tabela 14.

	Endereço	Telefone	Total
Registros Usados	200	130	
Receita Esperada	62.116,00	44.770,70	106.886,70
R Total (R\$)	(-) 1.574,84	(-) 4.966,96	(-) 6.541,84
U Total (R\$)	(-) 826,00	(-) 839,80	(-) 1.665,00
A Total (R\$)	0	0	0
<b>Total (R\$)</b>			<b>98.679,06</b>

Tabela 14 – Resultado da ação sem a aplicação de modelos. Elaborada pelo autor.

### 7.9. ANÁLISE E COMPARAÇÃO DAS SIMULAÇÕES DOS MODELOS

Para o caso estudado, o modelo ampliado permitiu inferir que as altas receitas esperadas pelo uso de ambos os atributos se comparados ao custo de atualizar a informação, fez com que o modelo apresentasse como saída a decisão *atualizar* para um grande número de registros (90%). Através do modelo ampliado, identificou-se como sendo 1 ano a idade máxima da informação para a qual o modelo apresenta a decisão de *não atualizar*, ou seja, para todos os registros cuja informação não foi atualizada há 2 ou mais anos, o modelo traz como saída a decisão *atualizar*. Portanto, o modelo ampliado permite concluir que, para o caso estudado e em termos do menor custo total, é recomendável atualizar a informação dos atributos Endereço e Telefone, em uma única ação, a cada 2 anos.

Observou-se, no caso do Banco B aplicação do modelo ampliado uma forte tendência para a decisão baseada no atributo Telefone (de maior volatilidade). Devido ao coeficiente

de declínio maior, o modelo apresentou a decisão *atualizar* em todos os registros em que a informação do atributo Telefone tinha mais de 2 anos de idade. Simulou-se através da planilha de cálculo elaborada (Apêndice A), uma situação em que o atributo Telefone tivesse, hipoteticamente, a mesma volatilidade que Endereço. Neste cenário, os percentuais das decisões resultantes foram de 38% para *atualizar* e 63% para *não atualizar* para o total de registro, valores bastante diferentes dos apresentados mediante os coeficientes reais.

Aplicando o modelo original de Görz (2011), o teste do atributo Telefone (etapa 2), apresentou como saída a decisão de *atualizar* para 89,2% dos registros, percentual semelhante ao apresentado pelo modelo ampliado, para o conjunto de atributos. Na etapa 1, por outro lado, ao testar o atributo Endereço, a quantidade de registros para o qual o modelo recomendou a decisão *atualizar* foi menor que para *não atualizar*. A diferença entre os percentuais de cada decisão em cada etapa ocorreu face a diferença entre a volatilidade da informação contida nestes atributos. Somente ao observar os coeficientes de declínio dos atributos (0,629%, Endereço; 2,957%, Telefone), já se pode inferir que a informação Endereço permanece por muito mais tempo atualizada se comparada a Telefone, entretanto, para que se possa inferir a idade limite que a informação de cada atributo pode alcançar, é indispensável contemplar as receitas esperadas pelo uso da informação e o custo da ação de atualizá-la. No caso estudado, o modelo de Görz (2011) recomenda atualizar registros cuja informação do atributo Telefone tenha 2 ou mais anos de idade, enquanto que para o atributo Endereço a idade máxima tolerada é de 7 anos. Isso significa que é recomendável atualizar a informação Endereço a cada 8 anos, e a informação Telefone a cada 2 anos.

Realizou-se ainda uma simulação para a situação de nenhum dos modelos ser aplicado (i.e. sem gestão da QI). Neste caso, nenhuma ação de atualizar a informação foi executada.

Os resultados obtidos podem são apresentados lado a lado na Tabela 15, a fim de facilitar a comparação. Como já era esperado, o resultado apresentado na simulação de realizar o a ação de venda de seguros sem a aplicação de nenhum modelo (i.e. sem gestão da QI) foram os piores dos três, situando-se 3% abaixo do resultado apresentado pelo modelo ampliado. O modelo original de Görz (2011) apresentou resultados próximos dos

obtidos através do modelo ampliado (99,9%), o que valida o modelo ampliado, mas ilustra que podem ser pequenos os ganhos em eficiência, numa primeira análise.

	<b>Modelo Ampliado</b>	<b>Modelo Original (Görz, 2011)</b>	<b>Sem modelo (i.e. sem gestão da QI)</b>
<b>Receita Esperada</b>	106.886,70	106.886,70	106.886,70
<b>R Total (R\$)</b>	(-) 179,43	(-) 1.403,75	(-) 6.541,84
<b>U Total (R\$)</b>	(-) 1.665,00	(-) 1.665,00	(-) 1.665,00
<b>A Total (R\$)</b>	(-) 3.317,40	(-) 2.151,99	0
<b>Total (R\$)</b>	<b>101.724,07</b>	<b>101.665,17</b>	<b>98.679,06</b>
<b>% (ampliado = 100%)</b>	<b>100%</b>	<b>99,9%</b>	<b>97%</b>

Tabela 15 - Comparação do resultado dos modelos e da ação sem aplicação de modelo para o ano 1. Elaborada pelo autor.

Realizou-se ainda uma simulação complementar considerando a mesma amostra de registros, porém tendo como instante de aplicação o mês Maio de 2013. Nesta simulação considerou-se que as decisões recomendadas pelos modelos no ano anterior haviam sido levadas a cabo, ou seja, a idade da informação para os registros cuja decisão haveria sido, em Maio de 2012, de atualizar foram zeradas. Novamente o modelo ampliado, o modelo original de Görz (2011) e a realização da ação sem a aplicação de nenhum modelo (i.e. sem gestão da QI), foram comparados. Os resultados obtidos, que são apresentados na Tabela 16, validam a funcionalidade do modelo ampliado de apoiar a melhor decisão considerando os custos no tempo futuro, já que o mesmo apresentou resultado tanto em valor absolutos como em valores relativos superior aos obtidos no período anterior.

	<b>Modelo Ampliado</b>	<b>Modelo Original (Görz, 2011)</b>	<b>Sem modelo (i.e. sem gestão da QI)</b>
<b>Receita Esperada</b>	106.886,70	106.886,70	106.886,70
<b>R Total (R\$)</b>	(-) 1.526,52	(-) 2.678,36	(-) 8.069,53
<b>U Total (R\$)</b>	(-) 1.665,00	(-) 1.665,00	(-) 1.665,00
<b>A Total (R\$)</b>	(-) 350,17	(-) 132,75	0
<b>Total (R\$)</b>	<b>103.344,21</b>	<b>102.196,71</b>	<b>97.151,37</b>
<b>% (ampliado = 100%)</b>	<b>100%</b>	<b>98,9%</b>	<b>94%</b>

Tabela 16 - Comparação do resultado dos modelos e da ação sem aplicação de modelo para ano 2. Elaborada pelo autor.

Finalmente, ao realizar uma comparação em termos de eficiência dos resultados da aplicação de ambos os modelos, os resultados mostram uma sensível superioridade do

modelo ampliado sobre o modelo de original de Görz. A ação realizada sem aplicação de nenhum modelo (i.e. sem gestão da QI), como era esperado, mostrou resultados inferiores, devido ao fato de que a inexistência de custos da ação de informação não compensa a custo originado pela perda de receita por problemas relacionados à QI.

Outra abordagem de comparação entre os modelos que vale a pena ser feita diz respeito à facilidade de uso dos mesmos. O modelo original de Görz (2011) apresenta duas vantagens claras sobre o modelo ampliado: o fato de ser matematicamente mais simples, que facilita o entendimento de quem o deseja aplicar; e a possibilidade de ser aplicável através de códigos de linguagem estruturada de consulta a bancos de dados (do inglês, *Structured Query Language - SQL*). Uma vez que se concluiu a elaboração das planilhas de cálculos (Apêndice A e B), no entanto, a aplicação tornou-se igualmente simples e rápida para ambos os modelos. A ligeira manipulação dos parâmetros e a inserção das datas de atualização dos registros permitiu extrair as saídas indicadas por ambos os modelos instantaneamente.

Ao analisar a dificuldade em construir dos modelos para o caso estudado, embora o modelo ampliado pareça matematicamente mais complexo que o original, os esforços dedicados à identificação dos parâmetros necessários à sua construção foram praticamente os mesmos que para o modelo original.

Em última análise, salienta-se que as simulações realizadas possibilitaram validar a aplicação de ambos os modelos enquanto ferramentas para a gestão da QI. As vantagens provenientes da aplicação dos modelos são claras ao comparar os resultados da ação de venda de renovação de seguros com e sem aplicação dos modelos. Para o caso estudado, o modelo ampliado apresentou melhores resultados em termos de eficiência de custos. No entanto, este é somente um caso prático de muitos outros nos quais os modelos podem ser testados e terem seus resultados comparados.



## 8 CONSIDERAÇÕES FINAIS

O presente estudo buscou apresentar uma contribuição à Qualidade da Informação, com foco na gestão da dimensão volatilidade. O modelo ampliado proposto apresenta-se como uma ferramenta de gerenciamento da qualidade da informação em nível operacional, pois propicia às organizações realizar ações de atualização da informação de forma eficiente.

A simulação de aplicação do modelo ampliado, baseada em dados reais de uma empresa privada, permitiu validar a aplicabilidade do mesmo. Através da análise do processo de vendas de renovações de seguros do Banco B, construiu-se um modelo capaz de apoiar a melhor decisão entre *atualizar* ou *não atualizar* a informação dos atributos Telefone e Endereço, do banco de dados de clientes da empresa.

A aplicação simulada do modelo original de Görz (2011), juntamente à simulação de o processo ser executado sem a aplicação de modelo algum (i.e. sem gestão da QI), permitiu realizar a comparação entre os resultados apresentados pelos dois modelos. Através desta comparação, o modelo ampliado apresenta resultados ligeiramente superiores em relação do modelo de Görz (2011) em termos de eficiência de custos, enquanto que a simulação sem aplicação de nenhum modelo apresentou, como já era esperado, um resultado inferior (6% menor que do modelo ampliado). Indica-se, pois, a utilização dos modelos matemáticos como ferramenta de apoio a gestão da QI, testando-os previamente a fim de averiguar qual é capaz de apresentar o resultado mais eficiente para o caso em questão.

Faz-se importante mencionar que, embora a empresa e o processo escolhidos para o estudo tenham atendido a todas as premissas levantadas durante o projeto desta pesquisa, o processo de construção do modelo e identificação dos parâmetros apresentou algumas dificuldades, são elas: a falta de conhecimento prévio sobre a volatilidade das informações dos atributos para os quais se deseja construir o modelo ou mesmo da frequência de erros nos bancos de dados, o que faz do processo de identificação dos coeficientes de declínio um dos principais desafios na construção dos modelos; o extenso trabalho dedicado para analisar o processo em questão, sobretudo se não há prévio mapeamento do custo das

atividades de atualizar e usar a informação, onera em custos relevantes e que não são considerados pelo modelo.

No caso estudado, embora se tenha percebido a preocupação por parte dos entrevistados quanto aos impactos gerados por problemas de QI, não há ferramentas que permitam mensurá-los nem tampouco gerenciar a QI. Ainda que a informação de Telefone e Endereço seja essencial para o sucesso de ações comerciais da empresa, a sua atualização não é apoiada por um processo recorrente, o que corrobora com as afirmações de que hoje muito dinheiro é perdido pelas empresas problemas na QI (ECKERSON, 2002). Neste contexto, o modelo ampliado proposto por este estudo, seria uma ferramenta capaz de incrementar a efetividade geral da empresa estudada.

Finalmente, a fim de incitar o seguimento da pesquisa científica no campo em que este trabalho focou seus esforços – a gestão da QI – sugere-se que esforços futuros sejam conduzidos na direção de:

- elaborar um modelo de programação linear que permita a escolha do menor custo entre vários conjuntos de atributos ou até mesmo, de atualizá-los de maneira independente;
- aprofundar o estudo a cerca da probabilidade de usar os atributos no futuro;
- comprovar através de técnicas estatísticas há a existência de relação entre múltiplos atributos de informação, de modo que seu custo de atualizar o conjunto seja menos custoso que atualizar os atributos independentemente;
- contemplar, no modelo proposto, valor do dinheiro no tempo inserindo uma taxa de desconto a fim de balancear os custos no instante da aplicação com os custos que incidem no tempo futuro;
- considerar a evolução dos custos e métricas envolvidos ao longo do tempo no modelo a fim de torná-lo mais robusto;
- conduzir estudo de caso que contemple mais de dois atributos de informação.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BALLOU, D. P.; PAZER, H. L. Designing information systems to optimize the accuracy-currency tradeoff. **Information Systems Research**. v. 6, n. 1, p. 51-72., 1995.
- BATINI, C.; SCANNAPIECO, M. Data Quality: Concepts, Methodologies and Techniques. **Data-Centric Systems and Applications**. Berlin, 2006.
- BATINI, C.; CAPPIELLO, C.; FRANCALANCI, C.; MAURINO, A. Methodologies for data quality assessment and improvement. **ACM Computing Surveys**. July, 2009, v. 41, n. 3, p. 1-52.
- BATINI, C.; BARONE, D.; MASTRELLA, M.; MAURINO, A.; RUFFINI, C. A framework and a methodology for data quality assessment and monitoring. **Proceedings of the 12th International Conference on Information Quality**, 2007.
- BRODBECK, Angela Freitag. Avaliação da Qualidade da Informação nos Sistemas de Informação e de Apoio à Decisão: um Estudo Introdutório. In: **Anais do XIX Congresso Nacional da ANPAD**. João Pessoa: 1995.
- CHEN, P. The Entity-Relationship Model: Toward a Unified View of Data. **ACM Transactions on Database Systems**. v. 1, n. 1, p. 9 - 36., 1976.
- DRUCKER, Peter. The Information Executives Trully Need. **Harvard Business Review**. Janeiro-Fevereiro 1995, v. 73, n. 1, p. 54-62.
- DRUCKER, Peter F. **Managing for the Future: The 1990s and Beyond**. Truman Talley Books/Dutton (Penguin Group), N.Y., 1992.
- ECKERSON, W. Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data. **The Data Warehousing Institute**, 2002.
- ENGLISH, Larry. **Applying Deming's 14 Points to Information Quality: The 14 points of Total Information Quality Management**. In: Information Impact International, 2011.
- EVEN, A.; KAISER, M. A Framework for Economics-Driven Assessment of Data Quality Decisions. In: **Proceedings of the Fifteenth Americas Conference on Information Systems (AMCIS 09)**, August 2009.
- FISHER, C. W.; CHENGALUR-SMITH, I. N.; BALLOU, D. P.; The Impact of Experience and Time on the Use of Data Quality Information in Decision Making. **Information Systems Research**, v. 14, n. 2, p. 170-188., 2003.

GÖRZ, Quirin. An Economics-Driven Decision Model for Data Quality Improvement - A Contribution to Data Currency. In: **Proceedings of the 17th Americas Conference on Information Systems**, AMCIS, 2011.

HARRIS INTERACTIVE. **Information Workers Beware: Your Business Data Can't Be Trusted**, 2006. Disponível em: [http://www.sap.com/about/newsroom/businessobjects/20060625\\_005\\_028.epx](http://www.sap.com/about/newsroom/businessobjects/20060625_005_028.epx), Acesso em 13/10/2008.

HEINRICH, B.; KAISER, M.; KLIER, M. DQ metrics: a novel approach to quantify currency and its application in CRM. **Proceedings of the 12th International Conference on Information Quality (ICIQ 07)**, p. 431-445., 2007

HEINRICH, B.; KAISER, M.; KLIER, M. A Procedure to Develop Metrics For Currency And Its Application. In: **CRM, ACM Journal of Data and Information Quality**, 1, 1, 5:1-5:28., 2009.

HEINRICH, B.; KLIER, M. A Novel Data Quality Metric For Currency Considering Supplemental Data. In: **Proceedings of the 17th European Conference on Information Systems (ECIS 09)**, June 2009, Verona, Italy, 2701- 2713.

HELPERT, M.; FOLEY, O; GE, M.; CAPIELLO, C. Limitations of Weighted Sum Measures for Information Quality. In: **Americas Conference on Information Systems**, 2009

HELLMAN, C. Address Management for Mail-Order and Retail. Directions Magazine, 2004. Disponível em <http://www.directionsmag.com/articles/address-management-for-mail-order-and-retail/123775>> Acesso em 15/10/2011

JAYNES, E.T. Probability Theory: The Logic of Science. Cambridge University Press, 1995.

JURAN, J.M. **A Qualidade desde o Projeto**. São Paulo: Pioneira, 1992.

JIANG, Z.; SARKAR, S.; DE, P.; DEY, D. A Framework for Reconciling Attribute Values from Multiple Data Sources. **Management Science**. v. 53, n. 12, p. 1946-1963., 2007.

KAHN, B.K.; STRONG, D.M.; WANG, R.Y. Information Quality benchmarks: Product and Service performance. **Communications of the ACM**. Apr 2002, v. 45, n. 4, p. 184-192

KAISER, M.; KLIER, M.; HEINRICH, B. A Procedure to Develop Metrics for Currency and its Application in CRM. **Journal of Data and Information Quality (JDIQ)**. v. 1, n. 1, p. 1-26., 2009.

KLEIN, B.D; CALLAHAN, T.J. A comparison of information technology professionals' and data consumers' perceptions of the importance of the dimensions of information quality. **International Journal of Information Quality**, v. 1, n. 4, p. 392-411., 2007.

LEE, Y.W.; STRONG, D.M.; KAHN, B.K.; WANG, R.Y. AIMQ: a methodology for information quality assessment. **Information & Management**, v. 40 (2), p. 133-146., 2002.

LIMA, Luis Francisco Ramos; MAÇADA, A. C. G. Modelo para Qualidade da Informação na Indústria Bancária - O caso dos bancos públicos. In: **XXXI Encontro Nacional da Anpad**. Rio de Janeiro. XXXI Encontro Nacional da Anpad, 2007. v. 1. p. 1-16., 2007

LIMA, Luis Francisco Ramos; MAÇADA, A. C. G. Pesquisa em Qualidade da Informação: um Estudo sobre o Estado da Arte no Tema e sua Consolidação. In: **XXXII EnANPAD**. Rio de Janeiro. XXXII EnANPAD, 2008. p. 1-14., 2008

McGEE, James; PRUSAK, Laurence. **Gerenciamento Estratégico da Informação**. São Paulo: Campus, 1994.

O'BRIEN, James A. **Management Information Systems**. McGraw-Hill, 1999.

OZ, Effy. **Management Information Systems**. Cambridge: Course Technology Thomson Learning, 2000.

PIPINO, Leo L.; LEE, Yang W.; WANG, Richard Y. Data Quality Assessment. **Communications of the ACM**. April 2002, v. 45, n. 4

REDMAN, Thomas C. **Data Quality for the Information Age**. Artech House Computer Science Library, 1996.

REDMAN, Thomas C. The impact of poor data quality on typical enterprise. **Communications of the ACM**. February 1998, v. 41, n. 2.

REDMAN, Thomas C. **Measuring Data Accuracy** – A Framework and Review in “Information Quality” Richard Wang, Elizabeth Pierce, Stuart Madnick, and Craig Fisher. M.E. Sharpe, 2005.

PWC, PRICEWATERHOUSECOOPERS 2004. Data Quality 2004 Survey. Accessed 2009/03/08. <http://www.pwc.com/extweb/ncpressrelease.nsf/docid/A0C678E7D21B7EB3CA256F490008D3E5>

STRONG, Diane M.; LEE, Yang W.; WANG, Richard Y. Data Quality in Context. **Communications of the ACM**. May 1997, v. 40, n. 5, p. 103-110. (a)

THE ECONOMIST. Data, data everywhere. London, Economist Newspaper Ltd. 25-Fev-2010 Edição Impressa.

THE ECONOMIST. All too much - Monstrous amounts of data. London, Economist Newspaper Ltd.. 25-Fev-2010 Edição Impressa.

WAND, Yair; WANG, Richard Y. Anchoring Data Quality Dimensions in ontological foundations. **Communications of the ACM**. November 1996, v. 39, n. 11, p. 86-95.

WANG, Richard Y. KON, Henry B. MADNICK, Stuart E. Data Quality requirements analysis and modeling. In: 9th International Conference on Data Engineering. April, 1993.

WANG, R. Y.; STRONG, D. M. Beyond accuracy: What data quality means to data consumers. **Journal of Management Information Systems**, v. 12, n. 4, p. 5-34, 1996.

WIXOM, Barbara H.; TODD, Peter A. A Theoretical Integration of User Satisfaction and Technology Acceptance. **Information Systems Research**. March 2005, v. 16, n. 1, p. 85-102.

YU, Z. and WANG, Y. An empirical research on non-technical factors related to statistical data quality in China's enterprises. **International Journal of Information Quality**, v. 1, n. 2, p. 193-208., 2007.

**APÊNDICE A – PLANILHA ELETRÔNICA PARA AMPLICAÇÃO DO MODELO AMPLIADO.**

Parâmetros do Modelo			Dados								
Parâmetro	Atributo	Valor	#	Cod. Registro	Data Info Endereço	Data Info Telefone	Idade Endereço	Idade Telefone	CT Atualizar	CT Não Atualizar	Melhor Decisão
Coef. Declínio	Endereço	0,62%	1	161852	12/02/2001	12/02/2001	11	11	30,10	130,58	Atualizar
Coef. Declínio	Telefone	2,957%	2	117900	18/01/2007	18/01/2007	5	5	30,10	71,84	Atualizar
U	Endereço	R\$ 4,13	3	237246	20/01/2010	20/01/2010	2	2	30,10	38,95	Atualizar
U	Telefone	R\$ 8,48	4	189470	19/03/2009	19/03/2009	3	3	30,10	50,19	Atualizar
R	Endereço	R\$ 310,58	5	158982	13/02/2008	13/02/2008	4	4	30,10	61,15	Atualizar
R	Telefone	R\$ 344,39	6	182803	19/12/2008	19/12/2008	3	3	30,10	50,19	Atualizar
A	-	R\$ 18,45	7	124231	08/03/2007	08/03/2007	5	5	30,10	71,84	Atualizar
Fu(0)	Endereço	100%	8	106784	06/03/2006	06/03/2006	6	6	30,10	62,25	Atualizar
Fu(0)	Telefone	67%	9	196117	11/02/2009	11/02/2009	3	3	30,10	50,19	Atualizar
			10	189035	04/11/2008	04/11/2008	4	4	30,10	61,15	Atualizar
Data de Aplicação		15/05/2012	11	106756	14/02/2006	14/02/2006	6	6	30,10	62,25	Atualizar
			12	160721	26/02/2008	26/02/2008	4	4	30,10	61,15	Atualizar
			13	239485	10/02/2010	10/02/2010	2	2	30,10	38,95	Atualizar
			14	279682	18/01/2011	18/01/2011	1	1	30,10	27,41	Não Atualizar
			15	284549	15/02/2011	15/02/2011	1	1	30,10	27,41	Não Atualizar
			16	44556	03/03/2004	03/03/2004	8	8	30,10	102,31	Atualizar
			17	243493	08/03/2010	08/03/2010	2	2	30,10	38,95	Atualizar
			18	107233	04/04/2008	04/04/2008	4	4	30,10	61,15	Atualizar
			19	237964	27/02/2008	27/02/2008	4	4	30,10	61,15	Atualizar
			20	157834	31/01/2008	31/01/2008	4	4	30,10	61,15	Atualizar
			21	238626	03/02/2010	03/02/2010	2	2	30,10	38,95	Atualizar
			22	106196	09/03/2006	09/03/2006	6	6	30,10	62,25	Atualizar
			23	242295	05/03/2010	05/03/2010	2	2	30,10	38,95	Atualizar
			24	236633	28/03/2008	28/03/2008	4	4	30,10	61,15	Atualizar
			25	158795	12/03/2008	12/03/2008	4	4	30,10	61,15	Atualizar
			26	251359	17/11/2009	17/11/2009	2	2	30,10	38,95	Atualizar
			27	278104	29/12/2010	29/12/2010	1	1	30,10	27,41	Não Atualizar
			28	243060	04/08/2008	04/08/2008	4	4	30,10	61,15	Atualizar
			29	59585	01/03/2005	01/03/2005	7	7	30,10	92,41	Atualizar
			30	289659	11/02/2011	11/02/2011	1	1	30,10	27,41	Não Atualizar
			31	232226	28/07/2008	28/07/2008	4	4	30,10	61,15	Atualizar
			32	193699	14/01/2009	14/01/2009	3	3	30,10	50,19	Atualizar
			33	125990	01/03/2007	01/03/2007	5	5	30,10	71,84	Atualizar
			34	107421	24/02/2011	24/02/2011	1	1	30,10	27,41	Não Atualizar
			35	41319	13/03/2004	13/03/2004	8	8	30,10	102,31	Atualizar
			36	57180	08/03/2005	08/03/2005	7	7	30,10	92,41	Atualizar
			37	298149	07/01/2010	07/01/2010	2	2	30,10	38,95	Atualizar
			38	186503	17/02/2008	17/02/2008	5	5	30,10	50,19	Atualizar
			39	184257	31/08/2006	31/08/2006	6	6	30,10	62,25	Atualizar
			40	58778	18/03/2005	18/03/2005	7	7	30,10	92,41	Atualizar
			41	196975	21/02/2009	21/02/2009	3	3	30,10	50,19	Atualizar
			42	161806	26/10/2006	26/10/2006	6	6	30,10	62,25	Atualizar
			43	105963	21/03/2006	21/03/2006	6	6	30,10	62,25	Atualizar
			44	156717	12/02/2008	12/02/2008	4	4	30,10	61,15	Atualizar
			45	239304	30/10/2009	30/10/2009	3	3	30,10	50,19	Atualizar
			46	125217	16/03/2007	16/03/2007	5	5	30,10	71,84	Atualizar
			47	241759	03/03/2010	03/03/2010	2	2	30,10	38,95	Atualizar
			48	186447	14/02/2009	14/02/2009	3	3	30,10	50,19	Atualizar
			49	156208	18/12/2000	18/12/2000	11	11	30,10	130,58	Atualizar
			50	124510	09/03/2007	09/03/2007	5	5	30,10	71,84	Atualizar
			51	244127	21/03/2010	21/03/2010	2	2	30,10	38,95	Atualizar
			52	289727	11/02/2011	11/02/2011	1	1	30,10	27,41	Não Atualizar
			53	107467	30/03/2006	30/03/2006	6	6	30,10	62,25	Atualizar
			54	237008	28/01/2004	28/01/2004	8	8	30,10	102,31	Atualizar
			55	238017	13/06/2009	13/06/2009	3	3	30,10	50,19	Atualizar
			56	240271	27/06/2002	27/06/2002	10	10	30,10	121,99	Atualizar

Resultado da Aplicação		
Total Registros	200	100%
Ord. Atualizar	180	90%
Ord. Não Atualizar	20	10%
R Total	R\$ 107.747,48	
U Total	R\$ 5.419,79	
A Total	R\$ 548,17	

## APÊNDICE B – PLANILHA ELETRÔNICA PARA APLICAÇÃO DO MODELO ORIGINAL.

Parâmetros do Modelo			Dados								
Parâmetro	Atributo	Valor	R	Cod. Registro	Data Info Endereço	Data Info Telefone	Idade Endereço	Idade Telefone	CT Atualizar	CT Não Atualizar	Melhor Decisão
Coef. Decisão	Endereço	0,629%	1	151852	12/02/2001	12/02/2001	11	11	30,10	130,58	Atualizar
Coef. Decisão	Telefone	2,857%	2	117900	18/01/2007	18/01/2007	5	5	30,10	71,84	Atualizar
U	Endereço	R\$ 4,13	3	237246	20/01/2010	20/01/2010	2	2	30,10	38,95	Atualizar
U	Telefone	R\$ 6,46	4	198470	19/03/2009	19/03/2009	3	3	30,10	50,19	Atualizar
R	Endereço	R\$ 310,58	5	158982	13/02/2008	13/02/2008	4	4	30,10	61,15	Atualizar
R	Telefone	R\$ 344,39	6	192801	19/12/2008	19/12/2008	3	3	30,10	50,19	Atualizar
A	-	R\$ 16,43	7	114231	08/03/2007	08/03/2007	5	5	30,10	71,84	Atualizar
Pui(D)	Endereço	100%	8	106784	06/03/2006	06/03/2006	6	6	30,10	82,25	Atualizar
Pui(T)	Telefone	67%	9	196117	12/02/2009	12/02/2009	3	3	30,10	50,19	Atualizar
Data de Aplicação			10	189035	04/11/2008	04/11/2008	4	4	30,10	61,15	Atualizar
			11	106758	14/02/2006	14/02/2006	6	6	30,10	82,25	Atualizar
			12	146723	26/02/2006	26/02/2006	4	4	30,10	61,15	Atualizar
Resultado da Aplicação			13	238485	10/02/2010	10/02/2010	2	2	30,10	38,95	Atualizar
Total Registros	200	100%	14	279682	18/01/2011	18/01/2011	1	1	30,10	27,41	Não Atualizar
Qtd. Analisar	180	90%	15	284549	15/02/2011	15/02/2011	1	1	30,10	27,41	Não Atualizar
Qtd. Não Analisar	20	10%	16	44396	09/03/2004	09/03/2004	8	8	30,10	102,31	Atualizar
R Total			17	243463	08/03/2010	08/03/2010	2	2	30,10	38,95	Atualizar
U Total			18	107233	04/04/2008	04/04/2008	4	4	30,10	61,15	Atualizar
A Total			19	237984	27/02/2008	27/02/2008	4	4	30,10	61,15	Atualizar
			20	157834	31/01/2008	31/01/2008	4	4	30,10	61,15	Atualizar
			21	238628	03/02/2010	03/02/2010	2	2	30,10	38,95	Atualizar
			22	106188	09/03/2006	09/03/2006	6	6	30,10	82,25	Atualizar
			23	242205	05/03/2010	05/03/2010	2	2	30,10	38,95	Atualizar
			24	296653	28/03/2008	28/03/2008	4	4	30,10	61,15	Atualizar
			25	156799	12/02/2008	12/02/2008	4	4	30,10	61,15	Atualizar
			26	211559	17/11/2009	17/11/2009	2	2	30,10	38,95	Atualizar
			27	278104	29/12/2010	29/12/2010	1	1	30,10	27,41	Não Atualizar
			28	243360	04/08/2008	04/08/2008	4	4	30,10	61,15	Atualizar
			29	55885	01/03/2005	01/03/2005	7	7	30,10	92,41	Atualizar
			30	289659	11/02/2011	11/02/2011	1	1	30,10	27,41	Não Atualizar
			31	232226	28/07/2008	28/07/2008	4	4	30,10	61,15	Atualizar
			32	193698	14/01/2009	14/01/2009	3	3	30,10	50,19	Atualizar
			33	123390	01/03/2007	01/03/2007	5	5	30,10	71,84	Atualizar
			34	107421	24/02/2011	24/02/2011	1	1	30,10	27,41	Não Atualizar
			35	41519	15/03/2004	15/03/2004	8	8	30,10	102,31	Atualizar
			36	57190	08/03/2005	08/03/2005	7	7	30,10	92,41	Atualizar
			37	236148	07/01/2010	07/01/2010	2	2	30,10	38,95	Atualizar
			38	196503	17/02/2009	17/02/2009	3	3	30,10	50,19	Atualizar
			39	194257	31/08/2006	31/08/2006	6	6	30,10	82,25	Atualizar
			40	58778	18/03/2005	18/03/2005	7	7	30,10	92,41	Atualizar
			41	196975	21/02/2009	21/02/2009	3	3	30,10	50,19	Atualizar
			42	161806	25/10/2006	25/10/2006	6	6	30,10	82,25	Atualizar
			43	105963	21/03/2006	21/03/2006	6	6	30,10	82,25	Atualizar
			44	156717	12/02/2008	12/02/2008	4	4	30,10	61,15	Atualizar
			45	235304	30/10/2009	30/10/2009	3	3	30,10	50,19	Atualizar
			46	125227	16/03/2007	16/03/2007	3	3	30,10	71,84	Atualizar
			47	241759	03/03/2010	03/03/2010	2	2	30,10	38,95	Atualizar
			48	196447	16/02/2009	16/02/2009	3	3	30,10	50,19	Atualizar
			49	156206	18/12/2000	18/12/2000	11	11	30,10	130,58	Atualizar
			50	124510	09/03/2007	09/03/2007	5	5	30,10	71,84	Atualizar
			51	244127	22/03/2010	22/03/2010	2	2	30,10	38,95	Atualizar
			52	283727	11/02/2011	11/02/2011	1	1	30,10	27,41	Não Atualizar
			53	107487	30/03/2006	30/03/2006	6	6	30,10	82,25	Atualizar
			54	237008	28/01/2004	28/01/2004	8	8	30,10	102,31	Atualizar
			55	238657	13/06/2009	13/06/2009	3	3	30,10	50,19	Atualizar