

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**

Rodrigo Reno Strey

**INTEGRAÇÃO DE MÉTODOS ESTATÍSTICOS À
METODOLOGIA SIEGO**

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação

Orientador: Prof. Marcelo Menezes Reis, Dr.

Florianópolis, Abril/2006

INTEGRAÇÃO DE MÉTODOS ESTATÍSTICOS À METODOLOGIA SIEGO

Rodrigo Reno Strey

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação Área de Concentração de Sistemas de Conhecimento e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Banca Examinadora

Raul Sidnei Wazlawick, Dr.
Coordenador do Curso de Pós-Graduação

Marcelo Menezes Reis, Dr. (orientador)

José Leomar Todesco, Dr.

Sílvia Modesto Nassar, Dra.

Oscar Dalfovo, Dr.

DEDICATÓRIA

Este trabalho de dissertação
é dedicado à minha esposa
Francine Hartmann Strey,
aos meus pais Reno e Dilva e
também à minha irmã Raquel.

AGRADECIMENTOS

Em primeiro lugar e antes de quaisquer outras coisas agradeço por mais esta conquista ao meu Deus que tem me abençoado a cada nascer e pôr do sol, não me deixando desanimar, nem enfraquecer, não deixando de estender sua mão e seu consolo nos momentos de angústia e também de fraqueza quando tive momentos em que os desafios pareciam ser insuperáveis e as dificuldades encobriam o caminho da vitória. A Ele, devo mais uma vez eterna gratidão pelas bênçãos derramadas.

Aos meus pais Reno e Dilva Maria Laurindo Strey, pelo apoio, incentivo, compreensão e dedicação que sempre tiveram comigo em todos os momentos de minha vida, onde também sempre serviram de exemplo para determinação, honestidade e vitória. Pelo carinho que sempre demonstraram por mim proporcionando-me um lar onde pude buscar forças quando minhas dificuldades apareceram.

A minha irmã Raquel Strey pelo carinho demonstrado e também pelo apoio a mim dado.

A minha amada e estima esposa Francine, pela compreensão que teve durante os momentos em que não pude lhe dar a atenção mais que merecida e tive que escolher entre o nosso lazer e os estudos para que tudo ocorresse bem e que eu pudesse hoje estar escrevendo este agradecimento. A você meu amor, o meu muito obrigado, você merece esta vitória tanto quanto eu.

Ao meu opa Roland Strey (em memória) e ao meu avô Bento Laurindo (em memória) que serviram como exemplo de honestidade, dignidade e de homens que lutaram para vencer na vida, mostrando a mim o quão importante é ser um homem de bem.

Ao professor, orientador, e amigo Dr. Marcelo Menezes Reis, pela atenção e auxílio dispensados na elaboração deste trabalho.

Aos demais amigos que contribuíram em diversos momentos com conhecimentos, apoio, incentivo e compreensão nos diversos momentos da elaboração deste.

A todos meus sinceros agradecimentos e que Deus vos abençoe.

RESUMO

Este trabalho de pesquisa visa ao desenvolvimento de um Sistema de Informação onde o objetivo principal é realizar a integração da metodologia SIEGO (Sistema de Informação Estratégico para o Gerenciamento Operacional) aos Métodos Estatísticos (Séries Temporais, Análise de Regressão e Distribuições de Frequências) para que através desta associação empresas possam tomar decisões visando melhorias relacionadas ao tripé Custo, Tempo e Qualidade com embasamento científico.

Para esta integração foi desenvolvido um sistema que aponta dados relativos ao tripé acima mencionado e posteriormente os exporta para o aplicativo Microsoft Excel, onde os dados são analisados.

Como resultados desta integração são apontados locais onde podem ser realizadas melhorias de acordo com o proposto pela metodologia SIEGO.

Para realização do trabalho é mostrada uma aplicação a uma empresa do ramo eletromecânico, situada em Blumenau, SC.

A fundamentação teórica deste trabalho aborda conteúdos relacionados a Sistemas de Informação de modo geral, à metodologia SIEGO, à montagem de *Data Warehouses* e também conteúdos relacionados aos métodos estatísticos empregados, análise de regressão, séries temporais e distribuições de frequências.

ABSTRACT

This research work aims at to the development of a System of Information where the main objective is to carry through the integration of methodology SIEGO (Strategical System of Information for the Operational Management) to the Statistical Methods (Temporal Series, Analysis of Regression and Distributions of Frequencies) so that through this association companies related improvements can take decisions aiming at to the tripod Cost, Time and Quality with scientific basement.

For this integration a system was developed that points given relative to the above-named tripod and later it exports them to the applicatory Microsoft Excel, where the data is analyzed.

As results of this integration they are pointed places where improvements can in accordance with be carried through the considered one for methodology SIEGO.

For accomplishment of the work it is shown an application to a company of the electromechanical branch, situated in Blumenau, SC.

The theoretical recital of this work approaches contents related the Systems of Information in general way, to methodology SIEGO, the assembly of Date Warehouses and contents also related to the statistical methods employees, analysis of regression, temporal series and distributions of frequencies.

LISTA DE ILUSTRAÇÕES

FIGURA 1 – AMBIENTE DE UM SISTEMA EMPRESARIAL	22
FIGURA 2 – TRIPÉ DA ORGANIZAÇÃO – SUGESTÃO DA METODOLOGIA	23
FIGURA 3 – FLUXOGRAMA REPRESENTATIVO DAS FASES E PASSOS DA METODOLOGIA SIEGO	25
FIGURA 4 - FLUXOGRAMA REFERENTE AOS PASSOS PARA CONSTRUÇÃO DE UM DATA WAREHOUSE	28
FIGURA 5 – EXEMPLO DE REDE NEURAL - ADAPTADO DE SANTOS (2005).....	30
FIGURA 6 – SÉRIE TEMPORAL DE VALORES DE VENDAS DE UMA EMPRESA X EM FUNÇÃO DO TEMPO - TENDÊNCIA.....	36
FIGURA 7 – EXEMPLO GRÁFICO DE VARIAÇÃO SAZONAL	36
FIGURA 8 – SÉRIE VIRTUAL DE QUANTIDADES VENDIDAS ANUAIS DO PRODUTO X.....	38
FIGURA 9 – SÉRIE VIRTUAL DE QUANTIDADES VENDIDAS ANUALMENTE DO PRODUTO X ASSOCIADA A UMA TENDÊNCIA ARTIFICIAL – MÉDIA MÓVEL	39
FIGURA 10 – DIAGRAMA DE DISPERSÃO - ADAPTADO DE LEVINE ET AL (2000).....	42
FIGURA 11 – EXEMPLO DE REGRESSÃO LINEAR – ADAPTADO DE LEVINE ET AL (2000).....	44
FIGURA 12 – EXEMPLO FICTÍCIO DE RESÍDUOS PADRONIZADOS.....	46
FIGURA 13 – STATISTICA (INTERFACE PRINCIPAL)	61
FIGURA 14 – INTERFACE MICROSOFT EXCEL COM ARQUIVO ABERTO.	62
FIGURA 15 – TELA INICIAL DO SISTEMA APRESENTADA AO INICIAR EXECUÇÃO	76
FIGURA 16 – REALIZAÇÃO DA CARGA DE DADOS	76
FIGURA 17 – TELA DE PARÂMETROS PARA UTILIZAÇÃO NA ANÁLISE DE QUALIDADE	77
FIGURA 18 – APRESENTAÇÃO DOS CLIENTES CLASSIFICADOS DE ACORDO COM DEFINIÇÕES DA TELA DE PARÂMETROS.....	78
FIGURA 19 – TELA PARA VERIFICAÇÃO DO ITEM CUSTO DA METODOLOGIA SIEGO	79
FIGURA 20 – TELA DE CONSULTA DE FATURAMENTO EM RELAÇÃO AO TEMPO E POR EMPRESA.....	80
FIGURA 21 – APRESENTAÇÃO DE VALORES COM ACRÉSCIMO DE UMA EMPRESA, CHAPECÓ.....	80
FIGURA 22 – UTILIZAÇÃO DO GRÁFICO DE PIZZA PARA APRESENTAÇÃO DOS DADOS	81
FIGURA 23 – UTILIZAÇÃO DO GRÁFICO DE BARRAS PARA VALORES TOTAIS MÊS A MÊS	81
FIGURA 24 – UTILIZAÇÃO DO GRÁFICO DE LINHA PARA VISUALIZAR O DESEMPENHO DE CADA EMPRESA EM RELAÇÃO AOS SEUS FATURAMENTOS.....	82
FIGURA 25 – APLICAÇÃO DA TÉCNICA DE CUBO DE DECISÃO	82
FIGURA 26 – HETEROCEDASTICIDADE NA VARIÁVEL INDEPENDENTE PREÇO DE REPOSIÇÃO	86
FIGURA 27 – HETEROCEDASTICIDADE NA VARIÁVEL INDEPENDENTE PREÇO DE PAUTA	87
FIGURA 28 – HETEROCEDASTICIDADE NA VARIÁVEL INDEPENDENTE PREÇO MÍNIMO	87
FIGURA 29 – HETEROCEDASTICIDADE NA PLOTAGEM DOS RESÍDUOS PADRONIZADOS VERSUS VALORES PREVISTOS	88
FIGURA 30 – PLOTAGEM DA PROBABILIDADE NORMAL (FALTA DE NORMALIDADE)	88
FIGURA 31 – PREÇO DE REPOSIÇÃO APÓS TRANSFORMAÇÃO LOGARÍTMICA	89
FIGURA 32 – PREÇO DE PAUTA APÓS TRANSFORMAÇÃO LOGARÍTMICA.....	90
FIGURA 33 – PREÇO MÍNIMO APÓS TRANSFORMAÇÃO LOGARÍTMICA.....	90
FIGURA 34 – INDEPENDÊNCIA DOS RESÍDUOS NA PLOTAGEM DOS VALORES PREVISTOS VERSUS RESÍDUOS PADRONIZADOS.....	91
FIGURA 35 – GRÁFICO DE PROBABILIDADE NORMAL (NORMALIDADE COMPROVADA)	91
FIGURA 36 – RESÍDUOS DO LOG DE REPOSIÇÃO	93
FIGURA 37 – RESÍDUOS DO LOG DE PAUTA	94
FIGURA 38 – RESÍDUOS DO LOG DO PREÇO MÍNIMO	94
FIGURA 39 – PROBABILIDADE NORMAL DO LOG DO PREÇO DE VENDA	95
FIGURA 40 – VALORES PREVISTOS VERSUS RESÍDUOS PADRONIZADOS APÓS TRANSFORMAÇÃO LOG.....	95
FIGURA 41 – RESÍDUOS DO LOG DO PREÇO DE REPOSIÇÃO	97
FIGURA 42 – RESÍDUOS DO LOG DO PREÇO DE PAUTA	97
FIGURA 43 – RESÍDUOS DO LOG DO PREÇO MÍNIMO	98
FIGURA 44 – PROBABILIDADE NORMAL DO LOG DO PREÇO DE VENDA	98
FIGURA 45 – VALORES PREVISTOS VERSUS RESÍDUOS PADRONIZADOS APÓS TRANSFORMAÇÃO LOG.....	99
FIGURA 46 – DISTRIBUIÇÃO DE FREQUÊNCIA DOS CLIENTES DE ACORDO COM SUA CLASSIFICAÇÃO	101
FIGURA 47 – VALORES DE VENDA MATRIZ BLUMENAU (TENDÊNCIA LINEAR DECRESCENTE).....	106
FIGURA 48 – VALORES SEM O EFEITO DA SAZONALIDADE	107
FIGURA 49 – VARIAÇÕES CÍCLICAS MATRIZ BLUMENAU.....	108
FIGURA 50 – VARIAÇÕES CÍCLICAS/IRREGULARES APÓS APLICAÇÃO DAS MÉDIAS MÓVEIS DE 12 PERÍODOS POSTERIORMENTE CENTRADA	109
FIGURA 51 – VALORES PREVISTOS DE FATURAMENTO PARA MATRIZ BLUMENAU EM 2006.....	110
FIGURA 52 – TENDÊNCIA LINEAR CRESCENTE PARA VALORES DE VENDAS DA FILIAL RIO DO SUL	111

FIGURA 53 – MOVIMENTO DA SÉRIE SEM AS VARIAÇÕES SAZONAIS	112
FIGURA 54 – VARIAÇÕES CÍCLICAS/IRREGURARES	113
FIGURA 55 – VARIAÇÕES CÍCLICAS/IRREGULARES APÓS APLICAÇÃO DAS MÉDIAS MÓVEIS DE 12 PERÍODOS POSTERIORMENTE CENTRADA	114
FIGURA 56 – VALORES PREVISTOS PARA 2006 DA FILIAL RIO DO SUL	115
FIGURA 57 – TENDÊNCIA LINEAR CRESCENTE DOS VALORES DE FATURAMENTO DA FILIAL CHAPECÓ	116
FIGURA 58 – VALORES DE FATURAMENTO SEM AS VARIAÇÕES SAZONAIS.....	117
FIGURA 59 – VARIAÇÕES CÍCLICAS/IRREGULARES SOBRE VALORES DA FILIAL CHAPECÓ	117
FIGURA 60 – VARIAÇÕES CÍCLICAS/IRREGULARES APÓS APLICAÇÃO DAS MÉDIAS MÓVEIS	118
FIGURA 61 – VALORES PREVISTOS DE FATURAMENTO PARA FILIA CHAPECÓ EM 2006.....	119

LISTA DE TABELAS

TABELA 01 – FAIXAS PARA CLASSIFICAÇÃO DOS CLIENTES COM RELAÇÃO À QUALIDADE.....	77
TABELA 02 – $O R$ AJUSTADO.....	84
TABELA 03 – VERIFICAÇÃO DO VALOR P	84
TABELA 04 – r AJUSTADO E VALOR P	85
TABELA 05 – SOMA QUADRÁTICA DOS ERROS E MÉDIA QUADRÁTICA DOS ERROS.....	85
TABELA 06 – QUADRADO MÉDIO.....	89
TABELA 07 – RESULTADOS DAS REGRESSÕES NA SEGUNDA ANÁLISE.....	92
TABELA 08 – RESULTADO DAS REGRESSÕES NA TERCEIRA ANÁLISE.....	96
TABELA 09 – DISTRIBUIÇÃO DE FREQUÊNCIA DOS CLIENTES.....	100
TABELA 10 – TABELA DINÂMICA REFERENTE AO VALOR TOTAL COMPRADO (%).....	102
TABELA 11 – MARGEM MÉDIA DE VENDAS (%).....	103
TABELA 12 – PRAZO MÉDIO (DIAS).....	104
TABELA 13 – ÍNDICES SAZONAIS MATRIZ BLUMENAU.....	107
TABELA 14 – VALORES PREVISTOS DE FATURAMENTO PARA 2006.....	110
TABELA 15 – ÍNDICES SAZONAIS ENCONTRADOS PARA FILIAL RIO DO SUL.....	112
TABELA 16 – VALORES PREVISTOS PARA 2006 DA FILIAL RIO DO SUL.....	114
TABELA 17 – ÍNDICES SAZONAIS ENCONTRADOS PARA FILIAL CHAPECÓ.....	116
TABELA 18 – VALORES PREVISTOS DE FATURAMENTO PARA FILIAL CHAPECÓ.....	118

SUMÁRIO

1	INTRODUÇÃO.....	12
1.1	MOTIVAÇÃO E JUSTIFICATIVA.....	12
1.2	OBJETIVO GERAL.....	17
1.3	OBJETIVOS ESPECÍFICOS.....	17
1.4	ORGANIZAÇÃO DO TRABALHO.....	18
2	FUNDAMENTAÇÃO TEÓRICA.....	20
2.1	SISTEMA DE INFORMAÇÃO GERENCIAL.....	21
2.2	METODOLOGIA SIEGO.....	23
2.2.1	<i>SIEGO versus Seis Sigma</i>	26
2.3	DATA WAREHOUSE.....	26
2.3.1	<i>Data warehouse – uma introdução</i>	27
2.4	REDES NEURAIS ARTIFICIAIS.....	29
2.5	MÉTODOS ESTATÍSTICOS.....	32
2.5.1	<i>Estatística e método estatístico</i>	32
2.5.2	<i>Séries temporais</i>	34
2.5.2.1	Fatores de influência em Séries Temporais.....	35
2.5.2.2	Médias Móveis e Ajuste Exponencial.....	38
2.5.2.3	Mínimos Quadrados.....	40
2.5.2.4	Modelagem Auto-Regressiva.....	40
2.5.3	<i>Análise de regressão e correlação</i>	41
2.5.3.1	Diagrama de Dispersão.....	42
2.5.3.2	Regressão linear simples.....	42
2.5.3.2.1	Análise de resíduos.....	45
2.5.3.3	Regressão linear múltipla.....	47
3	MATERIAL E MÉTODOS.....	49
3.1	A POPULAÇÃO.....	50
3.2	AS VARIÁVEIS.....	50
3.3	METODOLOGIAS UTILIZADAS.....	50
3.3.1	<i>Associação das abordagens</i>	50
3.3.2	<i>Fases da metodologia SIEGO</i>	51
3.3.2.1	Fase I - preparação do projeto SIEGO.....	51
3.3.2.1.1	Passo 1 - planejamento do grupo de trabalho.....	51
3.3.2.1.2	Passo 2 - definição dos processos.....	51
3.3.2.1.3	Passo 3 - motivação e instrução ao grupo de trabalho.....	52
3.3.2.1.4	Passo 4 - planejamento da implementação das idéias.....	52
3.3.2.1.5	Passo 5 - preparação para acompanhamento das idéias.....	52
3.3.2.2	Fase II - determinação e avaliação das ações de melhorias.....	52
3.3.2.2.1	Passo 1 - montagem do banco de dados.....	52
3.3.2.2.2	Passo 2 - desenvolvimento e avaliação das idéias de melhorias.....	53
3.3.2.2.3	Passo 3 - seleção das idéias em potencial.....	53
3.3.2.3	Fase III - implementação das idéias.....	54
3.3.2.3.1	Passo 1 - planejamento da implantação.....	54
3.3.2.3.2	Passo 2 - implementação e rastreamento das ações de melhorias.....	54
3.3.3	<i>A construção de um Data Warehouse</i>	55
3.3.3.1	As etapas propostas por Kimball.....	55
3.3.4	<i>Utilizando as séries temporais</i>	58
3.3.5	<i>Utilizando análise de regressão</i>	59
3.3.6	<i>Distribuição de Freqüências</i>	59
3.4	TECNOLOGIAS E FERRAMENTAS UTILIZADAS.....	60
3.4.1	<i>Statistica</i>	60
3.4.2	<i>Microsoft Excel</i>	61
3.4.3	<i>Banco de dados firebird</i>	62
3.4.4	<i>Dataflex</i>	63
3.4.5	<i>Ambiente Delphi</i>	64
4	RESULTADOS.....	67
4.1	O AMBIENTE DE PESQUISA.....	67

4.1.1	<i>Apresentação da empresa</i>	67
4.2	APLICAÇÃO DA METODOLOGIA SIEGO	71
4.3	APLICAÇÃO DO DATA WAREHOUSE	73
4.4	APRESENTAÇÃO DO SISTEMA	75
4.5	APLICAÇÃO DOS MÉTODOS ESTATÍSTICOS	83
4.5.1	<i>Aplicação da análise de regressão</i>	83
4.5.2	<i>Aplicação da distribuição de frequências</i>	100
4.5.3	<i>Aplicação das séries temporais</i>	105
4.6	DISCUSSÃO.....	120
5	CONCLUSÕES	122
5.1	CONSIDERAÇÕES FINAIS	122
5.2	LIMITAÇÕES DO TRABALHO.....	124
5.3	SUGESTÕES PARA TRABALHOS FUTUROS	124
	REFERÊNCIAS BIBLIOGRÁFICAS	126

1 INTRODUÇÃO

O capítulo introdutório deste trabalho de pesquisa visa situar o leitor com relação à motivação para o trabalho, quanto à justificativa para esta pesquisa e também quanto aos objetivos que se pretende atingir com este estudo.

1.1 MOTIVAÇÃO E JUSTIFICATIVA

O ser humano tem por natureza a capacidade e necessidade de evoluir à medida que suas ambições vão aumentando. Em virtude desta constante e crescente evolução, as organizações das quais as pessoas fazem parte também tem por consequência uma evolução, que hoje já não é mais uma opção, mas sim uma necessidade. A velocidade dos avanços tecnológicos descrita por Oliveira (2002) confirma esta necessidade.

Por volta de 1970, as empresas estavam lutando para conseguir implantar e gerar informações em computadores. Hoje isso já é realidade e as empresas a cada dia mais tem necessidade de utilizar estas informações como parte de seu patrimônio e sua estratégia de sobrevivência no mercado. Assim, Oliveira (2002) afirma que as empresas que estiverem atentas para esta situação, seguramente, já consolidarão uma vantagem competitiva interessante.

De acordo com Cautela (1996), Sistema é um conjunto de elementos independentes que atuam em conjunto para atingir um objetivo comum. Oliveira (2002) complementa a definição de Cautela dizendo que um Sistema é composto por várias partes que estão interagindo mas que mantém uma independência de modo que atuando de maneira conjunta visam atingir objetivos que acabam resultando nas execuções de diversas atividades.

A definição exposta por Dalfovo (2001) coloca Sistemas de Informação como sendo elementos agrupados que interagem através da coleta (que seria a entrada) de dados, de sua manipulação (que seria o processamento) e através de sua disseminação (que é a saída), podendo gerar posteriormente um mecanismo de *feedback*, que é o impacto gerado por este processo. Para Oliveira (1998), o objetivo dos Sistemas de Informação está focado no suporte ao executivo, de modo que o mesmo esteja sempre preparado, através de informações, para atuar em todos os pontos que precisam de melhorias em sua organização.

Baseado nestas afirmações coloca-se que a utilização de Sistemas de Informação pode trazer ao executivo dados importantes e até imprescindíveis em seu processo decisório diante das situações adversas que o mercado oferece. Sendo assim, para uma organização obter um bom desempenho e poder estar à frente no mercado ela precisa de informações precisas, claras, concisas e também que estas estejam disponíveis de forma rápida. Rezende (2003) diz que os benefícios trazidos pelos sistemas de informação são inúmeros, mas cita principalmente que estas informações personalizadas e oportunas facilitam o gerenciamento das organizações. Baseado neste oportunismo e personificação dos sistemas citado por Rezende (2003) nota-se que os executivos precisam algo mais que apenas as informações agrupadas ou disponíveis, precisam de informações que de alguma forma estejam correlacionadas e resultem em dados úteis para que se possa traçar uma estratégia ou tomar uma decisão. Colocação esta que motiva o estudo nesta área.

A crescente evolução tecnológica das empresas e a aquisição de equipamentos a cada mês que passa não significam que estas empresas estão utilizando as informações que tem disponíveis da melhor maneira possível. É preciso fazer com que as informações críticas cheguem às mãos de quem decide, é preciso deixar claro que Sistemas de Informações são mais do que apenas computadores. A questão é que muitas destas informações estão armazenadas em computadores e bancos de dados que em muitos casos não se comunicam e oferecem respostas apenas a questões de fácil entendimento, não abrangendo a complexidade das informações que poderiam ser obtidas. Dalfovo (2000) afirma que a não utilização das informações como recursos estratégicos, leva o executivo, muitas vezes, a administrar por impulsos, ou baseado em modismos.

De acordo com Oliveira (2002) toda empresa possui informações que proporcionam a sustentação para as suas decisões. Entretanto, apenas algumas têm um sistema estruturado de informações estratégicas que possibilita otimizar o seu processo decisório. Portanto, para administrar esta gama de informações contida nas empresas e se extrair daí dados que dêem sustentação para os executivos tomarem decisões existe metodologias estratégicas, sendo que dentre elas, será abordada neste trabalho a metodologia SIEGO que através de suas próprias contribuições servirá como parte deste para abordagem dos objetivos que serão apresentados.

De acordo com Dalfovo (2001) criador da metodologia, “SIEGO (Sistema de Informação Estratégico para o Gerenciamento Operacional) é o processo de transformação de dados em informações que são utilizadas na estrutura decisória da organização, e que proporcionam a sustentação administrativa para otimizar os resultados esperados.” “A metodologia SIEGO deve ser muito bem desenvolvida, implementada, gerenciada e ter uma efetiva colaboração na adequação das organizações perante os pontos inerentes a um cenário provável para a economia globalizada.”

Esta metodologia visa um trabalho onde ocorre o envolvimento de toda a organização. Envolvendo desde o nível estratégico da empresa (diretoria) até o nível operacional (gerentes de operações e usuários de produção), cujo objetivo principal é alcançar em um espaço curto de tempo, melhorias em relação ao tripé (custo, tempo e qualidade), substanciais e sujeitas às regras e limites de investimentos bem definidos.

Estando claro que um Sistema de Informações oferece diferenciais para os tomadores de decisão da empresa, é necessário também apresentar uma tecnologia que disponibiliza a estes executivos uma forma fácil, flexível, efetiva e eficiente de utilizar estas informações através da análise dos dados. Esta tecnologia é denominada *Data Warehouse*. Segundo Inmon (1997), um *Data Warehouse* (que pode ser traduzido como armazém de dados) é um banco de dados que armazena dados resultantes das operações da empresa (vendas, compras, etc), extraídos de uma fonte única ou múltipla, e transforma-os em informações úteis, táticas, estratégicas, oferecendo um enfoque histórico, para permitir um suporte efetivo à decisão. Uma outra definição para *Data Warehouse* pode ser obtida através de Groth (1997) que afirma que a tecnologia em torno de um *Data Warehouse* envolve recursos de inteligência artificial e análises estatísticas, busca correlações de dados dentro de um ambiente de Banco de Dados, fornecendo informações consideradas relevantes para o negócio. Assim este conceito provê aos executivos da organização uma forma de buscar informações e ter acesso a dados que antes estavam em fontes não correlacionadas, buscando então respostas para as mais diversas questões de negócios.

Inmon (1997 p.45) apresenta a granularidade como sendo o ponto mais importante em uma projeto de *Data Warehouse*. A granularidade está voltada para o nível de detalhamento que os dados contidos dentro do *Data Warehouse* terão.

Inmon (1997 p.45) coloca que a razão para esta importância da granularidade está relacionada com o fato de que o nível de granularidade vai afetar o volume de dados

residentes no *Data Warehouse*, pois a medida que a quantidade de detalhes aumenta, mais baixo será o nível de granularidade. Além disso, as consultas que poderão ser realizadas através do *Data Warehouse* também vão variar de acordo com a granularidade.

Outro aspecto a se considerar também em um Sistema de Informação é o Cubo de Decisão. Segundo Inmon (1997), o Cubo de Decisão (Decision Cube) refere-se a um conjunto de componentes de suporte a decisões, que podem ser utilizados para cruzar tabelas de um banco de dados, gerando visões através de planilhas ou gráficos. Envolve o cálculo, quando da carga do *Data Warehouse*, de dados que o usuário virá a solicitar, mas que podem ser derivados de outros dados. Quando o usuário solicita os dados, estes já estão calculados, agregados em um Cubo de Decisão.

Aproveitando a citação anterior de Groth (1997), apresenta-se aqui justamente o fator relacionado com métodos estatísticos, uma vez que estando o *Data Warehouse* construído, baseado na metodologia SIEGO, é importante que seja possível visualizar os dados sob uma análise estatística e que possibilite ao executivo uma visão diferenciada dos dados antes apresentados de forma isolada.

Baseado no que foi exposto anteriormente é possível notar que o executivo de uma organização precisa algo além das informações previamente separadas (mineradas). Precisa de ferramentas que possibilitem enxergar algo além das fronteiras das palavras e que não é possível através de sua simples apresentação. Aí então a inclusão de métodos estatísticos como parte deste trabalho faz com que estes dados tenham um significado mais claro para o executivo, uma vez que o mesmo terá um embasamento científico para tomar suas decisões. Os métodos estatísticos permitem uma visão mais precisa e direcionada das informações apresentadas pela empresa.

Segundo Kume (1993), os métodos estatísticos são ferramentas eficazes para a melhoria dos processos, entretanto, é preciso que se tenha em mente que as ferramentas estatísticas são apenas ferramentas: elas podem não funcionar, caso sejam aplicadas inadequadamente.

Em muitos casos os executivos tentam reduzir os problemas da organização rastreando a causa do defeito, porém isso muitas vezes ocorre de maneira empírica ou baseada em suposições ou possíveis experiências pessoais anteriores o que torna possível a abordagem de causas que não são verdadeiras, o que pode resultar em tentativas mal sucedidas e esforços desperdiçados. Em virtude disso Kume (1993) aborda que o primeiro passo na busca da verdadeira causa é a cuidadosa observação do

fenômeno do problema. Após tal observação atenciosa, a verdadeira causa torna-se mais clara se não evidente.

As ferramentas estatísticas conferem objetividade e exatidão à observação. Sendo assim Kume (1993) cita as máximas da forma estatística de pensar:

- Dar maior importância aos fatos do que aos conceitos abstratos;
- Não expressar fatos em termos de intuição ou idéias. Usar evidências obtidas a partir de resultados específicos da observação;
- Os resultados da observação sujeitos como são a erros e variações, são parte de um todo obscuro. A principal meta da observação é descobrir este todo obscuro.
- Aceitar o padrão regular que aparece em grande parte dos resultados observados como uma informação confiável.

Observando as máximas da forma estatística de pensar citadas por Kume (1993), e as questões citadas sobre sistemas de informação, é possível notar que em muitos casos o executivo realmente toma como base conhecimentos que são frutos de experiências passadas ou intuições e não resultados específicos de observações ou análises estatísticas.

Segundo H. G. Wells citado por Levine et al. (2000) “Raciocinar estatisticamente, será um dia tão necessário quanto à habilidade de ler e escrever”. Com base nesta citação, tem-se como objetivo macro, justamente realizar a integração da metodologia SIEGO aos Métodos Estatísticos (Séries Temporais, Análise de Regressão e Distribuição de Frequência) para que os resultados apresentados aos executivos possam estar embasados nas análises feitas através destes métodos.

E é então baseado neste pensamento que a proposta da utilização dos métodos estatísticos associados à metodologia SIEGO tem por objetivo auxiliar o executivo a visualizar os reais problemas existentes através de análises realizadas com o auxílio dos métodos estatísticos.

Assim, justifica-se o diferencial deste trabalho que é justamente eliminar possíveis decisões subjetivas e apresentar a possibilidade da tomada de decisão baseada em dados resultantes de análises onde foram utilizados os métodos estatísticos.

1.2 OBJETIVO GERAL

Com a integração citada objetiva-se que os usuários da informação possam obter bons resultados com relação à aplicação da metodologia SIEGO e conseqüentes melhorias no processo de tomada de decisão.

A aplicação deste refere-se a uma empresa do ramo eletromecânico da cidade de Blumenau/SC, sendo que a aplicação dos métodos utilizados neste trabalho pode ser estendida às organizações de um modo geral, respeitando obviamente as adaptações necessárias as realidades de cada organização, servindo esta, então, como ponto de referência.

Neste trabalho serão analisados dados referentes aos períodos de agosto de 1994 até junho de 2005, onde os mesmos abordarão questões relacionadas às vendas realizadas no período, objetivando a realização de previsões futuras, clientes da empresa, objetivando a análise da qualidade dos seus clientes e também a verificação da questão inadimplência sofrida pela empresa em questão.

É possível mencionar trabalhos como Schmitt (2001), Strey (2002) e Ghoddosi (2003), onde foram abordados os assuntos sobre Sistemas de Informação, *Data Warehouse* e SIEGO, porém em nenhum deles foi aplicada associação com métodos estatísticos, ficando então a tomada de decisão por parte do executivo baseada em sua experiência profissional (teoricamente empírica).

1.3 OBJETIVOS ESPECÍFICOS

O presente trabalho tem como objetivos específicos prover recursos para que o executivo possa aperfeiçoar sua análise em relação aos dados minerados e apresentados. Para isso então, este vem demonstrar um estudo que envolve os seguintes aspectos:

- Realizar uma pesquisa para uma fundamentação teórica e prática sobre o assunto;
 - Tal pesquisa visa levantar dados a cerca do que já foi escrito sobre o tema e direcionar a pesquisa visando fechar as possíveis lacunas que ficaram abertas em trabalhos anteriores, bem como preparar o autor deste para aplicação das técnicas estudadas;
- Realizar análises utilizando Séries Temporais, Análise de Regressão e Distribuições de Frequências;

- A aplicação de tais técnicas visa demonstrar a importância do uso destes métodos dentro da proposta de integração com a metodologia SIEGO;
- Verificar quais as contribuições obtidas pelo uso dos métodos estatísticos aplicados ao tripé da metodologia SIEGO;
 - Este item visa apresentar, após a análise anterior, quais os resultados de da aplicação dos métodos e onde isso poderá auxiliar o executivo, dentro da visão da metodologia SIEGO que observa melhorias em relação ao tripé tempo, custo e qualidade;
- Validar a integração dos métodos estatísticos com a metodologia SIEGO através da aplicação em uma organização com uma base de dados real;
 - Este item se refere à aplicação desta pesquisa à uma empresa do ramo eletromecânico que possui uma base de dados reais, cuja estrutura já é conhecida pelo autor da pesquisa, facilitando a implementação desta integração, bem como a obtenção de resultados que poderão ser aplicados na prática;
- Analisar os resultados.
 - Ao final da pesquisa, pretende-se analisar todos os resultados obtidos para que se possa fornecer um parecer sobre a pesquisa, demonstrando quais os pontos positivos de sua aplicação e os possíveis pontos de melhorias para futuras pesquisas.

1.4 ORGANIZAÇÃO DO TRABALHO

Para apresentar os conteúdos anteriormente mencionados, a presente dissertação é direcionada dentro da estrutura como segue:

Neste capítulo, o capítulo 1, apresenta-se uma introdução ao trabalho, abordando quais os assuntos expostos, bem como a descrição do objetivo macro deste e também os objetivos específicos que direcionam os demais capítulos.

O capítulo 2 aborda a questão da pesquisa para uma fundamentação teórica e prática a respeito do assunto apresentado no capítulo anterior. Pretende-se com este deixar o leitor ciente de um histórico sobre o assunto e sua evolução através das visões de diversos autores que foram considerados de fundamental relevância para compreensão e a evolução da pesquisa.

No capítulo 3 são apresentados de forma clara e objetiva, quais foram e como foram empregados os métodos, materiais e equipamentos de forma a tornar esta dissertação um modelo para o emprego de tais recursos em outras organizações, afim de, obterem-se resultados equivalentes para objetivos também equivalentes.

A abordagem do capítulo 4 refere-se à aplicação propriamente dita dos métodos aos dados da organização escolhida e à apresentação clara e objetiva dos resultados obtidos através da aplicação de tais métodos/metodologias.

Após a apresentação dos resultados, passa-se então para a uma análise em forma de discussão, no capítulo 5, onde se pretende questionar estes resultados, bem como apresentar conjecturas paralelas e algumas questões para as quais o trabalho não apresentou respostas e que necessitam de um estudo mais aprofundado sobre tal. Por último, tendo em vista todo o processo já concluído e discussões encerradas, o capítulo 6 apresenta as conclusões que são devidamente fundamentadas nos resultados e também na discussão já apresentada. Conclusões estas que se remetem aos objetivos propostos para esta dissertação.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo fazer um levantamento sobre os assuntos abordados durante a pesquisa e apresenta-los segundo citações de autores das respectivas áreas.

Os tópicos deste capítulo estão divididos da seguinte forma:

- Sistema de Informação Gerencial
 - Neste tópico, pretende-se apresentar os conceitos básicos de sistemas de informação, demonstrando sua importância e quais as suas relações com as empresas no enfoque competitivo;
- Metodologia SIEGO
 - Através deste tópico, tem-se o objetivo de apresentar, de modo breve, ao leitor o que é tal metodologia, bem como sua estrutura e seu objetivo principal;
 - Também se apresenta neste tópico as principais diferenças entre a metodologia SIEGO e a metodologia SEIS SIGMA, esclarecendo o leitor de que não seguem o mesmo caminho e tem propósitos semelhantes, mas não iguais;
- *Data Warehouse* – uma introdução
 - Como o próprio nome do tópico já apresenta, aqui o leitor deve ter uma breve apresentação sobre o que é *Data Warehouse*, bem como a estrutura para seu desenvolvimento de acordo com um autor consagrado da área;
- Redes Neurais
 - Este tópico tem a função de posicionar o leitor com relação à existência de tal técnica e também de apresentar o porquê de não ter sido utilizada neste trabalho, demonstrando que é importante, porém neste momento não foi empregada;
- Métodos Estatísticos
 - O assunto relatado neste tópico visa apresentar os métodos de séries temporais, análise de regressão e distribuição de frequência, através de conteúdos apresentados por outros trabalhos e por autores que são referência na área. Através deste tem-se uma

compreensão melhor dos métodos, contribuindo assim para um entendimento do todo deste trabalho.

Conforme observado, através dos tópicos abordados por este capítulo, tem-se um posicionamento quanto aos assuntos abordados, situando-se e fornecendo um embasamento para que nos tópicos mais a frente, possam ser compreendidas as aplicações de tais metodologias e técnicas apresentadas.

2.1 SISTEMA DE INFORMAÇÃO GERENCIAL

De acordo com O'brien (2003), sistemas de informação constituem-se da reunião de partes ou componentes interdependentes, que interagem entre si formando um todo e que trabalham em função de um objetivo comum.

De acordo com Oliveira (2002) define-se Sistema de Informação Gerencial (SIG) como sendo o processamento de dados em informações, sendo que estas informações são utilizadas para tomada de decisões que estão focadas em objetivos específicos que darão uma sustentação administrativa para a empresa de modo que a mesma se mantenha competitiva e obtendo os resultados esperados. Esse sistema é normalmente composto de diversos subsistemas de natureza conceitual idêntica à daquele que integram, mas com características específicas quanto à sua finalidade e justificação, quanto ao tipo de tecnologias utilizadas e quanto ao nível dos processos ou natureza das pessoas que envolvem.

A designação Sistema de Informação Gerencial (SIG) é indistintamente utilizada para referir cada um dos diferentes subsistemas de informação. Subsistemas de informações envolvem a utilização de computadores e correspondem à sua definição, também correntemente designados por “Sistema de Informação Baseado em Computadores”, ou simplesmente aplicações. O ambiente de um sistema de informação é exemplificado na figura 1 a seguir, onde é possível observar as diversas relações com as quais a empresa/organização obtém ou trocam informações, acrescentando assim seu volume de dados e conseqüente recurso para utilização estratégica futura.

A relação com a concorrência, por exemplo, pode acrescentar informações sobre preços praticados e mercados atingidos. Relações com o governo e sistema de

obrigações fiscais podem gerar informações de custos para a empresa bem como facilidades oferecidas pelo governo. Assim, todas as demais relações oferecem formas de captação de dados que poderão oferecer à empresa uma fonte a mais de informações para suas decisões estratégicas.

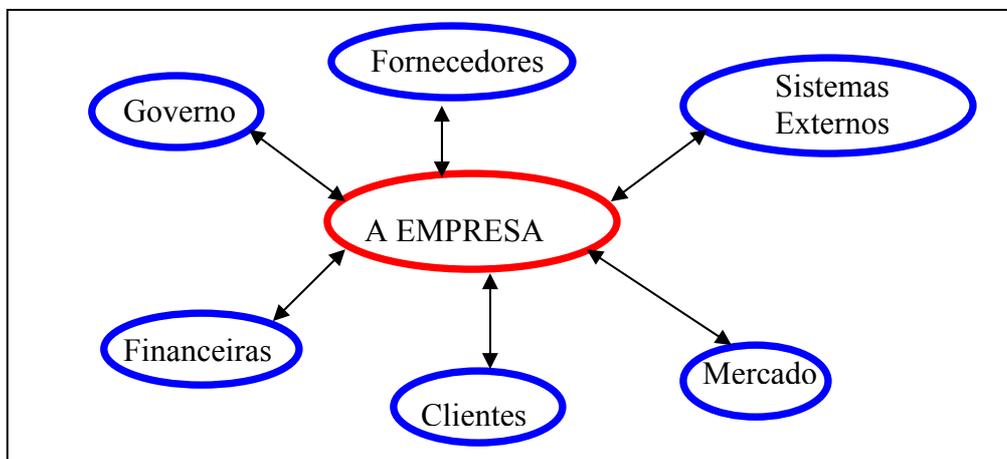


Figura 1 – Ambiente de um sistema empresarial

Fonte: Adaptado de Oliveira (2002) p. 25

Remetendo novamente a definição de Cautela (1996), tem-se que Sistema é um conjunto de elementos independentes que atuam em conjunto para atingir um objetivo comum. Oliveira (2002) complementa a definição de Cautela dizendo que um Sistema é composto por várias partes que estão interagindo mas que mantêm uma independência de modo que atuando de maneira conjunta visam atingir objetivos que acabam resultando nas execuções de diversas atividades.

A definição exposta por Dalfovo (2001) coloca Sistemas de Informação como sendo elementos agrupados que interagem através da coleta (que seria a entrada) de dados, de sua manipulação (que seria o processamento) e através de sua disseminação (que é a saída), podendo gerar posteriormente um mecanismo de *feedback*. O *feedback* é a saída que é usada para fazer ajustes ou modificações nas atividades de entrada ou processamento.

Dentro desta abordagem de sistemas de informação é possível diferenciar os sistemas de informação de acordo com Laudon (2004) em pelo menos quatros níveis onde os mesmo atuam, sendo eles:

- nível estratégico, servido aos gerentes seniores/executivos/diretoria;
- nível gerencial, servindo aos gerentes médios/apoio a decisão;
- nível de conhecimento, servindo aos trabalhadores do conhecimento;

- nível operacional, servindo aos trabalhadores direcionados a serviços operacionais.

Tendo em vista esta colocação, uma das metodologias importantes na implementação de um sistema de informação gerencial é a metodologia SIEGO (Sistema de Informação Gerencial para o Gerenciamento Operacional), que será apresentada a seguir.

2.2 METODOLOGIA SIEGO

Dalfovo (2001), criador da metodologia SIEGO, apresenta a metodologia utilizando um conceito simples de sistema de informação onde os dados são transformados de modo a constituir informações, que por sua vez irão dar embasamento para que os executivos da empresa possam tomar suas decisões. Ele coloca como diferencial o fato de que esta metodologia trabalha em cima das metas que a organização precisa atingir de modo a fornecer então ao executivo uma estrutura que demonstre quais as operações podem ser melhoradas ou quais processos podem ser melhorados de modo a produzir um estado atual melhor que o anterior onde a empresa se encontrava.

Para conseguir o que é proposto, Dalfovo (2001) diz que esta metodologia deve envolver a organização como um todo visando então melhorias com relação à estrutura citada anteriormente, estrutura esta que envolve três pontos fundamentais (citado por Dalfovo (2001) como tripé) que são o custo, o tempo e a qualidade.

Assim tem-se algo próximo ao que se vê na figura 2.

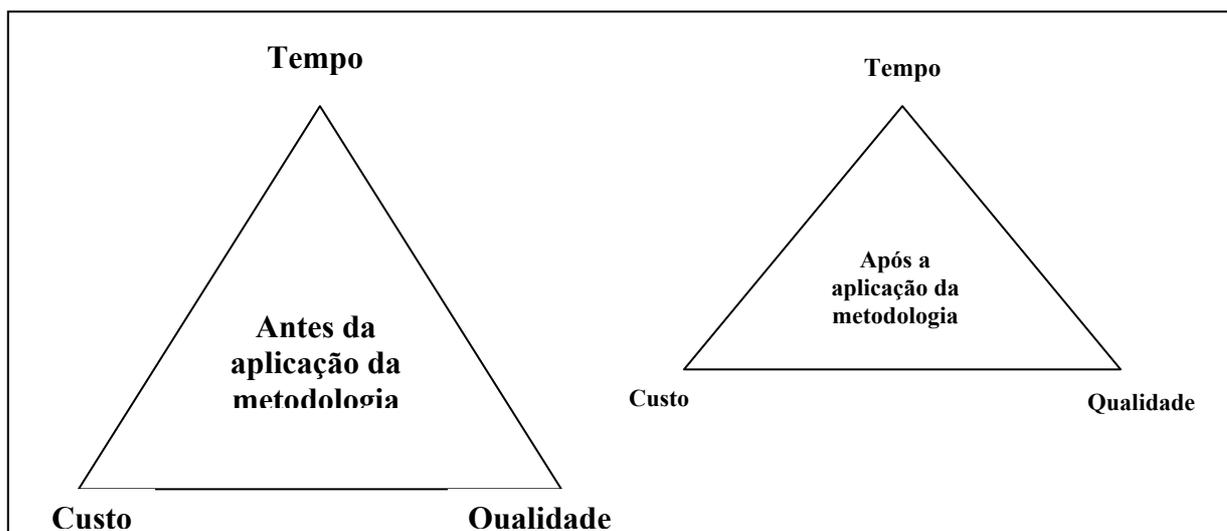


Figura 2 – Tripé da organização – Sugestão da metodologia

Fonte: Adaptado de Dalfovo (2001) p. 93

Observando a figura 2, é possível notar o foco da metodologia, que conforme mencionado anteriormente, objetiva otimizar processos relacionados ao tripé custo, tempo e qualidade, onde se tem um estágio atual e após a aplicação da metodologia pretende-se chegar à um estágio otimizado dos processos visando melhorias.

O fluxograma apresentado através da figura 3, na página a seguir, mostra de forma simplificada quais são as fases e passos que envolvem a metodologia SIEGO e que serão abordados de maneira mais detalhada no capítulo seguinte onde são abordados os métodos utilizados neste trabalho.

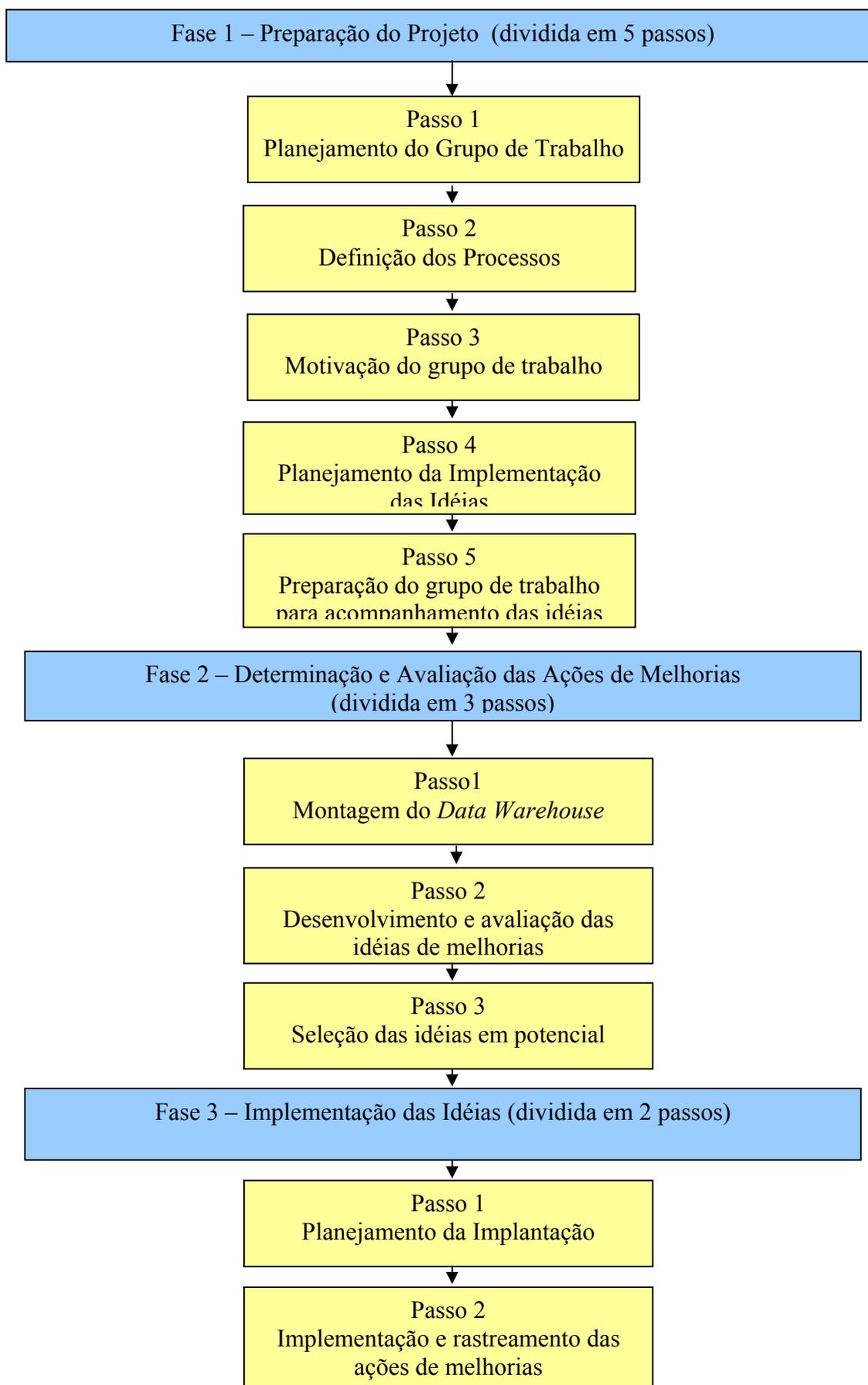


Figura 3 – Fluxograma representativo das fases e passos da metodologia SIEGO
Fonte: Adaptado de Dalfovo (2001) p. 99

2.2.1 SIEGO versus Seis Sigma

Ao abordar questões relativas à melhoria da qualidade, processos e outras melhorias referentes à organização, surgem outras questões relacionadas às metodologias e teorias a serem seguidas para implementação destes processos de melhorias. Assim, afim de esclarecer o leitor, este tópico pretende deixar clara a diferença principal entre a metodologia SIEGO e a filosofia Seis Sigma, demonstrando que não são a mesma coisa, mas possuem focos semelhantes.

Como observado no tópico anterior, SIEGO é uma metodologia que propõe melhorias relacionadas ao tripé constituído dos parâmetros Tempo, Custo e Qualidade, auxiliando assim no processo decisório da empresa. Para tal, a metodologia SIEGO se utiliza de várias fases e passos, justificando assim sua classificação como uma metodologia. Além disso, sua delimitação em cima dos três parâmetros relatados anteriormente também é um diferencial.

Segundo Wilson (2000), Seis Sigma é uma filosofia onde a melhoria dos processos é perpétua e existe uma busca pela redução de sua variabilidade na busca interminável de zero defeito. Wilson (2000) também afirma que o Seis Sigma pode significar várias coisas e pode ser utilizado de diversas maneiras, sendo elas por exemplo: como *Benchmark*, como Meta, como Medida, como Estatística, como Estratégia, como Valor e também como Visão, onde a mesma aponta para um horizonte onde a empresa seja a melhor do seu ramo. Wilson (2000) ainda frisa que apesar de o Seis Sigma ser tantas coisas, é importante saber o que o Seis Sigma não é, e isto o autor deixa bem claro afirmando que o Seis Sigma não é uma metodologia, ele é um fim, não um meio, o que já o diferencia da Metodologia SIEGO. Uma definição interessante apontada por Wilson (2000) é descrever o Seis Sigma como um nível otimizado de desempenho que se aproxima de um nível de 0 (zero) falhas em um processo de produção de algum produto, serviço ou transação.

2.3 DATA WAREHOUSE

Inmon (1997 p.33) conceitua *Data Warehouse* como sendo “um conjunto de dados baseados em assuntos, integrado, não-volátil, e variável em relação ao tempo, de apoio às decisões gerenciais.” Entre os pontos mais importantes a serem considerados, encontra-se a análise de consultas (visando evitar sobrecargas no sistema e definindo

prioridades), a identificação de perfis e o dimensionamento adequado, tanto de granularidade quanto de particionamento de dados.

Em um *Data Warehouse* (DW) o ponto crítico não é a resposta adequada ao grande número de transações, mas, a resposta adequada a consultas efetuadas pelos usuários e ao crescimento da base de dados.

Segundo Oliveira (1997) citado por Dalfovo(2001) “só as mais simples organizações não possuem uma Tecnologia de Gerenciamento da Informação e sua principal ferramenta para organizar as informações é o Banco de Dados”.

2.3.1 Data warehouse – uma introdução

Inmon (1997 p.68) afirma que “o *Data Warehouse* é o alicerce do processamento dos SADS”. O *Data Warehouse*, segundo Inmon (1997 p.33) é um Banco de Dados com base em assuntos (dados referentes a vendas, compras, financeiro, por exemplo), não-volátil (não está em atualização *on-line*) e variável com relação ao tempo.

Kimball (2002) coloca que um dos requisitos para um *Data Warehouse* de sucesso é que a comunidade de negócio deve aceita-lo, ou seja, a diretoria da empresa deve aprovar as informações contidas no nele.

Uma outra definição colocada por Kimball (2002), diz que o *Data Warehouse* deve funcionar como a base para uma melhor tomada de decisões.

Kimball (1996) e Kimball (1997) apresentam nove etapas de decisão para um projeto de banco de dados para o desenvolvimento de um *Data Warehouse*, então, através da figura 4, é apresentado um fluxograma simples para visualização destas nove etapas.

O capítulo três aborda de forma mais detalhada estas nove etapas.

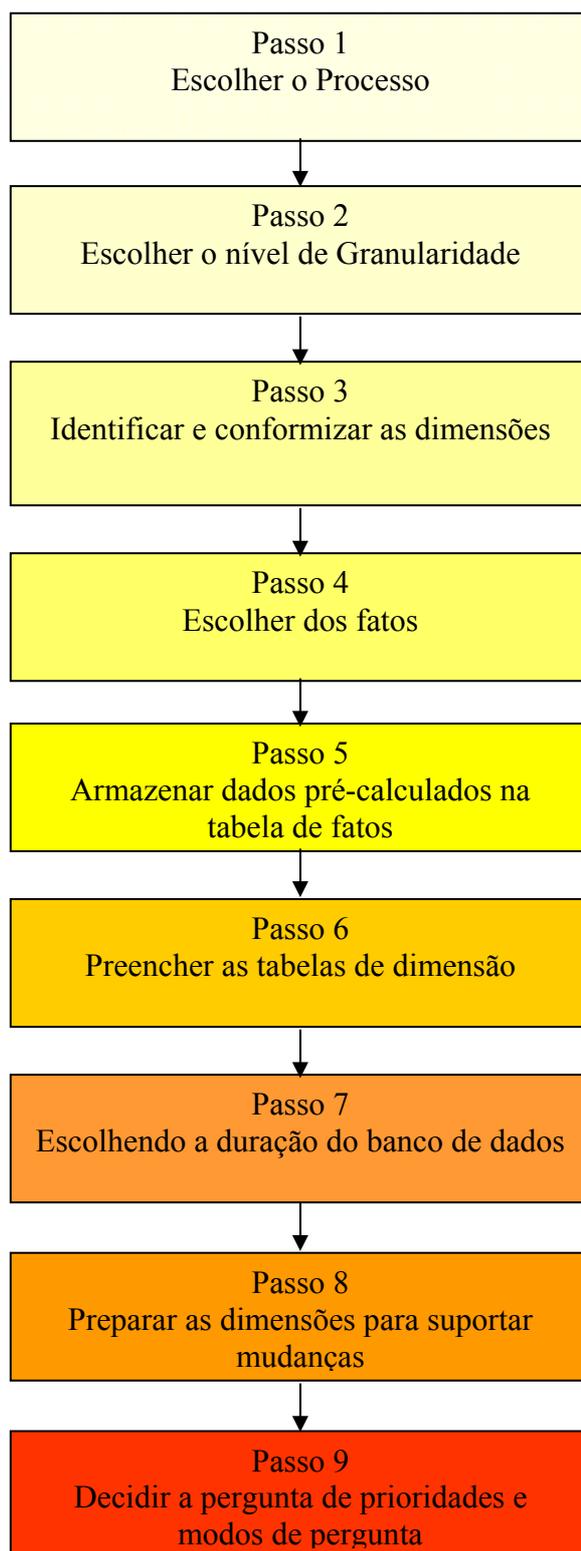


Figura 4 - Fluxograma referente aos passos para construção de um Data Warehouse

Fonte: Adaptado de Kimball (1996) e Kimball (1997)

2.4 REDES NEURAIS ARTIFICIAIS

Nascida de pesquisas que tentavam simular o funcionamento do cérebro humano, as Redes Neurais Artificiais (RNAs) estão sendo utilizadas com maior volume a partir da década de 80 e segundo Gonçalves (2000) o emprego das RNAs é bastante utilizado na extração de conhecimentos a partir de bases de dados.

Ainda segundo Gonçalves (2000) e também Santos (2005) tem-se que as RNAs podem ser aplicadas a um grupo bastante grande de situações, sendo que pode-se citar: o reconhecimento de padrões, classificação, previsão, controle, entre outros. Schenatz (2005) coloca que as RNAs são capazes de aprender sozinhas, por meio de treinamentos apropriados. Santos (2005) faz uma colocação importante, onde ressalta que uma desvantagem da utilização de RNAs é que os meios para se obter um resultado não se apresentam através de uma interpretação clara, ou seja, exige uma interpretação adicional para saber como a informação/resultado foi extraído não permitindo assim um entendimento imediato das razões que estão por trás de um resultado. Schenatz (2005) ainda coloca que justamente esta falta de explicações acaba por inibir a confiança, aceitação e aplicação dos resultados. Entretanto, Santos (2005) também coloca que uma vantagem em potencial da utilização de RNAs é que através de sua utilização é possível detectar qualquer relação não-linear entre a variável resposta e as variáveis explicativas.

A idéia básica de RNAs é construir um modelo constituído por um grande número de unidades de processamento (como se fossem os neurônios dos seres humanos) que são chamadas de neurônios de forma que estes sejam interconectados entre si e as informações sejam trocadas entres os mesmos. A figura 5 a seguir representa esta idéia.

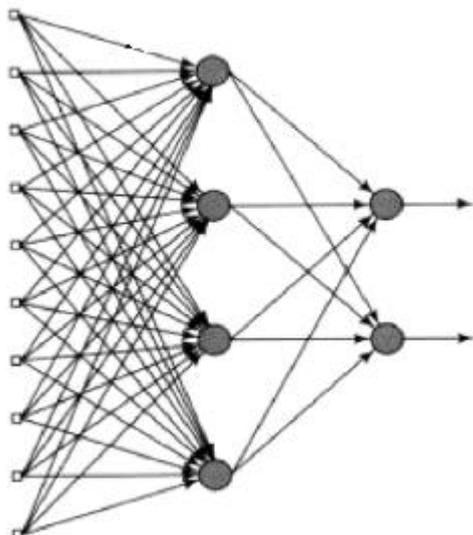


Figura 5 – Exemplo de Rede Neural - Adaptado de Santos (2005)

Santos (2005) apresentou uma pesquisa onde as RNAs apresentaram resultados 5% mais precisos que métodos estatísticos para previsão, porém a autora coloca que a escolha das variáveis relevantes na implementação das redes neurais deve ser realizada de maneira cuidadosa, pois a inclusão de variáveis não relevantes possivelmente prejudicará os resultados apresentados pelas redes neurais artificiais.

Dentre outras aplicações de redes neurais, além de Santos (2005), pode-se apresentar, apud Gonçalves (2005), duas aplicações visando previsões e predições:

- Paz e Borges (1999) que se utilizaram de RNAs para a previsão do consumo de energia elétrica (com dados de 1956 – 1995, para prever 1996);
- Rautenberg (1998) se utilizou de redes neurais para predição de receitas de cores em uma estamperia para que o desperdício e o retrabalho fossem minimizados;

Além da utilização das redes neurais para previsão, também pode ser citada a utilização da mesma na identificação de perfis de clientes, como no estudo apresentado por Júnior e Perez (2006) onde os mesmos além de aplicar redes neurais, também fizeram um comparativo desta técnica com o método estatístico árvore de decisão, chegando a conclusão de que os dois métodos apresentam dados/resultados importantes, porém os dois também apresentam problemas, sendo sua sugestão a utilização dos dois métodos como sendo um complemento do outro.

Sarle (2004) apresenta um estudo sobre a utilização de redes neurais e modelos estatísticos onde o mesmo coloca suas conclusões de acordo com Júnior e Perez (2006), dizendo que é improvável que as redes neurais irão substituir a metodologia estatística visto a necessidade de conhecimentos específicos que ela apresenta, nem tampouco a metodologia estatística criará um modelo que se torne um sistema *expert*. Assim, Sarle (2004) coloca que redes neurais e métodos estatísticos não estão competindo, mas sim há uma relação considerável entre os dois campos, onde um pode fornecer subsídios para o outro. Sarle (2004) acredita que uma comunicação mais intensa entre os dois campos poderia beneficiar ambos.

Sendo assim, este tópico apresentou dentro deste trabalho que as redes neurais também estão sendo utilizadas por outros autores em diversas aplicações no ambiente científico e poderiam ser utilizadas nesta pesquisa, porém este não era o foco do trabalho. Aproveitando a citação de Júnior e Perez (2006) onde os mesmos colocam que a utilização de métodos estatísticos necessita de um conhecimento aprofundado dos métodos e também do assunto/negócio ao qual se está abordando, o autor desta pesquisa justifica sua opção pelos métodos estatísticos por justamente ter uma maior familiaridade com os métodos estatísticos utilizados e um entendimento estruturado onde o mesmo possui conhecimento de várias variáveis preditoras que podem influenciar uma variável resposta e também tem razoável consciência de como isso acontece, possibilitando assim a utilização de métodos estatísticos, tais como análise de regressão múltipla, análise de séries temporais e análise de distribuição de frequências. Aliada a esta justificativa, também está a citação de Santos (2005) sobre a escolha das variáveis quando da utilização de redes neurais, onde se mal feita, esta escolha pode prejudicar os resultados apresentados pelas RNAs. Sendo assim o autor da pesquisa optou pela utilização dos métodos estatísticos por estar mais seguro em sua utilização e porque as bibliografias demonstram que tanto redes neurais quanto métodos estatísticos podem ser utilizados para os fins desta pesquisa obtendo bons resultados.

Assim, desprezando a necessidade de se comprovar quais dos métodos seria melhor para esta pesquisa, mas apenas relatando a existência de trabalhos na área de redes neurais, apresentam-se a seguir os métodos estatísticos que foram utilizados nesta pesquisa

2.5 MÉTODOS ESTATÍSTICOS

Conforme mencionado anteriormente de acordo com Kume (1993), os métodos estatísticos são ferramentas eficazes para a melhoria dos processos através das análises que elas permitem, entretanto, é preciso que se tenha em mente que as ferramentas estatísticas são apenas ferramentas: elas podem não funcionar, caso sejam aplicadas inadequadamente.

Segundo Bussab (2003), o pesquisador obrigatoriamente passa pelo problema de ter que analisar e entender uma massa de dados que é relevante para seu objeto de estudos e para isso, necessitará resumir ou minerar os dados para que estes se tornem informativos, quer seja, para compará-los com outros resultados, quer seja para julgar a sua adequação a alguma teoria. Parte deste processo, já é realizado, utilizando a metodologia SIEGO anteriormente apresentada.

De acordo com Bussab (2003), a essência da Ciência é a observação, e que seu objetivo básico é a inferência, ou seja, a parte da metodologia da Ciência em que se pretende tirar por conclusão ou consequência resultados que possam explicar algum problema ou conjectura sobre tal. Desta forma, entende-se que métodos estatísticos também podem fornecer resultados sob uma análise que possam dar subsídios aos executivos de uma organização para suas decisões.

2.5.1 Estatística e método estatístico

Segundo H. G. Wells citado por Levine et al. (2000) “Raciocinar estatisticamente, será um dia tão necessário quanto à habilidade de ler e escrever”. Observando esta citação, é possível verificar que já se passa por este momento. Sendo assim, se nota que atualmente os gerentes das organizações não enfrentam mais o problema da falta de informações, mas sim a dificuldade de lidar com uma gama enorme de informações e aproveitá-las para obter vantagens competitivas.

Levine et al. (2000) apresentam como justificativa para a necessidade de conhecimento estatístico os seguintes fatos:

- Os executivos precisam saber como apresentar e descrever informações de forma adequada;

- Os executivos precisam saber como melhorar os processos, daí a estatística lhe aponta o foco de atuação;
- Os executivos precisam saber como obter previsões confiáveis a partir de variáveis de interesse;

Segundo Moreira (1975), a Estatística pode ser considerada ou como ciência ou como método.

Como ciência, a Estatística é parte da Matemática que tem por objeto conjuntos quantificados pela contagem dos elementos que os compõem e pela mensuração de um ou mais atributos desses elementos. Pode ainda a Estatística ser considerada como ciência aplicada quando o investigador procura utilizar os processos da Estatística Matemática em campos específicos de outras ciências ou em áreas particulares do conhecimento.

A Estatística pode ser ainda considerada como método. Esta se distingue da Estatística Matemática por oferecer somente meios e subsídios, baseados na lógica matemática, para a obtenção, classificação, resumo, seleção, apresentação e interpretação dos dados quantitativos que caracterizam um conjunto a ser estudado.

Milone (2004) apresenta uma definição de estatística que contempla as subdivisões apresentadas acima por Moreira (1975). Segundo Milone (2004), conceituar a estatística é um tanto difícil devido à sua amplitude, por isso ele coloca uma definição bastante ampla que diz que a estatística é o estudo das maneiras possíveis de obter, coletar, organizar, processar e analisar as informações que forem relevantes e que permitam fazer uma quantificação, qualificação, coleção de fenômenos ou populações para que se possa concluir, deduzir ou fazer previsões de atributos, acontecimentos ou estados futuros.

Estes meios e subsídios variam de ciência para ciência e é por isso que serão aplicados aqui os métodos estatísticos mais apropriados para os pontos a serem abordados e objetivos previamente apresentados.

A seguir são apresentados os métodos estatísticos conforme introduzido no início deste capítulo.

2.5.2 Séries temporais

Uma vez que as condições econômicas e negócios são instáveis e variam ao longo do tempo, os executivos precisam se manter a par dos efeitos que estas mudanças irão causar em seus negócios. Tendo em vista esta condição, Levine et al. (2000) dizem que uma técnica que pode ser utilizada pelo executivo para ajudá-lo no seu planejamento futuro é a previsão. Faz-se isso então utilizando séries temporais, que segundo Stevenson (1986) é um conjunto cronológico (ordenado no tempo) de observações. Milone (2004) além de concordar com Stevenson (1986), ainda complementa dizendo que este conjunto de dados pode ser absoluto ou relativo e discreto ou contínuo. De acordo com Stevenson (1986), estas análises dos dados cronológicos têm por objetivo determinar se eles apresentam algum padrão não-aleatório, podendo então ser utilizados para as tais previsões quanto ao futuro. Novamente Milone (2004) complementa a definição acima dizendo que além do acompanhamento das variáveis visando descreve-las ao longo do tempo, pode também se utilizar as séries temporais para analisar possíveis relacionamentos histórico-funcionais entre determinadas variáveis, como por exemplo na análise de preços e quantidade de peças vendidas, pode-se analisar a elasticidade da demanda do produto em análise. Farias (2003) ainda contribui dizendo que a essência deste método não se limita a identificar padrões mas também separa-los dos ruídos (possíveis interferências nas observações individuais) e aí então realizar previsões de valores futuros .

Segundo Levine et al. (2000), existem basicamente, dois métodos de previsão: quantitativo e qualitativo.

Sendo assim o método de previsão qualitativo é mais empregado quando não se tem conhecimento de dados históricos, como a previsão de vendas de um novo produto, por exemplo. O método de previsão quantitativo, ao contrário, já é indicado para previsões baseadas em acontecimentos ocorridos no passado. E este será o método empregado nesta dissertação, por envolver dados históricos.

Os métodos de previsão quantitativos podem ser subdivididos em dois tipos: série temporal e série causal. O método de previsão causal envolve a determinação de fatores que se relacionam com a variável a ser prevista, como análise de regressão múltipla, com variáveis observadas no passado, análise do indicador principal e outros índices. Por outro lado, Levine et al. (2000) apresentam os métodos de previsão de séries temporais

como métodos que envolvem a projeção de valores futuros de uma variável, com base, inteiramente em observações do presente e do passado desta variável. Fonseca (1995) ainda apresenta uma série temporal com sendo uma série de observações em instantes distintos e sucessivos de tempo, concordando com Levine et al. (2000). Wild (2004), ainda coloca que o termo “série temporal” é utilizado quando a observação de pontos em uma seqüência de tempo não necessariamente é independente e a distribuição não necessariamente é estável.

De acordo com Moreira (1975), quando os dados são cronologicamente dispostos eles constituem uma série temporal e neste caso a natureza das informações a serem coletadas dependerá da finalidade da pesquisa. Milone (2004) coloca que além da finalidade da pesquisa, também é necessário levar em consideração possíveis relações de causa e efeito entre acontecimentos e levar em conta casos condicionados ao tempo. Neste trabalho serão apresentados dados referentes às vendas ocorridas na organização respeitando o período de tempo apresentado na introdução desta dissertação. Segundo Stevenson (1986), Farias (2003), Milone (2004) e Wild (2004) o pressuposto básico da análise de séries temporais é de que fatores que influenciaram padrões da atividade no passado e no presente, continuarão a fazê-lo no futuro. Assim, o tópico a seguir retrata alguns fatores de influência em séries temporais que de acordo com os autores acima, precisam ser considerados.

2.5.2.1 Fatores de influência em Séries Temporais

Segundo Wonnacott (1981), a principal característica de uma série temporal é que suas observações acusam certa forma de dependência em relação ao tempo. Um dos problemas é que esta dependência pode assumir qualquer forma. Wonnacott (1981) apresenta as três principais componentes de dependências, que são: a tendência, as variações sazonais e caminhos aleatórios ou variações irregulares. Stevenson (1986), Milone (2004), Wild (2004) além das três componentes acima citadas, apresenta mais uma forma que são as variações cíclicas.

A tendência pode ser linear e acusa crescimento constante, em cada período de tempo ou exponencial, acusando aumento percentual constante, em cada período de tempo. Caso a tendência não for linear, se aplicam técnicas usadas para estatísticas não-paramétricas.

Para dar uma idéia mais clara, segue abaixo um exemplo fictício de uma análise de tendência referente aos dados de valores de vendas em função do tempo.

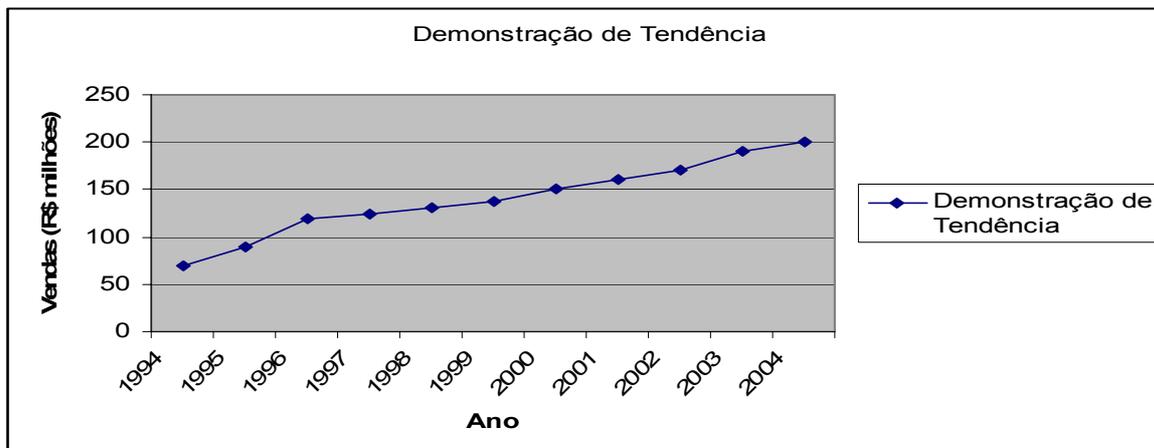


Figura 6 – Serie Temporal de Valores de vendas de uma empresa X em função do tempo - Tendência

Assim como o exemplo da figura 6, também é possível observar casos de tendência quando são observados outros casos práticos como, por exemplo, o crescimento vegetativo de uma população ou ainda o crescimento exponencial do número de usuários da internet ou até mesmo a queda do número de fumantes em torno do mundo.

A variação sazonal diz respeito justamente às variações que por algum fator (temperaturas, por exemplo) podem afetar a série temporal. Segundo Milone (2004), por se referir a eventos ligados às estações do ano, as variações sazonais também são ditas *estacionais*. Em casos onde se observa a possível existência de variações devido à sazonalidade deve-se considerar também este efeito sazonal.

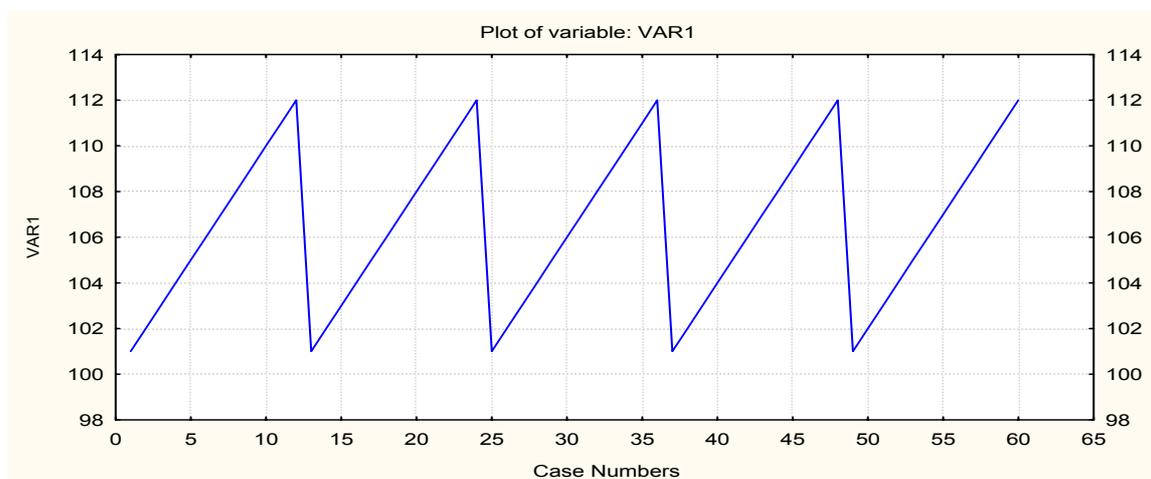


Figura 7 – Exemplo gráfico de variação Sazonal

A figura 7 mostra apenas um exemplo gráfico de uma variação Sazonal referente às vendas de um produto X em um período de observação de 60 meses. A partir deste gráfico então se faz um estudo para identificar os elementos que contribuem para esta sazonalidade.

Como mais um exemplo prático desta variação sazonal é possível relatar casos como o crescimento súbito da venda de ventiladores ocorridos no período entre novembro e fevereiro de cada temporada, dos quais a empresa em questão já pôde presenciar em suas vendas. Outro exemplo que também pode ser citado e que já ocorreu na empresa em questão é o aumento da venda de lavadoras no período de verão, quando muitos consumidores resolvem se utilizar deste equipamento. O aumento da venda de roupas de lã no inverno é outro exemplo da variação sazonal, demonstrando assim que durante praticamente todo o ano ocorrem variações deste tipo.

As variações cíclicas são variações periódicas de amplitude superior a um ano e segundo Stevenson (1986), ordinariamente, tais variações não podem ser separadas das variações irregulares ou caminhos aleatórios, de modo que são analisadas conjuntamente. Variações como estas são variações de longo prazo, o que as torna um tanto quanto difíceis de serem visualizadas. Segundo Milone (2004), é possível citar como exemplo prático o acontecimento de fatores climáticos como furações ou enchentes ou ainda mais radicalmente casos como erupções de vulcões ou fatores econômicos como crises de recessão e crescimento em uma determinada economia. Assim sendo, sua visualização fica muito prejudicada, pois dependem de ciclos que não necessariamente seguem um cronograma de tempo determinado para acontecer.

A terceira dependência diz respeito aos caminhos aleatórios onde já analisados a tendência e a variação sazonal, resta analisar os resíduos onde se consiste em relacionar o valor do erro no tempo atual com o valor do erro em um período $t-1$, adicionando então uma perturbação ou um erro aleatório que possa haver. Este tipo de variação está relacionada a eventos que são imprevisíveis e que podem alterar a forma como a economia ou a forma de viver das pessoas é conduzida. Exemplos práticos para estes tipos de eventos podem ser citados ao lembrar o atentado de 11/09 às “torres gêmeas” nos EUA ou ao terremoto que paralisou as fábricas de memórias RAM em Taiwan por algumas semanas.

2.5.2.2 Médias Móveis e Ajuste Exponencial

Em muitos casos, os dados que são examinados constituem-se de valores que possuem variações intensas de ano para ano e então, torna-se muito difícil a visualização de uma tendência ou movimentos de tendência gerais observados ao longo prazo através de um gráfico, por exemplo. Wild (2004) coloca que neste tipo de situação, avaliar de fato se existe alguma tendência nos dados observados exige o emprego do método de *médias móveis* ou o método de ajuste exponencial, que servem para ajustar (suavizar) a série com o intuito de permitir uma melhor impressão de movimento nos dados ao longo do período.

Segundo Levine et al (2000), o método de médias móveis é muito subjetivo e isso leva a uma dependência do período que é selecionado para construir as médias. As médias móveis, ainda de acordo com Levine et al (2000), são uma série de medidas aritméticas, que são calculadas ao longo de um período, de maneira que cada uma destas médias seja calculada para uma seqüência de valores observado que tenham uma extensão L .

A seguir um exemplo de um grupo de valores que correspondem às observações de quantidades de peças vendidas durante um período de 100 anos. A série foi gerada através do *MS Excel* com a seguinte função: $P_{anual} = 1500 + aleatorio() * 200$.

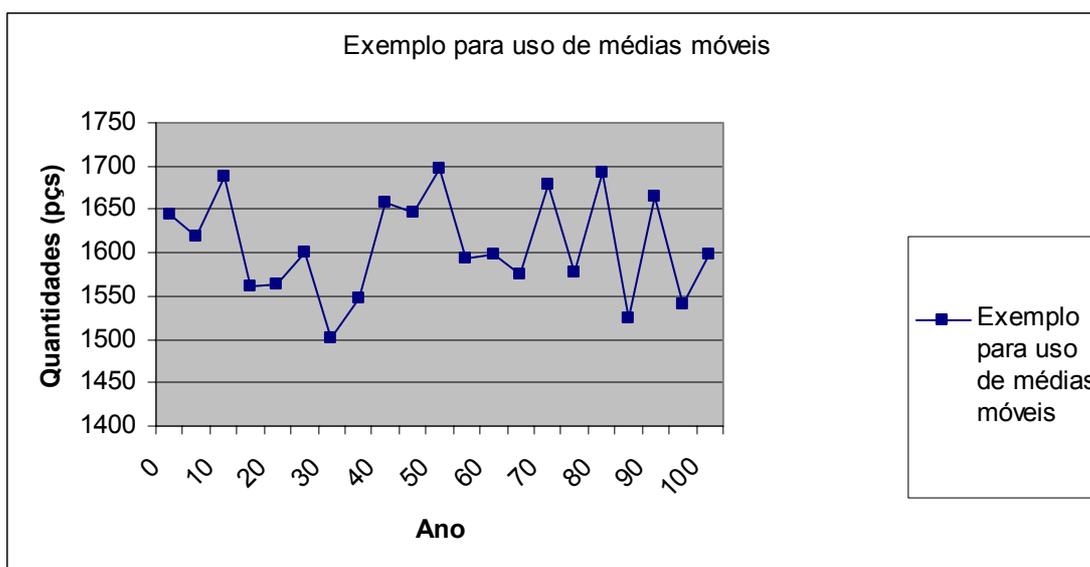


Figura 8 – Série Virtual de Quantidades vendidas anuais do produto X

A figura 9 demonstra a eficiência da análise de tendência utilizando-se o método de *médias móveis*, onde é possível verificar possíveis ciclos e reduzir a variabilidade mostrada na série, ficando clara a tendência, até mesmo observando apenas o gráfico. Foram utilizados ciclos de cinco anos para este exemplo.

A equação utilizada para este exemplo foi:
$$MM(L) = \frac{\sum_{t=(1-L)/2}^{(L-1)/2} Y_{i+t}}{L}$$

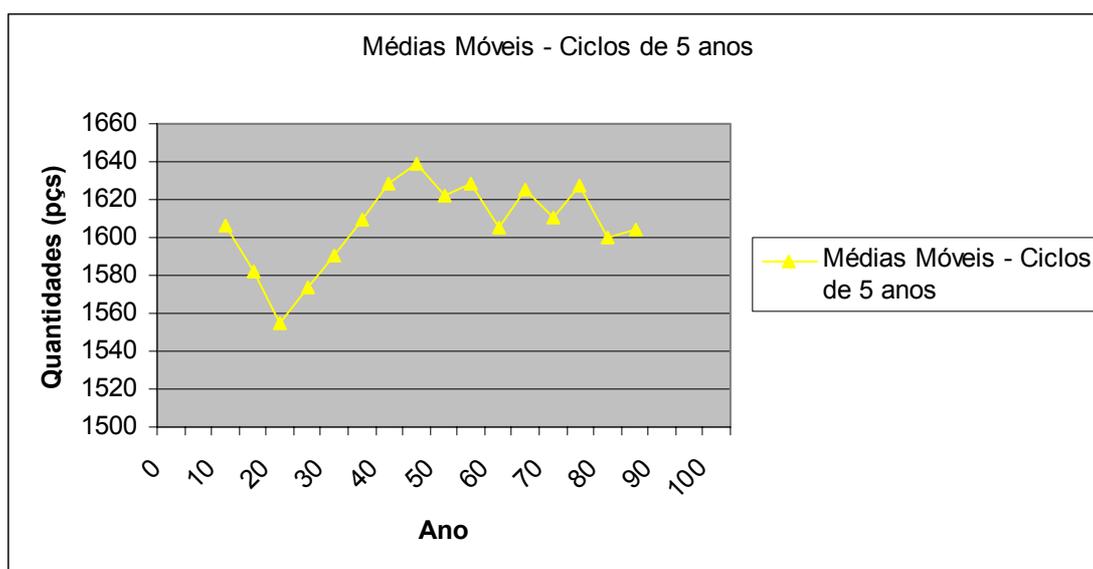


Figura 9 – Série Virtual de Quantidades vendidas anualmente do produto X associada a uma tendência artificial – média móvel

Quando se torna difícil a visibilidade de uma tendência ao longo prazo, também é possível utilizar um outro método denominado de *ajuste exponencial*. Este, além de proporcionar uma melhoria no sentido de visualização de tendência ao longo prazo, também possibilita a obtenção de previsões de curto prazo, utilizado para séries temporais onde o efeito da tendência de longo prazo seja um tanto questionável.

Segundo Levine et al (2000), o nome de *ajuste exponencial* tem sua origem devido ao fato de fornecer uma média móvel *exponencialmente ponderada* ao longo da série temporal. Este fato também constitui uma outra vantagem em relação ao método de *médias móveis*, pois considera toda a série, ou seja, todos os valores que compõem a série. O *ajuste exponencial* trabalha com o conceito de se estabelecer pesos para os valores das séries, de modo que o valor mais recente observado recebe o maior peso e o valor mais antigo observado recebe o menor valor. Uma desvantagem é que a escolha

dos pesos é um tanto subjetiva. Levine et al (2000) coloca que se o objetivo for a eliminação de variações cíclicas e irregulares, deve-se selecionar um valor entre 0 e 1, porém mais próximo de 0, entretanto, se o objetivo for a previsão propriamente dita, deve-se escolher um valor mais próximo de 1.

2.5.2.3 Mínimos Quadrados

Um método bastante útil para se fazer previsões a partir de dados de séries temporais anuais de um ciclo de negócios de modo geral é o método dos mínimos quadrados, onde uma vez que se tenha o objetivo de ajustar uma linha reta a um determinado conjunto de dados, basta então encontrar a equação da reta que melhor se ajuste aos dados e que possa promover as melhores previsões possíveis de uma variável dependente Y , o que não é tarefa fácil.

Segundo Freund (2000), o método dos mínimos quadrados exige que a reta a qual se ajusta os dados seja tal que a soma dos quadrados das distâncias verticais dos pontos à reta seja mínima.

Quando se utiliza o método dos mínimos para ajustar uma linha reta, tem-se o *modelo de tendência linear*. Ao utilizar este modelo é possível que ainda assim, possam ser observados dados que graficamente apontam para um modelo de tendência curvilínea. Deste modo, é possível também utilizar o *modelo de tendência quadrática* ou *polinômio de segundo grau*, como um dos mais simples modelos utilizados para tratar de casos curvilíneos.

Um outro modelo ainda pode ser utilizado quando uma série aparenta estar com uma taxa de crescimento cuja diferença percentual de uma observação para outra é constante. Este é o *modelo de tendência exponencial*. A equação deste modelo é linear, portanto é possível trabalhar com o método dos mínimos quadrados.

2.5.2.4 Modelagem Auto-Regressiva

A modelagem auto-regressiva também é um método bastante interessante para se fazer previsões com base em séries temporais anuais. Neste método são observadas as

autocorrelações entre os dados que permitem obter um grau de associação que possa existir entre os dados e então se utilizar das características potenciais destas autocorrelações referentes aos dados e realizar um melhor ajustamento histórico dos dados. As autocorrelações, segundo Downing (2002), demonstram a possibilidade dos erros não serem independentes, ou seja, se um erro acontecer em um tempo x , existe uma chance maior do erro se repetir em um tempo $x+1$.

Segundo Levine et al (2000), o modelo auto-regressivo pode trabalhar em primeira ordem, tornando-se semelhante ao modelo de regressão linear simples ou trabalhar em p -ésima ordem se assemelhando ao modelo de regressão linear múltipla, que será abordado no próximo tópico. A diferença entre os modelos auto-regressivos de primeira ordem e de p -ésima ordem está basicamente na observação das relações entre os valores das séries de acordo com a sua ordem de seqüência nas mesmas. Portanto primeira ordem refere-se à relação entre valores consecutivos e p -ésima ordem refere-se aos valores consecutivos e também separados por p períodos.

Sendo assim, a escolha do modelo apropriado não é muito fácil, porém conforme Levine et al (2000), deve-se escolher o modelo levando-se em conta o princípio da parcimônia que propõe a escolha do modelo mais simples que atenda o objetivo ao qual se deseja chegar.

2.5.3 Análise de regressão e correlação

Segundo Barbetta (2002), o termo regressão foi empregado pela primeira vez por Francis Galton no século XIX num estudo da relação entre as características dos pais e filhos.

De acordo com Neter (1996), Downing (2002) e Wild (2004) o objetivo de modelos de regressão é modelar o relacionamento entre diversas variáveis preditoras e uma variável resposta. Sendo assim, este relacionamento pode se dar através de uma equação linear ou uma função não linear.

Levine et al (2000) apresentam a análise de correlação como uma forma de avaliar a força de associação entre variáveis numéricas. Neste caso o objetivo não é prever uma outra variável, mas sim avaliar o quanto uma variável pode influenciar nos valores de outra. Segundo Downing (2002), o *coeficiente de correlação* é um valor que varia entre -1 para correlação perfeita negativa à +1 para correlação perfeita positiva e é

dado por $r = \sqrt{r^2}$ onde r^2 será apresentado a seguir no tópico de regressão linear simples.

2.5.3.1 Diagrama de Dispersão

Segundo Barbetta (2002), Farias (2003) e Wild (2004), uma forma interessante de se observar correlações entre duas variáveis é a utilização do diagrama de dispersão onde é possível verificar graficamente esta relação através de um sistema cartesiano onde são plotados os valores de x e y.

De acordo com Milone (2004), um diagrama de dispersão é um sistema de coordenadas em um sistema cartesiano em que os dados podem se agrupar em torno de uma linha imaginária ascendente ou descendente, apontando uma tendência, forma ou andamento de um fenômeno. Assim, quanto mais próximo desta linha imaginária os dados se agruparem, mais clara fica a correlação existente entre os valores destas variáveis. A seguir um exemplo com dados fictícios que demonstra o caso de uma *correlação positiva ascendente*. Os dados representam as quantidades de clientes e valores de vendas semanais de 20 empresas quaisquer.

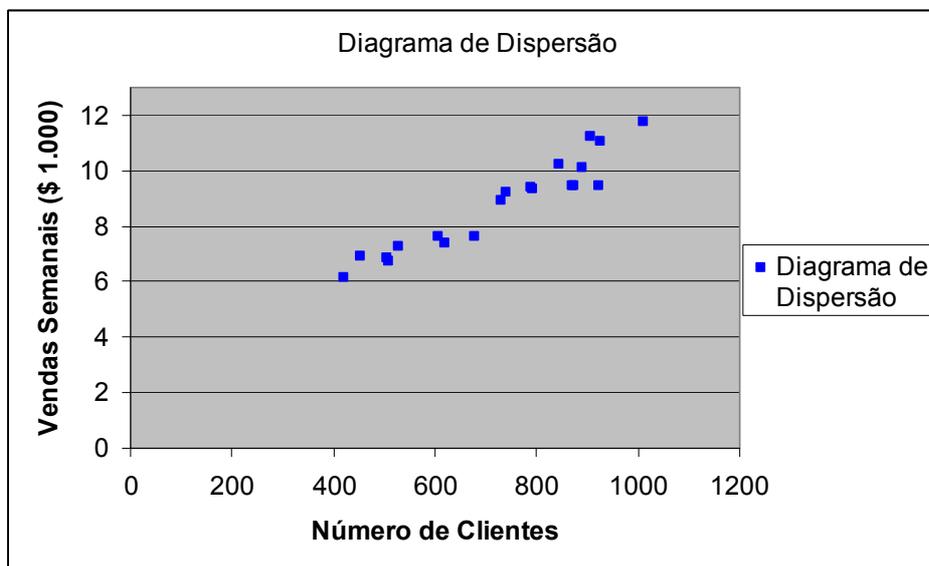


Figura 10 – Diagrama de dispersão - Adaptado de Levine et al (2000)

2.5.3.2 Regressão linear simples

Barbetta (2002) apresenta o modelo estatístico-matemático mais simples, como sendo aquele onde ocorre o relacionamento de uma variável Y, que se chama variável resposta ou dependente, com uma variável X, denominada de variável explicativa ou

independente. Milone (2004) complementa esta definição como sendo uma função do primeiro grau de uma única variável independente. Este modelo chama-se *o modelo da regressão linear simples*. Como exemplo, Barbetta (2002) cita que é possível mostrar a relação entre a renda de uma pessoa e o consumo dela (em termos de gastos) ou a memória *ram* de um computador e o tempo de resposta de um sistema. Assim, da mesma forma como acontece em um estudo de correlações, a análise de regressão, segundo Farias (2003), também toma como base um conjunto de observações relativas às variáveis X e Y. Assim sendo, apresenta-se aqui um modelo adequado para um par de observações:

$$y = \alpha + \beta x + \varepsilon$$

Onde ε está representando o efeito aleatório que na verdade podem ser inúmeros fatores que agem sobre a observação Y, α representa o valor médio de Y quando X for igual a 0 e β representa uma variação esperada em Y por cada variação unitária de X.

Com a análise de regressão é possível então avaliar quanto (em percentual) uma variável pode explicar relativo à variação da variável dependente, porém para isso é necessário realizar o levantamento de diversas medidas de variação. Segundo Levine et al (2000), Farias (2003), Montgomery (2003) e Hair (2005) pode-se iniciar este levantamento através da *soma total dos quadrados (STQ)* que é uma medida de variação dos valores de Y_i em torno da sua média aritmética \bar{Y} . Na análise de regressão a STQ pode ser subdividida em variações explicadas ou *soma dos quadrados devida à regressão (SQReg)* que refere-se à relação entre a variável dependente e a independente, e as variações inexplicadas ou *soma de quadrados dos resíduos (SQR)* que se refere a outros fatores que não a relação entre as variáveis.

Assim,
$$\mathbf{STQ = SQReg + SQR}$$

Hair (2005) e Montgomery (2003) também colocam que além destas três variações definidas, pode-se também definir um outro fator que é o *coeficiente de determinação (r^2)* que serve para medir o quanto da variação da variável dependente pode ser explicada pela variável independente. Por exemplo, suponha-se que o valor hipotético de SQReg é 45,12 e o valor hipotético de STQ é 50,68 resultando em um r^2 igual a 0,8902, pois $r^2 = \frac{SQReg}{STQ}$, que significa que 89,02% da variação da variável

dependente pode ser explicada pela variabilidade da variável independente. Assim apenas 10,98% da variabilidade da variável dependente é explicada por outros fatores ou variáveis que não aquela analisada. Assim como Levine et al (2000) e Barbetta (2002), Neter et al (1996) também trabalha com um r^2 ajustado que reflete tanto o número de variáveis explicativas quanto o tamanho da amostra ou população se for o caso.

A seguir é possível notar um gráfico que utilizou os mesmos dados do gráfico representado pela figura 10, onde então é traçada a linha de regressão e também o r^2 que foram obtidos através do assistente de gráfico do *MS Excel*:

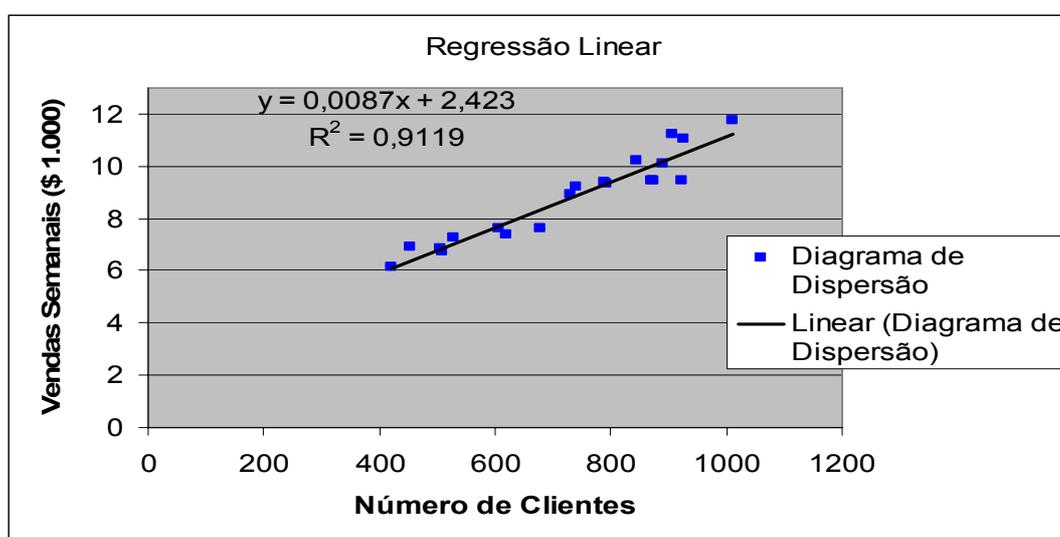


Figura 11 – Exemplo de regressão linear – Adaptado de Levine et al (2000)

Ao se estudar a aplicação de modelos estatísticos como auxílio de análise sobre variáveis também é possível notar que a aplicação correta destes modelos depende de uma série de premissas. Para análise de regressão e correlação Levine et al (2000), assim como Downing (2002) apresentam alguns pressupostos que devem ser observados, são eles: normalidade, homocedasticidade, independência de erros e linearidade.

Estes pressupostos significam respectivamente:

- Os valores precisam estar distribuídos de acordo com uma normal, o que significa dizer que esta distribuição, quando plotados em um plano cartesiano seguem o formato de um sino, o que indica a normalidade;
- A proporção da variabilidade entre os valores de X e Y deve se manter constantes, o que significa dizer que o modelo apresenta bons resultados para todo o intervalo de valores de X e Y;

- Os erros achados devidos à diferença entre os valores de Y e os valores previstos de Y devem ser independentes para cada valor de X ;
- A relação entre as variáveis deve ser linear.

2.5.3.2.1 Análise de resíduos

Este é um ponto importante em uma análise de regressão porque um dos pressupostos básicos que já foi citado é a independência dos erros ou resíduos e quando os dados são coletados ao longo de um período de tempo, existe uma tendência de que os erros obtidos num determinado espaço possam ser semelhantes aos erros obtidos através de observações em pontos próximos à observação anterior. Bussab (2003) coloca que para observar as possíveis discrepâncias entre os valores observados e os valores ajustados pelo modelo de regressão é necessário fazer a *análise de resíduos*. Segundo Farias (2003) os resíduos apontam a diferença entre o que foi realmente observado e o que foi predito pelo modelo de regressão, demonstrando aquilo que o modelo não foi capaz de explicar. Quando existe este padrão nos resíduos, dá-se o nome de *autocorrelação*, conforme citado no item 2.4.2.4, o que pode tornar a validade do modelo de regressão escolhido, fortemente comprometida, ou seja, se houver uma forte correlação no conjunto de dados, que pode ser observada através do coeficiente de correlação r , a validade do modelo de regressão fica prejudicada.

Conforme Downing (2002), para observar então se existe alguma correlação entre os resíduos, é possível plotar estes resíduos em um plano cartesiano em função do tempo, o que permitirá uma visão de um possível padrão dos resíduos.

Quando da análise de regressão observa-se os resíduos e caso houver autocorrelações ou heterocedasticidade, pode-se tentar a aplicação de algumas transformações logarítmicas ou aplicação da raiz quadrada sobre os dados, por exemplo, conforme cita Downing (2002).

A seguir, na figura 12 está um exemplo fictício da plotagem de resíduos em um plano cartesiano onde é possível observar um padrão cíclico dos resíduos demonstrando então uma possível violação do pressuposto de que os resíduos precisam ser independentes.

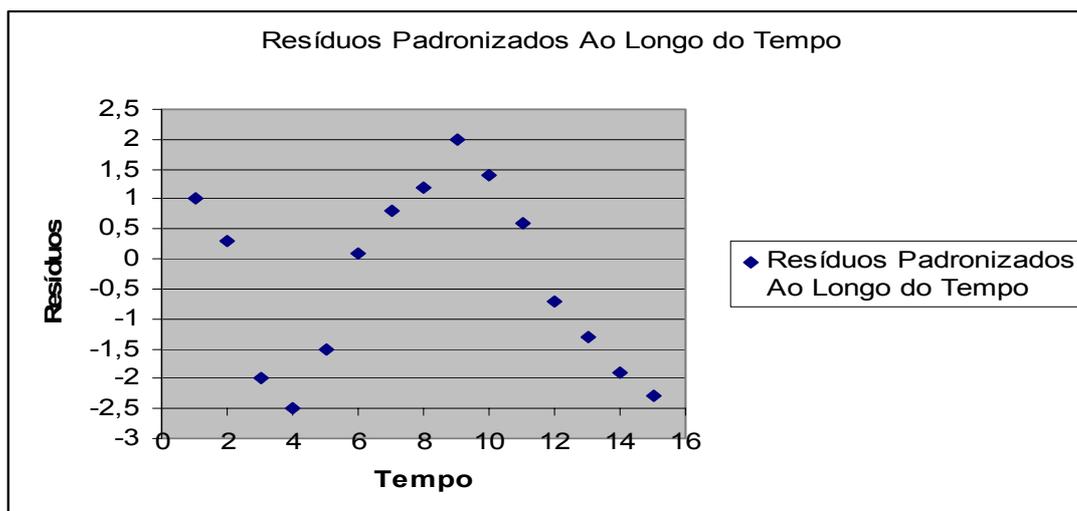


Figura 12 – Exemplo fictício de resíduos padronizados

Hoffmann (1987), Levine et al (2000) e Downing (2002) também apresentam um outro modo de avaliar se existe autocorrelação nos resíduos. Este modo é chamado de *teste de Durbin-Watson* e é baseado na seguinte equação:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=2}^n e_i^2}$$

onde e_i são os desvios da regressão ajustada pelo método dos mínimos quadrados. Assim, quando o valor de d se aproximar de 0, significa que os resíduos estão positivamente autocorrelacionados e quando o valor de d se aproximar de 2 significa que os resíduos não estão correlacionados. Se o valor de d for maior que 2 significa que há uma autocorrelação negativa. Uma dificuldade na utilização deste método é saber o quanto abaixo de 2 o valor de d está para se considerar um problema de autocorrelação. Como os valores de d dependem do tamanho da amostra e do número de variáveis independentes, foram tabelados, para diferentes valores da amostra e variáveis independentes, aos níveis de significância de 1% e 5% (unilaterais), intervalos que contêm o valor crítico de d para se determinar a autocorrelação.

Caso seja observado através dos resíduos que um dos pressupostos foi violado, é possível utilizar métodos alternativos de regressão dos mínimos quadrados ou modelos de mínimos quadrados alternativos como regressão curvilínea ou regressão múltipla dependendo da violação observada, ou aplicar transformações aos dados para permitir

que eles atendam às suposições, como por exemplo, transformações logarítmicas ou aplicação da raiz quadrada sobre os dados.

Maiores informações podem ser encontradas em Barbetta (2002), Hoffmann (1987), Levine et al (2000), Downing (2002) e Milone (2004).

2.5.3.3 Regressão linear múltipla

Segundo Milone (2004), funções simples, com apenas uma variável muitas vezes não conseguem explicar adequadamente os fenômenos, assim, a regressão linear múltipla tem como diferencial da regressão linear simples o fato de utilizar-se de mais variáveis independentes visando a uma melhor compreensão do comportamento da variável dependente. Barbetta (2002) faz esta afirmação ao dizer que através da regressão linear múltipla pretende-se construir um modelo estatístico-matemático para se estudar objetivamente a relação entre as variáveis independentes e a variável dependente. Hair (2005) diz que objetivo da análise de regressão múltipla é justamente utilizar as variáveis independentes, cujos valores são conhecidos para prever ou explicar as mudanças ocorridas na variável dependente, pois segundo Downing (2002), em muitas situações, o valor da variável dependente é afetada pelos valores das variáveis independentes. Assim, a partir deste modelo pode-se conhecer a influência causada por cada variável independente, bem como também realizar previsões sobre a variável dependente, baseadas no que se conhece sobre as variáveis independentes.

Segundo Barbetta (2002), a análise de regressão está baseada em um grupo de observações $(x_1, x_2, x_3, \dots, x_n, y)$ que refletem às variáveis X_1, X_2, \dots, X_n e Y , onde é possível afirmar que um valor y é dependente, parcialmente, dos correspondentes valores x_1, x_2, \dots, x_n e também de outros fatores representados por ε , formando assim a seguinte equação:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Onde os parâmetros α e β_i devem ser estimados e ε representa os demais fatores que influenciam a variável dependente, também denominado *efeito aleatório*. A equação acima também é citada por Downing (2002) e Montgomery (2003).

Ao realizar uma análise de regressão, como a regressão múltipla, é possível encontrar um modelo mais bem ajustado para explicar a variável dependente, afinal, observa-se mais de uma variável explicativa o que proporciona um entendimento mais abrangente sobre quais as influências das variáveis independentes.

Levine et al (2000) ainda coloca que é interessante atentar para o fato de que a análise de regressão e também correlação estão entre as técnicas estatísticas mais aplicadas em negócios ou na economia e este ponto também está intimamente ligado ao fato de serem técnicas muito mal utilizadas, pois em muitos casos os pressupostos não são observados, quando o são, em muitos casos não se sabe quais métodos alternativos devem ser utilizados, enfim, questões que podem prejudicar a boa utilização de tais técnicas.

Assim sendo, procurou-se observar questões como as mencionadas, para que a sua utilização neste trabalho realmente pudesse atender ao objetivo de demonstrar ao executivo as relações entre variáveis que influenciam os resultados do seu negócio.

3 MATERIAL E MÉTODOS

Do ponto de vista da abordagem do problema, a pesquisa é do tipo quantitativa, empregando-se a técnica de levantamento ou também conhecida como “*survey*”, abordada através de um delineamento descritivo.

Observando a natureza dos dados, notou-se que os dados apresentam variáveis quantitativas a respeito dos valores de vendas e também dos valores relacionados com clientes, classificando assim o problema como sendo de uma pesquisa quantitativa, pois neste caso tudo pode ser quantificado e assim utilizando os recursos apresentados no início desta pesquisa, que são os métodos estatísticos.

Os dados são levantados levando-se em conta seus conteúdos históricos que estão armazenados na base de dados da empresa em questão, que será apresentada no item 3.3, e que dizem respeito aos valores obtidos do período de 1994 a 2005, utilizando para isso a técnica de “*survey*”.

De acordo com Silva e Menezes (2005, p.21) esta pesquisa se enquadra como pesquisa descritiva uma vez que visa descrever as características de determinada população ou fenômeno ou ainda o estabelecimento de relações entre variáveis, o que realmente ocorre nesta pesquisa, quando a partir de variáveis encontradas na base de dados da empresa em questão, pretende-se relacioná-las para se obter informações significativas e úteis para a tomada de decisões dos executivos da organização.

Quanto a sua natureza também é enquadrada em pesquisa aplicada, uma vez que visa gerar novos conhecimentos para a aplicação prática na organização em questão para combater o problema de tomadas de decisões baseadas em suposições empíricas sem embasamento científico.

Tendo em vista esta necessidade de embasamento científico, já exposta exaustivamente anteriormente, utiliza-se aqui então a associação de três abordagens científicas para trazer ao executivo um modo mais direcionado e comprovado para que se possa tomar uma decisão.

Estas três abordagens são respectivamente, as metodologias SIEGO e *Data Warehouse* e também métodos estatísticos. Tal associação permite realizar as melhorias, (no sentido informativo), previstas pela metodologia SIEGO utilizando *Data Warehouse* e ainda apresentar relações e visões mais conclusivas a partir da aplicação dos métodos estatísticos sobre os dados já minerados. A seguir será abordada esta associação em um tópico específico.

3.1 A POPULAÇÃO

Os dados coletados fazem parte da base de dados da empresa em questão e são constituídos das vendas realizadas pela empresa e pelos movimentos realizados pelos clientes que ela atende. Estes dados dizem respeito a movimentos realizados desde agosto de 1994 até junho de 2005. Correspondem ao movimento da empresa matriz e também outras quatro filiais. Sendo esta a base de dados em questão, foi analisada toda a população, constituindo 36.305 clientes de 33 zonas de vendas, 119 ramos de atividades e 469.249 títulos.

3.2 AS VARIÁVEIS

As variáveis utilizadas neste trabalho estão todas relacionadas com o tripé Custo, Tempo e Qualidade, sendo que para isso, foram construídas tabelas de fatos, conforme menciona uma das etapas do *Data Warehouse* onde foram consideradas variáveis relativas aos clientes, valores de faturamento, data das vendas, itens vendidos, preços dos itens, margens das vendas e prazo médio de pagamentos.

Todas estas variáveis foram utilizadas para os cálculos efetuados pelo sistema e também para as análises feitas com os métodos estatísticos.

3.3 METODOLOGIAS UTILIZADAS

Esta seção tem o propósito de apresentar quais os métodos utilizados para extração e análise dos dados bem como os benefícios de suas utilizações, proporcionando assim o conhecimento necessário para possíveis aplicações posteriores em casos semelhantes.

3.3.1 Associação das abordagens

Os tópicos a seguir mostram de forma detalhada como são aplicadas as metodologias anteriormente mencionadas e também em quais pontos cada uma delas interage com as demais, afinal, como já mencionado anteriormente, através da associação destas três abordagens (*Data Warehouse*, SIEGO e Métodos Estatísticos), o trabalho em questão visa à melhoria de atuação da metodologia SIEGO a fim de obter

melhores resultados ou apresentar uma forma melhor de o executivo poder visualizar e analisar os dados de sua organização.

É importante ressaltar que foram usados os dados da empresa da aplicação, maiores detalhes no Capítulo 4, mas a metodologia é geral, podendo ser implementada em qualquer organização, com os devidos ajustes.

3.3.2 Fases da metodologia SIEGO

No capítulo em que foi apresentada a fundamentação teórica, mostrou-se a metodologia SIEGO, porém agora apresentamos os passos a serem seguidos para a implementação propriamente dita desta metodologia segundo Dalfovo (2001).

A metodologia SIEGO é dividida em três fases. A primeira fase é a preparação do projeto. Na segunda fase é a determinação e avaliação das ações de melhorias. Na terceira fase é a implantação das idéias.

3.3.2.1 Fase I - preparação do projeto SIEGO

De acordo com Dalfovo (2001), a contemplação da Fase I depende da execução de cinco passos que constituem a preparação do projeto SIEGO. O passo 1 envolve o planejamento do grupo de trabalho. O passo 2 é onde ocorre a definição dos processos. No passo 3 é preciso realizar a motivação e instrução ao grupo de trabalho que irá atuar na implementação da metodologia. No passo 4 ocorre o planejamento da implementação das idéias. Finalmente no passo 5 ocorre a preparação do grupo de trabalho para que haja um acompanhamento das idéias.

3.3.2.1.1 Passo 1 - planejamento do grupo de trabalho

Neste passo Dalfovo (2001) coloca que “deve-se planejar o trabalho do ciclo no que se refere à definição de processo e Líderes de processo; fornecer apoio ao Comitê de Liderança no desenvolvimento da comunicação; motivar o grupo de trabalho a participar na implantação da metodologia; repassar informações e esclarecer dúvidas; ser o elo de ligação entre a organização e os consultores; levantar informações da situação atual da organização”.

3.3.2.1.2 Passo 2 - definição dos processos

Aqui Dalfovo (2001) apresenta que “devem-se definir os processos e sub-processos e alocação de recursos; mapear os processos e sub-processos; identificar problemas junto com o grupo de trabalho; organizar os próximos passos”.

3.3.2.1.3 Passo 3 - motivação e instrução ao grupo de trabalho

O autor da metodologia apresenta que neste passo “deve-se instruir os participantes; ser responsável pela qualidade do trabalho e a observância dos prazos; estimular os participantes na busca de idéias inovadoras; priorizar as oportunidades de melhoria que tem maior impacto no desempenho do processo em conjunto com seu grupo de trabalho; estimular o intercâmbio de idéias; apresentar e requerer aprovação das idéias junto às equipes; fazer a apresentação ao Comitê de Liderança”.

3.3.2.1.4 Passo 4 - planejamento da implementação das idéias

A proposta de Dalfovo (2001) diz que neste passo “deve-se planejar a implementação das idéias aprovadas; assegurar que as medidas resultem em redução de custos; estimular o esclarecimento de idéias que ainda possuam algum questionamento pendente; preencher os formulários de acordo com a metodologia estabelecida”.

3.3.2.1.5 Passo 5 - preparação para acompanhamento das idéias

Com os passos anteriores já contemplados, deve ocorrer neste quinto passo, um acompanhamento para implantação das idéias, sendo que para isso utiliza-se de relatórios, faz-se o monitoramento e captura dos resultados para garantir que tudo está ocorrendo como o planejado.

3.3.2.2 Fase II - determinação e avaliação das ações de melhorias

Na Fase II, sua contemplação ocorre a partir da satisfação de três passos:

- 1 - definir a montagem do Banco de Dados
- 2 - desenvolver e avaliar as idéias de melhorias.
- 3 - selecionar as idéias em potencial.

3.3.2.2.1 Passo 1 - montagem do banco de dados

Neste passo devem-se compreender os aspectos econômicos ligados à empresa onde se está aplicando a metodologia. Deve-se:

- 1 – Definir os custos aceitáveis.
- 2 – Definir quais os fluxos que ocorrem na empresa.
- 3 – Convidar o gerente/líder da empresa/departamento onde a análise está ocorrendo e também lhe mostrar relatórios dos ocorridos anteriormente.

O banco de dados deve refletir o organograma da empresa/departamento onde se está aplicando o método, deve-se montar a base de custos, definir missões, atividades e subatividades, montar a estimativa de quanto as atividades irão precisar de recursos, fazer um levantamento dos fluxos das informações e montar uma análise dos indicadores de desempenho.

É neste ponto que entra a associação com a metodologia de *Data Warehouse*, cuja implementação será descrita na seção 3.3.3.

3.3.2.2.2 Passo 2 - desenvolvimento e avaliação das idéias de melhorias

Segundo Dalfovo (2001), “neste passo devem-se estabelecer as reuniões de *Brainstorming*. A técnica *Brainstorming* é atribuída a A. Osborn sendo criada por volta de 1938, onde o objetivo geral é conduzir um grupo de pessoas que tentam resolver um problema específico coletando todas as idéias da interação do grupo. Nesta fase também se desenvolvem as ações de melhorias. Desenvolver as idéias de melhoria. Calcular seus impactos e calcular seus riscos. Neste passo deve-se fazer a identificação de oportunidades de melhoria; as questões típicas para geração de idéias; as fontes típicas de idéias; a árvore para procura sistemática de melhorias; a geração de idéias; os fatores críticos de sucesso para prática do *Brainstorming*; as regras básicas para reunião de *Brainstorming*; as frases mortíferas do *Brainstorming*; a avaliação das sugestões de melhoria; as atividades e formulários”.

3.3.2.2.3 Passo 3 - seleção das idéias em potencial

O passo 3 tem como missão básica verificar quais idéias foram levantadas e escolher aquelas que podem gerar bons resultados, bem como identifica aquelas que ainda precisam se melhor estudadas e alteradas. Após esta verificação, demonstra-se os resultados da avaliação para o Comitê de Liderança.

Dentre outros, deve-se realizar nesta fase, atividades e formulários; visão geral do documento para apresentação ao Comitê de Liderança; disposições a serem tomadas na reunião do Comitê de Liderança; papel do Facilitador nas reuniões do Comitê de Liderança.

3.3.2.3 Fase III - implementação das idéias

A fase III é contemplada através de dois passos:

- 1 - Planejamento da Implantação
- 2 - Implementação e Rastreamento das ações de melhorias.

3.3.2.3.1 Passo 1 - planejamento da implantação

Dalfovo (2001) coloca que “neste passo procura-se desenvolver os planos de capturar as economias geradas pelas idéias de potencial. Aprofundar análise das idéias críticas. Definir um responsável para acompanhamento da implantação. Também neste passo devem-se implantar as atividades de delineamento das linhas gerais, do planejamento da implantação, do levantamento das implicações, da determinação dos itens de controle para acompanhamento e da revisão do plano com o responsável da unidade”.

3.3.2.3.2 Passo 2 - implementação e rastreamento das ações de melhorias

Por último procura-se verificar se as idéias estão sendo implantadas conforme os objetivos propostos, observando as economias geradas. O estímulo do nível operacional é importante aqui para que a implantação ocorra sem problemas neste nível.

Com este último passo, finaliza-se o roteiro para implementação da metodologia SIEGO, porém como esta metodologia trabalha baseada no *Data Warehouse*, segue adiante a apresentação deste método.

Mais detalhes sobre a metodologia SIEGO podem ser encontrados em Dalfovo (2001).

3.3.3 A construção de um *Data Warehouse*

Conforme mencionado no item 3.4.2.2.1, deve existir um banco de dados que contenha os dados pré-calculados que seja de fácil acesso e apropriado para consultas rápidas, tornando o trabalho do executivo mais ágil. Sendo assim, este tópico mostra quais os passos para se montar um *Data Warehouse* que possa contemplar este aspecto.

Kimball (1997) descreve que para construir um *Data Warehouse* há um processo de combinação das necessidades de informação de uma comunidade de usuários com os dados que realmente estão disponíveis. De acordo com Kimball (1996) e Kimball (1997), o projeto fundamenta-se em nove pontos de decisão que são direcionados pelas necessidades do usuário e pelos dados disponíveis.

3.3.3.1 As etapas propostas por Kimball

Kimball (1996) e (1997) descreve nove etapas que devem ser seguidas quando da criação e desenvolvimento de um *Data Warehouse*. Assim, a proposta deste trabalho segue estas etapas de acordo com o que segue:

Passo 1 – Escolha do Processo

Entende-se por processo o assunto ao qual um *Data Mart* (parte de um *Data Warehouse*) se refere. O primeiro *Data Mart* que se constrói deve ser o mais robusto com relação aos dados nele agregados. Ele deve responder simultaneamente às mais importantes questões de negócios e ser o mais acessível do ponto de vista da extração de dados. O melhor ponto para se iniciar na maioria das empresas, é construir um *Data Mart* que contenha informações oriundas de notas fiscais de clientes ou de movimentações mensais. Esta fonte de dados é provavelmente de fácil acesso e de alta qualidade.

Uma das leis de Kimball é que a melhor fonte de dados em qualquer empresa é o registro de “quanto dinheiro os clientes nos devem”. A menos que custos e utilidade sejam facilmente acessíveis, é melhor evitar a inclusão destes dados na construção do primeiro *Data Mart*.

Segundo Kimball (1996), nada derruba mais rapidamente a implementação de um *Data Mart* que a missão impossível de provir atividades baseadas em custos como parte do seu primeiro *Data Mart*.

Passo 2 – Escolha do Nível de Granularidade

Este segundo passo parece um detalhe técnico precoce, mas ele é atualmente o segredo para o progresso na construção de um projeto. A escolha da granularidade significa definir exatamente o que um registro da tabela de fatos representa.

Somente quando você tem escolhida a granularidade, você pode ter uma coerente discussão sobre qual é a dimensão da tabela de fatos do *Data Mart*.

Passo 3 – Identificar e conformizar as dimensões

As dimensões são os manipuladores do *Data Mart*. As dimensões são plataformas para pesquisar os valores limites possíveis e aplicar estes valores. As dimensões delimitam os “encabeçadores” de fileiras no relatório final do usuário. Um grupo de dimensões bem arquitetado torna o *Data Mart* entendível e fácil de usar. Um grupo de dimensões incompleto ou mal definido torna o *Data Mart* inútil.

Passo 4 – Escolha dos fatos

A granularidade da tabela de fatos determina quais fatos pode-se usar no *Data Mart*. Todos os fatos devem ser expressos em um nível uniforme subentendido pela granularidade. Conforme citado anteriormente, os fatos devem ser tão aditivos quanto possível.

Note que fatos podem ser adicionados à tabela de fatos a qualquer momento, desde que sejam consistentes com a granularidade da tabela de fatos. Estes fatos adicionais não invalidam a funcionalidade inicial da aplicação. Esta compatibilidade é um exemplo da enorme dimensão do projeto de um banco de dados.

Passo 5 – Armazenando dados pré-calculados na tabela dos fatos

Um exemplo comum da necessidade de armazenar dados pré-calculados ocorre quando os fatos compreendem um ganho e perda de valores. Esta situação ocorre muitas vezes quando a tabela de fatos está baseada em uma conta do cliente.

Passo 6 – Preenchendo as Tabelas de Dimensão

Neste ponto a tabela de fatos está completa, e pode-se entender os papéis das tabelas de dimensão em fornecer entradas para a tabela de fatos diretamente de atributos dimensionais. A granularidade definida no passo 2 também determinou a granularidade

de cada uma das tabelas de dimensão. No passo 3, deve-se identificar as dimensões em detalhe suficiente para descrever coisas tais como clientes e produtos no grau correto, e deve-se compreender onde as dimensões chaves devem ser obtidas.

Neste passo, se pode voltar às tabelas de dimensão e adicionar exaustivamente tantos textos como descrições para as dimensões. Até inerentemente pequenas dimensões como os tipos de transação devem ser honrados com boas descrições de textos na qual cada transação significa um tipo. Oportunidades são aquelas transações que podem ser arranjadas em grupos. O Grupo de transação deve ser outro atributo de texto.

Todos os atributos de texto devem consistir em palavras reais. Abreviaturas enigmáticas são extremamente indesejáveis. Lembre que estes atributos de textos são usados tanto na interface da aplicação como em linhas e colunas de relatórios. Deve-se fazer um trabalho profissional de qualidade garantida nos atributos da tabela de dimensão.

Passo 7 – Escolhendo a Duração do Banco de Dados

O limite de duração diz qual o período de tempo anterior que a tabela de fatos terá. Em muitos negócios, há uma necessidade natural em observar o tempo em um período de um ano atrás. Esta necessidade normalmente ocorre com os cinco últimos trimestres de dados. Próximo ao fim de um ano civil implica, portanto, em dois anos cheios de valor de dados. Estes argumentos podem ser repetidos com menor intensidade para dados de dois anos, dados de três anos, e assim por diante.

Passo 8 – É preciso preparar as dimensões para suportar mudanças

Para contemplar este passo, pode-se utilizar de campos adicionais nas tabelas para que as modificações lentas possam ser contempladas na medida em que forem acontecendo, por exemplo em uma mudança de unidade de produto.

Passo 9 – Decidindo a Pergunta de Prioridades e Modos de Pergunta

Depois dos primeiros oito passos, tem-se um projeto lógico completo do data mart. Agora é possível voltar atenção para questões físicas do projeto. Neste passo restringe-se a atenção para o grande projeto físico afetando a percepção do usuário final

do data mart: o tipo físico classificado na ordem das tabelas de fatos no disco e a presença de resumos pré-armazenados ou agregações.

As duas decisões mais importantes de projetos que podem ser tomadas dizem respeito à granularidade e ao particionamento dos dados.

Para acesso as informações contidas nestas dimensões será utilizado um Cubo de Decisão que nada mais é que uma técnica onde em um banco de dados multi-dimensional simula-se um cubo com n dimensões. Este cubo é utilizado para cruzar dados de tabelas específicas que possam gerar dados importantes para suporte à decisões.

Em relação à amplitude do tempo, implica em indicar quanto tempo de informações serão carregadas para as tabelas escolhidas para o armazenamento dos dados. A preocupação também será em qual a periodicidade de carga dos dados (diariamente, semanalmente, mensalmente, anualmente).

3.3.4 Utilizando as séries temporais

Conforme abordado no capítulo de fundamentação teórica e também de acordo com Levine et al (2000), a previsão pode ser realizada através do método quantitativo e séries temporais, quando da observação de dados históricos, objetivando projeções futuras.

A metodologia SIEGO propõe melhorias em relação ao tempo e em virtude disso a utilização de séries temporais se encaixa nesta visão, pois apresenta uma maneira de se visualizar projeções futuras baseadas em dados históricos, que neste caso dizem respeito aos valores de vendas realizadas pela empresa matriz e suas filiais em um tempo passado.

Assim, de acordo com Wonnacott (1981) procurou-se verificar uma dependência em relação ao tempo. Para tanto, observando Milone (2004), foi aplicado aos dados um índice para inflacionar os dados em respeito aos anos que se passaram e para refletir a realidade atual.

Segundo Spiegel (2004), um número índice é utilizado como uma medida estatística para demonstrar variações de uma variável ou grupo de variáveis ao longo do tempo, localização geográfica ou a outras características. Neste caso utilizou-se um

número índice justamente para compensar as variações que ocorreram durante os anos no mercado do atacado, varejo e construção civil.

O número índice utilizado neste caso foi o IGP-M/FGV (Índice Geral de Preços de Mercado – calculado pela Fundação Getúlio Vargas), que analisa a variabilidade nos preços dos produtos relacionados aos ramos de atacado, varejo e construção civil, sendo o mesmo composto percentualmente de 60%, 30% e 10% respectivamente e que foi extraído do site do Banco Central do Brasil através do endereço <http://www4.bcb.gov.br/tmp/seriesdownload/ST2005072516224655.csv> . É importante deixar claro que a utilização do IGP-M é indispensável para casos em que há séries temporais de valores monetários.

Após a aplicação do índice, os dados foram plotados em um gráfico onde foi possível observar uma tendência linear em relação aos valores de vendas.

3.3.5 Utilizando análise de regressão

A análise de regressão vem a ser utilizada neste trabalho em um ponto onde metodologia SIEGO propõe mudanças para melhorias.

Este ponto onde a análise de regressão é utilizada neste trabalho diz respeito à melhoria que pode ser realizada com relação ao custo onde são observados dados relativos aos produtos mais vendidos na empresa e com margem abaixo do custo do produto, também citado neste trabalho como preço mínimo ou L.A.I (Limite Aceitável Inferior). O objetivo é saber quais variáveis contribuíram de algum modo para que estes produtos fossem vendidos abaixo do custo e então poder traçar ações de melhorias para que nas próximas vendas este custo possa ser minimizado, aumentando a lucratividade da empresa em relação a este fator.

3.3.6 Distribuição de Frequências

Este trabalho se utiliza de distribuições de frequências e tabelas dinâmicas disponibilizadas pelo aplicativo Microsoft Excel no momento em que são feitas análises com relação ao ponto Qualidade abordado pela metodologia SIEGO.

Através destas distribuições o executivo fica sabendo, com relação aos resultados obtidos pelo sistema desenvolvido, quais e quantos clientes estão em cada classificação e

através das tabelas dinâmicas, pode visualizar outras informações referentes a cada variável envolvida no cálculo do índice que podem lhe esclarecer dúvidas e vislumbrar ações de melhorias para que mais clientes possam se enquadrar futuramente em classificações mais elevadas.

3.4 TECNOLOGIAS E FERRAMENTAS UTILIZADAS

Para um melhor entendimento da pesquisa e do trabalho como um todo, pretende-se com este capítulo realizar uma apresentação dos conceitos e algumas características do Software Statistica®, do Microsoft Excel®, do banco de dados Firebird®, da ferramenta DataFlex e também do ambiente Delphi®.

3.4.1 Statistica

O STATISTICA é uma aplicativo que disponibiliza ao usuário uma forma abrangente de análise de dados integrada, gráficos, gerenciamento de base de dados, dispondo de um vasto número de procedimento analíticos básicos e avançados voltado para negócios, mineração de dados, ciência e aplicações de engenharia. A versão utilizada neste trabalho de pesquisa foi a versão 6.0 *trial*.

Todas as ferramentas oferecidas pelo STATISTICA estão dispostas em formas de pacotes de aplicativos integrados e que podem ser selecionados através de uma interface gráfica facilitando assim a utilização por parte do usuário.

O STATISTICA foi utilizado neste trabalho como ferramenta de apoio ao pesquisador com o intuito auxiliá-lo quando da necessidade de verificações mais detalhadas, disponibilizando assim uma forma mais clara para que as conclusões fossem tomadas.

A seguir está uma tela exemplo da aplicação:

The screenshot shows the main interface of the STATISTICA software. The title bar reads 'STATISTICA - [Data: Qualidade.sta* (3v by 36999c)]'. The menu bar includes File, Edit, View, Insert, Format, Statistics, Graphs, Tools, Data, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main window displays a data table with the following columns: '1 OBS', '2 Previsto(a) Índice', and '3 Resíduos padrão'. The data rows are numbered 1 through 17. The status bar at the bottom shows 'Display help for clicked on buttons, menus and windows' and 'CSV3 1,15547813307394 Set:OFF | Weight:OFF CAPS NUM REC'.

	1 OBS	2 Previsto(a) Índice	3 Resíduos padrão
1	1	4,84734306	4,94761807
2	2	5,53815995	2,16701253
3	3	4,14741797	1,59019819
4	4	4,1656493	1,55619392
5	5	5,59850619	2,0544574
6	6	3,7477665	0,843485871
7	7	4,1528979	1,57997727
8	8	4,63949162	0,672404232
9	9	5,18049238	1,15547813
10	10	5,34900045	0,841184364
11	11	4,61174471	0,724156544
12	12	4,14477069	0,103011253
13	13	5,40092027	-0,747778698
14	14	5,54035952	-1,00785461
15	15	5,73366005	0,12373434
16	16	5,45121422	-0,841584745
17	17	6,18582936	0,959008269

Figura 13 – Statística (interface principal)

3.4.2 Microsoft Excel

O Microsoft® Excel é um aplicativo que faz parte da suíte de aplicativos da Microsoft denominado Microsoft® Office e é um programa de aplicação de planilhas de cálculo que segundo Levine et al (2000) é um dos mais adequados para a manipulação interativa de dados numéricos.

O Microsoft® Excel trabalha com planilhas de cálculo eletrônicas que nada mais são que linhas e colunas dispostas de forma retangular nas quais são possíveis introduzir valores, textos, fórmulas, gráficos e demais componentes.

Um fator que contribuiu para a utilização do Microsoft® Excel neste trabalho a sua ampla utilização dentre as empresas em um aspecto global e a facilidade com que os próprios gerentes e tomadores de decisão já trabalham com tal ferramenta. Assim sendo, a empresa em questão também possui o aplicativo em seu ambiente de trabalho, confirmando a afirmação acima. O Microsoft® Excel também possui uma série de ferramentas estatísticas que contribuem para análises realizadas no dia a dia e principalmente nas análises que foram utilizadas neste trabalho. Neste trabalho foram utilizadas com mais assiduidade as funções Analisar Dados e também as funções relacionadas com gráficos e tabelas dinâmicas. A utilização do Microsoft® Excel ainda

permite que relatórios típicos de gerência sejam elaborados através da sua interface baseada em seus aplicativos incorporados. A versão utilizada neste trabalho foi a versão contida no suíte de aplicativos do Microsoft Office Professional 2003®. Maiores informações sobre esta ferramenta que está amplamente difundida no mercado pode ser encontrada em <http://www.microsoft.com>. Assim sendo está justificada a sua utilização neste trabalho.

A seguir está a interface principal do Microsoft® Excel com a apresentação de um arquivo aberto com valores incluídos na planilha e também um gráfico como exemplo.

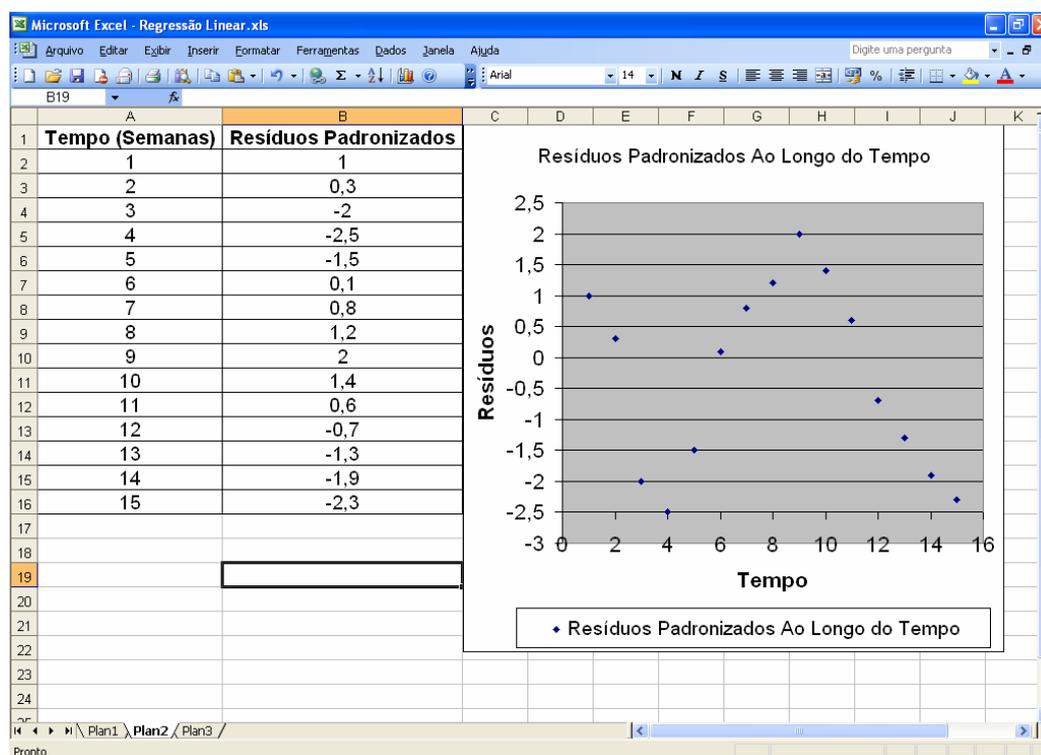


Figura 14 – Interface Microsoft Excel com arquivo aberto.

3.4.3 Banco de dados firebird

Segundo Firebird Project (2006) o Firebird® é um poderoso banco de dados Cliente/Servidor relacional que é compatível com SQL-ANSI, e foi desenvolvido para ser um banco de dados independente de plataformas e de sistemas operacionais, rodando tanto Linux® como em Windows ou outras plataformas Unix.

O Firebird® oferece alta performance, uma poderosa linguagem de suporte para armazenamento de dados e tem sido utilizado no desenvolvimento de sistemas sob uma variedade de nomes diferentes desde 1981.

Além das vantagens de performance e um longo período de atividade no mercado, o Firebird® também oferece algo muito importante que também foi visado para sua utilização neste trabalho que é o fato de ser um banco de dados livre, ou seja, sem a necessidade do pagamento de licenças e que pode ser utilizado em aplicações com fins comerciais ou não.

Este banco de dados dispensa maiores estruturas dentro da empresa, (DBA / Preparação), onde basta instalar o software e usá-lo, sem a interferência freqüente de profissionais especializados na manutenção do banco de dados de produção. O Firebird® também dispensa o uso de servidores extremamente potentes, usando pouco espaço em disco para sua instalação e utilizando pouca memória em situações normais de uso. Por isso a plataforma necessária para a sua instalação e utilização pode ser reduzida diminuindo consideravelmente os custos do projeto. Este é um dos fatores que também motivaram o uso do mesmo.

O Firebird® foi utilizado neste trabalho quando da criação do *Data Warehouse*, pois as características já citadas facilitaram sua criação e administração, tanto por parte do desenvolvedor quanto por parte da linguagem utilizada.

3.4.4 Dataflex

O DataFlex é uma ferramenta desenvolvida pela *Data Access Corporation* e foi lançado em 1981.

De acordo com *Data Access Corporation* (1992), define o DataFlex como sendo um conjunto de programas que auxilia o gerenciamento de dados armazenados no computador. Ele é mais orientado para acesso randômico de dados (tal como uma lista telefônica), do que para um acesso seqüencial de dados (tal como um texto de novela).

Segundo Souza (1988) todas as ferramentas e utilitários estão em um único e compacto bloco de programas e utilitários. O Dataflex é formado por:

- Gerenciador de Banco de Dados Relacional;

- Linguagem de 4ª geração, expansível com macrocomandos;
- Gerador de relatórios e entradas de dados automáticos;
- Query para geração imediata de relatórios individualizados com nove níveis de seleção;
- Compilador de programas;
- Teclas de função redefiníveis em tempo de programação;
- Editor de textos;
- Indexação de acesso seqüencial.

Uma parte importante do DataFlex é seu gerenciador de banco de dados, por essa razão DataFlex é algumas vezes chamado de SGBD (Sistema Gerenciador de Banco de Dados). Por o DataFlex não compreender somente um SGBD, mas também conter uma linguagem completa de desenvolvimento de aplicativos, ele também é conhecido como uma “Linguagem de 4ª geração”, segundo *Data Access Corporation* (1992).

3.4.5 Ambiente Delphi

A ferramenta de desenvolvimento de aplicações *Delphi*®, a qual é um dos produtos da *Borland Inprise Corporation*®, por ser uma linguagem de domínio do grande público, as considerações sobre o *Delphi*® só estão devidas em nível superficial neste trabalho.

A forma de trabalho com o *Delphi*® é a mesma, ou ao menos similar, a todas as ferramentas de desenvolvimento utilizadas com os sistemas operacionais da família *Windows*, onde uma aplicação é construída em torno de formulários, que por sua vez podem conter diversos componentes no padrão *Windows*®, e cada componente recebe uma série de atributos, denominados propriedades; estes componentes podem receber ainda linhas de código que desempenham determinadas tarefas, sendo estes códigos chamados de manipuladores de eventos.

O *Delphi*® possui um ambiente de desenvolvimento integrado onde estão dispostas as ferramentas necessárias para projetar, executar e testar uma aplicação. Entre estas ferramentas aparecem as janelas *main*, *code editor* e *object inspector*.

A janela *main* contém uma barra de menus com acesso a todas as opções do *Delphi*, uma *speedbar* com botões de atalho às funções mais utilizadas (configurável) e a paleta de componentes onde está o acesso aos componentes que podem ser usados nos programas.

A janela *code editor* (editor de código) é onde aparece o código de programa gerado automaticamente pelo *Delphi*® quando algum componente é inserido no programa, e é onde o programador complementa com sua própria programação o código gerado pelo *Delphi*®, customizando a aplicação e inserindo regras de validação.

A janela *object inspector* permite acessar e alterar de forma rápida as propriedades dos componentes e o conteúdo de código dos eventos do componente.

Segundo Cantù (2000), algumas características do *Delphi*® são:

- a) É baseado na linguagem *object pascal*;
- b) Permite a utilização de programação orientada a objetos;
- c) Linguagem RAD compilável e não interpretada;
- d) Usa padrão SQL para acesso a banco de dados;
- e) Possui conectividade via ODBC.

O *Delphi*® acessa um banco de dados via um núcleo de acesso denominado de *Borland Database Engine* (BDE), que permite criar e gerenciar as bases de dados.

A ferramenta *Delphi*® foi escolhida para desenvolver o aplicativo devido ao fato de já conhecê-la durante algum tempo, apresentada na disciplina programação no curso de Bacharel em Ciências da Computação da FURB, e por ser uma ferramenta que permite desenvolvimento rápido, gerando produtividade no nível desejado. É através desta ferramenta que o sistema de informação é desenvolvido, realizando a interface com o usuário, carregando os dados e também os exportando para o Excel para posterior análise.

Após esta breve apresentação das ferramentas utilizadas para realização deste trabalho de pesquisa, a seção a seguir apresenta os resultados que foram obtidos através da aplicação da associação das abordagens apresentadas nas seções anteriores, bem como uma apresentação da empresa onde a pesquisa foi aplicada.

4 RESULTADOS

Esta seção tem como objetivo apresentar o ambiente onde a pesquisa foi aplicada, bem como os resultados obtidos em relação à aplicação das metodologias citadas no capítulo 3 deste trabalho, demonstrando esta aplicação e as análises realizadas através das mesmas, construindo uma base para o tópico de discussão e considerações finais, onde serão levantadas questões que surgiram através da pesquisa.

4.1 O AMBIENTE DE PESQUISA

Este tópico tem como objetivo apresentar o ambiente de pesquisa demonstrando a área de atuação da empresa bem como as partes que a compõem.

4.1.1 Apresentação da empresa

A empresa em questão é uma empresa eletromecânica situada na cidade de Blumenau e que foi fundada em 12 de dezembro de 1910. Sua área de atuação está limitada aos produtos dos ramos de ferragens e materiais elétricos, dividindo-se então seus departamentos de vendas de acordo com estas áreas de atuação.

Sendo dividida atualmente em 5 empresas que fazem parte do mesmo grupo e estão alocadas como sendo uma matriz e mais quatro filiais situadas respectivamente em Blumenau, Gaspar, Indaial e Rio do Sul, a empresa conta com cerca de aproximadamente 80 funcionários que estão lotados nos seguintes departamentos: financeiro, compras, vendas (televendas, lojas – balcão, externo), assistência técnica, departamento de logística e centro de processamento de dados.

Desta forma, estes departamentos são representados por denominações específicas (criadas pela própria empresa) que serão apresentadas a seguir e passarão a ser consideradas a partir deste momento. São elas:

- DEFI (departamento financeiro)
- DECO (departamento de compras)
- DEVE (departamento de vendas)
 - TELEV (televendas)
 - LOJA (vendas balcão)
 - EXTERNO (vendas realizadas fora da empresa)
- AT (assistência técnica)

- DELOG (departamento de logística)
- CPD (centro de processamento de dados)

O DEFI é o departamento responsável por toda a parte que envolve dinheiro propriamente dito, pois ali estão alocados funcionários que lidam com contas a pagar, também existem funcionários ligados às contas a receber e outros que estão ligados aos serviços bancários diários, subdivididos em um setor dentro do departamento financeiro, denominado Tesouraria.

Dentro do DEFI – Contas a receber, os funcionários lidam com os mais diferentes tipos de clientes, sendo eles pessoas físicas ou jurídicas. O volume de vendas da empresa é bastante intenso, tornando inviável realizar as cobranças de forma manual ou sem o auxílio de tecnologia. Sendo assim o DEFI se utiliza de alguns convênios com bancos para troca de dados via sistema *on-line* tornando a tarefa de cobranças um trabalho mais rápido e seguro.

O DEFI – Contas a pagar também se utiliza destas parcerias com os bancos para tornar sua tarefa mais ágil e segura, efetuando seus pagamentos de dentro da empresa também utilizando sistemas *on-line*. Assim sendo, apenas as operações que não são efetuadas via sistema são então designadas a um funcionário denominado *Office Boy* que as efetua diretamente em cada banco ou instituição específicos. Este sistema também é alimentado por informações advindas do departamento de logística, quando da entrada de mercadorias.

Apesar dos diversos convênios com bancos, alguns clientes ainda preferem realizar seus pagamentos diretamente no caixa da empresa, podendo fazê-lo a qualquer momento, pois a empresa também possui um sistema integrado com seus caixas possibilitando tal operação e baixando as duplicatas automaticamente no ato do pagamento. Estes caixas também são responsáveis por parte dos faturamentos à vista e a prazo da empresa.

Devido ao imenso volume de vendas, o DEFI também faz um controle das vendas a prazo que são realizadas na empresa, tendo assim um controle melhor de sua inadimplência, pois avaliam cada venda realiza para seus clientes. Obviamente este controle seria muito trabalhoso se não houvesse algumas premissas que o tornam mais ágil. Para isso, cada cliente já cadastrado possui um pré-limite de crédito que muitas vezes dispensa a aprovação dos pedidos, facilitando e agilizando o processo de vendas.

Este limite de crédito então é decrescido à medida que o cliente vai comprando e acrescido quando ele vai efetuando seus pagamentos. Ainda para o auxílio à aprovação de pedidos, a empresa conta com parcerias com empresas como SCI (Serviço de Crédito e Informação) que permite consultas cadastrais *on-line* através da internet e também com o SPC (Serviço de Proteção ao Crédito) no caso de pessoas físicas.

O DECO é o departamento responsável por alimentar a empresa na questão dos produtos necessários para o atendimento perfeito aos seus clientes e pela busca de fornecedores parceiros que agilizem este processo. Observando a linha de atuação da empresa, que se constitui de materiais elétricos e ferragens, este departamento também se subdivide nestas áreas, tornando assim esta tarefa mais precisa e melhor gerenciada por cada membro responsável pela compra dos respectivos itens de cada área.

O cadastro de produtos e fornecedores fica a cargo dos compradores que possuem a sua disposição um sistema que lhes possibilita tal tarefa. Este sistema também está integrado com os outros sistemas da empresa, constituindo o sistema de gestão da empresa. Tal sistema também oferece aos compradores relatórios e sistemas de cálculos com fórmulas já implementadas que lhes auxiliam no momento de efetuar novas compras.

Outro departamento de fundamental relevância dentro da empresa é o DEVE, pois é através dele que a empresa realiza a venda de seus produtos.

Este departamento está subdividido em três partes sendo elas:

- TELEV: Esta divisão do DEVE é responsável pelo atendimento aos clientes que ligam para a empresa para efetuar uma compra ou apenas uma consulta, resultando na venda efetiva ou em um orçamento para uma possível venda futura. A empresa possui uma equipe qualificada para realizar este tipo de atendimento.
- LOJA: Esta divisão está presente na empresa matriz e também em todas as suas filiais. Quando um cliente se dirige diretamente ao espaço físico da empresa (loja) e se reporta ao vendedor que está a sua espera, então a venda é efetuada por este colaborador que faz parte da equipe de vendas LOJA.
- EXTERNO: A empresa possui ainda uma equipe de promotores de vendas que trabalham externos ao ambiente da empresa e que

possuem um parceiro de vendas dentro da empresa. Estes então atuam em conjunto, sendo que este promotor de vendas externo faz o trabalho de atender os clientes que necessitam de uma visita mensal ou semanal, bem como aqueles que têm suas instalações fora da área urbana ou em outras cidades do estado. Após atendê-lo este profissional pode então repassar as necessidades dos clientes para seu companheiro de equipe que atua dentro da empresa, prestando assim um atendimento personalizado ao cliente.

Além da equipe de vendas a empresa também possui um departamento de assistência técnica, denominado AT que apresenta um diferencial aos seus clientes, pois a empresa oferece assistência à parte dos produtos que comercializa, refletindo mais segurança aos seus clientes. Este departamento também possui um sistema que gerencia suas ordens de serviço e que está integrado aos outros sistemas que fazem parte do sistema de gestão da empresa.

O departamento denominado como DELOG atua em todos os momentos da empresa, pois a logística atua desde a entrada (compra) das mercadorias até a saída (venda) das mesmas. É através deste departamento que as notas são faturadas e as mercadorias despachadas para seus respectivos clientes. Os pedidos fechados pelo DEVE acabam sendo encaminhados para o DELOG para que então chegue até o cliente.

A empresa possui ainda um departamento responsável pela área de informática da empresa, que é o CPD. É neste departamento que é desenvolvido o sistema de gestão da empresa. Todos os sistemas anteriormente mencionados fazem parte do grupo de sistemas desenvolvidos pelo CPD e que trabalham de maneira integrada. Todas as alterações, manutenções e novas criações em termos de sistemas são desenvolvidas neste departamento. Toda a parte relacionada ao hardware da empresa também é de responsabilidade do CPD, eximindo-se apenas os assuntos relacionados ao conserto e reposição de peças. Neste departamento estão alocados dois analistas programadores.

A empresa ainda possui um site institucional que também é de responsabilidade do CPD.

Sendo assim, fica caracterizada a empresa em função dos departamentos envolvidos juntamente com suas funcionalidades.

Devido ao seu imenso número de clientes (mais de 36.000) e a imensa quantidade de diferentes itens de venda (mais de 30.000) o seu movimento de vendas

conseqüentemente é muito grande gerando assim também uma dificuldade na análise dos dados da empresa em função da necessidade da tomada de alguma decisão.

Como citado anteriormente, a empresa possui um sistema de gestão interno que é desenvolvido pela própria empresa e este sistema por conseqüência captura dados de todas as movimentações realizadas através dele. Estes dados são então armazenados em uma base de dados geral.

O trabalho em questão vem através dos recursos apresentados, mostrar ao executivo um modo de extrair as informações desta base de dados utilizando técnicas que lhe permitam uma melhor análise de tais dados. O motivo deste desenvolvimento em específico se dá justamente pelo fato de que os sistemas específicos da empresa podem acessar seus dados, porém existem muitos sistemas específicos para cada informação, o que torna a ação do executivo um tanto exaustiva na busca pelas informações desejadas e dificulta uma visão geral da realidade dos negócios e a conseqüente tomada de decisões. Um sistema que englobe já as informações necessárias em um único local de acesso facilita a ação do executivo.

Em virtude da dificuldade citada, foi utilizada a associação das abordagens descritas na próxima seção.

4.2 APLICAÇÃO DA METODOLOGIA SIEGO

Para a aplicação desta metodologia seguiram-se fielmente os passos apresentados no capítulo 3 de acordo com o que se segue:

Na fase I que relata a preparação do projeto, foram seguidos os cinco passos para sua contemplação onde no primeiro passo foi definido o grupo de trabalho que foi composto pelo acadêmico Rodrigo Reno Strey como sendo analista/programador da empresa em questão, também pelo professor Dr. Marcelo Menezes Reis atuando como orientador deste trabalho e para finalizar a equipe, o gerente comercial da empresa em questão, com o qual foram realizadas reuniões para definição de algumas variáveis e também para sanar algumas dúvidas.

O passo seguinte cuja definição de processos é seu objetivo, foi contemplado através da utilização das informações levantadas no passo anterior que permitiram escolher os processos que vão desde a carga de dados até quais seriam as áreas ou pontos avaliados para que os dados fossem coletados corretamente e que pudessem atingir os

objetivos da metodologia SIEGO que se relaciona com o tripé (custo, tempo e qualidade). Assim definiu-se que o processo de carga de dados seria feito através da varredura de dados no banco de dados original da empresa, onde também seriam feitos alguns cálculos após os dados já coletados, montando assim o *Data Warehouse* apenas com os dados pertinentes as necessidades do executivo.

O terceiro passo que atua na motivação e instrução do grupo de trabalho foi facilmente contemplado, pois os processos já estavam definidos e então o grupo começou a trocar as primeiras idéias sobre o que poderia ser feito com os mesmos para produzir algo útil ao executivo, onde então as diversas opiniões e experiências puderam ser colocadas.

O quarto passo da fase I que se refere ao planejamento da implantação das idéias discutidas no terceiro passo, foi contemplado através de discussões em que o gerente da empresa em acordo com o acadêmico definiu a possibilidade de implementação na própria empresa.

No quinto passo desta primeira fase, ocorreu a preparação para o acompanhamento das idéias onde foi definido que o professor orientador, juntamente com o acadêmico, estaria acompanhando em conjunto a implantação das idéias.

Com a fase I completada, o trabalho passa à fase II onde ocorre a determinação e avaliação das ações de melhorias. Nesta fase, que é constituída de 3 passos, o primeiro passo retratou a montagem do banco de dados, que após conversas com a equipe de trabalho foi decidido que os dados viriam do banco de dados original em Dataflex e então seriam filtrados e calculados de acordo com as necessidades do executivo para compor uma nova base de dados, constituindo o *Data Warehouse* (segunda abordagem deste trabalho).

O segundo passo desta fase, que se refere ao desenvolvimento e avaliações das ações de melhorias, foi contemplado através de reuniões com o gerente da empresa e também com o professor orientador onde foram filtradas alguma idéias e seguidas apenas aquelas que fossem factíveis.

A partir do segundo passo foi então no terceiro passo que as idéias em potencial puderam ser mais bem avaliadas pela equipe de trabalho e complementadas na medida

do necessário. Assim, contemplando mais esta fase, foi possível passar ao grupo de trabalho uma visão mais exata das implicações do projeto e sua validade.

A última fase do SIEGO que retrata a implementação das idéias foi contemplada através de seus dois passos, onde o primeiro deles refere-se ao planejamento da implantação onde então foi determinado o que seria feito em relação ao tripé (custo, tempo e qualidade), sendo que com relação ao Custo, foi decidido que seriam apresentadas informações sobre os 50 maiores clientes da empresa de acordo com o ano escolhido, onde seriam mostrados de cada uma destas empresas quais os itens haviam sido vendidos a um preço abaixo do limite aceitável inferior, conhecido na base de dados como preço mínimo.

Com relação ao Tempo ficou decidido que seriam apresentadas informações relativas aos valores de faturamento dentro do período determinado na introdução deste trabalho e também segmentado por filiais, onde poderiam ser feitas também visualizações gráficas para estes valores.

O item qualidade da metodologia SIEGO é contemplado através da verificação da qualidade dos clientes, baseada em suas compras e em uma série de parâmetros que foram definidos com a gerência da empresa e também cálculos baseados nestes parâmetros que gerariam um índice classificatório para cada cliente. Assim o executivo pode visualizar a classificação de seus clientes, baseado em suas próprias definições de qualidade. A apresentação destes resultados segue no item mais abaixo que trata a apresentação do sistema desenvolvido.

O segundo passo que trata a questão da implementação e rastreamento das ações de melhorias foi então associada com os métodos estatísticos para que pudessem ser mostrados os pontos onde o executivo poderia visualizar melhorias com base na informação coletada. Assim ao trabalhar a implementação da metodologia SIEGO, aplica-se o *Data Warehouse* e posteriormente os métodos estatísticos, conforme proposto inicialmente.

4.3 APLICAÇÃO DO DATA WAREHOUSE

Neste tópico pretende-se apresentar de forma concisa como foi a implementação do *Data Warehouse* neste trabalho, esclarecendo sua importância inclusive.

Conforme apresentado no item 3.3.3.1, a montagem do Data Warehouse, segundo Kimball (1996) e Kimball (1997) é feita em nove etapas, então se seguiu da seguinte forma:

A primeira etapa (escolha do processo) já foi contemplada na fase de preparação do projeto da metodologia SIEGO. Na segunda etapa (escolha do nível de granularidade) foi decidido que o nível de granularidade permitiria ao executivo realizar uma busca por ano (sendo este o nível do grão) quando relacionado ao item custo e realizar consultas por mês/ano (sendo este o nível do grão) e também por filiais (sendo este o grão) quando relacionado ao item tempo, contemplando assim um nível de granularidade aceitável (busca por ano/mês e filiais) por parte do gerente da empresa. A terceira etapa (identificar e conformizar as dimensões) foi contemplada facilmente ao definir quais seriam as colunas (dimensões) que apresentariam dados que seriam visualizados pelo executivo ao executar o aplicativo e visualizar os relatórios, por exemplo. Exemplos disso são descrições de produtos, classificações dos clientes, etc.

Ao entrar na quarta etapa (escolha dos fatos) da implementação do *Data Warehouse* então foi decidido, conforme reuniões prévias da metodologia SIEGO, quais seriam os dados relevantes que seriam captados do banco de dados original e que fariam parte do *Data Warehouse*. A quinta etapa trata exatamente de pré-calculado alguns dados e já carregar seus resultados para consultas posteriores. Na sexta etapa são preenchidas as dimensões identificadas na terceira etapa (alimentação dos campos). Na sétima etapa (escolha da duração do banco de dados) definiu-se que a tabela de fatos contemplará um período de 12 anos para valores de faturamento e 5 anos para registro de compras de clientes. Ao implementar a oitava etapa, foi decidido que seriam preparadas as dimensões com a inserção de campos adicionais para a necessidade de modificações lentas que pudessem vir a ocorrer.

Na nona e última etapa (frequência de extração e carga de dados) foi decidido que a carga de dados seria realizada mensalmente e apenas uma vez no dia, no horário que excede as 18:00h e vai até as 23:00h do mesmo dia.

Através da contemplação destas nove etapas, o *Data Warehouse* está montado e pronto para ser utilizado através da aplicação desenvolvida. Sua utilização facilita bastante na hora em que o executivo precisa realizar suas consultas, pois conforme já citado, os dados já foram pré-calculados/filtrados ao serem inseridos nas tabelas de fatos

do *Data Warehouse*, evitando assim que em cada consulta do executivo o mesmo tenha que ficar aguardando o aplicativo realizar os cálculos e filtros. Este é um benefício extremamente significativo oferecido pela adoção do *Data Warehouse*. Outro fator interessante que é importante frisar é justamente a questão da granularidade onde o executivo mesmo pode influenciar sugerindo o nível de granularidade que seria interessante para ele (seria qual o nível de detalhe das informações que ele deseja), conforme consta na implementação da segunda etapa. A questão do cubo de decisão também é interessante ressaltar pois o executivo também pode influenciar na questão de quais dimensões o mesmo gostaria de poder consultar e realizar as alterações entre as mesmas.

Assim, depois de contempladas os passos para implementação da metodologia SIEGO e também as 9 (nove) etapas para aplicação do Data Warehouse, segue na próxima seção a apresentação do sistema.

4.4 APRESENTAÇÃO DO SISTEMA

Este tópico visa apresentar a implementação do sistema para demonstrar a interface com o usuário e também os pontos abordados nos itens anteriores que foram implementados na prática através deste sistema.

Inicialmente o sistema apresenta apenas uma *splashscreen* que é uma tela onde aparecem informações sobre este trabalho, como a instituição, título do mesmo e acadêmico, entre outras, que apenas servem de caráter informativo. Veja conforme figura 15 a seguir:

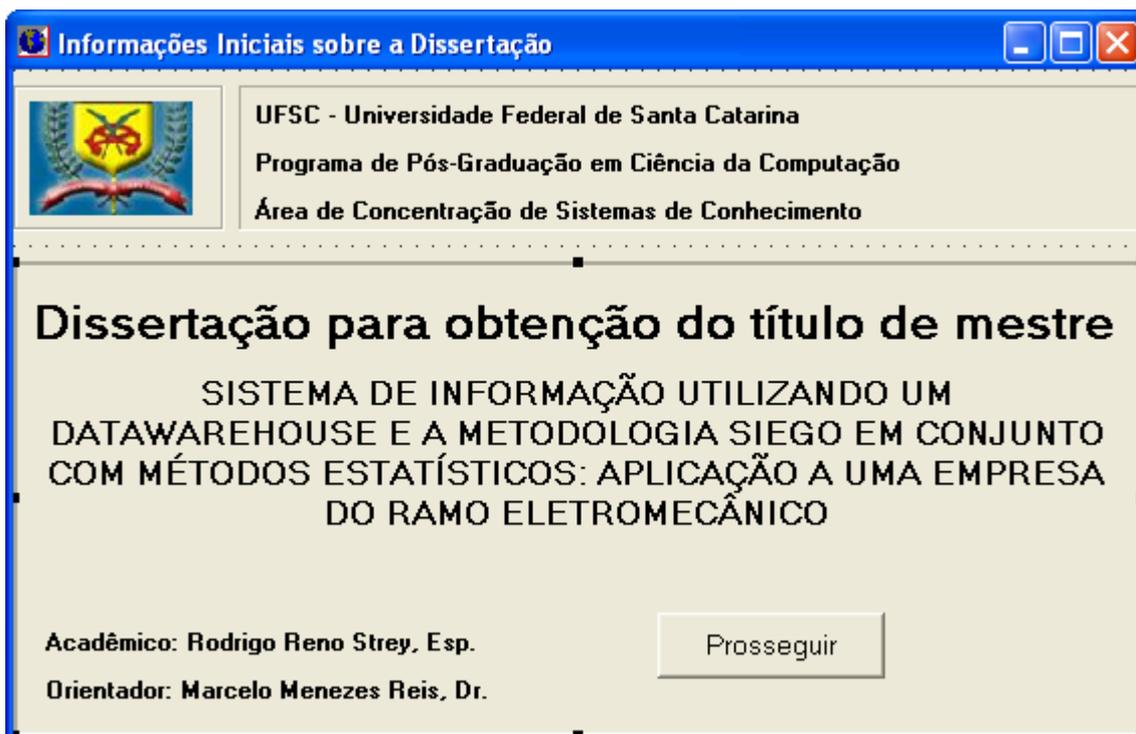


Figura 15 – Tela inicial do sistema apresentada ao iniciar execução

A tela apresentada através da figura 16 demonstra como é realizada a carga de dados no sistema, onde é possível preencher campos referentes ao período em que a pesquisa no banco de dados original se dará para depois realizar alguns pré-cálculos em cima destes dados pesquisados e então carregar os dados nas tabelas de fatos definidas na implementação do *Data Warehouse*, conforme item 4.3 deste trabalho.

Como a carga de dados não é algo instantâneo em função do volume de dados a pesquisar e também em virtude dos filtros a serem realizados, o sistema apresenta ao usuário uma barra de progressão e também um painel indicativo de quais tabelas de fatos estão sendo carregadas para que o usuário possa ter noção de qual é a fase da carga de dados em que o sistema se encontra.

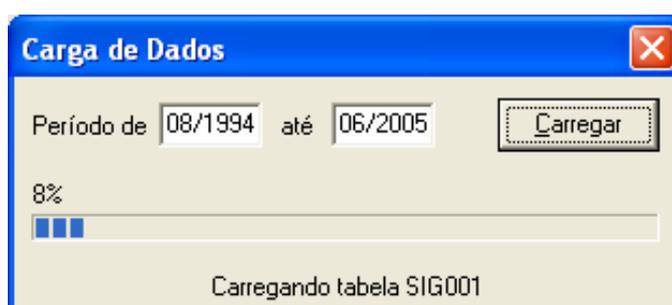


Figura 16 – Realização da Carga de Dados

Ao realizar os cálculos para apresentação dos dados referentes ao item Qualidade, o sistema adota alguns parâmetros, que são mencionados pela metodologia SIEGO, conforme item 3.3.2.3.1 como sendo tela de controles, onde o executivo define os quesitos para os limites de classificação dos clientes como sendo ótimos, bons, ruins ou regulares. Aí então estes dados são gravados e levados em consideração na rotina que classifica os clientes. Esta tela de parâmetros pode ser visualizada através da figura 17 a seguir:

Classificação	Valor de Venda (%)		Margem de Venda (%)		Atraso Médio (dias)	
	Máximo	Mínimo	Máximo	Mínimo	Máximo	Mínimo
Ótimo	100,00	90,00	25,00	20,00	0	1
Bom	89,99	65,00	24,99	20,00	7	1
Regular	64,99	35,00	19,99	15,00	30	8
Ruim	34,99	0,00	14,99	0,00	31	0

Gravar

Figura 17 – Tela de parâmetros para utilização na análise de Qualidade

Após o executivo então definir estes parâmetros, que foram previamente discutidos nas reuniões que foram mencionadas pelo item 4.2 deste trabalho que relata a aplicação da metodologia SIEGO, o sistema realiza alguns cálculos envolvendo pesos para cada um dos parâmetros e então aponta um índice para cada cliente de acordo com as suas compras, as margens de vendas nestas compras e também o atraso médio para o pagamento destas contas. Assim, ficou decidido que a classificação dos clientes ficaria condicionada segundo as faixas apresentadas na tabela 1 abaixo:

Classificação	Faixa do índice
Ótimo	entre 10 e 08
Bom	entre 7,99 e 6
Regular	entre 5,99 e 4
Ruim	entre 3,99 e 0

Tabela 01 – Faixas para classificação dos clientes com relação à Qualidade

Assim sendo, quando o executivo verifica o item Qualidade apresentado pelo sistema, lhe é apresentada uma tela onde é possível visualizar os clientes que estão na base de dados e verificar quais as suas classificações de acordo com os parâmetros previamente estabelecidos. Veja a seguir a apresentação desta tela, representada na figura 18:

Nome	Valor Total Comprado (%)	Margem Média (%)	Atraso Médio (Dias)	Índice	Classificação
▶ PORTOBELLO S/A	100,00	11,30	0	7,50	Bom
SADIA S/A	60,30	16,73	0	6,67	Bom
METISA MET. TIMBOENSE S/A	60,12	-0,02	0	5,00	Regular
WEG INDUSTRIAS S/A. - TRA	57,07	-0,02	0	5,00	Regular
SOUZA CRUZ S/A	48,99	16,56	0	6,67	Bom
MALWEE MALHAS LTDA	48,10	-3,08	1	4,17	Regular
ELECTRO ACO ALTONA S/A	41,30	1,29	1	5,00	Regular
ZEN S/A IND. METALURGICA	40,04	7,04	1	5,00	Regular
NETZSCH DO BRASIL IND. E	39,70	10,86	0	5,83	Regular
WALTEC ELETRO-ELETRON. LT	38,02	12,84	0	5,83	Regular
TOALIA S/A IND. TEXTIL	32,64	3,55	0	5,00	Regular
INDL. REX LTDA	32,46	0,55	1	4,17	Regular
COOP. REG. ALFA LTDA	30,55	15,52	1	5,00	Regular
KARSTEN S/A	29,78	14,53	0	5,00	Regular
PALMASOLA S/A - MADEIRAS	29,48	16,77	0	5,83	Regular
WEG INDUSTRIAS LTDA. - MO	29,44	13,39	0	5,00	Regular
M G M - AGRO FLORESTAL I	27,72	22,11	0	6,67	Bom
CREMER S/A	27,35	12,15	1	4,17	Regular
SADIA S/A	26,83	14,98	0	5,00	Regular
ALBANY INTERNATIONAL FELT	23,80	22,74	0	6,67	Bom
FAVOR NAO USAR ESTE CODIG	22,69	24,58	0	6,67	Bom
WALTEC ELETRO-ELETRON. LT	21,69	29,09	0	7,50	Bom
Z.M. S/A	20,52	5,01	0	5,00	Regular
NATIONAL STARCH & CHEMICA	20,01	14,09	0	5,00	Regular
TEXTIL RENAUX S/A	19,91	24,78	1	5,83	Regular
MEC. STORRER LTDA	18,54	6,09	0	5,00	Regular

Figura 18 – Apresentação dos clientes classificados de acordo com definições da tela de parâmetros

Ao abordar o item Custo, mencionado pela metodologia SIEGO como sendo um dos itens do tripé, o sistema apresenta uma tela onde o executivo pode escolher o ano de pesquisa ou todos os disponíveis para realizar uma consulta dos clientes que mais compraram na empresa naquele período, relatando quais foram os itens comprados por estes clientes que foram vendidos abaixo de uma Limite Aceitável Inferior, apresentado aqui como Preço Mínimo, gerando um possível déficit para a empresa em relação à venda destes produtos.

O sistema lista os clientes em ordem crescente de compras e na parte inferior da tela mostra quais foram então os itens que foram vendidos abaixo do L.A.I., conforme mencionado anteriormente.

A figura 19 representa a seguir esta interface com o executivo onde o mesmo pode realizar a sua escolha.

Informe o ano da pesquisa: 2005

Carregar

Cientes

Nome do Cliente

- TODOS
- 2000
- 2001
- 2002
- ▶ METISA MET. TIMBOEN
- 2003
- 2004
- 2005
- ELECTRO ACO ALTONA S/A
- MALWEE MALHAS LTDA
- GELNEX IND. E COM. LTDA
- NETZSCH DO BRASIL IND. E COM. LTDA

Items abaixo do L.A.I

Número da Nota	Série	Data da Venda	Descrição	Prazo Médio (Dias)	Quantid
▶ 309710	1	18/1/2005	ELETROD.G.FOGO +LUVA NBR 5598 1"x3,0mt 2,65mm	45	
309847	1	19/1/2005	DISJUNTOR MOTOR 3RV10 11-1CA10 1,8-2,5A S/CONTATC	35	
309847	1	19/1/2005	PLUG INDL. TRIF. 32A 3P+T 380V N-4276	35	
309847	1	19/1/2005	TOMADA INDL. EMB. TRIF. 32A 3P+T 380V N-4246	35	
309847	1	19/1/2005	CONDULETE ALUMINIO C/ ROSCA E TAMPA C-10 1/2"	35	
309847	1	19/1/2005	CONDULETE ALUMINIO C/ ROSCA E TAMPA LB-10 1/2"	35	
309847	1	19/1/2005	ELETROD.G.FOGO +LUVA NBR 5598 1"x3,0mt 2,65mm	35	
309847	1	19/1/2005	ELETROD.G.FOGO +LUVA NBR 5598 3/4"x3,0mt 2,25mm	35	
310027	1	20/1/2005	FUSIVEL NH-000 3NA3 824 80 A.	35	
310027	1	20/1/2005	CONECTOR MACHO RECUPERAVEL AL. C.M.R.A 3/4" SPT	35	
310027	1	20/1/2005	CONDULETE ALUMINIO C/ ROSCA E TAMPA LB-15 3/4"	35	

Exportar para o Excel

Figura 19 – Tela para verificação do item Custo da metodologia SIEGO

Ao abordar o item Tempo, que também faz parte do tripé mencionado pela metodologia SIEGO, é possível fazer verificações dos valores de faturamento da empresa de acordo com o período escolhido pelo executivo. Também é possível realizar em tempo real a inclusão ou exclusão de uma das empresas da tela de consultas, proporcionando ao executivo controle sobre as dimensões que ele deseja visualizar, conforme relatado no item 4.3 que trata sobre o *Data Warehouse*.

A figura 20 apresenta a seguir a tela inicial do item Tempo que é apresentada ao executivo para que o mesmo possa escolher o período de consulta bem como a empresa que deseja visualizar os valores. O sistema então apresentará os valores de vendas individuais, por mês e ano de acordo com o período e as empresas escolhidas previamente e mais um valor somatório que agrega os valores de todas as empresas envolvidas na consulta, separados por mês. Veja a seguir a consulta de valores de venda realizada no período de janeiro de 1995 à dezembro de 2005, constando apenas a empresa Matriz e uma filial localizada em Rio do Sul. É importante frisar que os valores

estão atualizados de acordo com a inflação através da aplicação do índice IGP-M conforme já citado no item 3.3.4 deste trabalho. :

Mês	Ano	Rio do Sul (R\$)	Blumenau Matriz (R\$)	Total
1	1995	239.676,03	1.818.309,07	2.057.985,10
2	1995	248.446,90	1.623.505,12	1.871.952,01
3	1995	295.226,96	2.314.024,24	2.609.251,20
4	1995	211.781,58	1.760.231,22	1.972.012,80
5	1995	287.083,74	1.637.010,22	1.924.093,96
6	1995	246.425,66	1.482.134,26	1.728.559,92
7	1995	194.792,54	1.362.979,82	1.557.772,36
8	1995	216.498,69	1.360.728,80	1.577.227,49
9	1995	194.074,16	1.204.653,47	1.398.727,63
10	1995	181.936,24	1.339.809,02	1.521.745,26
11	1995	193.911,14	1.191.964,89	1.385.876,03
12	1995	211.162,26	1.205.753,54	1.416.915,80
1	1996	217.760,63	1.081.309,78	1.299.070,41
2	1996	176.529,18	1.210.809,30	1.387.338,48
3	1996	196.965,23	1.294.315,64	1.491.280,86

Figura 20 – Tela de consulta de faturamento em relação ao tempo e por empresa

A figura 21 apresentada adiante demonstra novamente a consulta apresentada anteriormente, porém demonstrando a capacidade do executivo poder acrescentar mais uma empresa na consulta sem a necessidade de carregar os dados novamente, uma vez que os mesmos já foram previamente lançados no *Data Warehouse*. Veja:

Mês	Ano	Rio do Sul (R\$)	Chapecó (R\$)	Blumenau Matriz (R\$)	Total
1	1995	239.676,03	119.211,33	1.818.309,07	2.177.196,42
2	1995	248.446,90	91.463,77	1.623.505,12	1.963.415,78
3	1995	295.226,96	124.109,04	2.314.024,24	2.733.360,25
4	1995	211.781,58	119.784,54	1.760.231,22	2.091.797,34
5	1995	287.083,74	127.680,10	1.637.010,22	2.051.774,05
6	1995	246.425,66	95.568,81	1.482.134,26	1.824.128,73
7	1995	194.792,54	105.982,48	1.362.979,82	1.663.754,84
8	1995	216.498,69	112.372,40	1.360.728,80	1.689.599,89
9	1995	194.074,16	128.297,85	1.204.653,47	1.527.025,48
10	1995	181.936,24	113.295,79	1.339.809,02	1.635.041,05
11	1995	193.911,14	156.818,50	1.191.964,89	1.542.694,53
12	1995	211.162,26	113.800,40	1.205.753,54	1.530.716,19
1	1996	217.760,63	139.444,82	1.081.309,78	1.438.515,23
2	1996	176.529,18	132.915,74	1.210.809,30	1.520.254,22
3	1996	196.965,23	155.630,95	1.294.315,64	1.646.911,82

Figura 21 – Apresentação de valores com acréscimo de uma empresa, Chapecó

No item Tempo, além de poder visualizar os valores de venda através da tela apresenta anteriormente, o executivo pode também visualizar estes dados através de gráficos que possibilitam uma melhor análise da situação, uma vez que os dados em forma numérica dificultam um pouco em virtude da quantidade de valores apresentados e a cada empresa que se acresce, esta dificuldade aumenta.

A figura 22 apresenta os dados através do gráfico de pizza:

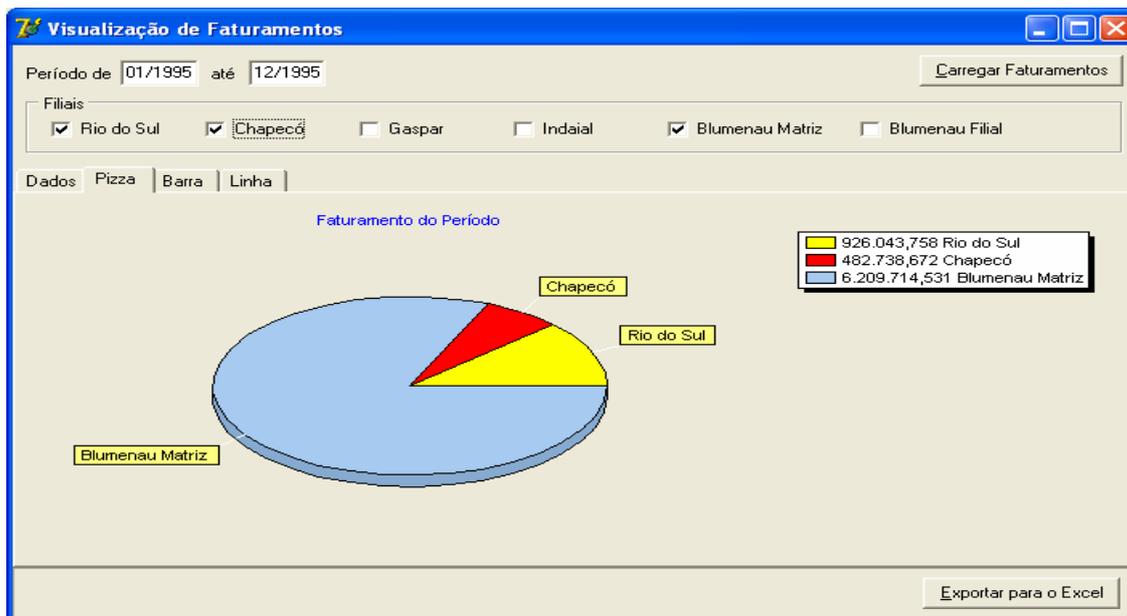


Figura 22 – Utilização do gráfico de pizza para apresentação dos dados

Além do gráfico de pizza, o sistema também proporciona uma visualização dos valores totais de vendas realizadas pelas empresas da consulta e que são apresentados através do gráfico de barras, conforme a figura 23 a seguir:

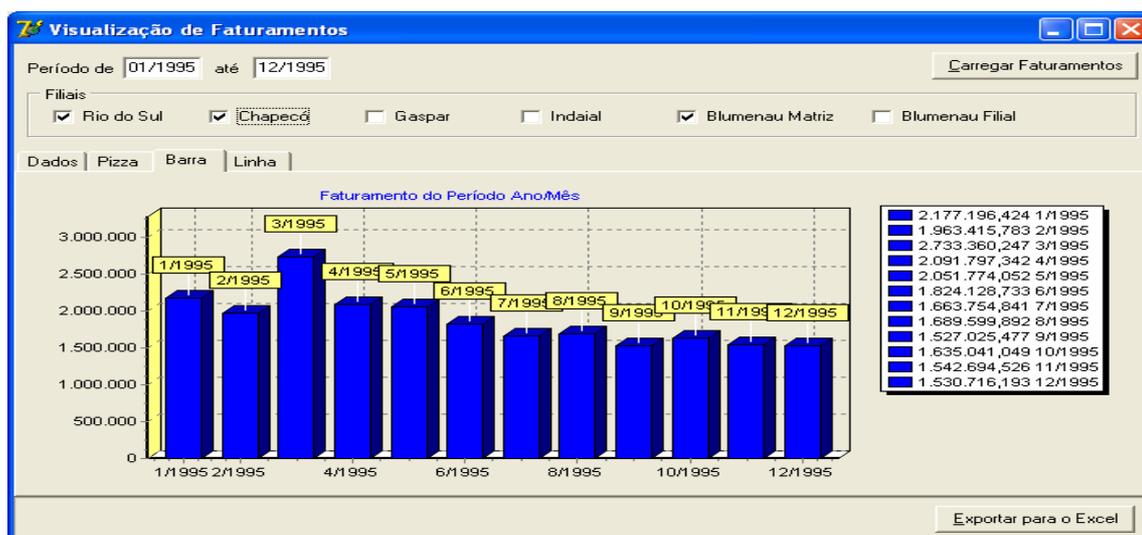


Figura 23 – Utilização do gráfico de barras para valores totais mês a mês

A figura 24 representa a tela onde podem ser visualizados os desempenhos de cada empresa da consulta em relação ao período pesquisado, onde através do gráfico de linhas fica mais fácil analisar o desempenho de cada empresa.

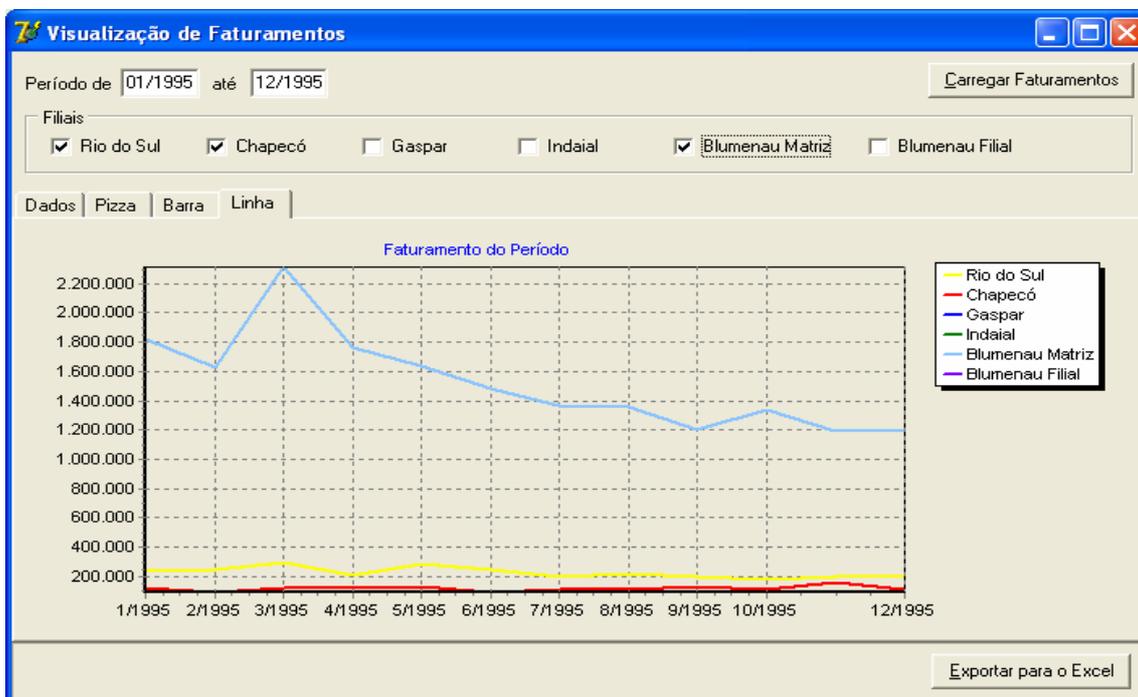


Figura 24 – Utilização do gráfico de linha para visualizar o desempenho de cada empresa em relação aos seus faturamentos

A figura 25 apresenta a implementação da técnica de cubo de decisão onde o executivo pode variar suas consultas em tempo de execução entre as variáveis cliente, prazo médio, data da venda obtendo os valores de venda para cada cliente

Cliente	Data da Venda	Prazo Médio	Valor
3 SEC SERVICOS L	2005	28	188,200000762939
		Sum	188,200000762939
	Sum		188,200000762939
43 S/A GRAF. E ED	2005	7	7,28000020980835
		28	24
		30	142
		Sum	173,280000209808
	Sum		173,280000209808
A. ANGELONI & CIA	2005	28	411,299999237061
		Sum	411,299999237061
	Sum		411,299999237061
A. ANGELONI & CIA	2005	28	143,800003051758
		Sum	143,800003051758

Figura 25 – Aplicação da técnica de Cubo de Decisão

A técnica de granularidade foi empregada nesta parte do sistema. Propiciou ao executivo visualizar os faturamentos de diversas empresas e também agrupados por valores mensais e anuais, contemplando assim o nível necessário de granularidade esperada pelo executivo.

Ao apresentar este sistema, todas as telas correspondentes aos itens do tripé (custo, tempo e qualidade) da metodologia SIEGO, continham em sua parte inferior um botão que tem a finalidade de exportar os dados apresentados pelo sistema para uma planilha do Microsoft Excel para que então possam ser aplicados os métodos estatísticos mencionados neste trabalho com a finalidade de se realizarem as análises necessárias para que posteriormente o executivo possa tomar decisões de melhorias conforme proposta da metodologia SIEGO.

A seção a seguir mostra as análises realizadas, bem como os métodos utilizados e também as conclusões sobre os dados exportados.

4.5 APLICAÇÃO DOS MÉTODOS ESTATÍSTICOS

Esta seção visa demonstrar como os métodos mencionados neste trabalho foram utilizados e quais resultados foram obtidos através de sua aplicação.

4.5.1 Aplicação da análise de regressão

Conforme mencionado no item 3.3.5 a regressão foi utilizada em um dos pontos deste trabalho. Este ponto está relacionado a um dos itens do tripé apontado pela Metodologia SIEGO que é o Custo.

Na aplicação em relação ao item Custo, foram observadas três empresas das 50 empresas listadas pelo sistema, onde foram realizadas análises semelhantes para cada empresa na busca por saber quais variáveis realmente tinham alguma influência sobre as vendas realizadas abaixo do L.A.I (preço mínimo) dentre todas apresentadas.

Os resultados obtidos são apresentados a seguir.

Ao realizar a pesquisa o aplicativo desenvolvido reportou 7007 itens que foram vendidos no período escolhido, do ano 2000 a 2005, a um valor abaixo do Limite Aceitável Inferior para a empresa que mais comprou no período. No entanto, observando apenas os dados apresentados não é possível, por parte do executivo, avaliar quais as variáveis que podem influenciar o resultado destes valores de venda tão baixos. Sendo

assim, utilizando o método de análise de regressão foram obtidos alguns resultados que podem auxiliar o executivo a tomar conhecimento de quais variáveis apresentam uma relação próxima com o valor de venda, permitindo agir sobre estas variáveis para obter melhores resultados nas vendas posteriores à análise.

De acordo com o apresentado no capítulo sobre regressão, foi observado primeiramente o seguinte valor, conforme tabela 2:

R-quadrado ajustado	0,997004
---------------------	----------

Tabela 02 – O r ajustado

O que significa que as variáveis independentes (Prazo Médio, Quantidade, Preço de Pauta, Preço de Reposição e Preço Mínimo) podem explicar 99,70% da variabilidade da variável independente (Preço de Venda), porém ainda é preciso analisar o valor p que vai indicar se todas as variáveis são confiáveis nesta representatividade.

Segundo StatSoft, Inc. (1995), o interesse na amostra reside na informação que ela pode prover sobre a população. Assim sendo o valor p é uma medida estimada do grau em que este resultado é "verdadeiro". Mais tecnicamente, o valor p representa um índice decrescente da confiabilidade de um resultado. Quanto mais alto (mais próximo de 1) o valor p , menos se pode acreditar que a relação observada entre as variáveis na amostra é um indicador confiável da relação entre as respectivas variáveis na população. Especificamente, o valor p representa a probabilidade de erro envolvida em aceitar o resultado observado como válido. Desta forma, pôde-se observar que tanto a variável Quantidade quanto a variável Prazo Médio, apresentam índices altos em relação ao valor p o que coloca a validade do resultado de sua relação com a variável independente em dúvida, pois existe um percentual bastante elevado de possibilidade de que este resultado possa ter acontecido por acaso, pois no trabalho de pesquisa foi utilizado um nível de significância de 5% e o valor p destas variáveis está bastante acima, significando que não há influência significativa destas variáveis.

Veja a tabela 3:

	Coeficientes	Valor p
Prazo Médio (Dias)	-0,00552	0,247316
Quantidade	-0,00194	0,307523

Tabela 03 – Verificação do valor p

Assim sendo, recomenda-se a realização de outra regressão apenas com as variáveis cujo valor p seja mais próximo de 0.

Aplicando-se então a regressão são obtidos os seguintes resultados, conforme tabela 4:

R-quadrado ajustado	0,997789
Preço de Reposição	Valor p 7,08E-07
Preço de Pauta	Valor p 4,11E-56
Preço Mínimo	Valor p 0

Tabela 04 – r ajustado e valor p

Assim, é possível notar que ao observar o r ajustado, pode-se verificar que mesmo utilizando menos variáveis (excluindo da análise as variáveis Quantidade e Prazo Médio), a variabilidade considerada pelo modelo através das variáveis independentes em relação a dependente, neste caso até aumentou passando para 99,78%, porém agora, conforme Montgomery (2003), ainda é preciso verificar a soma quadrática dos erros (SQR) e também a média quadrática dos erros (MQR), pois ao se retirar variáveis do modelo (que teoricamente não influenciariam nele) se deve observar que o grau de liberdade também diminui, o que influencia na soma quadrática dos erros e na média quadrática dos erros também. Sendo assim, Montgomery (2003) coloca que a menos que a soma quadrática dos erros (SQR) no novo modelo (sem as variáveis Quantidade e Prazo Médio) seja reduzida por uma quantidade igual à média quadrática dos erros (MQR) do modelo original, o novo modelo terá uma média quadrática dos erros maior do que o modelo antigo. Assim o novo modelo será pior do que o antigo, segundo a sua concepção. Assim, os valores para SQR e MQR se apresentam na tabela 5 a seguir:

	Modelo Anterior (gl 5)	Novo modelo (gl 3)
SQR	4332265	9804831
MQR	9,29166	9,303251

Tabela 05 – Soma Quadrática dos erros e Média Quadrática dos Erros

O que demonstra o relatado por Montgomery, contudo, aplicando-se o princípio da parcimônia, relatado por Levine et al (2000), foi mantido para análise neste trabalho o segundo modelo, pois a diferença apontada na média quadrática não justifica manter as duas variáveis para análise, em virtude de sua representatividade não ser significativa.

Decidido isto então, partiu-se para verificação dos pressupostos, onde foi possível notar que havia violações quanto à heterocedasticidade nos resíduos de cada variável independente e também nos resíduos dos valores previstos e violação também quanto à linearidade da probabilidade normal.

Veja as figuras 26, 27, 28, 29 e 30 comprovando o relatado:

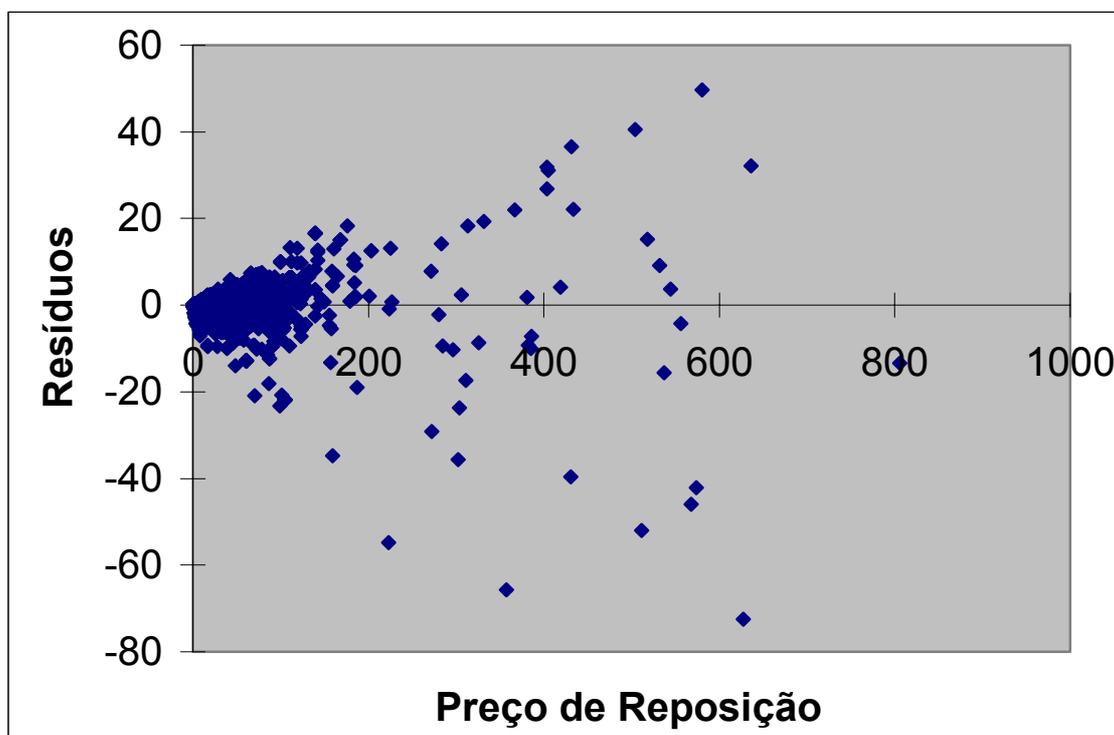


Figura 26 – Heterocedasticidade na variável independente preço de reposição

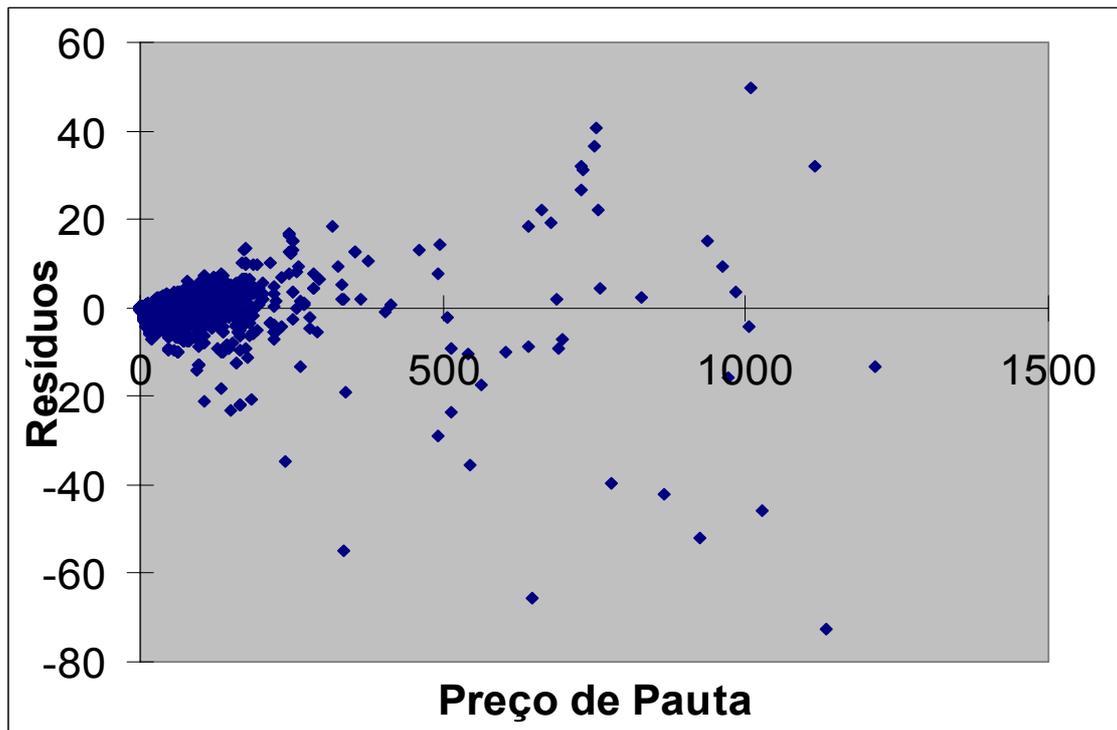


Figura 27 – Heterocedasticidade na variável independente preço de pauta

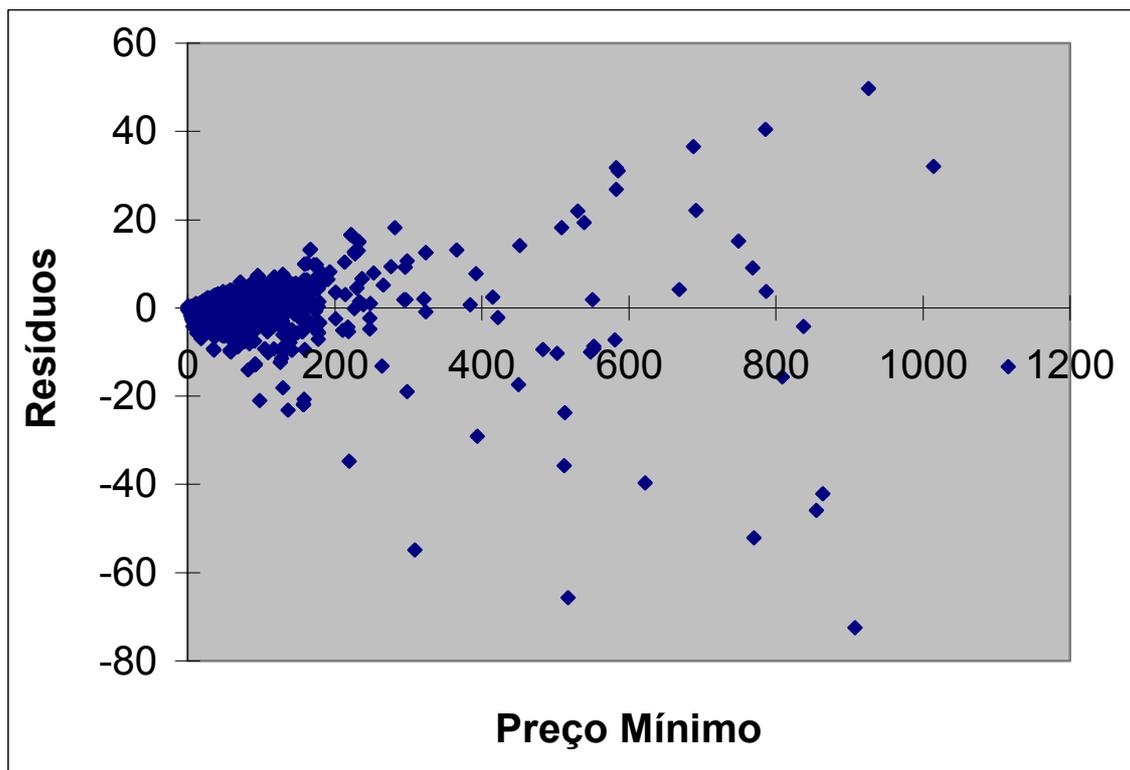


Figura 28 – Heterocedasticidade na variável independente preço mínimo

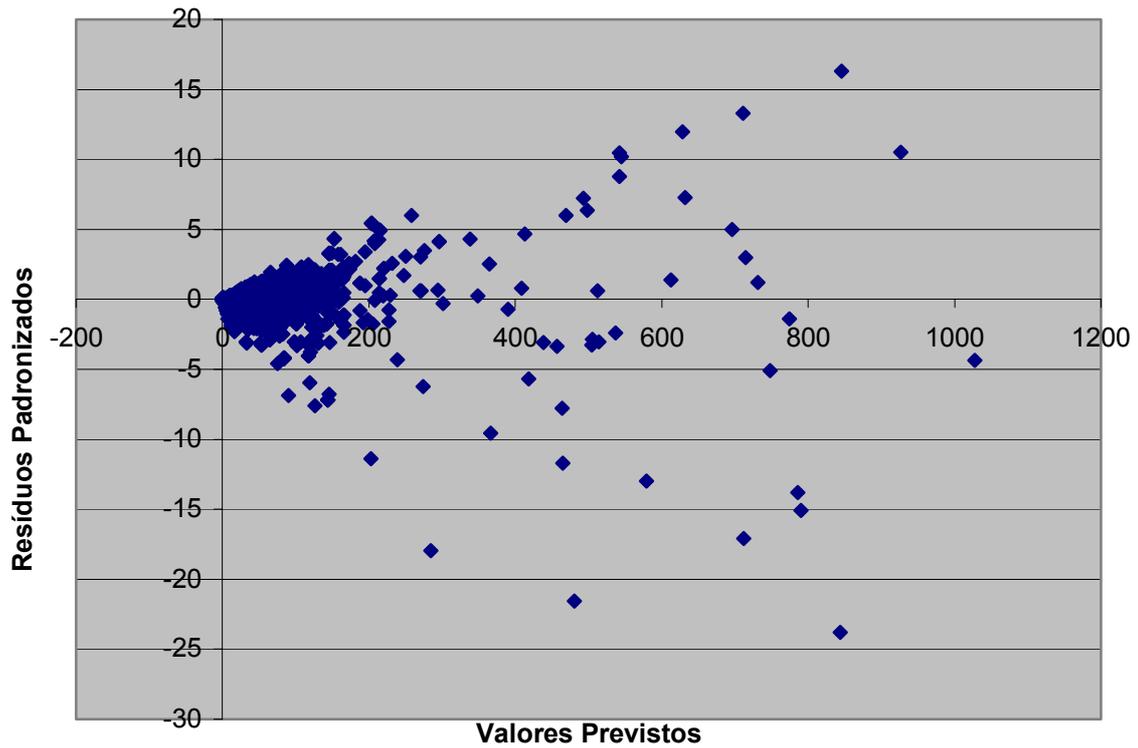


Figura 29 – Heterocedasticidade na plotagem dos resíduos padronizados versus valores previstos

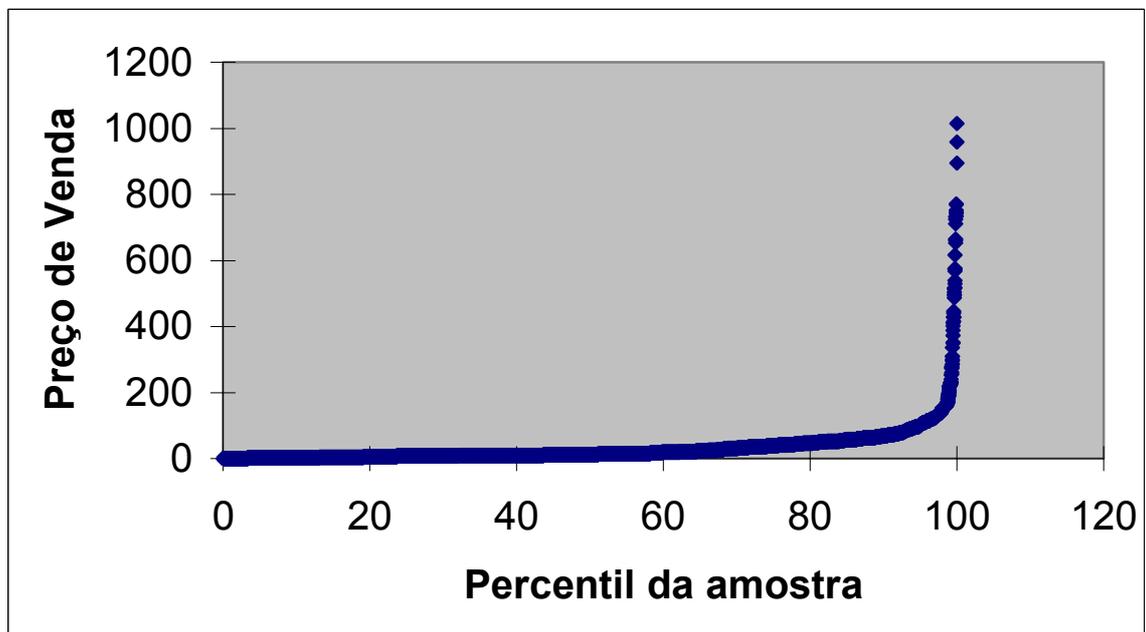


Figura 30 – Plotagem da probabilidade normal (falta de normalidade)

Sendo assim, foi necessário aplicar uma transformação logarítmica base 10, conforme sugerido no item 2.4.3.3 para tentar neutralizar as violações verificadas.

Ao realizar então uma nova regressão com base nas transformações realizadas, foi possível notar que os dados se comportaram de maneira apropriada e a transformação conseguiu neutralizar as violações dos pressupostos, tornando assim o modelo apropriado. Ao analisar o r ajustado, pôde-se observar que as variáveis independentes consideradas na análise podem explicar cerca de 99,76% da variabilidade da variável dependente, o que também é evidenciado ao analisar o Quadrado médio da regressão e dos resíduos onde é possível notar que praticamente toda variação é devido à regressão, veja a tabela 6:

	<i>MQ</i>
Regressão	788,7708
Resíduo	0,000816

Tabela 06 – Quadrado Médio

Veja abaixo as figuras 31, 32, 33 e 34 após a transformação logarítmica:

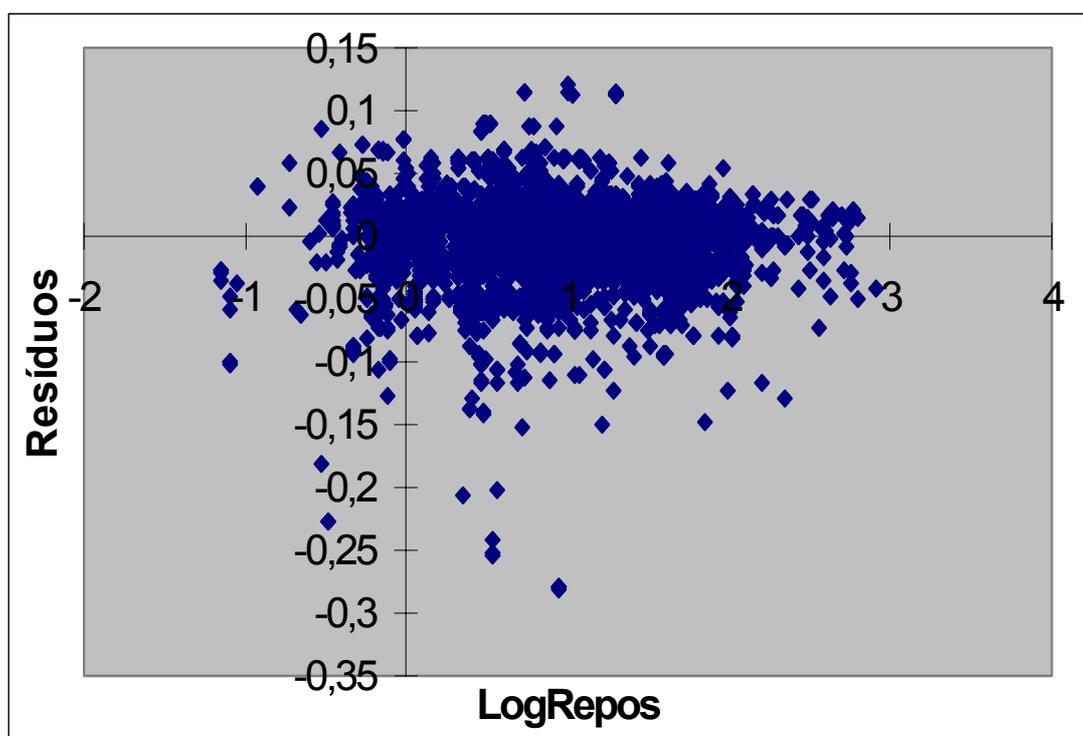


Figura 31 – Preço de Reposição após transformação logarítmica

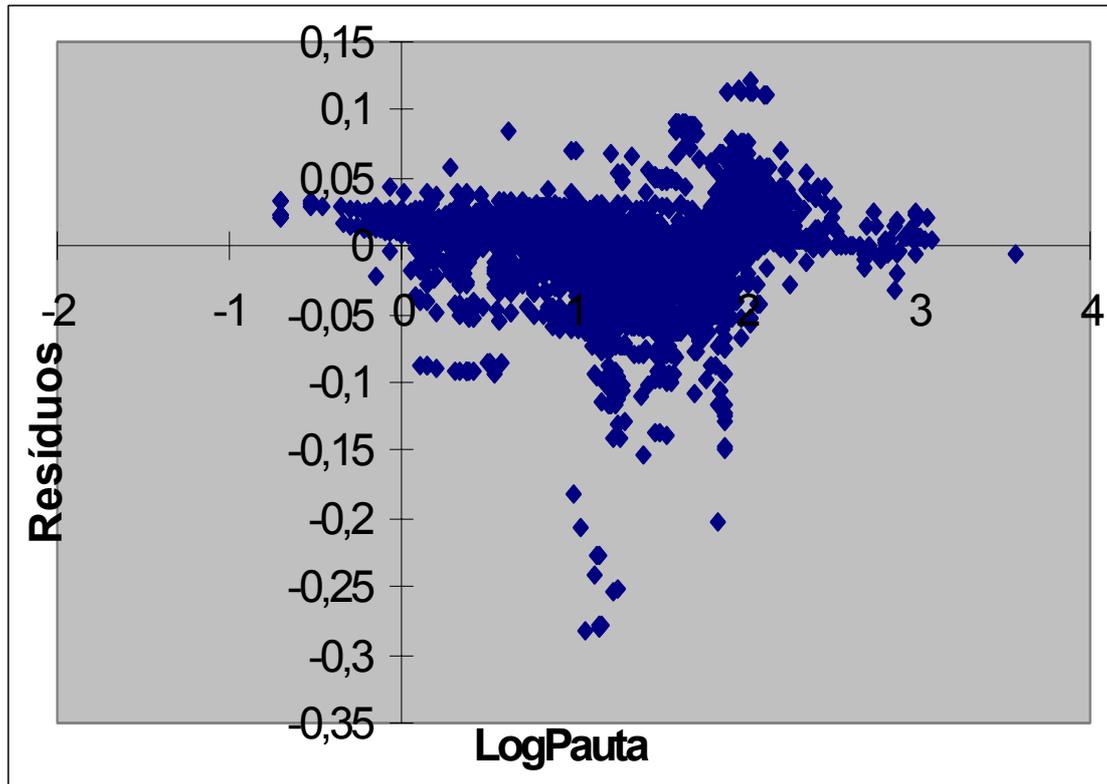


Figura 32 – Preço de Pauta após transformação logarítmica

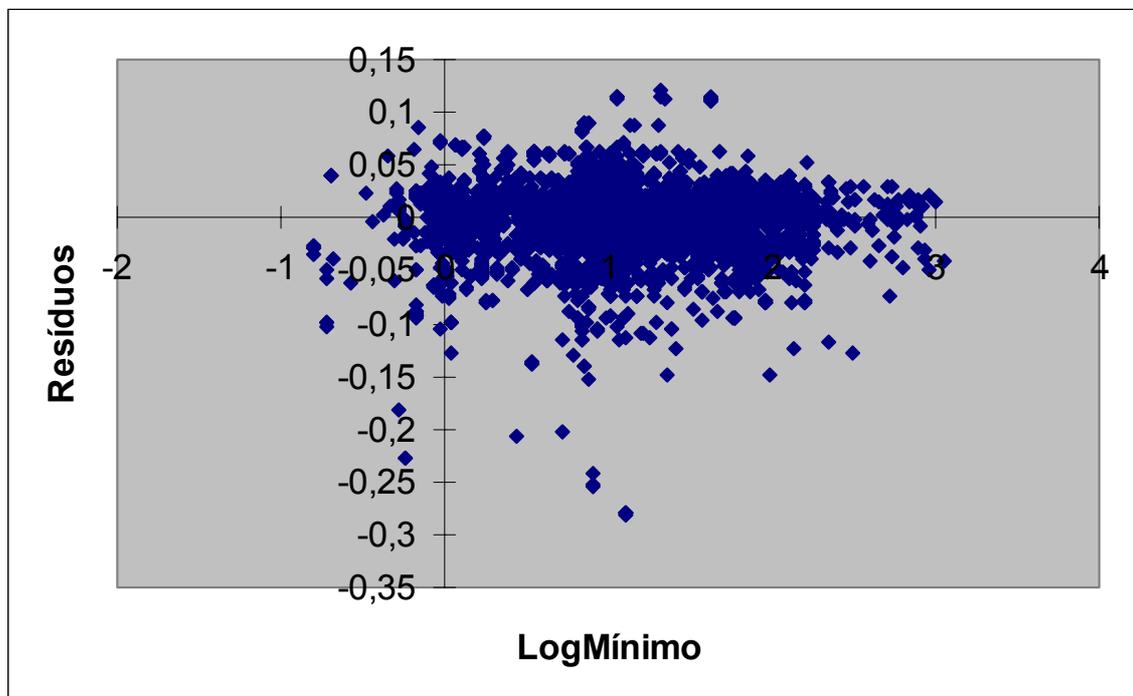


Figura 33 – Preço Mínimo após transformação logarítmica

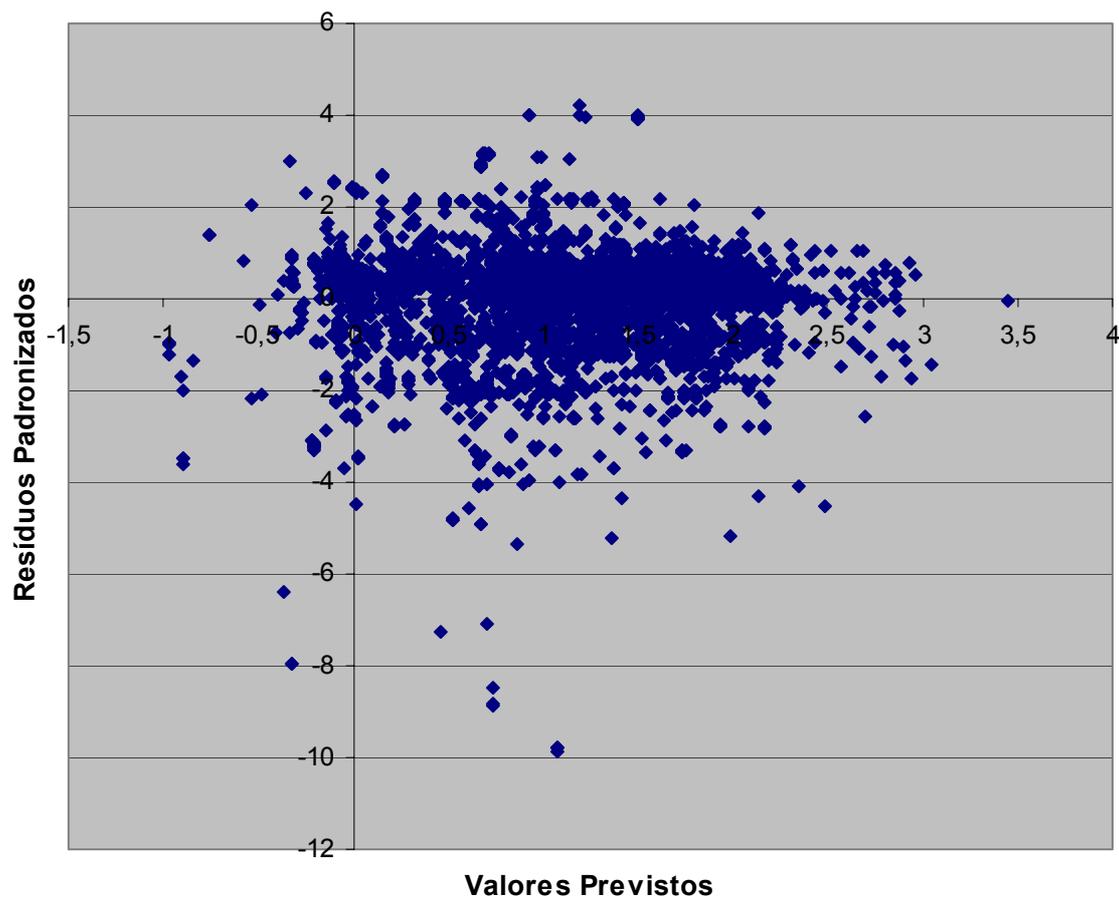


Figura 34 – Independência dos resíduos na plotagem dos valores previstos versus resíduos padronizados

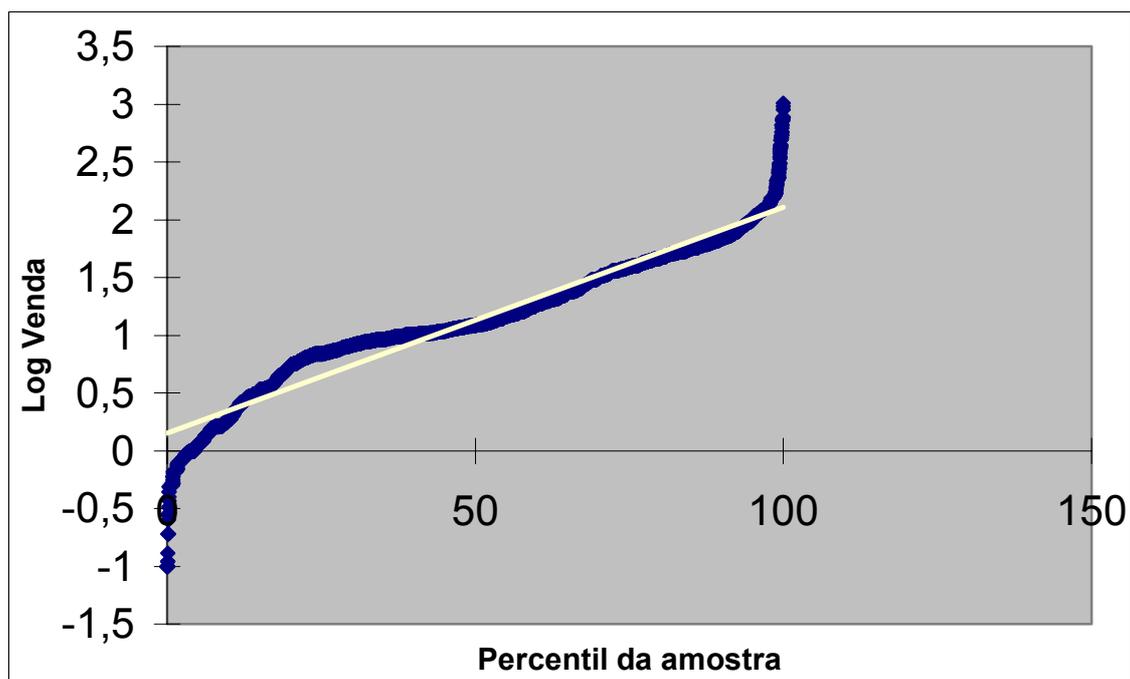


Figura 35 – Gráfico de probabilidade normal (normalidade comprovada)

É importante colocar que através da figura 35 é possível notar que o modelo demonstra a normalidade, porém ainda não é possível afirmar que este seria o modelo ideal, pois em suas extremidades ainda seria interessante um ajuste maior. Entretanto para esta pesquisa o modelo serve como exemplo de aplicação.

A segunda empresa que foi analisada ainda no item Custo, foi também a segunda empresa que mais comprou no período de pesquisa e o aplicativo reportou 2322 itens que foram vendidos abaixo do L.A.I. Assim, utilizando as mesmas considerações já apresentadas na análise da primeira empresa, segue baixo alguns dados obtidos ao realizar a regressão, observe a tabela 7:

Regressão 01 - Com todas as variáveis						
<i>r</i> ajustado	Variáveis	valor <i>p</i>	SQE	QME	SQReg	QMReg
0,997149	Prazo Médio (Dias)	0,373219	58945,27	25,45133	20659311	4131862
	Quantidade	0,065478				
	Preço de Reposição	1,01E-71				
	Preço de Pauta	3,49E-20				
	Preço Mínimo	6,3E-231				
Regressão 02 - Sem Var Quantidade e Prazo Médio						
<i>r</i> ajustado	Variáveis	valor <i>p</i>	SQE	QME	SQReg	QMReg
0,997146	Preço de Reposição	1,41E-71	59053,69	25,47614	20659203	6886401
	Preço de Pauta	4,38E-20				
	Preço Mínimo	2,6E-231				
Regressão 03 - Após transformação logarítmica						
<i>r</i> ajustado	Variáveis	valor <i>p</i>	SQE	QME	SQReg	QMReg
0,997049	Log Reposição	2,8E-180	2,709977	0,001169	916,6887	305,5629
	Log Pauta	0,03418				
	Log Mínimo	4,6E-133				

Tabela 07 – Resultados das regressões na segunda análise

Ao realizar a regressão 01 foi possível notar conforme apresentado na tabela 07 que o *r* ajustado era realmente significativo, porém assim como na primeira análise, foram verificados o valor *p* onde também foi identificado, assim como no primeiro caso, um valor considerável acima de 0, portanto retiraram-se estas variáveis e então foi feito a regressão 02 onde foi possível notar que o *r* ajustado permaneceu na faixa de 99% e assim como no primeiro caso, e também relatado por Montgomery (2003) o quadrado médio do erro aumentou, mas novamente seguindo o princípio da parcimônia, resolveu-se adotar o modelo. Aqui, novamente como na primeira empresa analisada, houve violações nos pressupostos, porém aplicando-se a transformação logarítmica também a este caso, foi possível neutralizar as violações.

Segue abaixo os gráficos, representados nas figuras 36, 37, 38, 39 e 40, após transformação logarítmica comprovando a não violação dos pressupostos conforme relatado acima:

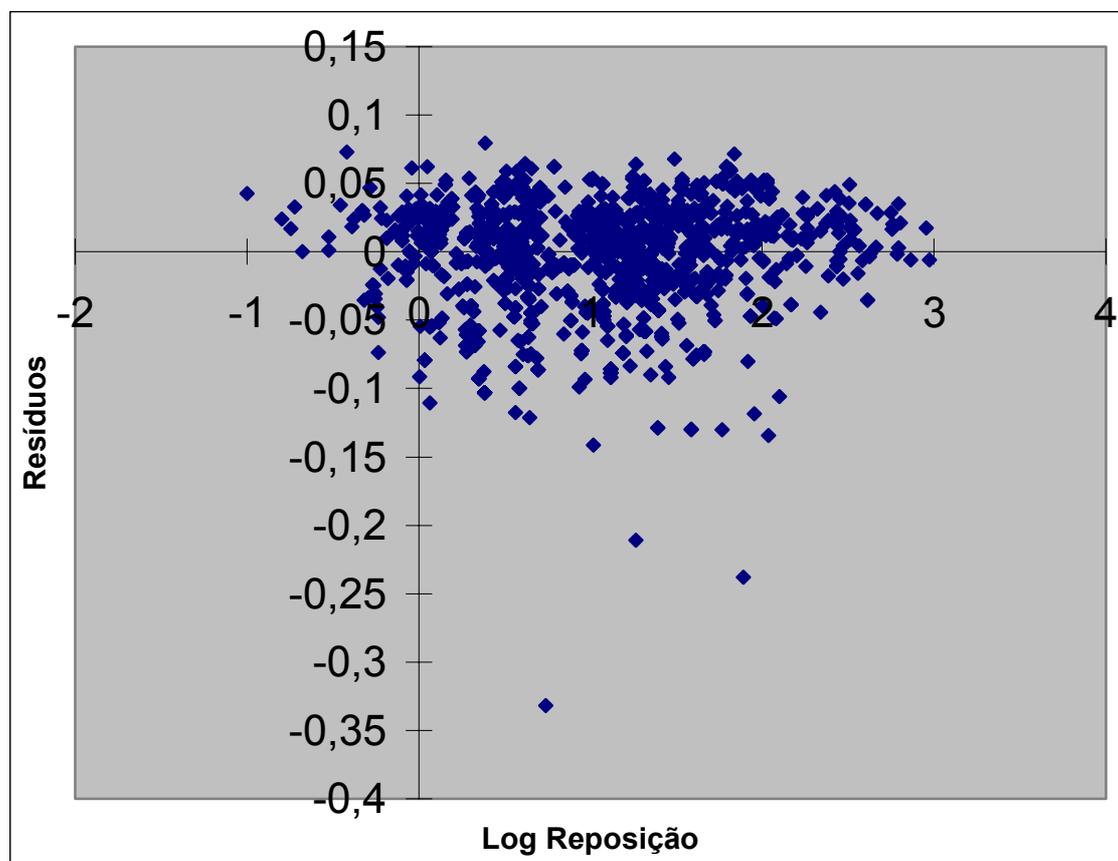


Figura 36 – Resíduos do Log de Reposição

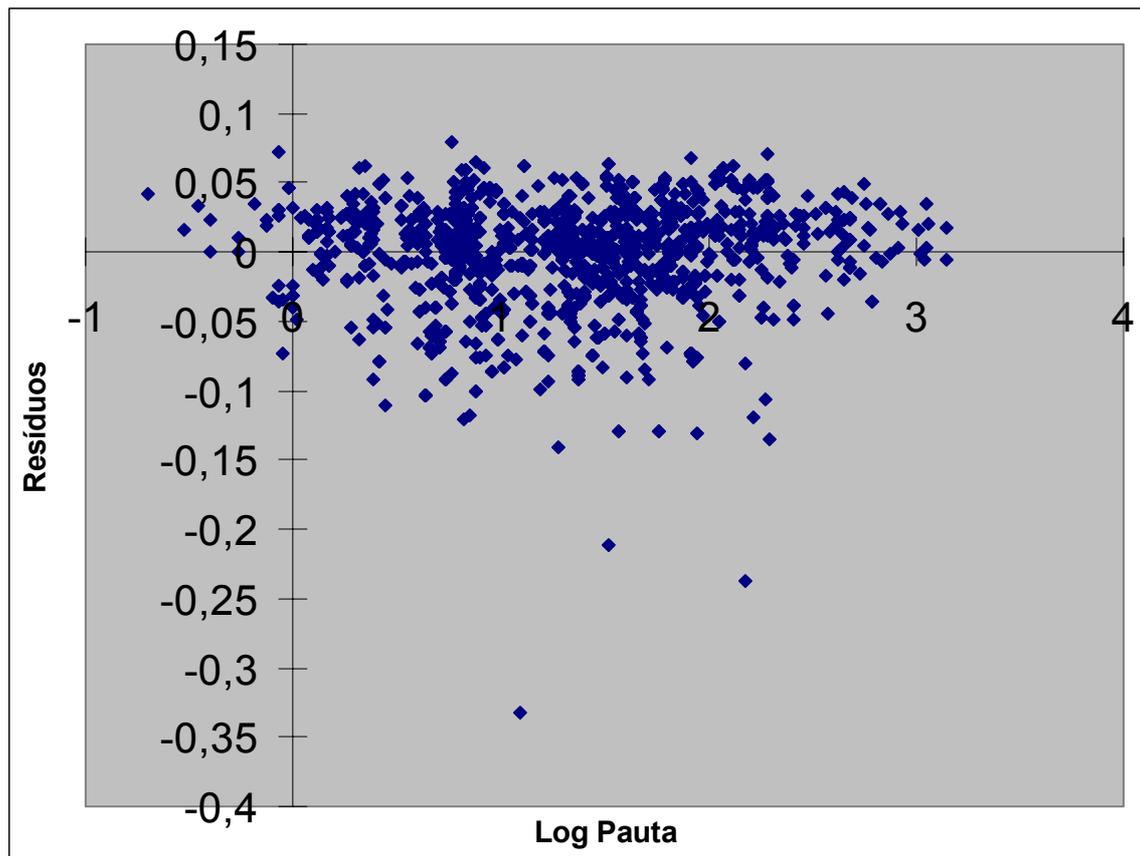


Figura 37 – Resíduos do Log de Pauta

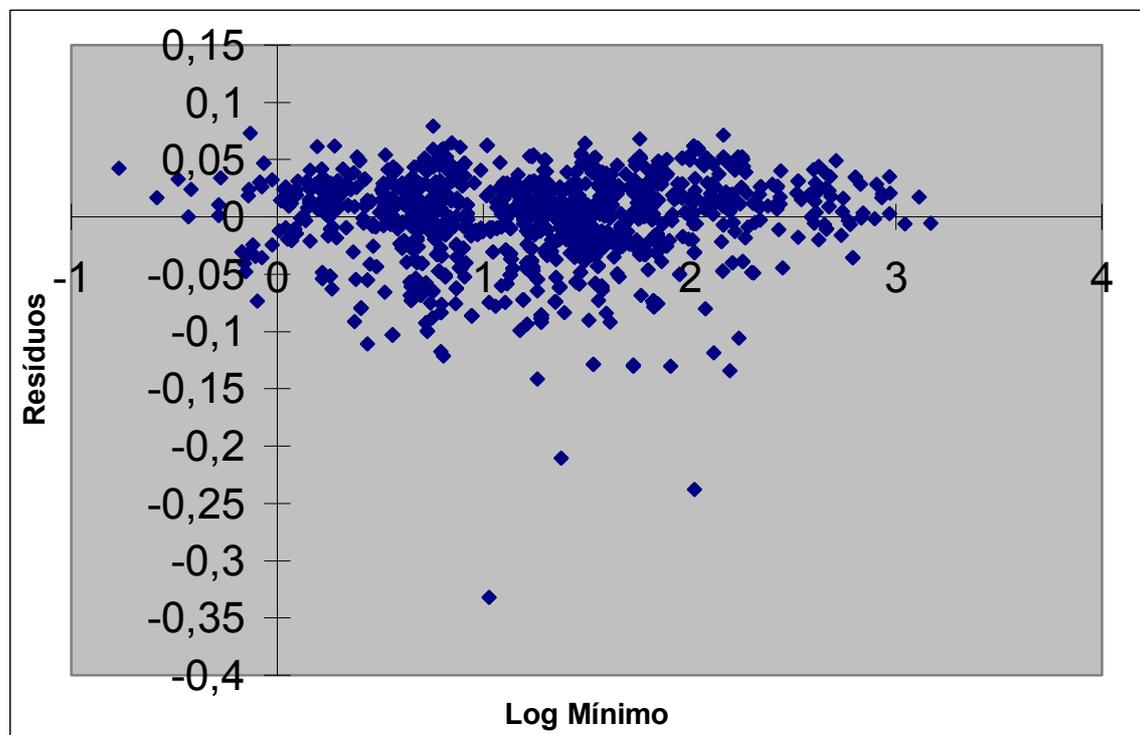


Figura 38 – Resíduos do Log do Preço Mínimo

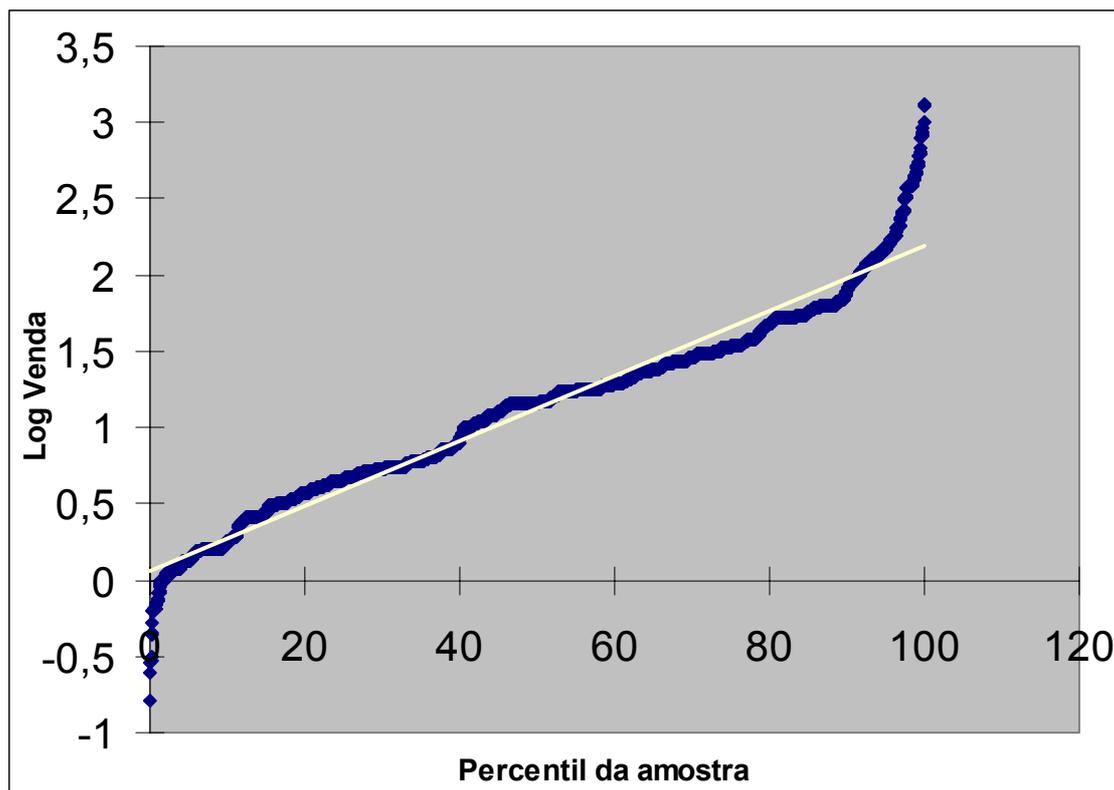


Figura 39 – Probabilidade normal do Log do Preço de Venda

O mesmo comentário realizado para a figura 35 serve para esta figura com relação ao modelo.

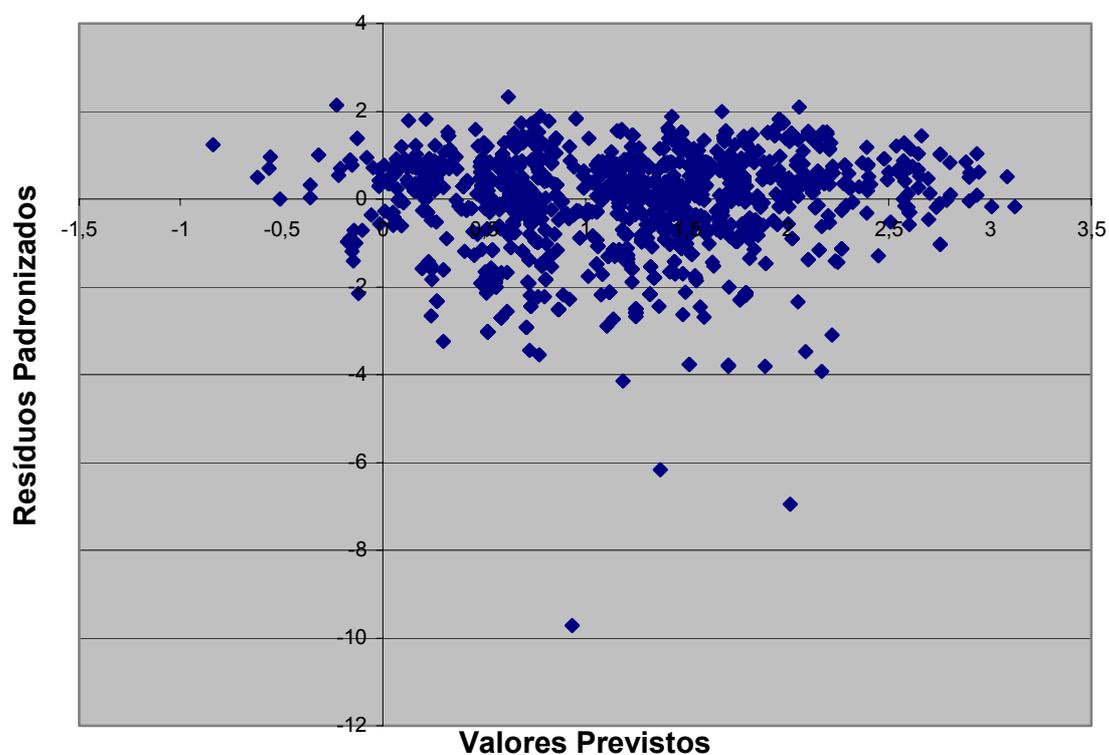


Figura 40 – Valores Previstos versus Resíduos Padronizados após transformação log

Da mesma forma como nas outras duas análises foi analisada uma terceira empresa que seria a 43ª empresa no ranking das 50 que mais compraram no período escolhido para análise. Também como nas outras duas empresas foram realizadas regressões com todas as variáveis, depois foram retiradas as variáveis com valor p relativamente maiores que 0, aí então foram observadas as mesmas violações nos pressupostos que a análise nas outras duas empresas apresentaram. Assim sendo, foi aplicado às estes dados a transformação logarítmica repetindo os procedimentos realizados nas outras duas empresas e obtendo igual sucesso em relação a neutralidade da violação dos pressupostos, obtendo um modelo adequado para o objetivo, que era saber a respeito variabilidade considerada pelo modelo através das variáveis independentes em relação a dependente, ou seja, quais seriam estas variáveis independentes mais significativas. A seguir podem ser vistos a tabela 08 que demonstra os valores obtidos nas regressões e após, os gráficos, representados nas figuras 41, 42, 43,44 e 45 com valores depois da transformação logarítmica, comprovando o relatado:

Regressão 01 - Com todas as variáveis						
<i>r</i> <i>ajustado</i>	Variáveis	valor <i>p</i>	SQE	QME	SQReg	QMReg
0,997918	Prazo Médio (Dias)	0,310027	33949,96	308,636	17014795	3402959
	Quantidade	0,339393				
	Preço de Reposição	0,046134				
	Preço de Pauta	0,000869				
	Preço Mínimo	3,05E-06				
Regressão 02 - Sem Var Quantidade e Prazo Médio						
<i>r</i> <i>ajustado</i>	Variáveis	valor <i>p</i>	SQE	QME	SQReg	QMReg
0,997916	Preço de Reposição	0,065091	34605,23	308,9753	17014140	5671380
	Preço de Pauta	0,000676				
	Preço Mínimo	1,49E-06				
Regressão 03 - Após transformação logarítmica						
<i>r</i> <i>ajustado</i>	Variáveis	valor <i>p</i>	SQE	QME	SQReg	QMReg
0,999266	Log Reposição	4,4E-28	0,047282	0,000422	66,07319	22,0244
	Log Pauta	2,06E-05				
	Log Mínimo	4,69E-09				

Tabela 08 – Resultado das regressões na terceira análise

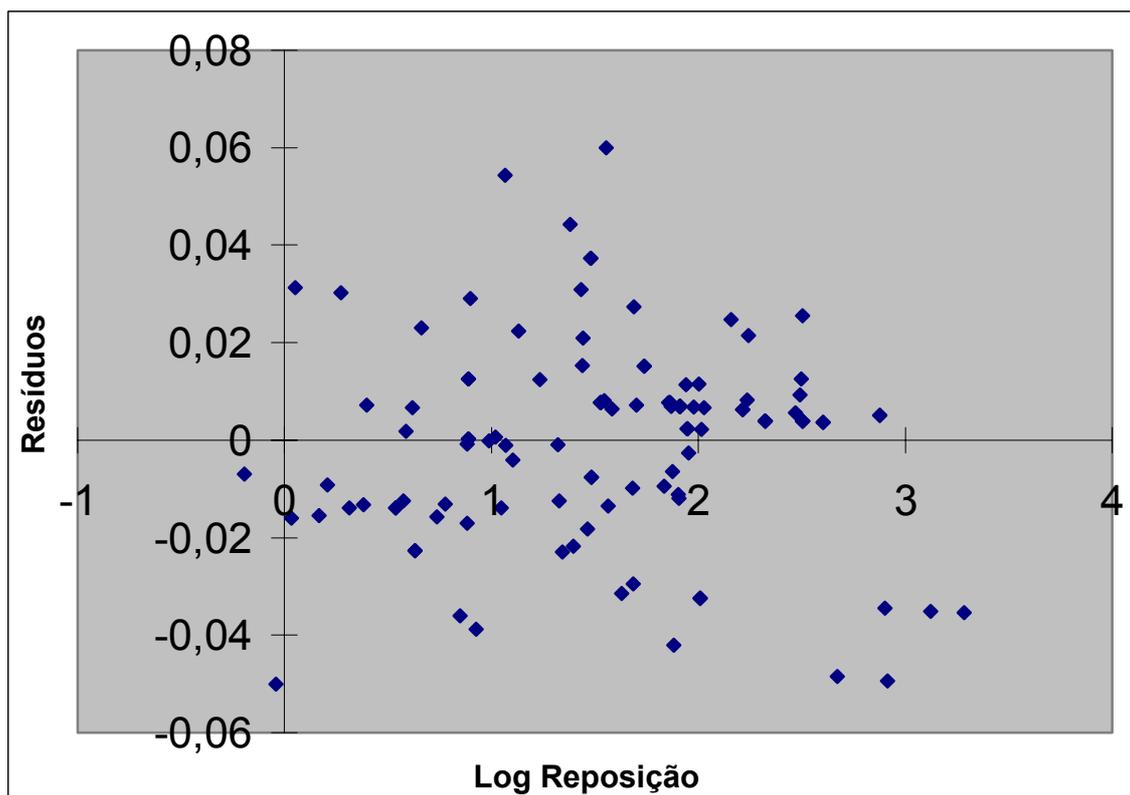


Figura 41 – Resíduos do Log do Preço de Reposição

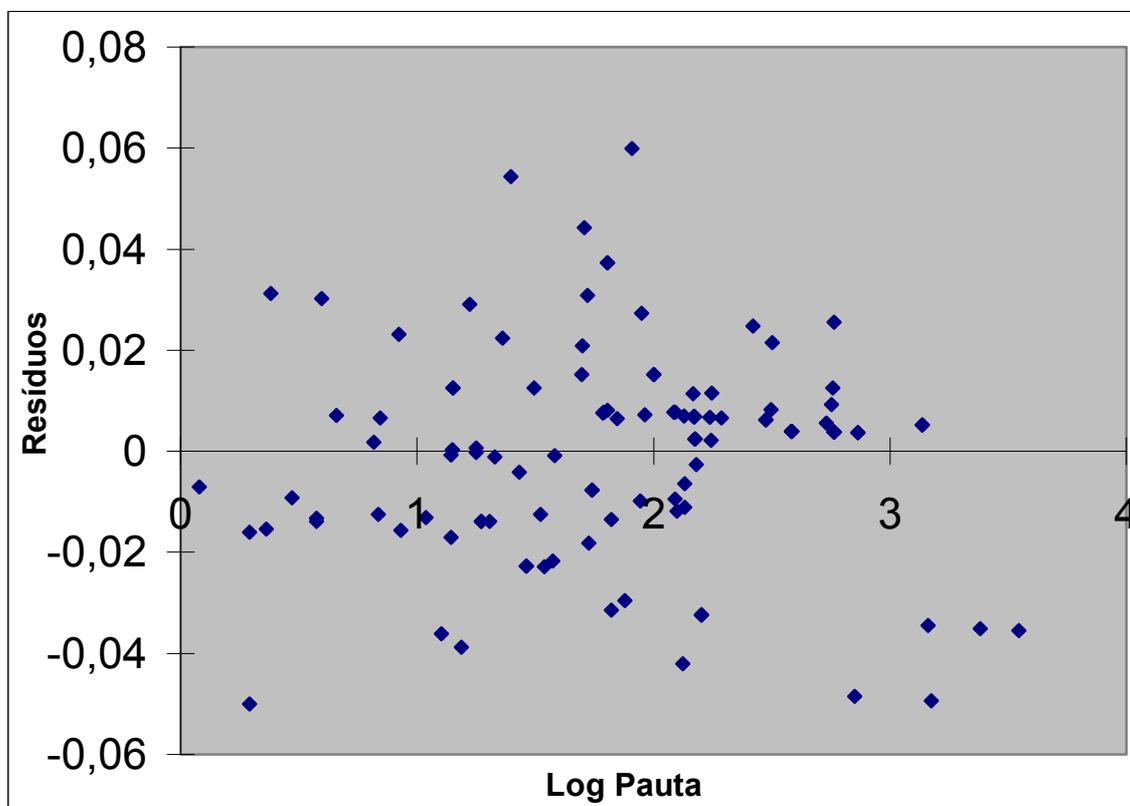


Figura 42 – Resíduos do Log do Preço de Pauta

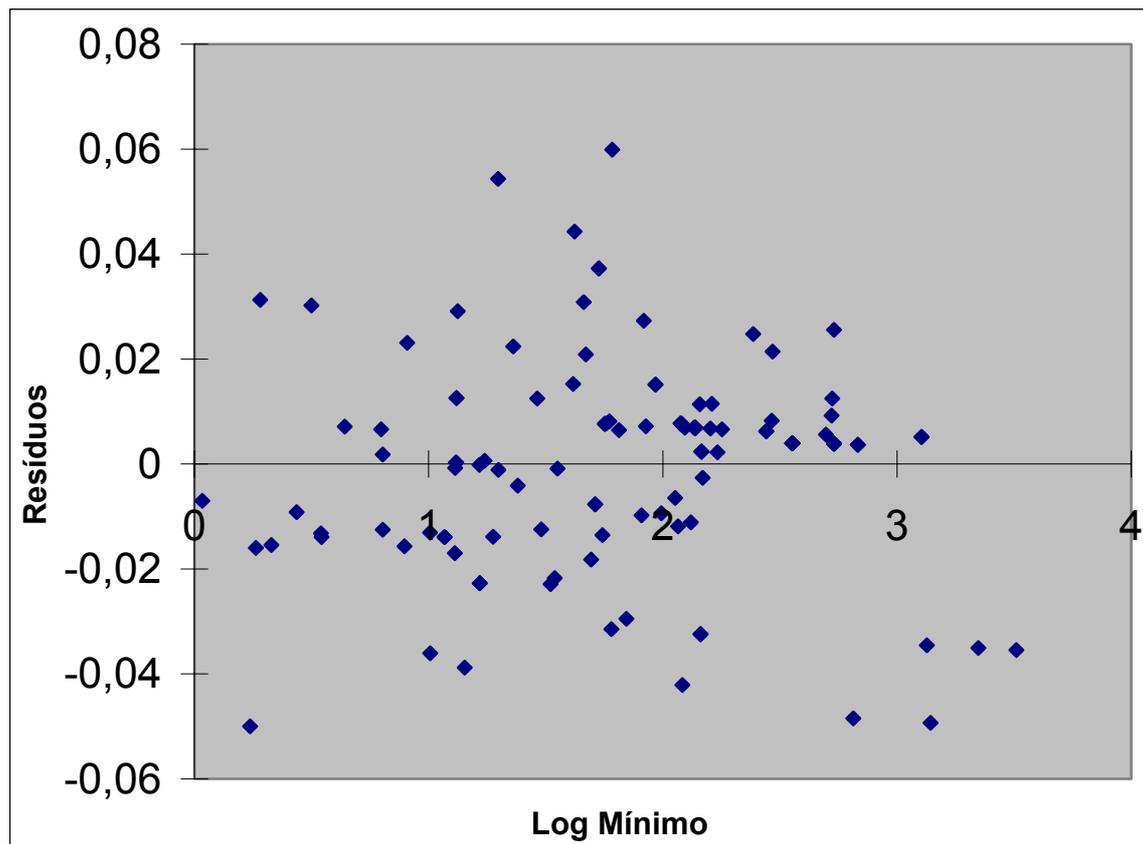


Figura 43 – Resíduos do Log do Preço Mínimo

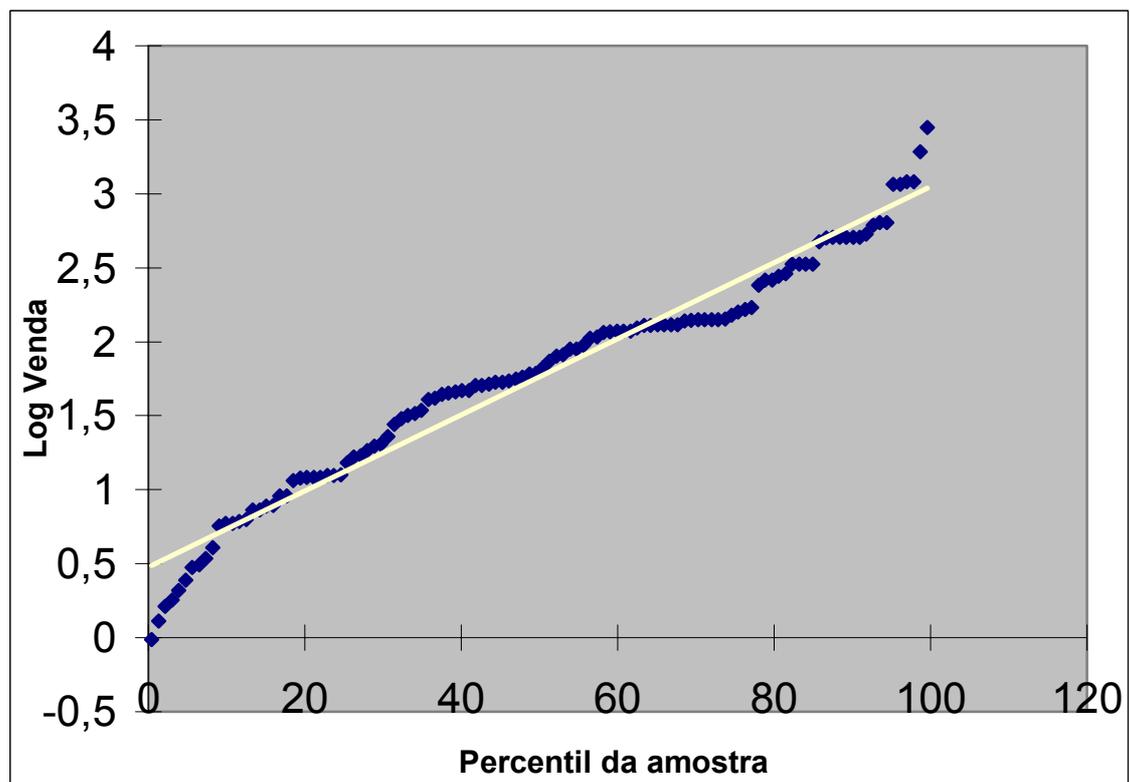


Figura 44 – Probabilidade normal do Log do Preço de Venda

O mesmo comentário realizado para as figuras 35 e 39 se aplica a figura 44.

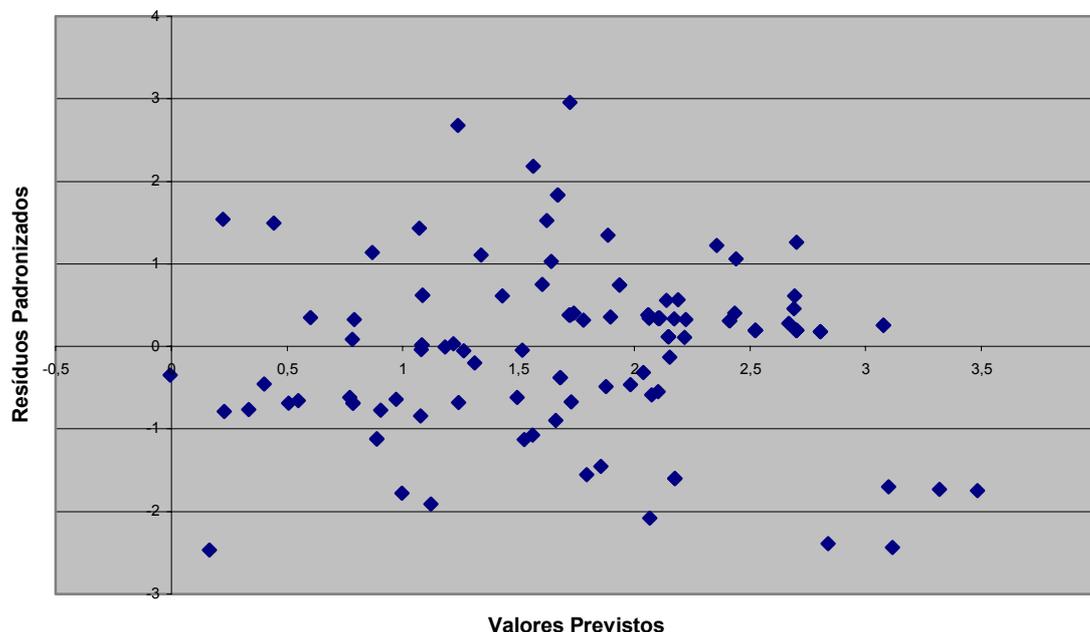


Figura 45 – Valores Previstos versus Resíduos Padronizados após transformação log

Após analisar estes três casos onde os acontecimentos se repetiram e o resultado dos testes também, é possível relatar ao executivo então que das variáveis apresentadas, a alteração destas três (Preço de Pauta, Preço de Reposição e Preço Mínimo) pode influenciar o Preço de Venda, podendo assim em suas decisões futuras levar em consideração apenas estas dentre as 5 apresentadas inicialmente para a análise, facilitando seu trabalho e minimizando suas dúvidas, afinal através desta análise foi possível observar que estas três variáveis consideram cerca de 99,70% da variabilidade do Preço de Venda.

É importante deixar claro também que o fator relacionado com a venda de itens abaixo do L.A.I. está ligado com negociações prévias que são feitas com estas empresas para que sejam fornecidas quantidades X de mercadorias por um período Y de tempo, fazendo assim com que o preço fique fixo ao longo do tempo, correndo o risco de seu custo aumentar e não haver a possibilidade de aumentar o preço de venda em virtude da negociação de fornecimento previamente realizada. Este é um fator que não há como prever ou mensurar, porém é importante levar em consideração na análise como um todo e na tomada de decisões.

4.5.2 Aplicação da distribuição de freqüências

Para a análise do item Qualidade, onde o sistema desenvolvido através da utilização de parâmetros (definidos pelo executivo) classificou os clientes da empresa de acordo com os valores comprados desde o ano inicial contido na base de dados, foram realizadas algumas verificações através da quantificação de clientes em suas respectivas classes (distribuição de freqüências), bem como através da observação analítica de algumas tabelas dinâmicas geradas a partir do aplicativo Microsoft Excel relativas a cada uma das variáveis integrantes do índice de classificação, conforme segue. Índice este cujo cálculo foi determinado pela gerência executiva da empresa, onde as classificações ótimo, bom, ruim e regular receberiam respectivamente os pesos 4,3,2,1 e as variáveis VR_VENDA, MARGEM_VENDA e PRAZO_MÉDIO seguiriam estes pesos para suas 4 faixas de valores, constituindo a fórmula utilizada para o cálculo que é a seguinte: Índice = (peso_VR_VENDA + peso_MARGEM_VENDA + peso_ATRASO MÉDIO)/1,2

Ao carregar os dados através do aplicativo e obter suas classificações, foram contabilizados 36.305 clientes que foram distribuídos de acordo com as classificações definidas pelo executivo e pelo índice criado para tal. Esta distribuição pode ser visualizada através da tabela 09 conforme segue:

Contar de Classificação	
Classificação	Total
Bom	19655
Regular	16505
Ruim	145
Total geral	36305

Tabela 09 – Distribuição de freqüência dos clientes

A visualização desta distribuição de freqüência também pode ser feita através do gráfico demonstrado pela figura 46.

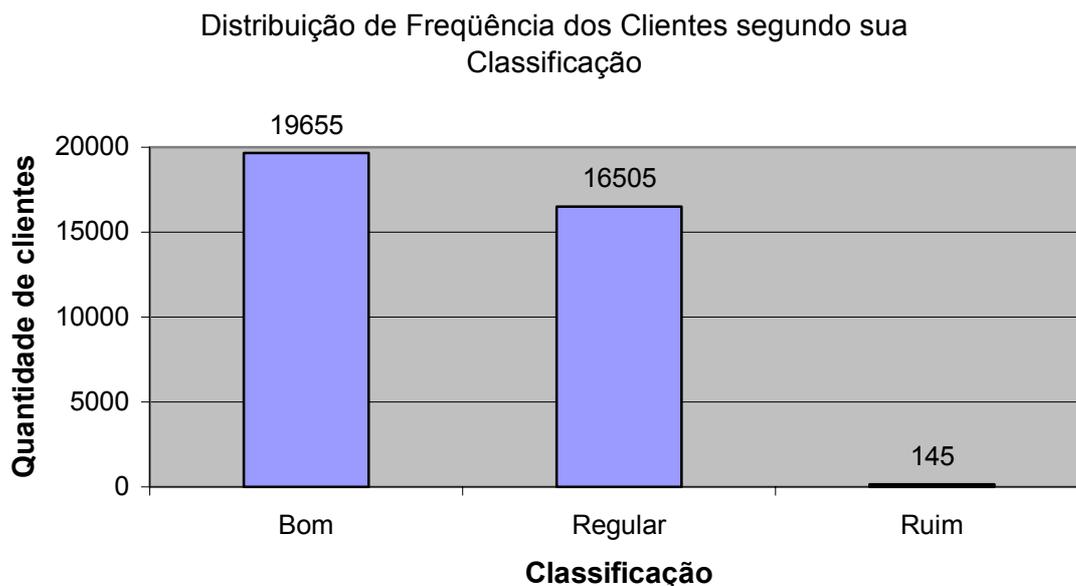


Figura 46 – Distribuição de frequência dos clientes de acordo com sua classificação

Observando a tabela de distribuição de frequência e o gráfico, é possível notar que não houve clientes classificados como clientes ótimos. Este resultado está totalmente atrelado aos parâmetros de qualidade definidos pelo executivo para que o sistema pudesse classificá-los.

Em um primeiro momento, isso não representa nenhum erro, porém a partir dos dados apresentados o executivo pode avaliar estes parâmetros iniciais configurados para perceber se estes realmente estão de acordo com a sua realidade.

A seguir serão mostradas tabelas dinâmicas geradas através do aplicativo Microsoft Excel onde podem ser vistos comparativos e informações relativas às três variáveis envolvidas para a classificação dos clientes onde o executivo pode ter uma melhor idéia sobre os resultados obtidos e o por quê das informações apresentadas. Além disso, ainda pode novamente rever seus parâmetros e tomar decisões de melhorias com base nestas informações.

A tabela 10 apresentada a seguir, mostra números relacionado à variável Valor Total Comprado (%) que considera como 100% o valor de maior compra realizada por um clientes, logo os demais percentuais são decorrentes de um comparativo com este maior. Esta também foi uma definição que partiu do executivo, ao determinar a forma de cálculo para esta variável, salientando que o mesmo é integrante da equipe de desenvolvimento. Mas, poderiam ser feitas mudanças no caso de aplicar o sistema em outra empresa, ou mesmo em outro momento na mesma organização.

Assim, segue:

Classificação	Dados	Total
Bom	Contar de Valor Total Comprado	19655
	Mínimo de Valor Total Comprado (%)	0
	Máx de Valor Total Comprado (%)	100
	Média de Valor Total Comprado (%)	0,090255915
	DesvPad de Valor Total Comprado (%)	1,04013232
Regular	Contar de Valor Total Comprado	16505
	Mínimo de Valor Total Comprado (%)	0
	Máx de Valor Total Comprado (%)	60,1
	Média de Valor Total Comprado (%)	0,200465314
	DesvPad de Valor Total Comprado (%)	1,460938425
Ruim	Contar de Valor Total Comprado	145
	Mínimo de Valor Total Comprado (%)	0
	Máx de Valor Total Comprado (%)	16,5
	Média de Valor Total Comprado (%)	0,654758621
	DesvPad de Valor Total Comprado (%)	2,265430448
Total Contar de Valor Total Comprado		36305
Total Mínimo de Valor Total Comprado (%)		0
Total Máx de Valor Total Comprado (%)		100
Total Média de Valor Total Comprado (%)		0,142613965
Total DesvPad de Valor Total Comprado (%)		1,257136163

Tabela 10 – Tabela dinâmica referente ao Valor Total Comprado (%)

Uma das informações interessantes demonstradas pela tabela e que vale a pena ressaltar, é referente à Média de Valor Total Comprado que é em relação ao valor máximo da variável Valor Total Comprado. Analisando este indicativo, é possível notar, por exemplo, que os clientes considerados Bons, compram cerca de 0,09% do valor que o maior cliente compra, ou seja, a maioria dos clientes bons não são clientes que compram valores muito altos, mas que talvez obtenham pesos melhores nas outras duas variáveis (margem de venda e prazo médio), afinal o índice é baseado nos pesos obtidos das três variáveis. Esta é uma informação importante para o executivo.

Assim partindo nesta mesma linha de raciocínio, nota-se que à medida que as compras vão aumentando em direção ao maior valor total comprado, a classificação dos clientes vai piorando, ou seja, é possível verificar que nem todos os clientes que compram valores mais altos são bons clientes, pois podem estar comprando a uma margem de venda muito baixa ou tendo um atraso médio muito elevado.

Com base no relatado anteriormente, faz-se necessária também a análise das outras duas variáveis, como segue através das tabelas 11 e 12.

A tabela 11 mostrada a seguir demonstra alguns números encontrados em relação à variável Margem de Venda, onde é interessante analisar a média da Margem Média.

Classificação	Dados	Total
Bom	Contar de Margem Média (%)	19655
	Mínimo de Margem Média (%)	11,3
	Máx de Margem Média (%)	100
	Média de Margem Média (%)	27,83314678
	DesvPad de Margem Média (%)	7,573394672
Regular	Contar de Margem Média (%)	16505
	Mínimo de Margem Média (%)	-9,93
	Máx de Margem Média (%)	37,3
	Média de Margem Média (%)	16,65814541
	DesvPad de Margem Média (%)	5,075939633
Ruim	Contar de Margem Média (%)	145
	Mínimo de Margem Média (%)	-9,95
	Máx de Margem Média (%)	14,7
	Média de Margem Média (%)	-
	DesvPad de Margem Média (%)	3,101310345
Total Contar de Margem Média (%)		36305
Total Mínimo de Margem Média (%)		-9,95
Total Máx de Margem Média (%)		100
Total Média de Margem Média (%)		22,62921085
Total DesvPad de Margem Média (%)		8,73952042

Tabela 11 – Margem Média de Vendas (%)

Ao analisar esta variável, foi possível notar algo muito interessante em relação ao valor do item Mínimo de Margem Média, onde é possível verificar um valor de 11,3% que é exatamente a margem média do cliente que tem o índice de 100% em relação ao maior valor da variável Valor Total Comprado, ou seja, fazendo com que este cliente obtenha um peso melhor nesta variável (margem média), e abaixando seu índice final de classificação. Outra informação valiosa ao executivo, para que possa tomar decisões de melhoria, afinal um de seus melhores clientes em termos de faturamento está incluso na classificação BOM e poderia tentar em um futuro enquadrá-lo em uma classificação acima. Agora o executivo já sabe uma informação importante para isso.

Assim, observando a tabela, constata-se algo mais natural também que é o fato de que à medida que a classificação dos clientes vai decrescendo, a margem média também cai. Agora é interessante notar também que a margem média dos clientes bons é de 27,83% e a margem média mínima para clientes ótimos seria 25%, ou seja, estes clientes estão sendo considerados apenas Bons, em virtude de estarem obtendo pesos melhores na outras variáveis, como analisado anteriormente na variável Valor Total Comprado.

Assim o executivo tem em suas mão informações valiosas para tomar decisões de melhorias ou até mesmo ajuste de seus parâmetros de qualidade.

A próxima tabela, a tabela 12, demonstra números em relação ao Prazo Médio (dias).

Classificação	Dados	Total
Bom	Contar de Atraso Médio (Dias)	19655
	Mínimo de Atraso Médio (Dias)	0
	Máx de Atraso Médio (Dias)	7
	Média de Atraso Médio (Dias)	0,219587891
	DesvPad de Atraso Médio (Dias)	0,695690944
Regular	Contar de Atraso Médio (Dias)	16505
	Mínimo de Atraso Médio (Dias)	0
	Máx de Atraso Médio (Dias)	27
	Média de Atraso Médio (Dias)	0,729536504
	DesvPad de Atraso Médio (Dias)	1,467878954
Ruim	Contar de Atraso Médio (Dias)	145
	Mínimo de Atraso Médio (Dias)	1
	Máx de Atraso Médio (Dias)	27
	Média de Atraso Médio (Dias)	2,544827586
	DesvPad de Atraso Médio (Dias)	3,743474527
Total Contar de Atraso Médio (Dias)		36305
Total Mínimo de Atraso Médio (Dias)		0
Total Máx de Atraso Médio (Dias)		27
Total Média de Atraso Médio (Dias)		0,460707891
Total DesvPad de Atraso Médio (Dias)		1,174226016

Tabela 12 – Prazo Médio (dias)

Através desta tabela foi possível observar, por exemplo, que dentre os clientes ruins o atraso médio máximo foi de 27 dias, o que os classificaria como clientes no mínimo regulares, porém assim como nas outras duas análises, os resultados das outras duas variáveis pode não ter sido muito bons, resultando em pesos menores e em consequência, índices de classificação também menores.

Observando as três tabelas, foi possível notar que o cliente com melhor desempenho, foi ainda classificado como um cliente classe BOM, pois apesar de apresentar o maior valor do total comprado e também um prazo médio de atraso no pagamento 0, sua margem de venda ficou muito aquém do objetivo para um cliente classe ÓTIMO. A partir destes dados, o executivo pode tomar decisões para que clientes como estes possam elevar a sua classificação ou ao menos permanecer assim.

Assim sendo, ao observar estar três tabelas o executivo poderá tomar decisões que reflitam melhorias na empresa migrando o número de clientes classificados como

ruins ou regulares para clientes bons ou até ótimos. Assim, fica clara a importância destas informações contribuindo para que o executivo possa projetar melhorias com relação ao item qualidade.

4.5.3 Aplicação das séries temporais

Ao analisar o terceiro e último item proposto na metodologia SIEGO, que é o item Tempo, foram observados dados referentes ao faturamento das empresas Matriz Blumenau, Filial de Rio do Sul e Filial de Chapecó, pois as outras três filiais oferecem um período de observação muito curto (inferior a 36 meses), dificultando a aplicação de séries temporais. De qualquer forma, a aplicação de séries temporais para estas três empresas iniciais que fazem parte do grupo, já possibilitou a visualização de informações importantes que poderão ser aplicadas às demais empresas quando em um período futuro.

Observando os resultados da análise com a empresa Matriz Blumenau, foi possível verificar que há uma tendência linear decrescente nos valores de vendas, baseado nos valores de vendas realizados de agosto de 1994 até junho de 2005. É importante frisar que os valores estão atualizados de acordo com a inflação através da aplicação do índice IGP-M conforme já citado no item 3.3.4 deste trabalho.

Veja a figura 47 que representa graficamente esta informação e apresenta a equação de tendência obtida pelo método dos mínimos quadrados a partir dos dados coletados e também médias móveis suavizando a linha de tendência para melhor visualização.

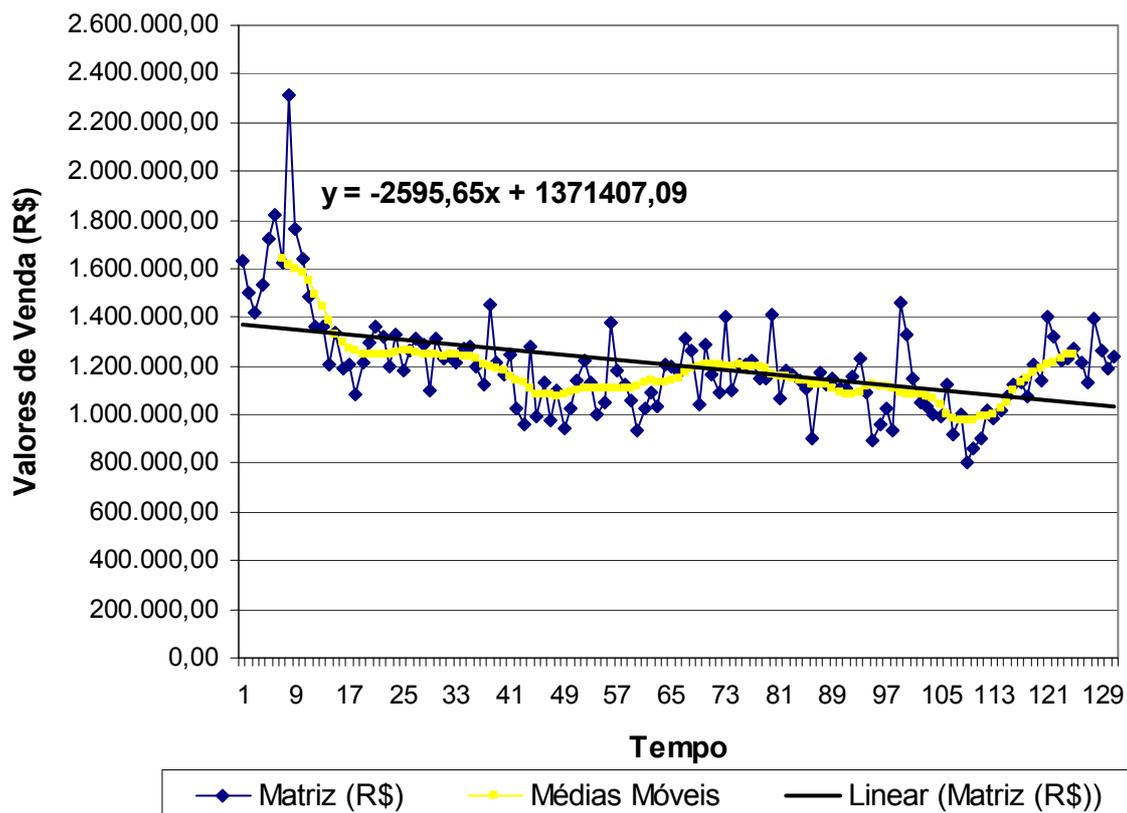


Figura 47 – Valores de venda Matriz Blumenau (tendência linear decrescente)

É interessante deixar claro que para esta e as demais análises relacionadas ao item tempo, foi utilizado neste trabalho de pesquisa o modelo clássico multiplicativo com 4 componentes, que segundo Stevenson (1986) é mais utilizado porque retrata melhor a experiência. As 4 componentes são a tendência, a sazonalidade e as variações cíclicas e irregulares, utilizando a seguinte equação: $y = T \times S \times CI$.

Após esta análise inicial buscou identificar se havia influência de sazonalidade na série, calculando os índices sazonais, para, em caso positivo, dessazonalizá-la. Aplicou-se então médias móveis de 12 meses, posteriormente centrando-as, resultando na série mostrada na figura 48, e os índices sazonais estão na tabela 13.

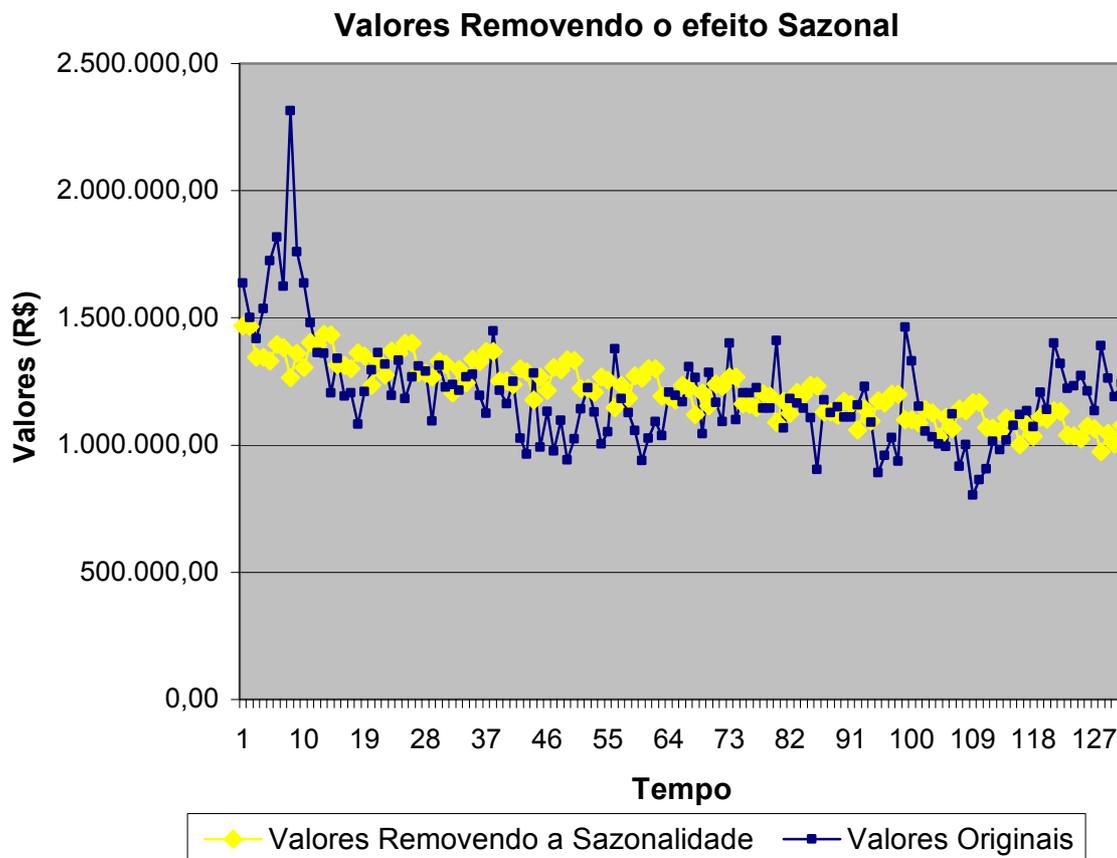


Figura 48 – Valores sem o efeito da sazonalidade

Foi possível observar que os valores sofrem o efeito das variações sazonais, pois seus índices apresentaram valores significativamente diferentes de 1 (chegando até quase 7% de diferença em alguns meses) portanto é preciso considerá-las ao realizar uma previsão. A tabela 13 demonstra os índices sazonais calculados.

Meses	Índices
Janeiro	0,971634
Fevereiro	0,979117
Março	1,06831
Abril	0,990334
Mai	1,029834
Junho	0,957648
Julho	0,962634
Agosto	0,932319
Setembro	0,931225
Outubro	1,01328
Novembro	1,012133
Dezembro	1,02176

Tabela 13 – Índices Sazonais Matriz Blumenau

Verificada a sazonalidade, também foi necessário verificar se as vendas sofrem o efeito das variações cíclicas e irregulares, para tanto, foram retiradas as variações de tendência (estimada por mínimos quadrados) e variações sazonais ($CI = \text{Valores originais} / (\text{Tendência} \times \text{Sazonalidade})$) e então foram plotadas em um gráfico apenas as variações cíclicas para tentar se visualizar variações sistemáticas em torno de 1, que caracterizariam os ciclos. Pode ser citado como exemplo desta aplicação aqui o caso do primeiro valor observado, agosto de 1994, cujo valor original é R\$ 1.636.157,20, o valor de tendência obtido pela equação $y = -2595,65x + 1371407,09$ é R\$ 1.368.811,44, o valor do índice sazonal encontrado para o mês de agosto é 0,932319 (obtido através das médias móveis de 12 períodos, posteriormente centrada e após através da mediana referente a cada mês envolvido na série) e aplicando-se a fórmula $CI = (\text{Valores originais} / (\text{Tendência} \times \text{Sazonalidade}))$ obteve-se o índice da variação cíclica igual a 1,2821. Este cálculo se repetiu para toda a série afim de se obter todos os índices das variações cíclicas/irregulares.

A figura 49 mostra o resultado.

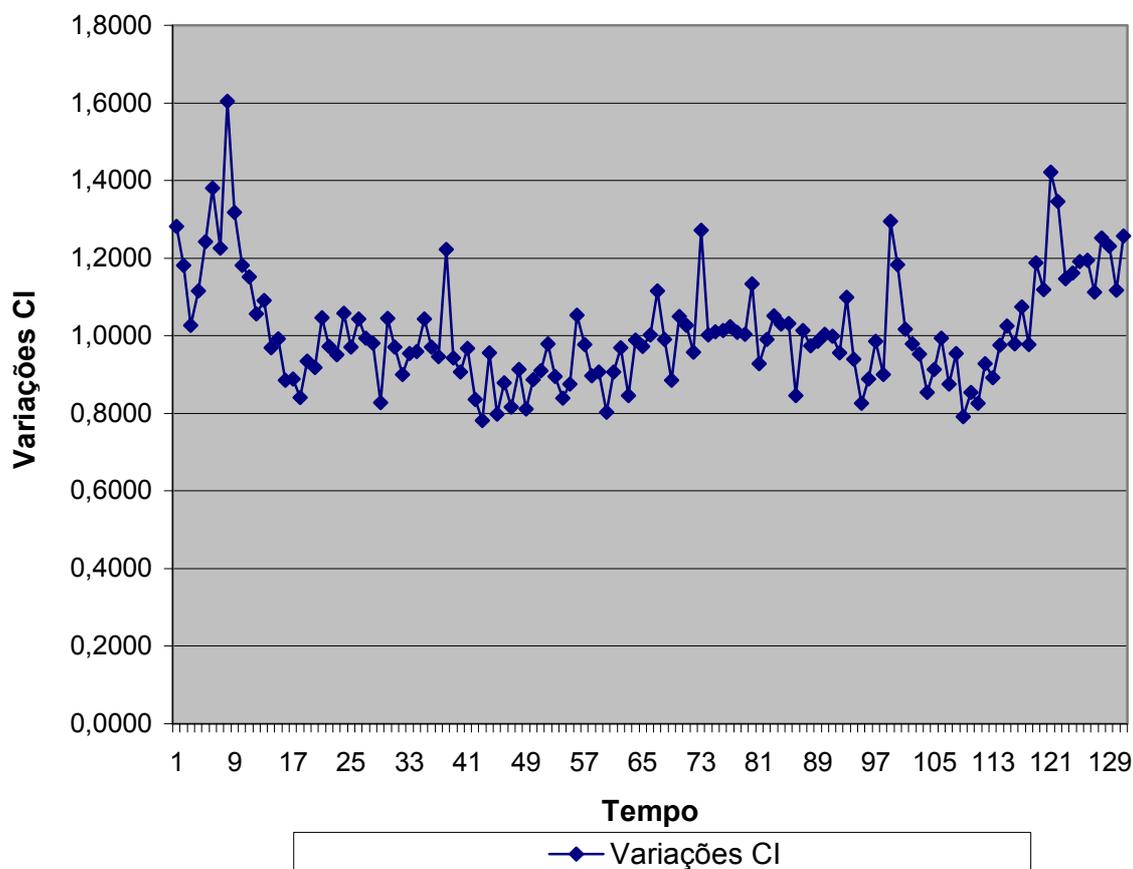


Figura 49 – Variações Cíclicas Matriz Blumenau

Observando a figura 49, não foi possível visualizar as variações sistemáticas em torno de 1 que caracterizam os ciclos, então se calculou as médias móveis de 12 períodos, posteriormente centrando-as para se tentar visualizar melhor possíveis variações cíclicas, conforme figura 50 a seguir:

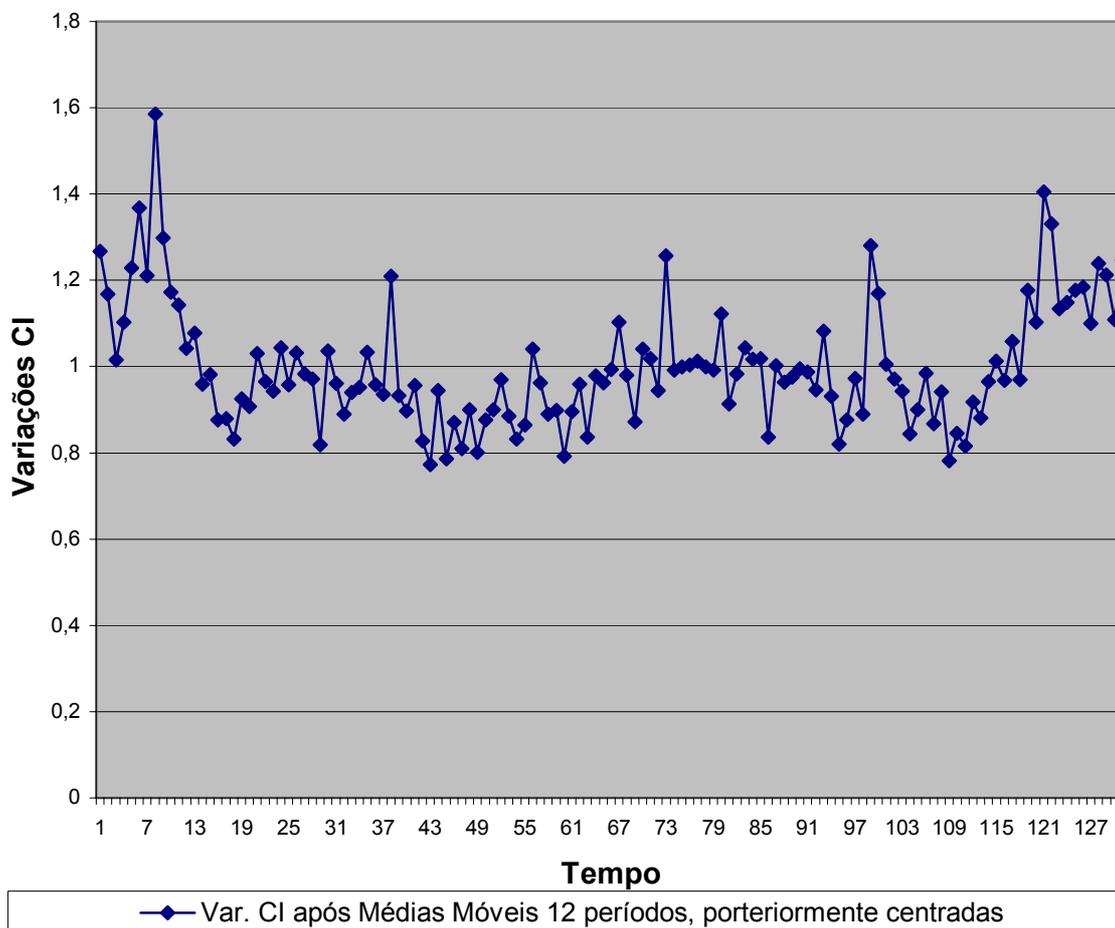


Figura 50 – Variações Cíclicas/Irregulares após aplicação das médias móveis de 12 períodos posteriormente centrada

Assim, foi possível observar que mesmo aplicando as médias móveis sobre as variações cíclicas/irregulares, não foi possível notar uma variação sistemática em torno de 1 caracterizando o efeito de ciclos. Desta forma, as variações cíclicas foram desconsideradas para efeitos de previsão deste modelo.

Após todas estas observações e critérios serem verificados, foi possível realizar alguns cálculos demonstrativos de valores de faturamento para o ano de 2006 observando seus 12 meses. Para o cálculo foi utilizada a equação:

$$y = ((-2595,65x) + 1371407,09) * \text{ind.sazonal}$$

que considera a tendência e as variações sazonais, obtendo os resultados demonstrados através da tabela 14 e da figura 51, como segue:

Períodos	Índices Sazonais	Meses	Previstos
138	0,971634159	jan/06	R\$ 984.466,91
139	0,97911657	fev/06	R\$ 989.506,70
140	1,068310391	mar/06	R\$ 1.076.874,06
141	0,99033402	abr/06	R\$ 995.702,07
142	1,029833767	mai/06	R\$ 1.032.742,83
143	0,957648074	jun/06	R\$ 957.867,51
144	0,962634392	jul/06	R\$ 960.356,31
145	0,932319166	ago/06	R\$ 927.692,85
146	0,931225229	set/06	R\$ 924.187,21
147	1,013279506	out/06	R\$ 1.002.991,21
148	1,012133241	nov/06	R\$ 999.229,44
149	1,021759825	dez/06	R\$ 1.006.081,17

Tabela 14 – Valores previstos de faturamento para 2006

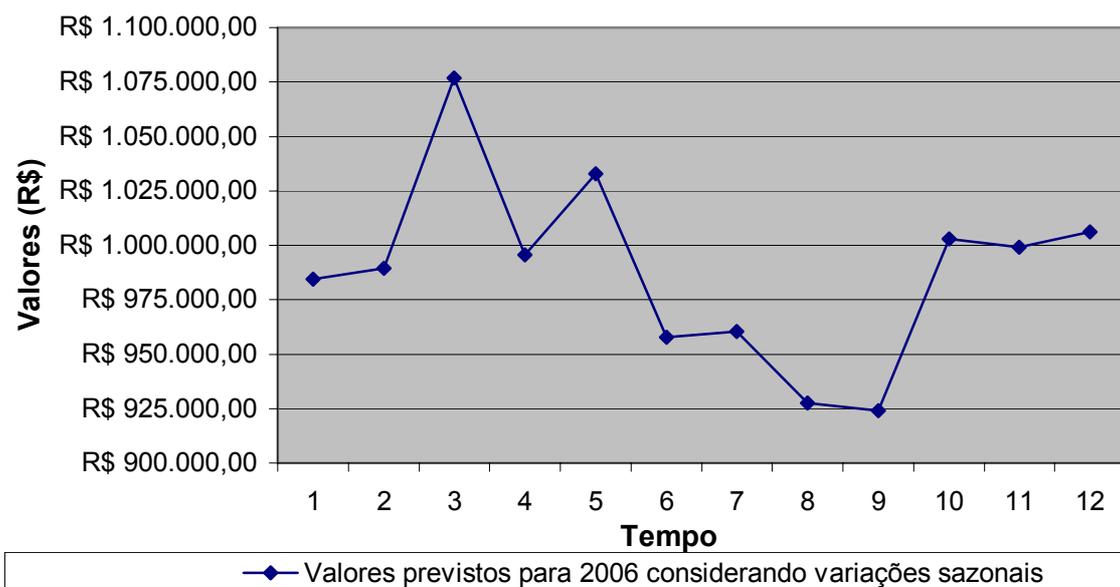


Figura 51 – Valores previstos de faturamento para Matriz Blumenau em 2006

Assim, através da utilização de séries temporais, o executivo pode traçar ações de melhorias visto que a análise demonstra que as vendas vêm caindo no decorrer dos anos na Matriz Blumenau e a previsão é que comece com determinados valores, oscile durante o ano e termine o ano muito próximo de como começou, não demonstrando avanço.

Num segundo momento foram realizadas análises com relação a filial de Rio do Sul, seguindo o mesmo roteiro utilizado para a análise na Matriz Blumenau.

Assim, ao verificar os valores de vendas desta filial que também compreende o período de agosto de 1994 até junho de 2005, foi possível notar uma tendência linear crescente. Novamente os valores foram atualizados de acordo com o IGP-M.

Desta forma, na figura 52 é possível notar a tendência linear e mais clara ainda após a aplicação de médias móveis de 12 períodos e depois centrada, que suaviza o modelo, demonstrando mais claramente o fator da tendência.

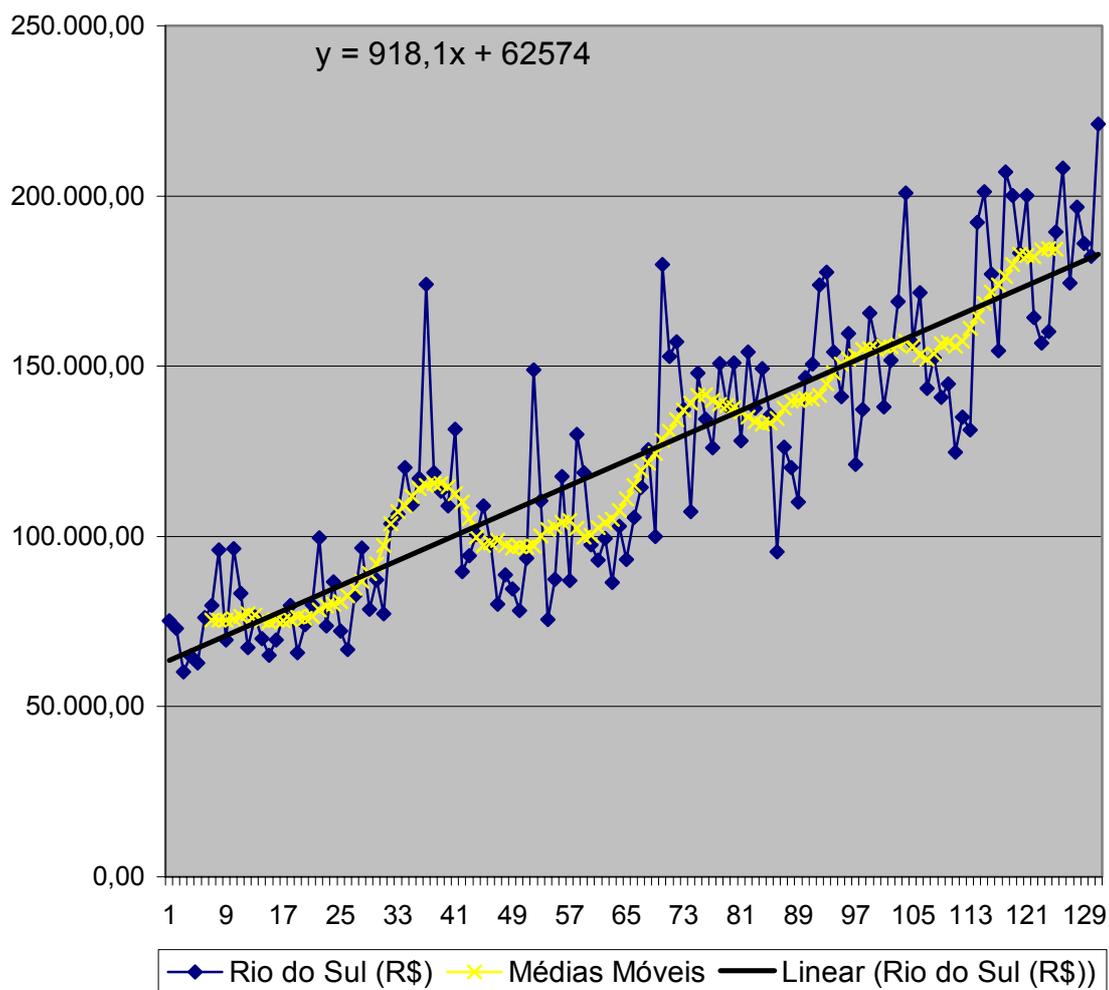


Figura 52 – Tendência linear crescente para valores de vendas da filial Rio do Sul

Após esta verificação, assim como na análise anterior, procurou-se observar a questão do possível efeito sazonal que pudesse estar sofrendo a série de valores observada. Para tanto então, seguindo o roteiro da análise da Matriz Blumenau, foram criados índices sazonais, baseados na mediana dos valores sazonais encontrados para cada mês em específico. Ao encontrar estes índices, os mesmos foram aplicados aos valores da tendência com a intenção de remover o seu efeito, pois seus índices apresentaram valores significativamente diferentes de 1 (chegando até quase 16% de

diferença em alguns meses) e demonstrar graficamente os valores sem o efeito sazonal.

A tabela 15 demonstra os índices sazonais alcançados;

Meses	Índices
Janeiro	0,974444
Fevereiro	0,978951
Março	1,063223
Abril	0,96642
Maio	1,157902
Junho	1,003135
Julho	1,013593
Agosto	0,94859
Setembro	0,894163
Outubro	0,942291
Novembro	0,950701
Dezembro	0,89376

Tabela 15 – Índices Sazonais encontrados para filial Rio do Sul

A figura 53 apresentada a seguir demonstra graficamente os valores de faturamento da filial Rio do Sul, sem o efeito sazonal onde é possível notar mais claramente a tendência linear crescente da série.

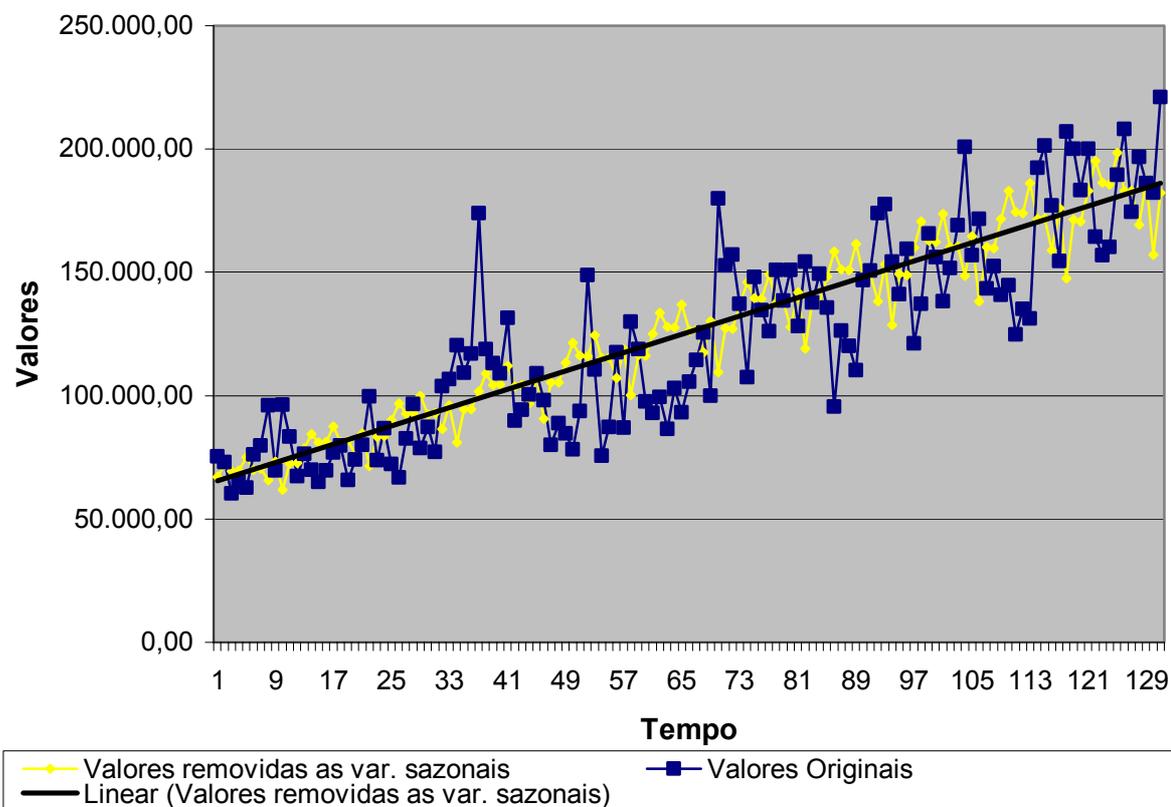


Figura 53 – Movimento da série sem as variações sazonais

Ao observar a tabela de índices e depois sua aplicação demonstrada graficamente, é possível notar o efeito sazonal que está sobre a série de valores de faturamento, assim, é importante frisar a necessidade de se considerar estas variações para efeitos de previsão.

Após esta primeira observação, novamente procurou-se observar também se a série sofre o efeito de variações cíclicas/irregulares e para tanto, foram retiradas a tendência e o efeito sazonal, ficando apenas as variações cíclicas que são demonstradas através da figura 54.

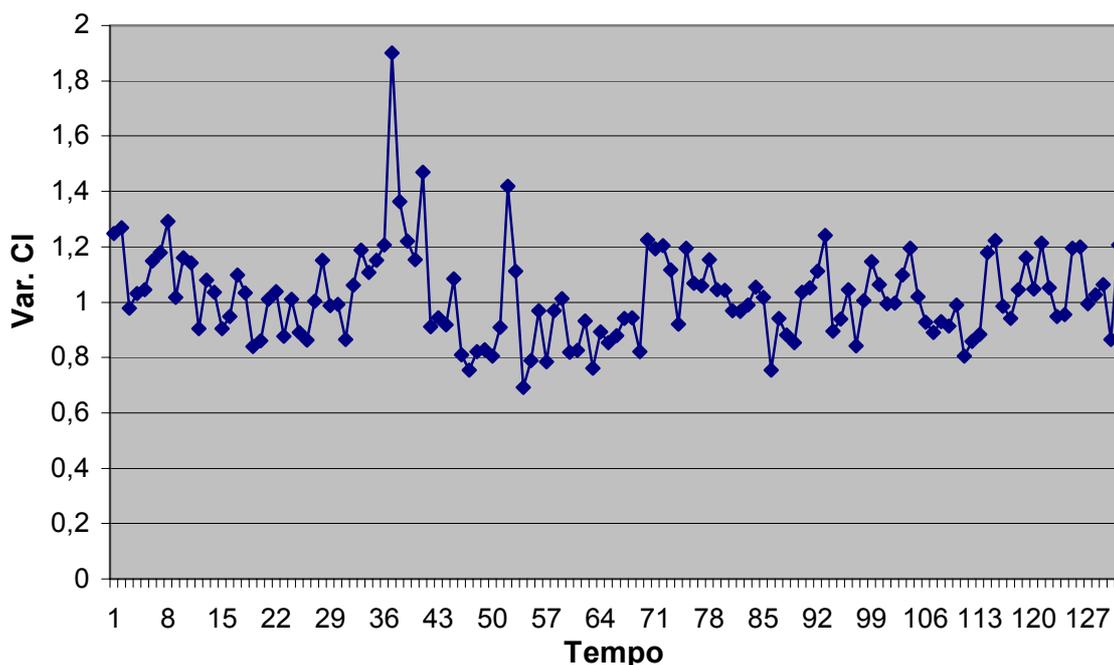


Figura 54 – Variações Cíclicas/Irregulares

Observando a figura 54 não é possível notar variações sistemáticas em torno do valor 1 que caracterizam os ciclos, porém, para se ter uma idéia mais exata disto, aplicou-se as médias móveis de 12 períodos e depois centrada para tentar suavizar a série e tentar perceber se havia realmente algum tipo de ciclo na série apresentada. Veja conforme figura 55.

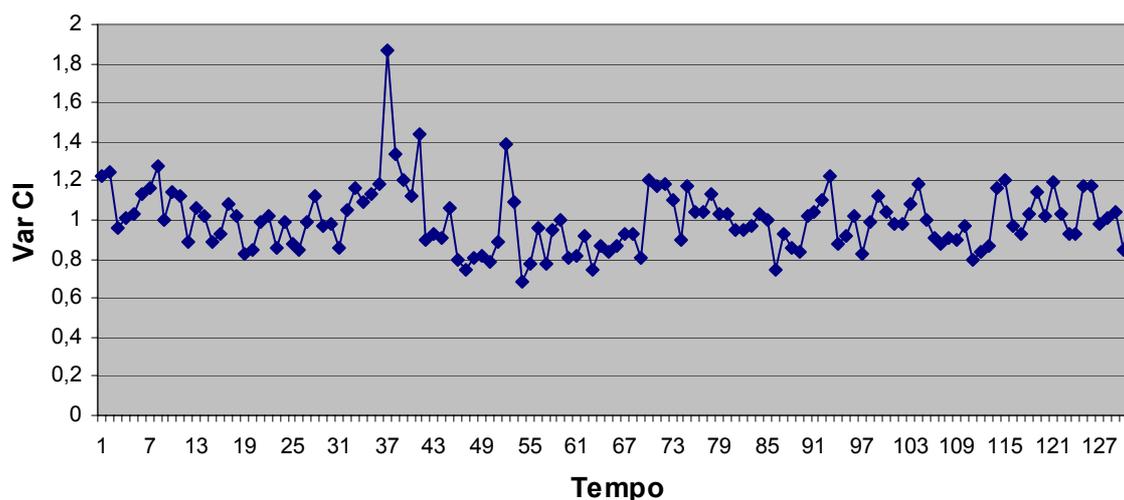


Figura 55 – Variações Cíclicas/Irregulares após aplicação das médias móveis de 12 períodos posteriormente centrada

Assim, mesmo após a aplicação das médias móveis não foi possível observar variações sistemáticas em torno do eixo 1 constituindo ciclos, portanto, também nesta análise estas variações serão desconsideradas do modelo para fins de previsão.

Desta forma, após realizar as verificações de tendência, sazonalidade e variações cíclicas/irregulares, foi possível através da equação $y = ((918,1x) + 62574) * \text{ind.sazonal}$, realizar a previsão de valores para os 12 meses de 2006, demonstrando ao executivo a importância da utilização das séries temporais e permitindo ao mesmo que possa traçar ações de melhoria para continuar fazendo os valores de faturamento crescer ainda mais ou no mínimo manter os valores atuais.

A tabela 16 mostra os valores previstos através da equação citada, apenas como exemplo de aplicação.

Períodos	Índices Sazonais	Meses	Previstos
138	0,974443793	jan/06	R\$ 184.434,73
139	0,978950733	fev/06	R\$ 186.186,54
140	1,063222899	mar/06	R\$ 203.190,40
141	0,966419709	abr/06	R\$ 185.577,81
142	1,157902171	mai/06	R\$ 223.410,51
143	1,003135412	jun/06	R\$ 194.470,14
144	1,013592762	jul/06	R\$ 197.428,00
145	0,948589788	ago/06	R\$ 185.637,60
146	0,894162838	set/06	R\$ 175.807,26
147	0,94229135	out/06	R\$ 186.135,24
148	0,950701121	nov/06	R\$ 188.669,30
149	0,893759999	dez/06	R\$ 178.189,74

Tabela 16 – Valores previstos para 2006 da filial Rio do Sul

Abaixo segue a visualização gráfica que é apresentada através da figura 56, onde é possível observar as oscilações que ocorrem durante os meses do ano.

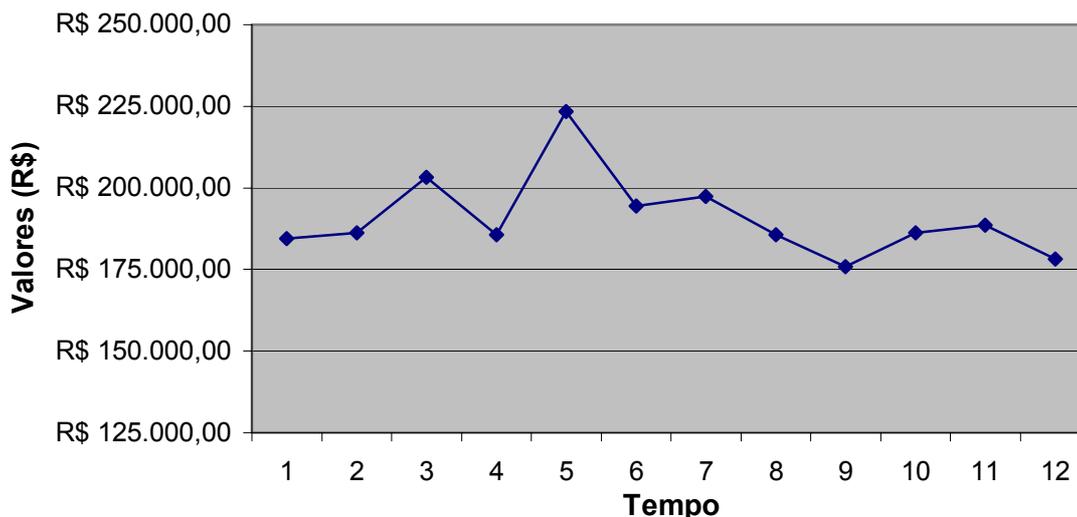


Figura 56 – Valores Previstos para 2006 da filial Rio do Sul

Para uma última demonstração de séries temporais como instrumento de suma importância para a tomada de decisões por parte do executivo em relação ao ponto Tempo abordado pela metodologia SIEGO, foi analisada a filial de Chapecó que atualmente, por motivos não pertinentes a este trabalho, já foi fechada, porém sua análise é importante para demonstrar ao executivo uma previsão de faturamento para o ano de 2006, por exemplo, se a empresa tivesse continuado sua atividade e permanecesse dentro do cenário em que vinha atuando. Assim, segue.

O período passado de análise compreende valores de faturamento desde agosto de 1994 até outubro de 2002 quando a empresa encerrou suas atividades. Assim, observando os valores passados, foi possível notar uma tendência linear crescente conforme apresentado na figura 57, que ficou ainda mais clara com a aplicação de médias móveis 12 períodos posteriormente centrada. Nota-se que após o período 76 há uma queda, mas que se justifica pelo fato da empresa ter reduzido seu quadro de funcionários após este período e também ter reduzido sua linha de atuação.

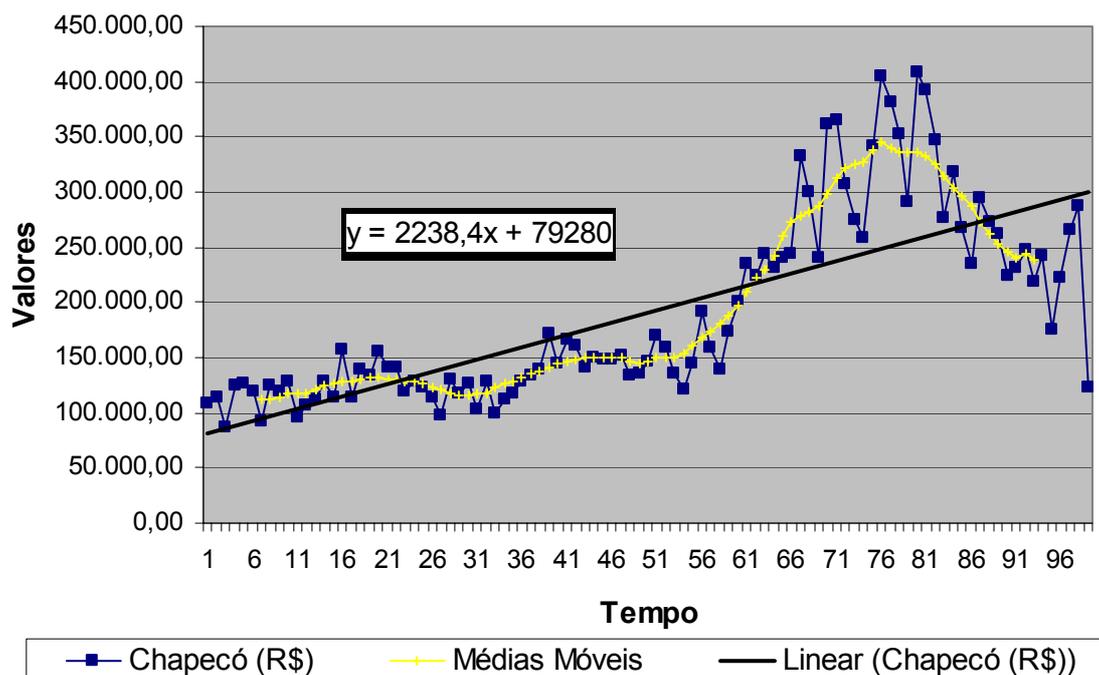


Figura 57 – Tendência linear crescente dos valores de faturamento da filial Chapecó

Após analisar a questão da tendência, assim como nas demais análises procurou-se observar também se a série sofre o efeito da sazonalidade. Para tanto, foram calculados os índices sazonais e posteriormente a mediana destes índices para cada mês de faturamento constituindo 12 índices, referentes a cada mês específico do ano.

Através destes índices e também da visualização gráfica dos valores sem o efeito sazonal, foi possível notar que a série apresenta influência sazonal sobre seus valores, pois os valores chegam a apresentar cerca de até 9% (mês de março, por exemplo) de diferença em relação ao valor 1 que indicaria a não presença de variações sazonais.

A tabela 17 apresenta os índices encontrados.

Meses	Índices
Janeiro	1,044618
Fevereiro	0,920128
Março	1,090204
Abril	0,952423
Maior	1,068658
Junho	0,928916
Julho	0,970572
Agosto	0,930399
Setembro	0,995892
Outubro	1,062973
Novembro	1,062558
Dezembro	1,006343

Tabela 17 – Índices Sazonais encontrados para filial Chapecó

A figura 58 demonstra os valores de faturamento sem o efeito sazonal onde é possível notar claramente uma tendência linear de crescimento, assim para efeitos de previsão é preciso considerar este efeito de sazonalidade.

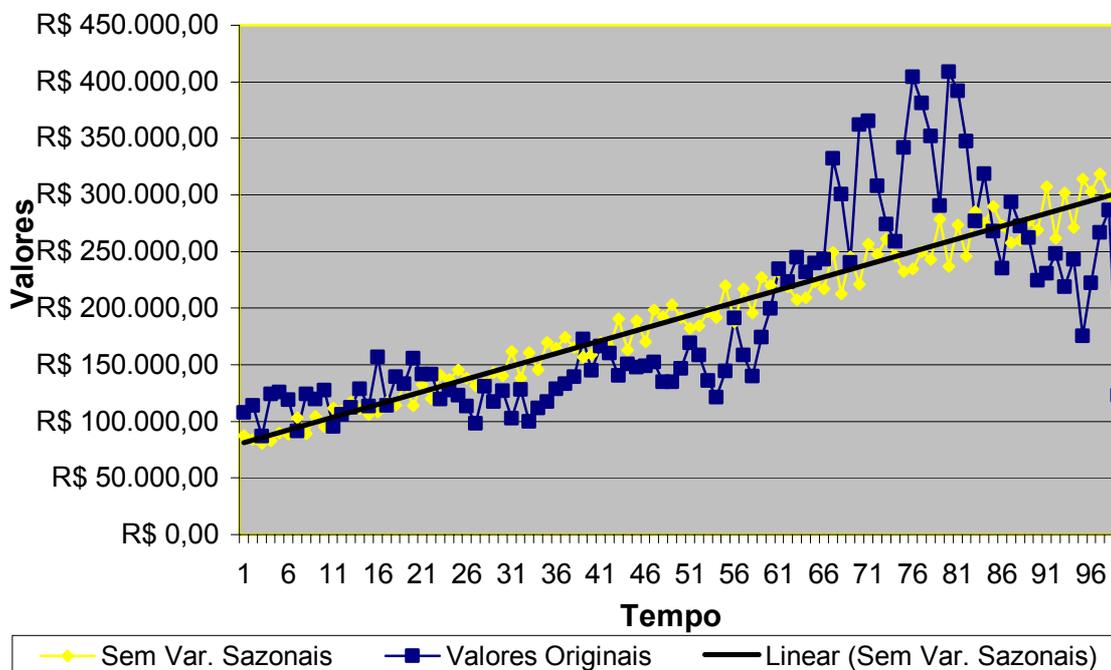


Figura 58 – Valores de faturamento sem as variações sazonais

Após a verificação do efeito sazonal sobre a série, procurou-se observar se a série também sofre os efeitos das variações cíclicas/irregulares e assim como nas demais análises foram retirados os efeitos sazonais deixando apenas as variações cíclicas/irregulares para verificação de possíveis variações sistemáticas em torno do eixo 01 constituindo ciclos.

Segue a figura 59 onde estão apenas as variações cíclicas/irregulares.

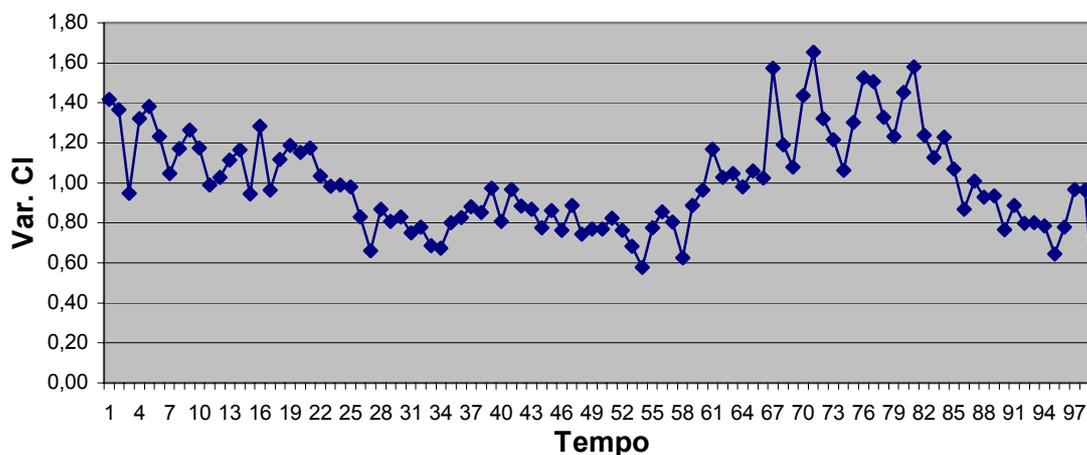


Figura 59 – Variações cíclicas/irregulares sobre valores da filial Chapecó

Ao observar a figura 59 parecem existir três períodos de oscilações em torno do valor 1, porém não ficando muito clara esta variação. Para tentar uma melhor visualização aplicaram-se as médias móveis de 12 períodos e posteriormente centrada para tentar uma suavização da série, conforme a figura 60.

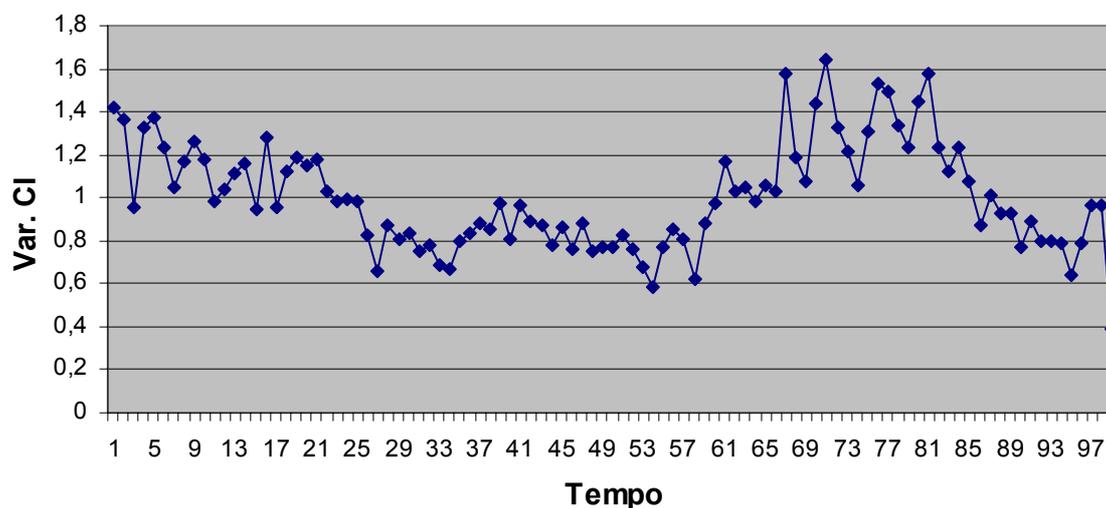


Figura 60 – Variações cíclicas/irregulares após aplicação das médias móveis

Mesmo após aplicação das médias móveis não é possível afirmar que existem variações sistemáticas com precisão, portanto neste modelo também foram desconsideradas estas variações.

Assim sendo, para efeitos de previsão foi utilizada a equação que leva em consideração a tendência e sazonalidade ficando assim:

$$y = ((2238,4x) + 79280) * \text{ind.sazonal}$$

Baseado então nesta equação, a tabela 18 apresenta valores de faturamento previstos para o ano de 2006 (períodos de 138 a 149).

Meses	Valores Previstos
jan/06	R\$ 405.498,94
fev/06	R\$ 359.234,21
mar/06	R\$ 428.074,95
abr/06	R\$ 376.106,49
mai/06	R\$ 424.398,97
jun/06	R\$ 370.982,36
jul/06	R\$ 389.791,02
ago/06	R\$ 375.739,81
set/06	R\$ 404.418,30
out/06	R\$ 434.038,23
nov/06	R\$ 436.247,10
dez/06	R\$ 415.419,97

Tabela 18 – Valores previstos de faturamento para filial Chapecó

A figura 61 demonstra graficamente as oscilações nos valores previstos de faturamento da filial Chapecó para o ano de 2006.

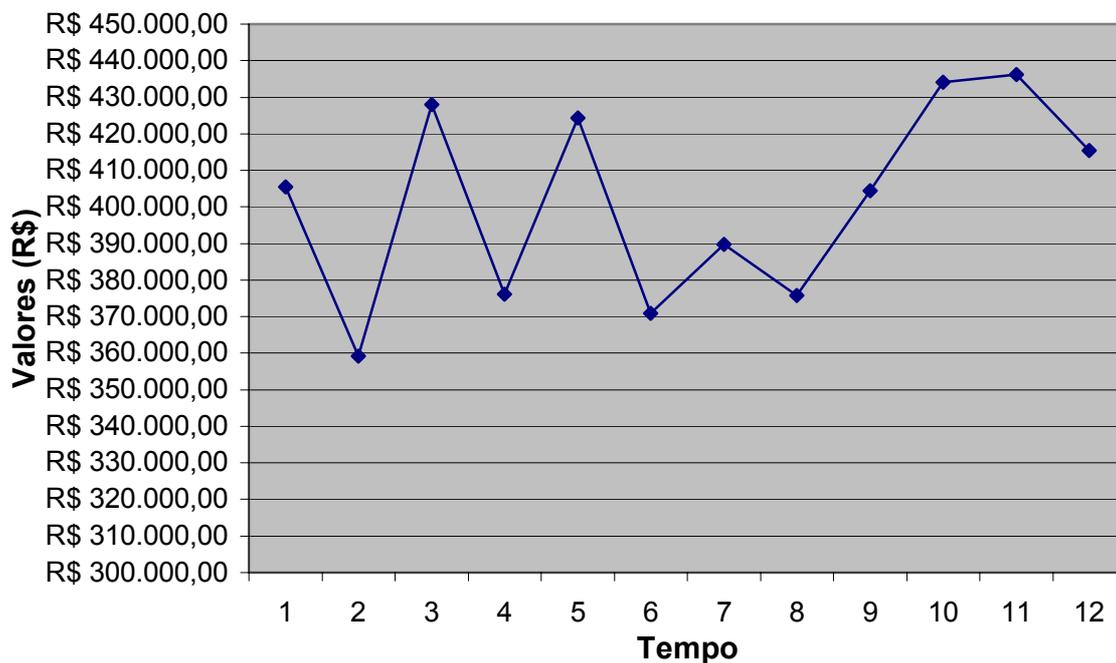


Figura 61 – Valores previstos de faturamento para filia Chapecó em 2006

Observando os valores de faturamento apresentados, foi possível notar que se a filial de Chapecó ainda estivesse em atividade e seus cenários de atuação bem como outros fatores como, por exemplo, a situação econômica do país tivesse mantido as reações que vinha tendo no passado, a filial de Chapecó poderia estar atuando com valores expressivos de faturamento. Obviamente que a questão da lucratividade é um outro assunto a ser abordado e que não é pertinente a este trabalho.

Sendo assim, fica mais uma vez demonstrado a importância da integração dos métodos estatísticos com as outras duas abordagens para fins de auxílio ao executivo em suas tomadas de decisões.

O próximo tópico visa discutir os resultados obtidos através das análises realizadas e o benefício trazido através da integração das abordagens.

4.6 DISCUSSÃO

Este tópico visa uma breve discussão sobre a integração realizada, observando seus resultados e os benefícios trazidos.

Como já mencionado no capítulo introdutório deste trabalho, uma grande questão a ser levantada quando da tomada de decisão por parte do executivo é justamente o fato de que em muitos casos o executivo resolve tomar algumas atitudes baseado “apenas” na sua experiência com relação ao assunto no qual deseja agir. Isso pode até dar certo, mas a proposta deste trabalho era justamente tentar eliminar esta questão da subjetividade na decisão e mostrar ao executivo com embasamento científico quais os pontos onde ele realmente pode atuar para que ocorram melhorias.

Dentro desta visão, a integração das três abordagens, SIEGO, *Data Warehouse* e métodos estatísticos, casou muito bem e tanto isso é verdade que através dos resultados obtidos foi possível demonstrar ao executivo que nem sempre a resposta mais lógica e aparentemente correta com base em sua experiência de negócios, é a resposta correta e que pode lhe levar a tomar as melhores ações de melhorias na empresa. Esta visão ficou bastante clara quando ao analisar o item custo da metodologia SIEGO através do método estatístico de análise de regressão foi possível observar que a variável quantidade não era tão significativa quanto o executivo imaginava antes que as análises fossem iniciadas. Em um primeiro momento o executivo tinha a visão de que muitos itens haviam sido vendidos abaixo do preço mínimo em função da quantidade comprada, o que ficou claro ser uma visão um pouco equivocada, após a análise. Assim, outros exemplos como este tornam claro a validade da integração das três abordagens resultando em informações valiosas e comprovadas cientificamente.

Este tipo de informação é de fundamental importância para que o executivo possa traçar suas ações de melhorias e que tenha uma chance de acertar muito maior. É claro que a decisão das ações de melhoria a serem tomadas a partir dos dados apresentados deve partir do executivo, uma vez que a função do “sistema” como um todo é informar o executivo dos caminhos que ele pode tomar.

Fica claro também que para se analisar valores com a intenção de previsão futura através de séries temporais, é necessário ter um período de avaliação passada bastante grande (pelo menos superior a 36 meses), segundo Montgomery (2003), o que não ocorreu ainda nas filiais de Blumenau, Gaspar e Indaial, sendo assim que um estudo

futuro poderá ainda contemplar estas análises, pois o período passado de análise será maior.

Dentro de uma visão global, vale comparar que inicialmente o executivo não tinha noção de nenhuma das informações apresentadas após a análise, sendo para ele de suma importância os valores apresentados. Outro ponto importante a ser colocado é que o executivo é peça fundamental no processo de construção do modelo de análise, o que já é previsto na metodologia SIEGO.

5 CONCLUSÕES

Esta seção visa discutir os resultados do trabalho de pesquisa bem como apresentar as conclusões obtidas a partir da pesquisa realizada, apresentando algumas limitações e também sugestões para trabalhos futuros.

5.1 CONSIDERAÇÕES FINAIS

O mercado em que a empresa foco deste estudo está inserida é um ambiente de extrema competitividade o que torna crucial cada passo tomado pela diretoria executiva da empresa, o que tornou a finalidade deste trabalho também de suma importância para os executivos desta empresa. Este tipo de informação apresentada não é uma necessidade que seja um privilégio da empresa em estudo, portanto, a aplicação desta dissertação não fica condicionada apenas ao ambiente onde ela foi inserida, mas sim sua aplicação é recomendada a todos os ramos de atividades onde a tomada de decisões seja um fator delimitador entre lucrar e perder dinheiro.

Ao analisar o objetivo macro deste trabalho é possível perceber que o mesmo foi contemplado suprimindo a questão do trabalho, que seria a tomada de decisões sem embasamento científico por parte do executivo. As análises apresentadas através da associação das abordagens refletem justamente o fator crucial que é a apresentação de informações ao executivo que demonstrem de maneira comprovada os pontos onde ele pode agir para trazer melhorias para seu negócio, o que demonstra a contribuição deste. Naturalmente os objetivos específicos também foram contemplados.

Conforme a sugestão da metodologia SIEGO, todos os pontos do tripé, Custo, Tempo e Qualidade foram abordados e também analisados de modo a apresentar ao executivo os locais de melhoria a serem focados.

Quando a pesquisa abordou a questão do Custo, ficou bastante claro para o executivo, após as análises, que em muitos casos decisões são tomadas a partir de conjecturas que podem não refletir o que os dados realmente demonstram. Inicialmente o executivo tinha forte convicção de que as variáveis Quantidade e Prazo médio, por exemplo, eram variáveis de suma importância ou extrema influência na variação do preço de venda, o que após as análises foi desmentido, dando uma nova visão ao executivo e demonstrando a importância de se ter um embasamento científico para a tomada de decisão.

Ao verificar a questão da Qualidade, o executivo também se surpreendeu ao verificar que o seu melhor cliente tinha sido classificado apenas como BOM, sendo que em sua “visão” o mesmo era com certeza um ótimo cliente. Mais uma vez, foi possível notar que em um primeiro momento o executivo se deixa levar por impulsos e apenas cifras, mas quando observa as demais variáveis nota que sem uma análise prévia não é adequado tirar conclusões apressadas. Assim, o executivo também pode verificar pontos onde o mesmo poderia agir para aumentar a quantidade de clientes classificados como bons e ótimos e também analisar novamente os parâmetros de qualidade definidos inicialmente.

Com relação ao item Tempo, as análises demonstraram ao executivo sua situação real em termos de faturamentos de modo que o mesmo pudesse ter uma idéia ao longo do tempo de como sua empresa vem atuando em termos de vendas e através das séries temporais aplicadas o mesmo pudesse já ter uma noção de como sua empresa poderia agir no período previsto. Também foi muito interessante o fato de se observar possíveis previsões de vendas para a filial de Chapecó que encerrou suas atividades em 2002, fazendo o executivo refletir sobre a atitude tomada no passado. Outro fato muito interessante apresentado através da análise foi o fator da tendência linear decrescente da empresa Matriz mostrando ao executivo que o mesmo precisa analisar bem o seu negócio para ver o que está acontecendo com a empresa para que o faturamento esteja caindo de tal forma, bem como ao analisar a filial de Rio do Sul o executivo pôde observar que a mesma vem crescendo seguindo uma tendência linear crescente o que pode ser reflexo de ações corretas tomadas na empresa ou de outros fatores que a partir de agora o executivo poderia tentar visualizar para quem sabe aplicar a empresa Matriz para reverter o seu quadro. Assim, mais uma vez ficou claro para o executivo que com o auxílio de informações com embasamento científico fica muito mais transparente a situação da empresa permitindo que o mesmo possa tomar decisões adequadas e atingir os verdadeiros focos que resultarão em melhorias.

Hoje, uma das maiores dificuldade enfrentadas pelo executivo é justamente o fato de que as informações estão presentes na empresa, porém o agrupamento delas de forma significativa e útil em muitos casos se torna algo inatingível por parte deste executivo, logo, as informações aqui apresentadas, foram totalmente ao encontro com as necessidades do executivo, demonstrando mais uma vez as contribuições práticas desta dissertação.

Foi a partir dos dados apresentados que o executivo passou a observar de modo diferenciado o seu negócio, podendo então tomar atitudes focadas nos pontos apresentado através das análises deste trabalho.

5.2 LIMITAÇÕES DO TRABALHO

Durante a elaboração deste trabalho de pesquisa pode-se definir como uma limitação a dificuldade encontrada na equipe de trabalho onde o executivo muitas vezes por não conhecer da área de sistemas tinha receio em aceitar as idéias apresentadas e também em aceitar as informações colocadas, o que acabou sendo superado após todas as análises feitas, pois o mesmo ficou convencido dos resultados apresentados.

Outra dificuldade encontrada foi o volume de informações levantadas em detrimento ao banco de dados utilizado, pois o mesmo em alguns casos apresentou dificuldades de processamento.

Observou-se que os gráficos apresentados quando da utilização de séries temporais não apresentaram resultados que possam ser considerados excelentes, então talvez os métodos estatísticos utilizados para este fim não possam ser considerados os melhores, apresentando limitações, logo a sua aplicação demonstrou certa limitação.

5.3 SUGESTÕES PARA TRABALHOS FUTUROS

Durante o desenvolvimento deste trabalho foram utilizadas três abordagens relativas aos processos de levantamento, desenvolvimento e análise que são a metodologia SIEGO, *Data Warehouse* e os métodos estatísticos. Esta associação permitiu ao executivo observar pontos onde poderiam ser realizadas melhorias.

Dentro desta visão, sugere-se que a partir do apresentado possam ser utilizadas outras técnicas como redes neurais, por exemplo, para auxiliar o executivo em que tipo de melhoria ele pode realizar com relação aos pontos de melhorias apresentados.

Uma outra sugestão ainda seria ampliar os objetos de estudo com relação ao tripé Tempo, Custo e Qualidade apontando outros pontos de melhorias e aplicando-se outras técnicas de análise de dados.

Como última sugestão, poderia ser realizado um trabalho onde a metodologia SIEGO pudesse ser associada com a técnica de *Balanced Scorecard* tentando criar uma

metodologia genérica onde a mesma pudesse ser utilizada para quaisquer áreas apenas identificando os pontos de melhorias dentro de cada etapa do *Balanced Scorecard* e focando no tripé mencionado na metodologia SIEGO.

REFERÊNCIAS BIBLIOGRÁFICAS

BARBETTA, Pedro A. **Estatística aplicada às ciências sociais**. 5 ed. Florianópolis: Editora da UFSC, 2002.

BUSSAB, W. O., MORETTIN, P. A. – **Estatística básica**. 5 ed. São Paulo: Editora Saraiva, 2003.

CAUTELA, Alciney L., POLLONI, Enrico G. F. **Sistemas de Informação na Administração de Empresas**. São Paulo: Atlas, 1996.

DALFOVO, Oscar. **Quem tem informação é mais competitivo**. Blumenau: Acadêmica, 2000.

DALFOVO, Oscar. **Metodologia sistema de informação estratégico para o gerenciamento operacional (SIEGO)**. Um modelo siego para universidade com aplicação na gestão ambiental baseado em data warehouse. 2001. 308 f. Tese (Doutorado em Ciência da Computação) – Centro Tecnológico de Computação - Universidade Federal de Santa Catarina, Florianópolis.

DOWNING, Douglas.;CLARK, Jeffrey. **Estatística aplicada**. 2 ed. São Paulo: Saraiva, 2002.

FARIAS, Alfredo A. de.; SOARES, José F.; CÉSAR, Cibele Comini. **Introdução à estatística**. 2 ed. Rio de Janeiro: LTC, 2003.

FIREBIRD PROJECT. **Firebird**: Relational database for the new millennium. New South Wales, Austrália, 2005. Disponível em: <<http://firebird.sourceforge.net/>>. Acesso em: 09 jan. 2006.

FONSECA, J. S.; MARTINS, G.A; TOLEDO, G.L. **Estatística Aplicada**. São Paulo: Atlas, 1995.

FREUND, John E.; SIMON, Gary A. **Estatística aplicada: economia, administração e contabilidade**. 9 ed. Porto Alegre: Bookman, 2000.

GONÇALVES, Alexandre Leopoldo. **Utilização de Técnicas de Mineração de Dados em Bases de C&T: Uma Análise dos Grupos de Pesquisa no Brasil**. 2000. Dissertação (Curso de Pós-Graduação em Engenharia de Produção) Universidade Federal de Santa Catarina.

GHODDOSI, Nader. **Sistema de Informação Estratégico para o Gerenciamento Operacional (SIEGO) como uma ferramenta de apoio de decisão utilizando Data Warehouse. 2003.** Dissertação (Curso de Pós-Graduação em Ciência da Computação) Universidade Federal de Santa Catarina.

GROTH, Robert. **Data Warehouse: a hands-on approach for business professionals conceitos e soluções.** New Jersey: Prentice Hall, 1997.

HAIR, J.F. Jr. et al. **Análise Multivariada de Dados.** Porto Alegre: Bookman, 2005.

HOFFMANN, Rodolfo; VIEIRA, Sônia. **Análise de regressão: uma introdução à econometria.** São Paulo: Hucitec, 1987.

INMON, William H. **Como construir o data warehouse.** Rio de Janeiro: Campus, 1997.

JÚNIOR, Adelir José Schuler.; PEREZ, Anderson Luiz Fernandes. Análise do perfil do usuário de telefonia utilizando técnicas de mineração de dados. **RESI**, [s.l],v.1, n1, 2006.

KIMBALL, Ralph. **Data warehouse architect: letting the users sleep, part1,** New York, dez. 1996. Disponível em: <<http://www.dbmsmag.com/9612d05.html>>. Acesso em: 04 abr. 2002.

KIMBALL, Ralph. **Data warehouse architect: letting the users sleep, part2,** New York, jan. 1997. Disponível em: <<http://www.dbmsmag.com/9701d05.html>>. Acesso em: 04 abr. 2002.

KIMBALL, Ralph.; ROSS, Margy. **The data warehouse toolkit: guia completo para modelagem dimensional.** Rio de Janeiro: Campus, 2002.

KUME, Hitoshi. **Statistical Methods for Quality Improvement.** São Paulo: Gente, 1993

LAUDON, Jane P., LAUDON, Kenneth C. **Sistemas de Informação Gerenciais – Administrando a Empresa Digital.** São Paulo: Prentice Hall, 2004.

LEVINE, D. M., BERENSON, M. L. e STEPHAN, D. – **Estatística: Teoria e aplicações usando o Excel.** Rio de Janeiro: LTC, 2000.

MILONE, GIUSEPPE. **Estatística Geral e Aplicada**. São Paulo: Pioneira Thomson Learning, 2004.

MONTGOMERY, Douglas C., RUNGER, George C. **Estatística Aplicada e Probabilidade para Engenheiros**. 2a ed. Rio de Janeiro: LTC, 2003.

MOREIRA, Délio. **Métodos Estatísticos para administradores e economistas**. São Paulo: Loyola, 1975.

NETER, John. et al. **Applied Linear Statistical Models**. 4a ed. Boston: McGraw-Hill, 1996.

O'BRIEN, James A. **Sistemas de Informação e as decisões gerenciais na era da internet**. São Paulo: Saraiva, 2003.

OLIVEIRA, Maurício. Luz no túnel para indústria têxtil. **Revista Empreendedor**, Santa Catarina: v. 29, n. 29, p. 12-23, fev. 1997.

OLIVEIRA, Adelise G. de. **Data Warehouse conceitos e soluções**. Florianópolis: SFO Gráfica e Editora Ltda, 1998.

OLIVEIRA, Djalma de Pinho Rebouças. **Sistemas de informações gerenciais**. São Paulo: Atlas, 2002.

PAZ, Salete do Bonfim; BORGES, Díbio Leandro. Uso de redes neuronais artificiais de função de base radial em previsão do consumo de energia elétrica. In: SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 19, Rio de Janeiro. **Anais do XIX Congresso Nacional da Sociedade Brasileira de Computação**. Rio de Janeiro: EntreLugar, 1999. 584 p. v. 4, p. 481-493.

RAUTENBERG, Sandro. **Predição de receitas de cores na estamparia têxtil através de redes neurais com função de base radial**. Florianópolis, 1998. Dissertação (Mestrado em Engenharia de Produção) Universidade Federal de Santa Catarina.

SARLE, Warren S. **Neural Networks and Statistical Models**. Proceedings of the Nineteenth Annual SAS Users Group International Conference, April, 2004.

SANTOS, Alcione Miranda dos.; et al. **Usando Redes Neurais Artificiais e Regressão Logística na Predição da Hepatite A**. São Luiz: UFMA, 2005.

SCHENATZ, Bianca Nardelli. **Utilização de *Data Mining* em um Sistema de Informação Gerencial para o Diagnóstico da Formação de Professores de Graduação.** 2005. Dissertação (Pós-Graduação em Engenharia de Produção) Universidade Federal de Santa Catarina.

SCHMITT, Sandro. **Aplicação do SIEGO na área financeira baseado em um *Data Warehouse*.** 2001. Monografia (Bacharelado em Ciência da Computação) – Departamento de Sistemas e Computação – Universidade Regional de Blumenau, Blumenau.

SILVA, Edna Lúcia da., MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação.** Florianópolis: UFSC, 2005.

SPIEGEL, Murray R. **Estatística.** São Paulo: Makron Books, 2004.

STATSOFT Inc. **STATISTICA for Windows: Electronic Manual.** Tulsa, Oklahoma, USA, 1995.

STEVENSON, Willian J. **Estatística aplicada à administração.** São Paulo: Harbra, 1986.

STREY, Rodrigo Reno. **Sistema de informação aplicado ao setor financeiro de uma empresa baseado no SIEGO utilizando *Data Warehouse*.** 2002. Monografia (Bacharelado em Ciência da Computação) – Departamento de Sistemas e Computação – Universidade Regional de Blumenau, Blumenau.

WILD, Christopher J.; SEBER, George A. F. **Encontros com o acaso: Um primeiro curso de análise de dados e inferência.** Rio de Janeiro: LTC, 2004.

WILSON, Mario Perez. **Seis Sigma: compreendendo o conceito, as implicações e os desafios.** Rio de Janeiro: Qualitymark, 2000.

WONNACOTT, Thomas H. **Estatística aplicada à economia e à administração.** Rio de Janeiro: Livros Técnicos e Científicos, 1981.