

ANDRÉ CHANG ALVIM

**PREVISÃO DE
DEMANDA NO
VAREJO**

Trabalho de Formatura apresentado
à Escola Politécnica da Universidade de
São Paulo para a obtenção do Diploma de
Engenheiro de Produção – Área Mecânica

São Paulo

2003

ff 2003
Ab 888

ANDRÉ CHANG ALVIM

**PREVISÃO DE
DEMANDA NO
VAREJO**

Trabalho de Formatura apresentado
à Escola Politécnica da Universidade de
São Paulo para a obtenção do Diploma de
Engenheiro de Produção – Área Mecânica

Orientador: Prof. Dr. Miguel Cezar Santoro

São Paulo

2003

Agradecimentos

Aos meus pais pelo total apoio, confiança e colaboração não somente na elaboração deste trabalho, mas também durante todos os anos do curso de Engenharia de Produção.

Ao Professor Miguel Santoro pela valiosa orientação durante a execução deste trabalho.

A todos na Procter & Gamble que de alguma forma contribuíram na elaboração deste trabalho, e em especial a Malú e ao Vicente pelo apoio e colaboração prestados.

Resumo

Este trabalho de formatura tem como proposta desenvolver um sistema de previsão de vendas no varejo. O trabalho está estruturado da seguinte forma: no capítulo 1 é feita uma breve apresentação da empresa onde foram realizados o trabalho e o estágio. O capítulo 2 esclarece o problema e o objetivo proposto. No capítulo 3 é feita uma abordagem dos principais métodos de previsão (temporais e causais) e outros pontos relevantes ao assunto. O desenvolvimento do tema é o foco do capítulo 4, que segue a seguinte lógica: análise dos dados levantados, seleção do modelo de previsão, desenvolvimento do modelo temporal, desenvolvimento do modelo causal e os testes e resultados do sistema de previsão. O capítulo 5 é dedicado às conclusões finais, e em seguida estão a bibliografia e os anexos relevantes.

Abstract

The proposal of this essay is to develop a retail's demand forecasting model. This work is based on the following structure: on chapter 1 there is a presentation of the company where the work and the internship took place. On chapter 2, the problem and the goal are defined. On chapter 3 the main forecasting methods (time-series and explanatory models) and other relevant topics are explored. The chapter 4 focuses on the development of the subject, according to the following sequence: data analysis, forecasting model selection, time-series model development, explanatory model development and last the tests and results of the demand forecasting model. The chapter 5 is dedicated to the final conclusions, and after that one can find the bibliography and the attachments.

Sumário

1. INTRODUÇÃO	6
1.1. EMPRESA.....	7
1.2. PRODUTOS.....	7
1.3. PRINCÍPIOS E VALORES FUNDAMENTAIS	7
1.4. ÁREA DE VENDAS	8
1.4.1. <i>As Equipes Multifuncionais</i>	9
1.4.1.1. Vantagens	9
1.4.1.2. Desvantagens.....	9
1.5. O ESTÁGIO.....	9
1.5.1. <i>Gerenciamento de Categoria</i>	10
2. O TRABALHO.....	11
2.1. O PROBLEMA	12
2.2. O OBJETIVO	12
2.3. RESUMO	13
2.3.1. <i>Revisão Bibliográfica</i>	14
2.3.2. <i>Levantamento dos Dados</i>	14
2.3.3. <i>Análise Preliminar dos Dados</i>	14
2.3.4. <i>Seleção do Modelo de Previsão</i>	15
2.3.5. <i>Desenvolvimento do Modelo de Previsão</i>	15
2.3.6. <i>Testes/ Análises dos Resultados</i>	15
2.3.7. <i>Sistema de Predição de Vendas</i>	15
3. REVISÃO BIBLIOGRÁFICA.....	16
3.1. INTRODUÇÃO.....	17
3.1.1. <i>Por que prever?</i>	17
3.2. MÉTODOS DE PREVISÃO	18
3.2.1. <i>Definição de Periodos</i>	19
3.2.2. <i>Medidas de Erro de Previsão</i>	20
3.2.3. <i>Métodos Qualitativos</i>	22
3.2.3.1. Método Delphi.....	23
3.2.3.2. Pesquisa Sobre Intenção de Compra	24
3.2.3.3. Consenso da Força de Vendas.....	24

3.2.3.4. Consenso de Executivos	25
3.2.4. <i>Métodos Quantitativos</i>	25
3.2.4.1. Métodos de Séries Temporais	27
3.2.4.1.1. Método da Média Simples	27
3.2.4.1.2. Método da Média Móvel	28
3.2.4.1.3. Método da Suavização Exponencial Simples	29
3.2.4.1.4. Método da Suavização Exponencial Simples com Taxa de Resposta Adaptativa	30
3.2.4.1.5. Método da Suavização Exponencial Dupla (Método de Brown)	31
3.2.4.1.6. Método da Suavização Exponencial Linear com Dois Parâmetros (Método de Holt)	32
3.2.4.1.7. Método da Suavização Exponencial Linear com Três Parâmetros (Método de Winters)	33
3.2.4.1.8. Métodos de Decomposição	34
3.2.4.1.9. Método de Box-Jenkins	34
3.2.4.2. Métodos Causais	35
3.2.4.2.1. Regressão Simples	35
3.2.4.2.2. Regressão Múltipla	40
4. DESENVOLVIMENTO	44
4.1. MODELO DE PREVISÃO ATUAL	45
4.2. HORIZONTE DE PREVISÃO	45
4.3. DADOS DISPONÍVEIS	46
4.3.1. <i>Vendas em Volume</i>	46
4.3.2. <i>Preços</i>	47
4.3.3. <i>Dados de Encartes</i>	48
4.3.4. <i>Pacotes Especiais</i>	48
4.3.5. <i>Rupturas de Gôndola</i>	48
4.3.6. <i>Eventos Especiais</i>	49
4.3.7. <i>Dia da Semana / Semana no Mês</i>	49
4.4. ANÁLISE PRELIMINAR DOS DADOS	50
4.4.1. <i>Vendas Totais da Categoria</i>	50
4.4.2. <i>Vendas do Produto 1</i>	54
4.4.3. <i>Vendas do Produto 2</i>	58

4.4.4. <i>Vendas do Produto 3</i>	60
4.4.5. <i>Vendas do Produto 4</i>	61
4.5. SELEÇÃO DO MODELO DE PREVISÃO	63
4.5.1. <i>Horizonte de Previsão</i>	63
4.5.2. <i>Disponibilidade de Dados</i>	64
4.5.3. <i>Comportamento dos Dados</i>	64
4.5.4. <i>Complexidade</i>	65
4.5.5. <i>Custos</i>	65
4.5.6. <i>Precisão</i>	65
4.6. DESENVOLVIMENTO DO MODELO – PARÂMETROS	66
4.6.1. <i>Modelo Temporal</i>	66
4.6.1.1. Previsão de Vendas da Categoria.....	69
4.6.1.2. Previsão de Vendas do Produto 1.....	73
4.6.1.3. Previsão de Vendas do Produto 2.....	75
4.6.1.4. Previsão de Vendas do Produto 3.....	77
4.6.1.5. Previsão de Vendas do Produto 4	78
4.6.2. <i>Modelo Causal</i>	79
4.6.2.1. Seleção das Variáveis	80
4.6.2.2. Desenvolvimento.....	81
4.6.2.3. Análise de Mercado.....	84
4.7. TESTES E RESULTADOS DO SISTEMA DE PREDIÇÃO DE VENDAS.....	84
5. CONCLUSÕES	87
6. BIBLIOGRAFIA	91
7. ANEXOS.....	93

Lista de Figuras

- Figura 3.1 – Categorias de Métodos de Previsão e Exemplos de Aplicações
- Figura 3.2 – Definição de Períodos
- Figura 4.1 – Vendas Diárias da Categoria
- Figura 4.2 – Vendas Diárias da Categoria com Média Móvel (k=7)
- Figura 4.3 – Detalhe do Gráfico da Categoria com Média Móvel (k=7)
- Figura 4.4 – Correlograma das Vendas da Categoria
- Figura 4.5 – Vendas Diárias do Produto 1 com Média Móvel (k=7)
- Figura 4.6 – Detalhe do Gráfico do Produto 1 com Média Móvel (k=7)
- Figura 4.7 – Correlograma das Vendas do Produto 1
- Figura 4.8 – Vendas Diárias do Produto 2 com Média Móvel (k=7)
- Figura 4.9 – Correlograma das Vendas do Produto 2
- Figura 4.10 – Vendas Diárias do Produto 3 com Média Móvel (k=7)
- Figura 4.11 – Correlograma das Vendas do Produto 3
- Figura 4.12 – Vendas Diárias do Produto 4 com Média Móvel (k=7)
- Figura 4.13 – Correlograma das Vendas do Produto 4
- Figura 4.14 – Definição de Períodos Aplicada aos Dados
- Figura 4.15 – Saída do *E-Views* para a Projeção de Vendas da Categoria
- Figura 4.16 – Previsão de Vendas da Categoria
- Figura 4.17 – Evolução do MAPE para as Previsões da Categoria
- Figura 4.18 – Saída do *E-Views* para a Projeção de Vendas do Produto 1
- Figura 4.19 – Previsão de Vendas do Produto 1
- Figura 4.20 – Evolução do MAPE para as Previsões do Produto 1
- Figura 4.21 – Saída do *E-Views* para a Projeção de Vendas do Produto 2
- Figura 4.22 – Previsão de Vendas do Produto 2
- Figura 4.23 – Evolução do MAPE para as Previsões do Produto 2
- Figura 4.24 – Previsão de Vendas do Produto 3
- Figura 4.25 – Previsão de Vendas do Produto 4
- Figura 4.26 – Sistema de Predição de Vendas

Lista de Tabelas

- Tabela 3.1 – Exemplo de Matriz de Correlação
- Tabela 4.1 – Resumo das Variáveis Existentes
- Tabela 4.2 – Participação das Vendas da Categoria por dia
- Tabela 4.3 – Participação das Vendas do Produto 1 por dia
- Tabela 4.4 – Participação das Vendas do Produto 2 por dia
- Tabela 4.5 – Participação das Vendas do Produto 3 por dia
- Tabela 4.6 – Participação das Vendas do Produto 4 por dia
- Tabela 4.7 – Erro para Diferentes Métodos
- Tabela 4.8 – Exemplo de Tabela Utilizada para Projeção das Vendas
- Tabela 4.9 – Erros de Previsão de Vendas da Categoria
- Tabela 4.10 – Erros de Previsão de Vendas do Produto 1
- Tabela 4.11 – Erros de Previsão de Vendas do Produto 2
- Tabela 4.12 – Erros de Previsão de Vendas do Produto 3
- Tabela 4.13 – Erros de Previsão de Vendas do Produto 4
- Tabela 4.14 – Coeficientes da Equação de Regressão do Produto 2
- Tabela 4.15 – Matriz de Correlação
- Tabela 4.16 – Resultado do Sistema de Predição de Vendas

1. INTRODUÇÃO

1.1. Empresa

A Procter & Gamble (P&G) é uma empresa multinacional americana de bens de consumo que possui filiais em mais de 60 países. Vende suas mais de 300 marcas em 140 países. Iniciou suas operações no Brasil em 1988, com a aquisição da empresa Perfumarias Phebo AS. Atualmente possui três fábricas no território nacional: uma na Bahia, onde são fabricadas as matérias-primas para o sabão em pó; uma em Anchieta (SP), onde é fabricado o sabão em pó; e uma em Louveira (SP), onde são fabricados os outros produtos.

1.2. Produtos

A Procter&Gamble, ao desenvolver as diversas estratégias de marketing de seus produtos, não associa em suas propagandas o nome da empresa com as marcas dos produtos, por isso seu nome não é muito familiar aos consumidores. A empresa comercializa no Brasil produtos das seguintes categorias:

- a) Cuidados com o Lar – sabões em pó.
- b) Cuidados com o Bebê – fraldas descartáveis.
- c) Proteção Feminina – absorventes.
- d) Cuidados com a Beleza – shampoos e condicionadores.
- e) Alimentos – salgadinhos*.
- f) Cuidados com a Saúde – remédios*.

* possui alguns produtos fabricados no exterior

1.3. Princípios e Valores Fundamentais

A P&G prega a todos os seus funcionários e recém-contratados alguns valores e princípios que considera fundamentais:

- Valores:
 - Liderança

- Propriedade
 - Integridade
 - Paixão por vencer
 - Confiança
- Princípios:
- Demonstramos respeito por todos os indivíduos
 - Os interesses da companhia e do indivíduo são inseparáveis
 - Temos foco estratégico em nosso trabalho
 - Inovação é a base do nosso sucesso
 - Temos foco externo
 - Valorizamos a competência pessoal
 - Procuramos ser os melhores
 - A interdependência mútua é uma forma de vida

A P&G sustenta, como declaração de propósito:

“Forneceremos produtos de qualidade e valor superiores que melhorem a vida dos consumidores em todo o mundo”.

Como resultado, os consumidores nos retribuirão com a liderança em vendas e crescimento em lucro, permitindo o progresso de nossa gente, de nossos acionistas e das comunidades em que vivemos e trabalhamos.”

1.4. Área de Vendas

A área de vendas da P&G, onde foi realizado o trabalho, é constituída de diversas equipes. Para alguns clientes maiores, há uma equipe exclusivamente dedicada a ele. Em cada equipe, além dos representantes de vendas e do líder, há pelo menos um profissional especialista em logística, um em marketing, um em sistemas e um em finanças, que são os chamados multifuncionais.

1.4.1. As Equipes Multifuncionais

1.4.1.1. Vantagens

Essa estrutura adotada pela empresa tem a vantagem de estreitar o relacionamento com seus clientes, onde o varejista passa de “cliente” a “parceiro”. O varejista colabora com a indústria fornecendo dados e informações, e em troca a indústria ajuda a aumentar as vendas do mesmo, contando com o esforço de suas equipes multifuncionais. Parte-se do princípio de que se o varejista vender mais e crescer, ocorrerá o mesmo com a indústria.

1.4.1.2. Desvantagens

Porém uma desvantagem que se torna evidente nesse tipo de estrutura é a inevitável competição interna entre as equipes. Há uma guerra constante de preços. Se a equipe de vendas que atende o hipermercado A conseguir colocar o produto X a um preço menor àquele conseguido pela equipe que atende o hipermercado B, está vencida uma batalha. E não é somente através do preço que as equipes de vendas competem entre si. Há diversos outros recursos para tentar aumentar suas vendas, tais como propagandas na televisão, anúncios em revistas e jornais, encartes, tablóides (folhetos com promoções que são distribuídos nos hipermercados), demonstradoras em lojas, pacotes promocionais, etc.

1.5. O Estágio

O estágio foi realizado em uma das equipes de vendas, junto à responsável de Marketing da equipe, que cuida de atividades que são realizadas dentro das lojas, como por exemplo algumas promoções de produtos, concursos, brindes, entre outras. Além das funções dessa área, essa pessoa ainda assumia a função de Gerente de Categoria. O estágio foi focado principalmente no suporte técnico em projetos e atividades de Gerenciamento de Categoria.

1.5.1. Gerenciamento de Categoria

O conceito de Gerenciamento de Categoria é relativamente recente, sendo definido pelo Comitê ECR (*Efficient Consumer Response* – Resposta Eficiente ao Consumidor) como:

"Uma nova forma de gerenciar com base nas necessidades do consumidor e em um relacionamento mais integrado entre o fornecedor e varejista, envolvendo muito mais que Volume x Preço".

Uma determinada categoria é vista e gerenciada como uma unidade estratégica de negócio. A indústria (fornecedor) detém o profundo conhecimento do consumidor – quanto ao hábito de uso e da categoria de produtos (posicionamento, imagem, etc). O varejo, por sua vez, conhece melhor do que ninguém o funcionamento do negócio, e assim ambos unem esforços para encontrarem oportunidades a fim de alavancar a categoria. Essa parceria passa a gerenciar não mais um produto ou seção, mas sim uma solução completa para o consumidor. O fornecedor escolhido para gerenciar a categoria geralmente é o que representa o maior volume em vendas.

2. O TRABALHO

2.1. O Problema

As equipes de vendas devem constantemente fazer previsões de quanto será vendido ao cliente varejista, para que o departamento de Planejamento de Mercado possa coordenar a produção dos produtos na fábrica. Atualmente, para se fazer essa previsão, há uma ferramenta que leva em consideração dados como:

- Vendas no ponto-de-venda nas últimas 4 semanas (vendas ao consumidor final)
- Estoques no cliente e em trânsito, e parâmetros como o estoque de segurança necessário.

O modelo que prevê as vendas no ponto-de-venda se baseia em dados de vendas das últimas quatro semanas e através de *inputs* do analista, não existindo um processo muito formal, o que prejudica um pouco sua acurácia.

Fatores que influenciam as vendas tais como: presença em encartes, pacotes especiais, promoções e preços, tanto dos próprios produtos quanto os da concorrência, não são considerados nesse modelo. Se não há uma previsão eficiente, podem ocorrer problemas como a falta de produtos na gôndola na ocasião de um anúncio em encarte.

O modelo atual pode ser uma solução simples, rápida e barata, porém depende muito do conhecimento adquirido e da experiência do gerente de conta. Existe nesse ponto um risco, devido à concentração de informações em uma só pessoa que conhece bem o processo. E há também o fator erro humano, que não pode ser descartado. E o fato de só considerar os dados de quantidade vendida nas previsões e “deixar de lado” os outros tipos de dados citados acima cria uma grande oportunidade de melhoramento.

2.2. O Objetivo

O objetivo desse trabalho é o de aperfeiçoar a ferramenta atual de previsão de vendas no ponto-de-venda (loja).

O produto final do trabalho será um “Sistema de Predição de Vendas”. Primeiro, será desenvolvido um modelo de série temporal de projeção de vendas baseando-se estritamente nos dados históricos de vendas. Paralelamente, será desenvolvido um modelo causal, onde serão considerados todos os dados disponíveis (preços, encartes, etc), e não somente os de vendas. Assim, a idéia é usar as informações do modelo causal para “alimentar” o modelo temporal e assim aumentar a precisão de suas previsões.

O estudo será baseado nas vendas de alguns produtos de uma categoria escolhida, sendo omitidos aqui por questão de confidencialidade. Pelo mesmo motivo os dados utilizados nesse trabalho não são os reais; foi aplicado um coeficiente neles.

O trabalho, sendo desenvolvido em um dos vários clientes da Procter & Gamble, poderá ser visto como um teste. Se apresentar bons resultados, poderá futuramente ser usado como exemplo para uma possível aplicação nos outros clientes da empresa.

2.3. Resumo

A primeira parte do trabalho foi dedicada a apresentar a empresa onde foram realizados o trabalho de formatura e o estágio, assim como a área de atuação e atividades desenvolvidas pelo autor. Em seguida foi definido o problema a ser resolvido e o objetivo do trabalho.

Então, para se atingir o objetivo final deste trabalho, faz-se necessária a definição de uma seqüência lógica a ser seguida ao longo do mesmo. Dessa forma, o autor sugere o seguinte:

1. Revisão bibliográfica
 - a. Explicação de modelos quantitativos e qualitativos
2. Levantamento dos Dados
3. Análise Preliminar dos Dados
4. Seleção do Modelo de Previsão
5. Desenvolvimento do Modelo Selecionado
6. Testes/ Análise dos Resultados
7. Sistema de Predição de Vendas

2.3.1. Revisão Bibliográfica

A etapa da revisão bibliográfica é de extrema importância porque é nela que o autor tomará conhecimento dos diversos modelos e técnicas de previsão existentes. Serão explicados objetivamente alguns modelos quantitativos e outros qualitativos, porém estes últimos serão explicados somente a título de curiosidade e conhecimentos gerais, não estando no escopo do trabalho. Outras técnicas mais avançadas de previsão serão incluídas nessa parte.

2.3.2. Levantamento dos Dados

Nessa etapa serão levantados os dados que o autor terá disponíveis. Segundo Hanke; Reitsch (1998), o levantamento de dados válidos e confiáveis é uma das tarefas mais difíceis e que mais consome tempo em um processo de previsão. É o caso típico para se usar a expressão “Entra lixo, sai lixo” (“garbage in, garbage out”). Uma previsão não consegue ser mais precisa do que os dados nos quais ela é baseada. Mesmo o modelo de previsão mais sofisticado irá falhar se for alimentado com dados não-confiáveis.

2.3.3. Análise Preliminar dos Dados

Nesta etapa são aplicadas e analisadas algumas ferramentas estatísticas para verificar e entender o comportamento dos dados, tal como a existência ou não de tendências significativas, se há alguma sazonalidade a ser considerada ou se há evidência da presença de um ciclo de negócio. Outro motivo é o de explicar, com auxílio dos conhecedores no assunto, pontos extremos que eventualmente aparecerão. Essas análises auxiliarão na seleção do método de previsão mais adequado.

2.3.4. Seleção do Modelo de Previsão

Com base no conhecimento – obtido na etapa anterior – do comportamento dos dados que serão utilizados para fazer as previsões, a busca por um modelo de previsão apropriado poderá ser restringida e se tornará mais rápida.

2.3.5. Desenvolvimento do Modelo de Previsão

Nessa etapa será feita a parametrização dos modelos, uma etapa de extrema importância, que se mal realizada pode comprometer os resultados finais.

2.3.6. Testes/ Análises dos Resultados

Com os modelos devidamente parametrizados, eles serão testados com alguns dados reais e os resultados serão analisados.

2.3.7. Sistema de Predição de Vendas

Ao final do trabalho o autor terá desenvolvido um Sistema de Predição de Vendas, que auxiliará na tomada de decisões da empresa.

3. REVISÃO BIBLIOGRÁFICA

3.1. Introdução

3.1.1. Por que prever?

Definição de “previsão” retirada de Santoro (2003): “Previsão aplica-se a coisas afetadas por eventos externos, não sob controle, e não a coisas afetadas por eventos internos controláveis. Refere-se ao que acontecerá se não ocorrer nenhuma alteração nos processos de decisão que podem afetar o previsto”.

Segundo O’Donovan (1983), a gestão de uma organização implica em tomar decisões frente a incertezas. O objetivo das previsões é exatamente o de reduzir o risco nessa tomada de decisões através de uma previsão eficaz dos valores futuros de variáveis importantes.

Para Makridakis; Wheelwright; Hyndman (1998), freqüentemente existe um *lead-time* entre a ciência do acontecimento de um determinado evento e a sua efetiva realização. Esse *lead-time* é a principal razão para se planejar e prever. Se esse *lead-time* fosse zero ou muito pequeno, não haveria necessidade de previsão. Mas quando o *lead-time* é longo, a previsão dos dados se torna importante pois permitirá tomar as decisões e ações apropriadas em tempo hábil. Nos negócios, os *lead-times* podem variar de alguns anos (p.ex. ampliação de uma fábrica) a alguns segundos (p.ex. em roteadores de telecomunicações).

As previsões de demanda têm um papel-chave na gestão de uma organização, pois auxiliam na tomada de decisões, sendo uma ferramenta importante no planejamento. Cada vez mais as organizações tentam diminuir sua dependência do “acaso” e serem mais racionais e lógicas ao lidarem com o ambiente em que estão envolvidas. Três exemplos de utilização de previsões em uma determinada empresa:

- A curto prazo: a partir das previsões, definir o que, quando e quanto produzir, a fim de não ter ruptura de gôndola (perda de vendas por falta de produtos) nem estoque em excesso.
- A médio prazo: aquisição dos recursos, tais como matérias-primas, mão-de-obra, máquinas, etc.
- A longo prazo: necessidade de investimentos na empresa a fim de aumentar sua capacidade de produção.

3.2. *Métodos de Previsão*

Várias técnicas de previsão foram desenvolvidas ao longo dos anos. Elas podem ser divididas basicamente em duas grandes categorias: quantitativas e qualitativas. O quadro a seguir fornece alguns exemplos disso:

Quantitativas: existe informação quantitativa suficiente. Exemplos:

- Séries Temporais: prever a continuidade de uma série histórica de vendas ou do PIB
- Causais: entender como variáveis explicativas como preço e propaganda afetam as vendas de um produto

Qualitativas: pouca ou nenhuma informação quantitativa está disponível, porém há suficiente informação qualitativa. Exemplos:

- Prever a velocidade das telecomunicações no ano de 2020
- Prever como um grande aumento no preço do óleo afetaria o consumo deste produto.

Imprevisível: pouca ou nenhuma informação está disponível.

- Prever os efeitos de viagens interplanetárias
- Prever a descoberta de uma nova forma de energia que seja muito barata e que não polua.

Figura 3.1 - Categorias de Métodos de Previsão e Exemplos de Aplicações

Retirado de Makridakis; Wheelwright; Hyndman (1998).

Será feita em seguida uma explicação breve e objetiva de alguns dos métodos de previsão mais comuns, observando-se a divisão por categorias abordada anteriormente e baseando-se na árvore a seguir:

➤ Métodos Qualitativos

- Delphi
- Pesquisa sobre Intenção de Compra
- Consenso da Força de Vendas
- Consenso de Executivos

➤ Métodos Quantitativos

- Séries Temporais
 - Média Simples
 - Média Móvel
 - Suavização Exponencial Simples
 - Suavização Exponencial Simples com Taxa de Resposta Adaptativa
 - Suavização Exponencial Dupla (Método de Brown)
 - Suavização Exponencial Linear com Dois Parâmetros (Método de Holt)
 - Suavização Exponencial Linear com Três Parâmetros (Método de Winters)
 - Decomposição
 - Box-Jenkins
- Causais
 - Regressão Simples
 - Regressão Múltipla

3.2.1. Definição de Períodos

Para se avaliar a previsão de um método são utilizadas medidas de erro de previsão, que serão detalhadas a seguir. Porém, se o analista deseja medir imediatamente a previsão do método que desenvolveu, teria que esperar alguns períodos para obter os valores das observações e então compara-las com as previsões que havia feito, obtendo evidentemente uma resposta não-imediata. Para contornar esse problema, um procedimento adotado no processo de previsão de demanda é a definição de um período de inicialização e outro de testes, ao longo da série histórica, para então fazer as projeções para o futuro desconhecido. O esquema abaixo representa a idéia:

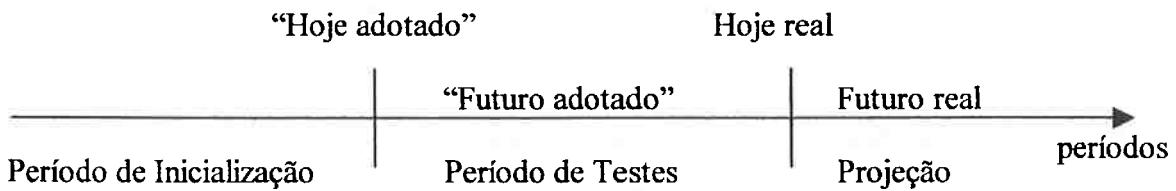


Figura 3.2 - Definição de Períodos

Esquema adaptado de Santoro (2000)

Por exemplo, se o analista se encontra no período 10 (hoje real), ele decide que o seu período de inicialização será do período 0 até 5 ($t=5$ é o “hoje adotado”), ficando então o período de $t=6$ até $t=10$ como o período de testes, sobre o qual serão calculados os erros de previsão. No período de inicialização são calculados os valores iniciais do modelo, assim como os coeficientes que forem necessários.

3.2.2. Medidas de Erro de Previsão

O erro de previsão é definido como sendo:

$$e_t = Y_t - F_t \quad (3.1)$$

Onde Y_t é uma observação e F_t é uma previsão em t . O erro percentual é definido por:

$$Pe_t = \frac{Y_t - F_t}{Y_t} \quad (3.2)$$

Para se medir a precisão de um método de previsão, são utilizadas algumas medidas de erro, dentre elas destacam-se:

Erro Relativo Médio (ME – Mean Error)

Nada mais é do que a média aritmética dos erros, que pelo fato destes poderem assumir valores positivos ou negativos, a média final poderá ter um valor pequeno. É dado por:

$$ME = \frac{1}{n} \sum_{t=1}^n e_t \quad (3.3)$$

Erro Absoluto Médio (MAE – Mean Absolute Error)

Como os erros de cada previsão são tomados em módulo, o problema da medida anterior é contornado:

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (3.4)$$

Erro Quadrático Médio (MSE – Mean Squared Error)

Aqui os erros muito grandes são “penalizados” ao elevá-los ao quadrado:

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad (3.5)$$

Erro Relativo Percentual Médio (MPE – Mean Percentage Error)

A medida de erro percentual permite uma melhor comparação entre diferentes séries de dados e com suas previsões. Porém se a série contiver valores nulos essa medida não poderá ser calculada.

$$MPE = \frac{1}{n} \sum_{t=1}^n Pe_t \quad (3.6)$$

Nesse caso a porcentagem assumirá valores positivos e negativos, o que poderá distorcer o resultado final quando feita a somatória.

Erro Absoluto Percentual Médio (MAPE – Mean Absolute Percentage Error)

$$MAPE = \frac{1}{n} \sum_{t=1}^n |Pe_t| \quad (3.7)$$

Da mesma forma que nas medidas anteriores, colocando-se os erros em módulo o problema citado anteriormente poderá ser resolvido.

Estatística Durbin-Watson (D-W)

Essa estatística testa a hipótese de haver ou não viés, positivo ou negativo, nos erros. É calculada pela seguinte fórmula:

$$D - W = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (3.8)$$

Seu valor varia entre 0 e 4. Valores de D-W próximo de 2 indicam não haver viés nos erros; valores menores que 2 indicam viés positivo e maiores que 2, viés negativo.

3.2.3. Métodos Qualitativos

Os métodos qualitativos de previsão não necessitam de dados numéricos como os métodos quantitativos. Os *inputs* necessários dependem do método específico a ser usado. Sua eficiência é mais difícil de ser medida, em relação aos quantitativos.

Segundo Makridakis; Wheelwright; Hyndman (1998) esses métodos podem ser usados separadamente, mas são mais comumente utilizados em combinação com outros métodos quantitativos. Eles são usados principalmente para fornecer dicas, para auxiliar o planejador e para suplementar previsões quantitativas, mais do que prover uma previsão numérica específica.

Geralmente são utilizados para fazer previsões de médio e longo prazo, tais como formular estratégias da organização e desenvolver novos produtos e tecnologias.

Um ponto-chave nas previsões por métodos qualitativos é exatamente a escolha do método mais adequado à situação, pois eles variam muito em precisão, extensão, horizonte de previsão e custo.

Uma observação pertinente de Hanke; Reitsch (1998) é que mesmo o método de previsão qualitativo mais puro ainda se baseará em dados históricos, pois apesar das previsões nesse caso serem baseadas no “julgamento” de um indivíduo, esses “julgamentos” na verdade serão resultados de sua manipulação mental dos dados históricos.

3.2.3.1. Método Delphi

Esse método consiste em organizar um grupo de especialistas que darão sua opinião e perspectiva para o futuro para assuntos como as atividades do negócio, tecnologias, desenvolvimento de produtos e mudanças no mercado. O grupo não precisa necessariamente ser constituído por especialistas de uma mesma área. Para assegurar que as opiniões sejam neutras e sem influências, eles não entram em contato físico. Após preencherem o questionário com suas opiniões, o coordenador do processo recolhe todas, elabora um ranking e reenvia a cada um dos participantes uma tabulação onde consta a opinião “média” e a sua específica. O participante pode então mudá-la. Isso é repetido até que nenhuma mudança significativa na sua perspectiva seja feita.

Para Jarrett (1987), o motivo dessas repetições é o de diminuir a amplitude das opiniões obtidas. Porém uma desvantagem disso é que um especialista pode ser influenciado a mudar sua opinião baseada nos erros dos outros.

Ainda segundo o autor, uma dificuldade nesse método é a de explicar o problema ou situação para os participantes. Se o problema não for bem explicado, o participante pode não entender direito o que ele deve fazer e o que esperam dele. Outras dificuldades são: ranquear as respostas dos participantes, uma vez que elas são qualitativas; e escolher um grupo de especialistas, que pode ser uma tarefa trabalhosa e cara.

A principal diferença entre o método Delphi e os outros qualitativos é a utilização de muitas opiniões independentes no processo de previsão.

3.2.3.2. Pesquisa Sobre Intenção de Compra

Esse método nada mais é do que uma pesquisa, junto aos clientes, de informações da quantidade e da data aproximada em que pretendem fazer a próxima compra. Esse tipo de pesquisa aplica-se às compras planejadas, tais como bens de consumo duráveis (devido ao seu alto valor, vida útil longa e baixa freqüência de compra) e para bens industriais, onde o produtor desses bens pode fazer uma previsão de suas vendas baseado nas intenções de compras dos seus clientes. Já para bens de consumo não-duráveis esse método não se aplica, pois além da freqüência de sua compra ser alta, o consumidor dificilmente faz um planejamento para tal.

A pesquisa consiste em entrevistar periodicamente uma amostra de consumidores em potencial e assim obter a intenção de compra do produto para um determinado período de tempo. A porcentagem de pessoas da amostra que pretende comprar o produto é o indicador de intenção de compra. Suas variações sobre períodos anteriores fornecem uma perspectiva das vendas futuras.

Segundo Gradia (1991), a incerteza implícita na intenção de compra do consumidor final será maior do que aquela da empresa, pois o primeiro é mais vulnerável às flutuações da conjuntura econômica.

3.2.3.3. Consenso da Força de Vendas

Esse método parte do princípio de que o departamento de Vendas de uma empresa é aquele que possui o melhor conhecimento e contato com o mercado onde opera, e assim pode ter uma perspectiva e sensibilidade mais apurada para perceber tendências e alterações no mercado. As previsões são feitas baseadas em questionários preenchidos pelas equipes de vendas, onde devem constar todas as políticas a serem adotadas pela empresa.

A maior vantagem desse método de previsão reside no fato de que os indivíduos que devem cumprir os objetivos participem do estabelecimento dos mesmos. Uma desvantagem é que, sendo a equipe de vendas incentivada a alcançar a meta, ela poderá subestimar propositalmente esse número para que ao final do período alcance-a facilmente. Além disso,

os vendedores podem não estar conscientes das tendências da economia nacional e do mercado.

3.2.3.4. Consenso de Executivos

Trata-se de um método de previsão muito comum nas empresas, onde um consenso entre a alta administração é obtido geralmente após muitas reuniões. Porém essas previsões não são feitas exclusivamente a partir da sensibilidade e do julgamento dos executivos, mas sim com o auxílio de grupos como o de Pesquisa de Mercado e de Análises Econômicas, que fornecem relatórios com perspectivas sobre a situação da empresa no que se refere à concorrência, seus produtos, conjuntura econômica, etc. Baseando-se nisso cada executivo elabora sua previsão e através das reuniões procura-se chegar a um consenso entre eles.

Uma desvantagem clara nesse processo são as pressões pessoais e hierárquicas que poderão surgir nas reuniões de consenso entre a alta administração, o que fatalmente poderá influenciar as opiniões individuais e assim distorcer o resultado final.

3.2.4. Métodos Quantitativos

Diferentemente dos métodos qualitativos, os quantitativos devem ser baseados em dados históricos da empresa/produto. Ajusta-se um modelo matemático formal aos dados e as previsões são obtidas através da projeção desse modelo no futuro.

Segundo Makridakis; Wheelwright; Hyndman (1998), existem três condições que devem ser respeitadas para se fazer uma previsão quantitativa:

1. Informações sobre o passado estão disponíveis;
2. Estas informações podem ser quantificadas;
3. Pode-se assumir que alguns aspectos do comportamento passado vão se repetir no futuro (Hipótese da Continuidade - *Assumption of Continuity*)

Os métodos quantitativos podem ser classificados em duas categorias:

MÉTODOS DE SÉRIES TEMPORAIS

Uma série temporal consiste em dados que são coletados ao longo do tempo em intervalos iguais.

Nesses métodos, a única variável explicativa é o tempo, por isso são chamados de univariados. Após ajustar um modelo matemático aos dados históricos, é feita uma extração para se obter as previsões desejadas. Pelo fato das previsões serem obtidas estritamente a partir dos dados históricos de vendas, temos que a premissa básica desse método é a de que o comportamento dos dados no passado se repetirão no futuro.

Segundo Hanke; Reitsch (1998), essas são técnicas estatísticas, que se focam estritamente no comportamento dos dados, suas mudanças e distúrbios.

Para Makridakis; Wheelwright; Hyndman (1998), existem duas razões principais para se tratar um sistema como uma “caixa preta” e não se tentar descobrir os fatores que afetam o comportamento dos dados, a saber:

1. O sistema pode não ser entendido, e mesmo se for, pode ser extremamente difícil de medir as relações assumidas e que afetam o seu comportamento.
2. O principal interesse pode ser prever o que vai acontecer e não o porquê disso.

MÉTODOS CAUSAIS

Esses métodos partem do princípio de que a variável a ser prevista (chamada dependente - por exemplo, as vendas de um produto) possui uma relação causal com uma ou mais variáveis independentes (por exemplo: preço, propaganda, dados da concorrência) e que as relações medidas no passado entre essas variáveis se manterão no futuro. Assim, qualquer mudança na entrada do sistema afetará sua saída de um modo previsível.

Segundo Hanke; Reitsch (1998), essas são técnicas determinísticas, que envolvem a identificação e determinação das relações entre a variável a ser prevista e as variáveis que a influenciam.

3.2.4.1. Métodos de Séries Temporais

Dados históricos podem ser suavizados de várias maneiras. O objetivo é utilizar os dados passados para desenvolver um modelo de previsão para os períodos futuros. A premissa embutida nessas técnicas é de que as flutuações nos dados passados representam variações aleatórias ao longo de uma determinada curva suavizada.

Em grande parte dos métodos utilizados aqui, parte-se do princípio de que os dados podem ser decompostos em tendência, ciclo, sazonalidade e irregularidade.

TENDÊNCIA:

É o componente de longo-prazo que representa o crescimento ou declínio da série temporal ao longo de um extenso período de tempo.

CICLO:

Consiste em uma flutuação ao redor da tendência, geralmente causada pelas condições econômicas presentes.

SAZONALIDADE:

Trata-se de uma mudança no comportamento dos dados ao longo do tempo que se repete em intervalos regulares. Por exemplo, sazonalidade anual ou mensal. Geralmente é causado pelas variações climáticas, feriados ou pela quantidade de dias em cada mês.

IRREGULARIDADE:

Mede a variabilidade da série temporal após os outros componentes terem sido removidos.

3.2.4.1.1. Método da Média Simples

Nesse método a previsão para um período é obtida através da média de todas as observações passadas:

$$F_{t+1} = \frac{1}{t} \sum_{i=1}^t Y_i \quad (3.9)$$

onde:

t= período atual, última observação obtida.

Assim que uma nova observação é obtida, a previsão para t+2 é a nova média entre os dados antigos mais essa última observação. À medida que o tempo passa e os dados vão sendo coletados, esse modelo se torna cada vez mais estável. Essa técnica é apropriada quando o comportamento dos dados não apresenta uma tendência aparente nem sazonalidade, ou seja, é um processo constante.

3.2.4.1.2. Método da Média Móvel

A técnica da média móvel consiste em utilizar os dados mais recentes para calcular a previsão para o próximo período. O número de dados utilizados nesse cálculo é sempre constante. A cada nova observação obtida, uma nova média (a previsão) é calculada descartando-se o dado mais antigo e incluindo esse mais recente. A previsão para o período t+1 é dada pela média móvel de ordem k:

$$F_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^t Y_i \quad (3.10)$$

Os pesos atribuídos a cada observação são iguais. Segundo Hanke; Reitsch (1998), a taxa de resposta a mudanças no comportamento dos dados depende do número de períodos k incluídos na média móvel. Quanto maior for o k, mais lenta essa resposta será. Ainda segundo esses autores, esse modelo funciona melhor com dados que apresentam uma certa permanência, mas não suporta muito bem tendências nem sazonalidades, apesar de fazê-lo melhor do que a média simples. Diferentemente desta última, esse método utiliza somente os últimos k dados para computar a média. E o número de pontos em cada média não muda ao longo do tempo.

Fica a cargo do analista definir o número de períodos k que será incluído na média móvel. No caso em que $k=1$, a previsão para o período seguinte seria o próprio valor atual da variável em questão. Para dados trimestrais, uma média móvel de $k=4$ representaria uma média de quatro trimestres. Para dados mensais, uma média móvel de $k=12$ eliminaria ou suavizaria os efeitos de uma sazonalidade.

Para Makridakis; Wheelwright; Hyndman (1998), um problema desse método pode ser o armazenamento dos k dados que compõem a média móvel. Caso o valor de k seja grande e o número de séries temporais a serem previstas também (como é o caso de inventários que possuem milhares de itens), esse método pode não ser o mais indicado. Na prática, a média móvel não é largamente empregada porque os métodos de suavização exponencial são geralmente superiores.

3.2.4.1.3. Método da Suavização Exponencial Simples

Ao contrário do que ocorre na média móvel, esse método não atribui pesos iguais às diferentes observações para o cálculo da previsão. Aqui é dada uma maior importância aos últimos dados. A observação mais recente recebe um peso α , a segunda mais recente $\alpha(1-\alpha)$, em seguida $\alpha(1-\alpha)^2$, e assim por diante (onde $0 < \alpha < 1$). Se esses pesos forem plotados em um gráfico, para qualquer valor de α , ficará claro que eles decrescem exponencialmente, daí a origem do nome suavização exponencial. Assim, o método pode ser representado pela seguinte equação:

$$F_{t+1} = \alpha Y_t + (1-\alpha) F_t \quad (3.11)$$

ou seja,

$$F_{t+1} = \alpha Y_{t+1} + (1-\alpha) F_{t+1}$$

$$F_{t+2} = \alpha Y_{t+2} + (1-\alpha) Y_{t+1} + (1-\alpha)^2 F_t$$

$$\text{NovaPrevisão} = [\alpha \times (\text{NovaObservação})] + [(1-\alpha) \times (\text{PrevisãoAntiga})]$$

$$F_{t+3} = \alpha Y_{t+3} + (1-\alpha) Y_{t+2} + (1-\alpha)^2 Y_{t+1} + (1-\alpha)^3 F_t$$

Uma outra equação utilizada para representar o método de suavização exponencial é dada por:

$$F_{t+1} = F_t + \alpha(Y_t - F_t) \quad (3.12)$$

Aqui, pode-se enxergar a previsão F_{t+1} como sendo simplesmente a soma da previsão antiga com um ajuste do erro ocorrido na última previsão. Quando α for próximo de 1, a nova previsão incluirá um ajuste substancial devido ao erro obtido na previsão anterior. Por outro lado, se for próximo de 0, esse ajuste será pequeno. Segundo Gradia (1991), quanto maior for o valor de α , mais rápida será a adaptação da previsão às oscilações da demanda. Porém, se for muito elevado, a previsão pode ficar sujeita às oscilações aleatórias das demandas observadas, diminuindo a precisão da estimativa.

Segundo Jarrett (1987), “esse método produz previsões autocorretivas com ajustes que regulam os valores das previsões através da mudança deles na direção oposta a dos erros antigos”. Para Makridakis; Wheelwright; Hyndman (1998), essa equação possui um princípio básico de *feedback* negativo. Esse é o mesmo mecanismo que direciona um piloto-automático para sua rota correta uma vez que tenha ocorrido um desvio na mesma.

3.2.4.1.4. Método da Suavização Exponencial Simples com Taxa de Resposta Adaptativa

Um problema da suavização exponencial simples pode ser a definição da constante α . Porém no método com taxa de resposta adaptativa a especificação de α não se faz necessária.

Segundo Jarrett (1987), este método torna-se atrativo quando a previsão a ser feita envolve uma enorme quantidade de itens. Pelo termo “adaptativa”, entende-se que este método pode alterar o valor de α a cada nova observação. Ainda segundo o autor, o método muda o valor de α quando há uma alteração no comportamento dos dados necessitando dessa forma uma constante de suavização diferente.

A equação deste método é muito similar àquela utilizada na suavização exponencial simples, exceto que nesse caso a constante α é substituída por um α_t variável no tempo:

$$F_{t+1} = \alpha_t Y_t + (1 - \alpha_t) F_t \quad (3.13)$$

onde

$$\alpha_{t+1} = \left| \frac{A_t}{M_t} \right| \quad (3.14)$$

$$A_t = \beta E_t + (1 - \beta) A_{t-1} \quad (3.15)$$

$$M_t = \beta |E_t| + (1 - \beta) M_{t-1} \quad (3.16)$$

$$E_t = Y_t - F_t \quad (3.17)$$

$$0 < \beta < 1$$

A_t = erro suavizado

M_t = erro suavizado absoluto

E_t = erro de previsão

Esse método é completamente automático, a variável α_t vai sendo alterada à medida que novos dados surgem. Porém deve-se ficar atento a essas flutuações. Para Makridakis; Wheelwright; Hyndman (1998), uma maneira de controlar as mudanças em α_t é através do valor de β : quanto menor for este valor, menores serão as mudanças em α_t . Outra maneira de se fazer esse controle é impondo um limite superior de quanto α_t é permitido mudar de um período para o seguinte.

3.2.4.1.5. Método da Suavização Exponencial Dupla (Método de Brown)

Esse método é utilizado para se fazer previsões de séries temporais que apresentam uma tendência linear. Essa técnica é simplesmente uma suavização exponencial de valores exponenciais simples, de onde vem o termo “exponencial dupla”. A seqüência lógica desse método é explicada por Ferrari (1996): “já que os valores obtidos pela suavização exponencial simples e dupla diferem dos dados reais quando existe uma tendência, a diferença entre os valores obtidos pela suavização exponencial simples e dupla poderá ser adicionada ao valor

obtido pela suavização exponencial simples e ajustada para uma tendência”. As equações para esse método são dadas por:

$$S_{t+1} = \alpha Y_{t+1} + (1 - \alpha) S_t \quad (3.18)$$

$$S'_{t+1} = \alpha S_{t+1} + (1 - \alpha) S'_t \quad (3.19)$$

onde

S_{t+1} é o valor obtido pela suavização exponencial simples

S'_{t+1} é o valor obtido pela suavização exponencial dupla

α é a constante de suavização

As diferenças entre os valores das suavizações exponenciais simples e dupla fornecem os ajustes, que são dados pelas seguintes equações:

$$a_t = 2S_t - S'_t \quad (3.20)$$

$$b_t = \frac{\alpha}{1 - \alpha} (S_t - S'_t) \quad (3.21)$$

então:

$$F_{t+m} = a_t + b_t m \quad (3.22)$$

onde:

m é o número de períodos a frente a ser previsto.

A previsão por este método é dada por F_{t+m} . O fator de ajuste a_t representa o ponto inicial da previsão, e o fator b_t representa a tendência existente, que é similar a uma medida de inclinação que pode mudar ao longo da série.

3.2.4.1.6. Método da Suavização Exponencial Linear com Dois Parâmetros (Método de Holt)

Esse método proposto por Holt em 1957 é similar ao de Brown. É utilizado para prever dados que apresentam uma tendência. Nesse caso, porém, são utilizadas duas constantes de suavização, α e β , a primeira correspondendo ao índice de estabilidade e a segunda ao da

tendência. Isso possibilita ao modelo uma maior flexibilidade no rastreamento do índice de tendência. Makridakis; Wheelwright; Hyndman (1998) consideram o método de Brown como sendo um caso particular do método de Holt quando $\alpha=\beta$.

As três equações são:

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (3.23)$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \quad (3.24)$$

$$F_{t+m} = L_t + b_t m \quad (3.25)$$

Aqui, L_t representa uma estimativa do nível (altura) da série temporal no período t e b_t representa uma estimativa da inclinação (tendência) da série em t . Na primeira equação, L_t é ajustado somando-se o último valor suavizado L_{t-1} com a tendência dos períodos passados (b_{t-1}). A tendência é atualizada na segunda equação, que vem da diferença entre os dois últimos valores suavizados. Uma possível aleatoriedade que possa existir nesse ponto é eliminada através da constante de suavização β . A previsão é dada por F_{t+m} .

3.2.4.1.7. Método da Suavização Exponencial Linear com Três Parâmetros (Método de Winters)

Quando uma série temporal apresenta uma sazonalidade nos dados o método de Holt não é apropriado para ela. Sendo uma extensão deste último, o método de Winters propõe que uma equação seja adicionada para tratar esse comportamento sazonal dos dados, onde consta um fator de sazonalidade:

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (3.26)$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \quad (3.27)$$

$$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s} \quad (3.28)$$

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \quad (3.29)$$

Na equação que define L_t (estimativa do nível), o efeito da sazonalidade da observação Y_t é removido através da subtração deste pelo fator de sazonalidade S_{t-s} , onde s corresponde ao intervalo de repetição desta (seja em meses, semanas, dias, etc). A estimativa da tendência linear é dada por b_t . F_{t+m} calcula a previsão desejada m períodos a frente. Os coeficientes de suavização α , β e γ devem ter valor entre 0 e 1.

3.2.4.1.8. Métodos de Decomposição

Neste método parte-se do princípio de que a série temporal é composta da seguinte forma:

Dados = padrões + erro

Dados = $f(\text{tendência, sazonalidade, ciclicidade, erro})$

Assim, esses componentes são separados, analisados e então projetados individualmente. A previsão é obtida pela recomposição deles. Makridakis; Wheelwright; Hyndman (1998) defendem a idéia de que os métodos de decomposição devem ser utilizados mais como uma ferramenta para se entender a série temporal do que para se fazer previsões. Isso é justificado pela dificuldade que pode haver para se projetar os componentes individuais da série. Os autores preferem utilizar esses métodos como um passo preliminar para então selecionar e aplicar um método de previsão.

3.2.4.1.9. Método de Box-Jenkins

Uma das mais completas e sofisticadas técnicas para análise e previsão de séries temporais é a metodologia desenvolvida por Box & Jenkins. Nessa metodologia são seguidos os seguintes procedimentos:

1. Identificação, dentro de um conjunto de modelos, qual é o mais adequado (ordem e tipo do modelo)
2. Estimativa dos parâmetros do modelo escolhido de modo a obter o ajuste de mínimo erro quadrático para os dados históricos.
3. Projeção da série de modo a minimizar o erro de previsão a cada ponto.

O método de Box-Jenkins possui uma introdução teórica extensa pela sua significativa complexidade, por isso será omitida nesse trabalho, sendo abordado somente a título de curiosidade.

3.2.4.2. Métodos Causais

Nos métodos causais ou explicativos o escopo passa ser outro, como o próprio nome já diz: estudar as causas da demanda. Uma previsão é expressa como uma função de um certo número de fatores que influenciam os valores que ela assume. Essas previsões não serão necessariamente dependentes do tempo. Um modelo explicativo que relate variáveis de entrada com uma variável de saída proporciona um melhor entendimento da situação e permite fazer testes com diferentes combinações de “entradas” a fim de quantificar e estudar seus efeitos na “saída” (a previsão).

Parte-se do princípio de que a relação entre a variável dependente (de “saída”) e a ou as variáveis independentes (de “entrada”) se manterão ao longo do tempo. A ferramenta básica para análise de modelos causais é a análise de regressão.

3.2.4.2.1. Regressão Simples

A regressão linear simples expressa a relação entre a variável dependente y e somente uma variável independente x , na forma de uma reta. Matematicamente, isso é representado pela seguinte equação:

$$Y = a + bX + e \quad (3.30)$$

onde e representa o erro randômico. Esse erro mede o desvio, na vertical, de uma observação à reta:

$$e_i = Y_i - \hat{Y}_i \quad (3.31)$$

Hanke; Reitsch (1998) definem a linha de regressão (a reta da equação) como sendo “a linha que melhor se adequa a uma coleção de pontos X,Y minimizando a soma das distâncias ao quadrado dos pontos a essa linha, na direção de Y (vertical)”. O método utilizado para se estimar os valores dos coeficientes a e b é chamado de Método dos Mínimos Quadrados (MMQ). Assim, para se encontrar esses valores deve-se minimizar a soma dos erros quadráticos das observações:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - a - bX_i)^2 \quad (3.32)$$

onde \hat{Y} representam os valores estimados de Y_i quando conhecidos somente os X_i .

Utilizando a derivada parcial e com alguns cálculos chega-se a:

$$b = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (3.33)$$

$$a = \bar{Y} - b\bar{X} \quad (3.34)$$

onde \bar{Y} e \bar{X} são as médias das observações Y e X respectivamente.

Atualmente, os cálculos para se encontrar os valores de a e b são raramente feitos a mão, sendo na grande maioria das vezes feitos por meio de calculadoras ou programas de computador.

É importante frisar aqui que a linha de regressão encontrada por esses cálculos (ou seja, os coeficientes a e b), faz parte de uma “população de linhas de regressão”, uma vez que foram obtidas a partir de uma amostra aleatória de pontos X,Y e não de todos os pontos X,Y da população. Uma outra amostra aleatória de pontos X,Y diferentes da primeira produziria uma linha de regressão diferente. Esse raciocínio é análogo ao de se obter diferentes médias para diferentes amostras aleatórias de pontos retirados de uma única população.

Na regressão linear, é interessante medir a “distância média” dos pontos Y_i em relação às estimativas \hat{Y}_i da linha de regressão, na direção de Y. Esse conceito de medir a dispersão é

similar à noção de desvio padrão usada para se medir a dispersão de dados quaisquer em volta de sua média. Na análise de regressão essa estatística é denominada de erro-padrão de estimativa, representada pela seguinte equação:

$$s_e = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}} \quad (3.35)$$

Porém, como explicado acima, os coeficientes a e b são randômicos. Dessa forma, tanto a dispersão dos pontos Y_i da amostra ao redor da linha de regressão (desvio-padrão dos erros, equação anterior) quanto a dispersão das muitas linhas de regressão ao redor da “verdadeira” linha de regressão da população deve ser considerada. Isso é denominado de erro-padrão de previsão, e ele mede a variabilidade do valor de \hat{Y}_0 previsto em relação ao verdadeiro valor de Y_0 para um dado X_0 . É obtido pela seguinte equação:

$$s.e.(\hat{Y}_0) = s_e \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}} \quad (3.36)$$

Uma outra estatística importante e que é utilizada para se medir o quanto duas variáveis estão relacionadas entre si é o coeficiente de correlação r . Ele é obtido pela fórmula:

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{n \sum X^2 - (\sum X)^2} * \sqrt{n \sum Y^2 - (\sum Y)^2}} \quad (3.37)$$

O valor de r pode variar entre 0 (o que indica nenhuma correlação) e ± 1 (o que indica uma perfeita correlação). Se r tem um valor positivo, as duas variáveis são positivamente correlacionadas, ou seja, movem-se na mesma direção (por exemplo, quando uma cresce a outra também cresce). Inversamente, quando r possui um valor menor que zero, diz-se que as variáveis são negativamente correlacionadas, ou seja, movem-se em direções opostas (por exemplo, quando uma cresce a outra decresce). Makridakis; Wheelwright; Hyndman (1998) fazem algumas observações importantes: o coeficiente de correlação é uma medida da associação linear entre duas variáveis. Se duas variáveis quaisquer forem relacionadas de uma

maneira não-linear, o coeficiente r não “fará justiça” à força dessa correlação. Outro ponto é que r é instável para amostras com poucos dados. E ainda, esse coeficiente pode ser seriamente influenciado por um valor que esteja muito distante dos outros pontos. Um valor muito extremo já é suficiente para prejudicar a medida do coeficiente.

Elevando-se r ao quadrado, obtém-se o coeficiente de determinação R^2 . Esse coeficiente pode ser interpretado como sendo a variação explicada de Y sobre a variação total de Y , ou ainda como preferem Hanke; Reitsch (1998) R^2 “mede a porcentagem da variabilidade de Y que pode ser explicada através do conhecimento da variabilidade da variável independente X ”. A equação de R^2 é:

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad (3.38)$$

Na equação da linha de regressão, um teste importante a ser feito é o de se verificar se os coeficientes calculados a e b são significativamente diferentes de zero. Isso é calculado através do Teste-t, aplicado a cada coeficiente:

$$\begin{aligned} t_a &= \frac{a}{s.e.(a)} \\ t_b &= \frac{b}{s.e.(b)} \end{aligned} \quad (3.39)$$

onde:

$$\begin{aligned} s.e.(a) &= \sigma_\varepsilon \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2}} \\ s.e.(b) &= \sigma_\varepsilon \sqrt{\frac{1}{\sum (X_i - \bar{X})^2}} \end{aligned} \quad (3.40)$$

que são os erros-padrão dos coeficientes a e b , que por sua vez apresentam uma distribuição normal com médias α e β respectivamente. Nas duas equações acima, o desvio -padrão dos erros σ_ε é desconhecido, porém podem ser estimados por s_e , definido anteriormente.

Em um Teste-t, por exemplo, quando a inclinação da linha de regressão b for significativamente diferente de zero, o valor de t_b será alto (em módulo). Este teste pode ser visto também como uma medida da estabilidade do coeficiente, uma vez em que no seu cálculo esse coeficiente é dividido pelo seu erro-padrão. Assim, um coeficiente que tenha um erro-padrão muito alto proporcionalmente ao seu valor terá como consequência um valor de t baixo.

É comum ser calculada, juntamente com o valor de t , uma probabilidade P associada a cada estatística t . Segundo Makridakis; Wheelwright; Hyndman (1998), cada valor de P representa a probabilidade de se obter um valor de $|t|$ tão alto quanto aquele encontrado para os “seus” dados se na verdade o coeficiente em questão (a ou b) fosse igual a zero. Logo, se P tem um valor baixo, pode-se concluir que o coeficiente calculado é significativamente diferente de zero.

Uma vez encontrada a equação da linha de regressão a partir dos pontos da amostra coletada, é natural se questionar se existe realmente uma relação entre Y e X . O teste-F permite testar a significância do modelo de regressão como um todo. Seu cálculo é feito da seguinte forma:

$$F = \frac{\sum (\hat{Y}_i - \bar{Y})^2 / (m-1)}{\sum (Y_i - \hat{Y}_i)^2 / (n-m)} \quad (3.41)$$

onde:

m = número de coeficientes na equação de regressão

Essa equação nada mais é do que a variância explicada pelo modelo dividida pela variância que não é explicada pelo modelo. Logo, quanto maior for o valor de F , mais significativa será a relação entre X e Y . A estatística F também está relacionada ao coeficiente de determinação R^2 , através da seguinte equação:

$$F = \frac{R^2 / (m-1)}{(1-R^2) / (n-m)} \quad (3.42)$$

3.2.4.2.2. Regressão Múltipla

A regressão simples estudada anteriormente pode ser entendida como um caso especial da regressão múltipla. Ao contrário da primeira, nesta última existe uma variável dependente e duas ou mais variáveis independentes (explicativas). A equação é da forma:

$$Y_i = b_0 + b_1 X_{1,i} + b_2 X_{2,i} + \dots + b_k X_{k,i} + e_i \quad (3.43)$$

onde $Y_i, X_{1,i}, \dots, X_{k,i}$ representam a i -ésima observação de cada uma das variáveis Y, X_1, \dots, X_k respectivamente; b_0, b_1, \dots, b_k representam os coeficientes das variáveis e e_i é o erro randômico da i -ésima observação, calculado de forma semelhante àquela da regressão simples. É importante frisar que se parte do pressuposto de que o modelo que melhor se adequa aos dados é linear, ou seja, trata-se de uma regressão linear múltipla.

De maneira similar à regressão simples, os coeficientes da equação são calculados minimizando-se a soma dos erros quadráticos (MMQ). Porém, devido ao maior número de variáveis e coeficientes, é quase imprescindível a utilização de um software para tais cálculos.

O teste-t para cada um dos coeficientes encontrados é feito de maneira similar àquela da regressão simples – dividindo-se o próprio coeficiente pelo seu erro-padrão. O coeficiente de correlação múltipla R^2 também é calculado da mesma forma. Seu valor é interpretado como sendo a porcentagem da variação que pode ser explicada pelas variáveis independentes. O cálculo do teste-F, por sua vez, leva em consideração o número k de variáveis explicativas, onde n é o número de observações:

$$F = \frac{\sum (\hat{Y}_i - \bar{Y})^2 / k}{\sum (Y_i - \hat{Y}_i)^2 / (n - k - 1)} \quad (3.44)$$

Uma etapa importante na regressão múltipla é a matriz de correlação. Nela, são representados os coeficientes de correlação r entre cada variável existente, tanto Y quanto X_i . Analisando essa matriz pode-se ter uma noção da existência de relação ou não entre a variável Y e uma determinada variável independente X . Ela tem o seguinte formato:

VARIÁVEIS	Y	X1	X2
Y	1,00	-0,74	0,87
X1		1,00	0,54
X2			1,00

Tabela 3.1 – Exemplo de Matriz de Correlação

No exemplo acima, fica evidente a correlação positiva existente entre a variável Y e X2, onde $r=0,87$. Não é necessário o preenchimento de metade da matriz, uma vez que a correlação entre (Y, X2) e (X2, Y) obviamente possuem o mesmo valor.

A matriz de correlação também é útil pois pode revelar eventuais problemas de multicolinearidade – situação na qual variáveis independentes, em uma regressão múltipla, estão altamente correlacionadas. De acordo com Hanke; Reitsch (1998), a presença de multicolinearidade em uma regressão múltipla pode causar os seguintes problemas:

- Uma determinada variável X1 pode ter um coeficiente de regressão positivo em uma regressão simples mas, quando introduzida uma outra variável independente X2, o coeficiente de regressão da primeira pode passar a ser negativo (ou vice-versa), caso X1 e X2 sejam altamente correlacionadas.
- Os coeficientes de regressão calculados podem variar demasiadamente de amostra para amostra porque os erros-padrão desses coeficientes são muito altos.
- A regressão múltipla é utilizada como uma ferramenta para analisar a importância relativa de variáveis independentes em Y. Caso houver duas variáveis X intercorrelacionadas, elas explicarão a mesma variação em Y. Por essa razão é extremamente difícil separar a influência individual de cada uma delas.

Ainda segundo Hanke; Reitsch (1998), quando existe uma colinearidade extrema, fazer previsões baseadas no modelo pode ser válido, porém os coeficientes de regressão estimados não são confiáveis. Os autores sugerem duas soluções caso o analista queira utilizar o modelo de regressão mesmo com a presença de multicolinearidade: a) remover da equação final uma das variáveis que estão altamente intercorrelacionadas ou b) criar e usar uma variável nova formada pela combinação dessas duas variáveis intercorrelacionadas.

Uma maneira de incluir dados qualitativos na análise de regressão múltipla é através das variáveis binárias (*dummy variables*). Sua criação é simples: quando ocorre o evento em questão seu valor é 1, caso contrário assume valor 0. Seu campo de aplicação é vasto, podendo ser utilizada para incluir no modelo dados como:

- sazonalidade (por exemplo, para dados de vendas diárias, cada dia da semana seria uma variável binária, devido às suas diferentes concentrações de vendas ao longo da semana)
- ocorrência de feriados (ainda no exemplo de vendas diárias)
- qualquer evento que possa influenciar as vendas de um produto (por exemplo, um anúncio na mídia)

No caso de se utilizar variáveis binárias para os dados de sazonalidade, deve-se incluir no modelo P-1 variáveis, onde P é o número de períodos em que ocorre uma sazonalidade completa. Por exemplo, caso os dados fossem mensais, as variáveis binárias seriam X_1, X_2, \dots, X_{12} . Isso deve ser feito para se evitar o problema da multicolinearidade. Aqui, para representar o mês de dezembro, todas as variáveis binárias assumiriam valor 0, englobando dessa forma todos os meses do ano no modelo.

Nessas variáveis, o coeficiente associado a cada uma delas representa a diferença média, no resultado da previsão, entre elas e a variável omitida. Por exemplo, o coeficiente associado ao mês de janeiro é uma medida do efeito desse mês na previsão comparado com o efeito nas vendas do mês de dezembro, que foi o omitido. Caso outro mês tivesse sido omitido, ao invés de dezembro, os coeficientes estimados seriam diferentes, mas os resultados seriam os mesmos.

No processo de seleção das variáveis que farão parte do modelo de regressão, um método intuitivo e direto é logo delineado:

- As variáveis vão sendo adicionadas ao modelo, e como consequência o coeficiente de determinação R^2 vai aumentando
- Em seguida são eliminadas ou modificadas as variáveis que apresentem multicolinearidade.

Em princípio esse processo pode parecer simples. Mas quando o modelo envolve muitas variáveis independentes a dificuldade aumenta. Existem alguns métodos que auxiliam nesse

processo, entre eles o *best subset regression* e o *stepwise regression*. No primeiro, todas as combinações de possíveis variáveis independentes são testadas, começando com um modelo contendo uma variável explicativa, depois duas e assim por diante. Este método é uma forma eficiente de se identificar bons modelos que utilizem o menor número possível de variáveis explicativas. Já pelo método *stepwise*, as candidatas à variável independente são adicionadas e removidas do modelo de forma a identificar a combinação de variáveis mais adequada. Isso é feito baseando-se no valor da estatística F que aquela combinação de variáveis independentes apresenta. Geralmente esses tipos de análises são realizadas através de um software específico.

4. DESENVOLVIMENTO

4.1. Modelo de Previsão Atual

Como explicado brevemente no capítulo 2, o modelo de previsão de demanda no ponto-de-venda atual se baseia em dados de vendas das últimas quatro semanas. Esses dados são plotados em um gráfico de controle que auxilia nessa previsão, não existindo assim um processo muito formal a ser seguido. Observa-se a média e as flutuações recentes e gera-se uma previsão.

O modelo atual é um pouco “míope”. Ele considera somente os dados históricos de vendas, e não considera fatores que certamente influenciam as vendas tais como: presença em encartes, pacotes especiais, promoções e preços, tanto dos próprios produtos quanto os da concorrência. A possível influência desses acontecimentos é inserida nas previsões mais qualitativamente e pela intuição do que quantitativamente por algum método de previsão específico.

4.2. Horizonte de Previsão

Devido ao dinamismo do mercado estudado no nível da loja, onde uma diminuição no preço ou um anúncio no encarte pode movimentar as participações de mercado das principais marcas, é importante fazer uma previsão de vendas de imediato ou curto-prazo, a fim de manter o nível mais baixo de estoque de produtos na loja mas que o mesmo seja suficiente para não ocorrer falta de produtos e consequente perda de vendas. Assim, o modelo a ser desenvolvido trabalhará com dados diários, que permita ao gerente de conta acompanhar de perto a performance de suas marcas na loja e alavancar suas vendas.

Assim que o pedido de um produto é feito pelo varejista, ele sai da fábrica e é levado ao centro de distribuição do cliente e desse ponto é transportado até a loja que efetuou o pedido. O *lead-time* total é de aproximadamente 3 dias. Assim foi decidido que o horizonte de previsão será de 5 dias.

4.3. Dados Disponíveis

Uma etapa de extrema importância na elaboração de qualquer modelo de previsão de vendas é o levantamento dos dados históricos. Sendo um dos primeiros passos a ser executado, um bom trabalho nesse ponto é condição necessária para se obter um resultado final eficiente. Vale ressaltar que quanto mais forem os dados disponíveis, melhores serão os resultados obtidos, pois o histórico de vendas será mais completo.

Para possibilitar um estudo mais real e próximo do que acontece em um hipermercado, optou-se por trabalhar com os dados de uma loja somente. Existem variações entre diferentes lojas para séries de dados como preço e ruptura de gôndola, o que causaria uma distorção nas análises caso fossem estudadas mais de uma loja em um mesmo modelo. E tendo desenvolvido uma metodologia de análise para uma loja, a mesma pode ser reaplicada em outra qualquer. A loja escolhida foi a que apresenta a média de vendas mais elevada para a categoria em questão.

Para realizar o trabalho, foram levantados diversos dados das últimas 104 semanas (2 anos), todos com periodicidade diária, o que totaliza aproximadamente 720 observações para cada série de dados obtida. Foram analisados no total 13 produtos, sendo 4 da P&G e 9 da concorrência. Os dados que não os de vendas são utilizados somente no modelo causal. A seguir uma breve explicação sobre eles:

4.3.1. Vendas em Volume

A partir do banco de dados, pode-se obter a quantidade de unidades vendidas de um determinado produto. Os dados chegam ao nível da embalagem, existindo uma série de dados para cada código de barras (ou seja, para cada tamanho de embalagem). Por exemplo: quantidade de unidades vendidas da embalagem de x gramas do produto y, por dia. Cada marca, normalmente, apresenta mais de um tamanho de embalagem. Eventualmente surgem embalagens menos usuais, com tamanhos diferentes.

Um “complicador” é a embalagem promocional, como por exemplo “Compre x gramas e leve grátis um produto y”. Esses itens são temporários, mas pelo banco de dados geram uma série de dados própria. Um dos seus objetivos é o de gerar maior visibilidade da marca e alavancar suas vendas.

Para calcular então as vendas de uma determinada marca como um todo, deve-se primeiro descobrir todos os itens/produtos/embalagens/promoções que existiram ao longo da série observada, obter esses dados e consolidá-los. Para os dados de vendas totais da categoria o processo é análogo. Essa consolidação nada mais é do que a multiplicação, em cada dia de venda, da quantidade de cada embalagem existente que foi vendida pela sua capacidade em volume. Dessa forma os dados de vendas estão todos em gramas vendidas por dia.

4.3.2. Preços

Os dados relativos ao preço do produto são os mais complicados de serem tratados. Como explicado anteriormente, para uma mesma marca podem existir diferentes tamanhos de embalagens, ou seja, diferentes preços. A idéia inicial seria ter todas as séries de preços por marca que existem (logo, uma série de preço por tamanho de embalagem existente). Porém, fazendo uma análise marca a marca ao longo do período estudado, verificou-se que era comum ter embalagens que existiam por espaços de tempo relativamente curtos, ou seja, somente em parte do período de 2 anos analisado que havia dados de vendas/preços. Quanto aos dados de vendas em volume não há problema nesse ponto, pois basta somar todo o volume vendido na loja no dia para obter assim um total geral. Mas para os dados de preços não faz sentido “somá-los”. Uma opção seria fazer um índice diário como \$/vol, mas isso poderia camuflar o real efeito da movimentação dos preços nas vendas, pois os diferentes tamanhos de embalagens possuem diferentes participações nas vendas. Em função disso tudo, foi decidido que a série de preços que seria levantada seria a da embalagem com participação mais significativa nas vendas da categoria, para todos os produtos.

4.3.3. Dados de Encartes

Uma atividade que certamente alavanca as vendas de qualquer produto em um hipermercado é o anúncio em encarte (folheto com fotos e preços de produtos distribuídos na entrada da loja). A partir de um cadastro de todos os encartes dos últimos dois anos esses dados foram gerados. Pelo fato de serem informações qualitativas, a série criada se trata de uma *dummy variable* (explicada no capítulo 3). Ela pode assumir, a cada dia, valor 1 (presença em encarte) ou valor 0 (não presença).

4.3.4. Pacotes Especiais

Pacotes especiais (ou promocionais) são embalagens que trazem algo diferente para incrementar a venda desse produto. Por exemplo, pacotes do tipo “leve 3 pague 2”, “leve o produto x e ganhe grátis o produto y”, etc. Isso seguramente incrementa as vendas da marca. O acréscimo gerado em vendas em volume é considerado na variável “vendas em volume”. Porém, partindo-se de outra perspectiva, um determinado produto que tenha um pacote especial à venda na loja vai ter um incremento nas suas vendas. Mas como “explicar” esse incremento no modelo de previsão olhando-se somente os dados de vendas? Para isso foi criada essa variável, que também é binária, sendo 1 quando há um pacote especial e 0 caso contrário.

4.3.5. Rupturas de Gôndola

Uma variável que não pode faltar em uma análise de previsão de demanda é aquela que mede a falta de produtos na gôndola. Como a base do modelo de previsão são os dados históricos de vendas, a falta de um determinado produto em um dia pode confundir o efeito real das variáveis independentes nas vendas desse produto. Por exemplo, se há uma diminuição do seu preço, mas por outro lado ocorre uma ruptura de gôndola, os dados de vendas daquele dia não refletirão a relação real entre o preço e o volume vendido. Essa variável também é binária: 1 para produto na loja e 0 para a falta dele.

4.3.6. Eventos Especiais

Essa variável engloba todo evento que ocorra na loja e que tenha como objetivo aumentar as vendas de um ou mais produtos. Exemplos disso são: aniversário do varejista (onde geralmente os produtos são vendidos muito mais baratos que o preço normal), distribuição de brindes com a compra de um produto, concursos relacionados com um determinado produto, etc. É uma variável qualitativa, por isso binária, e de fácil previsibilidade. Assume valor 1 quando há um evento no dia e 0 caso contrário.

4.3.7. Dia da Semana / Semana no Mês

Como será demonstrado adiante, os dados de vendas possuem duas sazonalidades, ao analisá-los no contexto de um mês: a primeira, relativa aos dias da semana – um dia de fim de semana vende mais do que um dia no meio da semana – e outra relativa à semana dentro do mês – as vendas nas primeiras semanas do mês são mais fortes do que nas últimas. Novamente, para “justificar” essas variações no modelo foram criadas essas variáveis, todas elas *dummy variables*. Por exemplo, a variável “quarta-feira” assume valor 0 para todos os dias da semana com exceção à própria quarta-feira. A variável “2ª semana no mês” assume valor 1 para os dias d do mês tal que $8 \leq d \leq 14$ e 0 para o resto.

Nem todas as variáveis existem para cada um dos 13 produtos. Em resumo, os dados existentes estão organizados da seguinte forma, lembrando que os produtos 5 ao 13 são os concorrentes:

	1	2	3	4	5	6	7	8	9	10	11	12	13
Vendas	x	x	x	x									
Preço	x	x	x	x	x	x	x	x	x	x	x	x	x
Encarte	x	x	x	x	x	x	x	x	x	x	x	x	x
Ruptura	x	x	x	x									
Pacotes Especiais		x	x		x	x	x	x	x				
Eventos	x	x	x	x									

Tabela 4.1 – Resumo das Variáveis Existentes

4.4. Análise Preliminar dos Dados

Antes de partir para o desenvolvimento de um modelo de previsão, é recomendável estudar e analisar a série temporal a ser prevista a fim de conhecer melhor o comportamento dos dados. Essa análise auxilia o analista a perceber possíveis tendências, sazonalidades ou ciclos na série. Estando familiarizado com ela, uma correta seleção do melhor método de previsão poderá ser feita.

As séries de dados de vendas (em gramas vendidas por dia) que serão pré-analisadas nessa etapa são:

- Vendas totais da categoria
- Vendas do Produto 1
- Vendas do Produto 2
- Vendas do Produto 3
- Vendas do Produto 4

Apesar da previsão das vendas da categoria não fazer parte do objetivo do trabalho, esses dados serão analisados e previstos da mesma forma a fim de proporcionar uma visão geral da situação da categoria. Nessa visão geral da categoria serão explicados todos os conceitos, dados e análises; nas quatro análises específicas das marcas não serão repetidas essas explicações, a abordagem será mais direta e objetiva com os resultados obtidos.

4.4.1. Vendas Totais da Categoria

Somando-se as vendas diárias de todas as marcas da categoria comercializadas nesse varejista, pode-se ter uma noção de como é a freqüência de compra dos consumidores na loja:

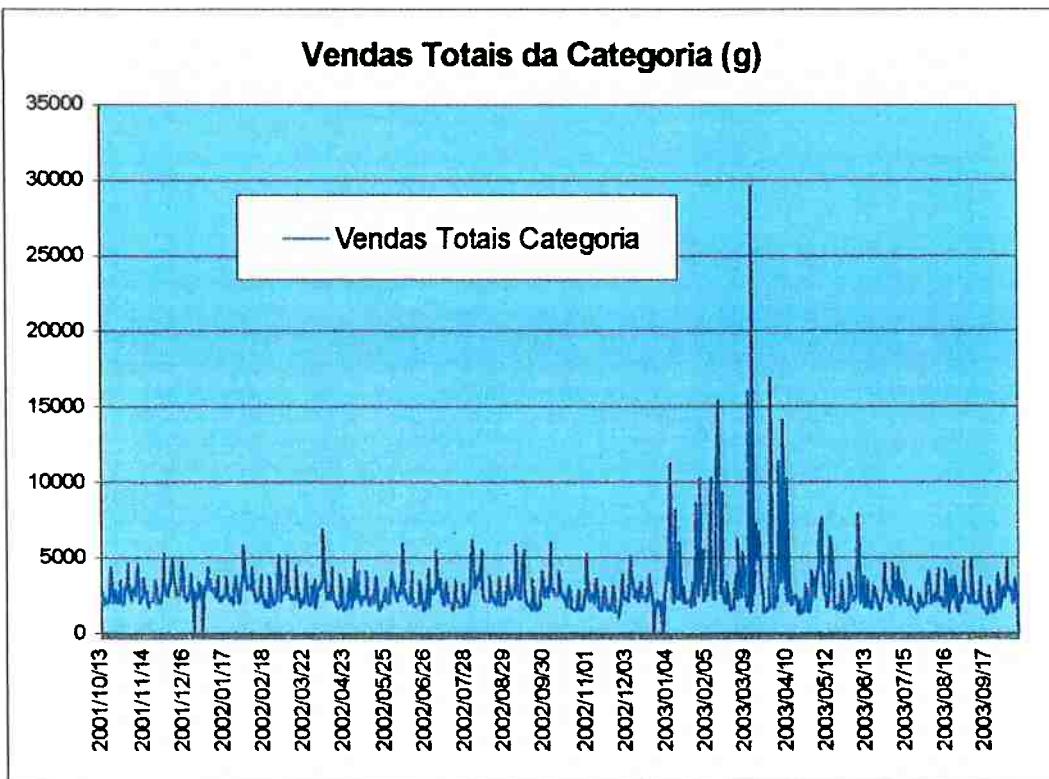


Figura 4.1 – Vendas diárias da Categoria

A princípio pode-se notar que existe uma sazonalidade nas vendas, aparentemente diária. Uma vez levantada a hipótese dessa sazonalidade entre os dias da semana, foi aplicada uma média móvel de $k=7$ nos dados para tentar suavizá-los. O resultado é mostrado a seguir:

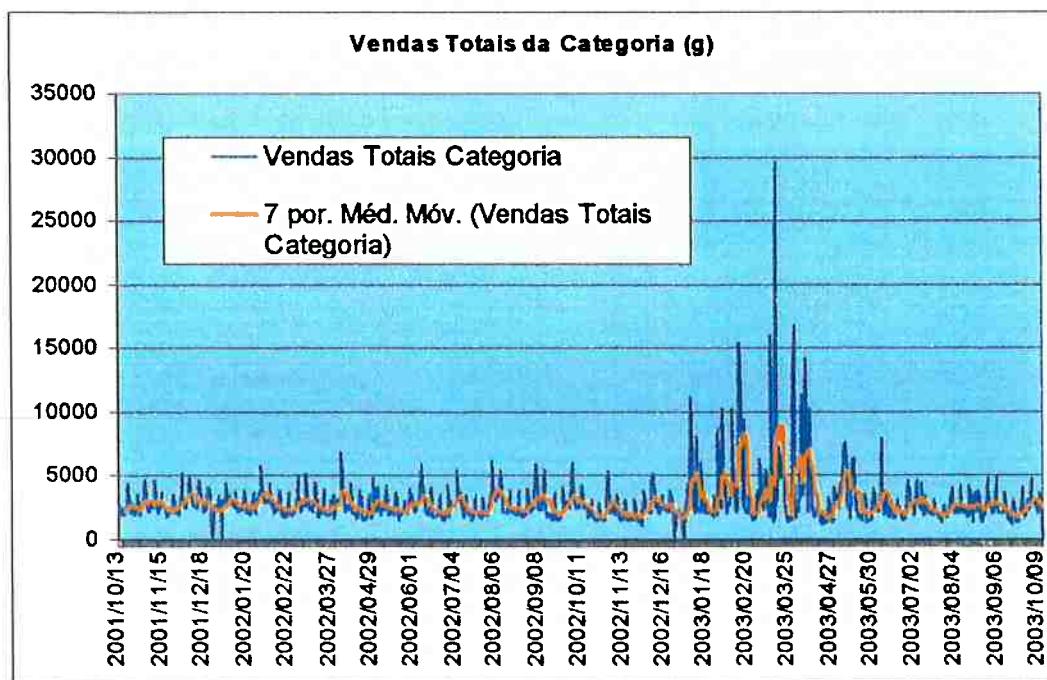


Figura 4.2 – Vendas diárias da Categoria com Média Móvel (k=7)

Os picos que o gráfico apresenta são devidos a eventos que ocorrem na loja, geralmente de fim-de-semana, que alavancam as vendas de um ou vários produtos. Um exemplo disso é a presença de uma promotora na loja distribuindo brindes para os consumidores que comprarem um determinado produto. São, portanto, eventos previsíveis qualitativamente.

Nota-se que a suavização dos dados aplicada ainda apresenta uma “ondulação”, ou seja, ainda há uma sazonalidade remanescente na série temporal. Observando esse mesmo gráfico mas com um detalhe maior, focando em 4 meses somente, temos:

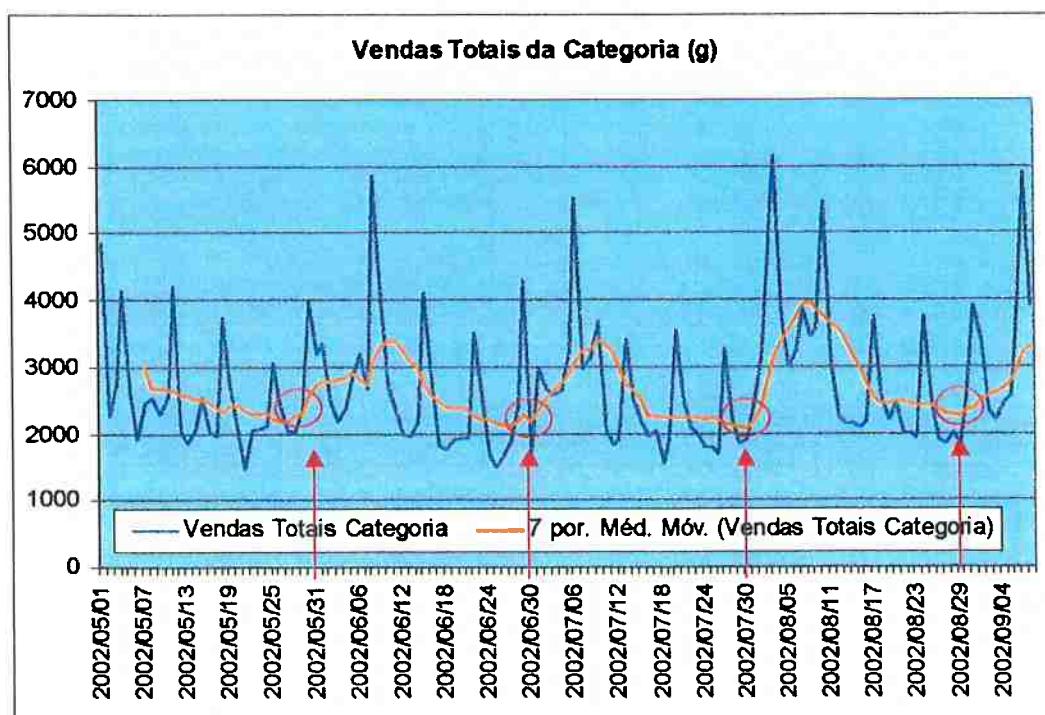


Figura 4.3 – Detalhe do Gráfico da Categoria com Média Móvel ($k=7$)

Analizando o comportamento dos dados no gráfico acima, percebe-se que existe uma sazonalidade dentro de um mês completo. No gráfico, a média móvel começa a se elevar aproximadamente no início de cada mês (círculos e setas em vermelho). Ou seja, quando o primeiro dia do mês começa a entrar no cálculo da média móvel (aproximadamente 3 dias antes de acabar o mês anterior), esta última “reage” e passa a se elevar. As participações de vendas por dias da semana e por semanas dentro de um mês estão representadas na tabela a seguir (dados calculados à partir de todas as vendas das 104 semanas analisadas):

Semana no Mês	Dia da Semana							Total
	DOM	SEG	TER	QUA	QUI	SEX	SAB	
1aSEM	3,6%	3,5%	3,3%	3,2%	3,0%	4,1%	6,2%	26,9%
2aSEM	3,7%	2,6%	3,1%	3,6%	2,7%	3,8%	5,2%	24,7%
3aSEM	3,5%	2,5%	2,7%	2,9%	2,8%	2,9%	5,9%	23,0%
4aSEM	2,9%	2,0%	2,2%	2,0%	2,1%	2,3%	3,9%	17,3%
5aSEM	1,6%	0,8%	0,8%	0,9%	0,9%	1,1%	1,9%	8,1%
Total	15,2%	11,3%	12,1%	12,6%	11,5%	14,2%	23,1%	100,0%

Tabela 4.2 – Participação das Vendas da Categoria por dia

A primeira semana de um mês é a que mais vende (26,9%), depois a segunda, e assim vai até o fim do mês (a quinta semana possui no máximo três dias). Isso se deve claramente ao fato de que o consumidor possui mais dinheiro no início do mês e ao longo do mesmo vai gastando cada vez menos. Nas primeiras semanas do mês ele está mais “disposto” a realizar maiores gastos. E entre os dias da semana, sábado é o mais forte (23,1%) e a segunda-feira o mais fraco (11,3%).

Porém, para verificar se existe uma sazonalidade nos dados, simplesmente observar um gráfico não é suficiente. Em função disso foram calculadas as autocorrelações na série temporal, com defasagens de 1 a 31 períodos (no caso, dias). A autocorrelação é uma estatística que mede a relação que os dados de uma série temporal têm com ela mesma, defasada de k períodos. Se for calculado, por exemplo, a correlação entre a observação em t com uma observação em $t-1$, pode-se verificar como observações consecutivas estão relacionadas (trata-se de uma autocorrelação de $k=1$). Diz-se que a observação em $t-1$ está defasada em 1 período. Os valores da autocorrelação normalmente decrescem a medida em que k aumenta. De acordo com Hanke; Reitsch (1998), se existe uma tendência em uma determinada série, Y_t e Y_{t-1} estão altamente correlacionados, sendo o valor da autocorrelação para $k=1$ muito alto (r_1), próximo de 1. Para $k=2$, o valor de r ainda será alto, e assim por diante, decrescendo o valor de r à medida em que a defasagem aumenta. Ainda segundo o autor, se na série temporal existe uma sazonalidade, um valor de autocorrelação significativo irá ocorrer a cada k períodos, onde k é a quantidade de períodos da sazonalidade. Calculando-se então os valores das autocorrelações para $k=1$ até $k=31$, tem-se o seguinte correlograma, onde são plotados os valores de r em função de k :

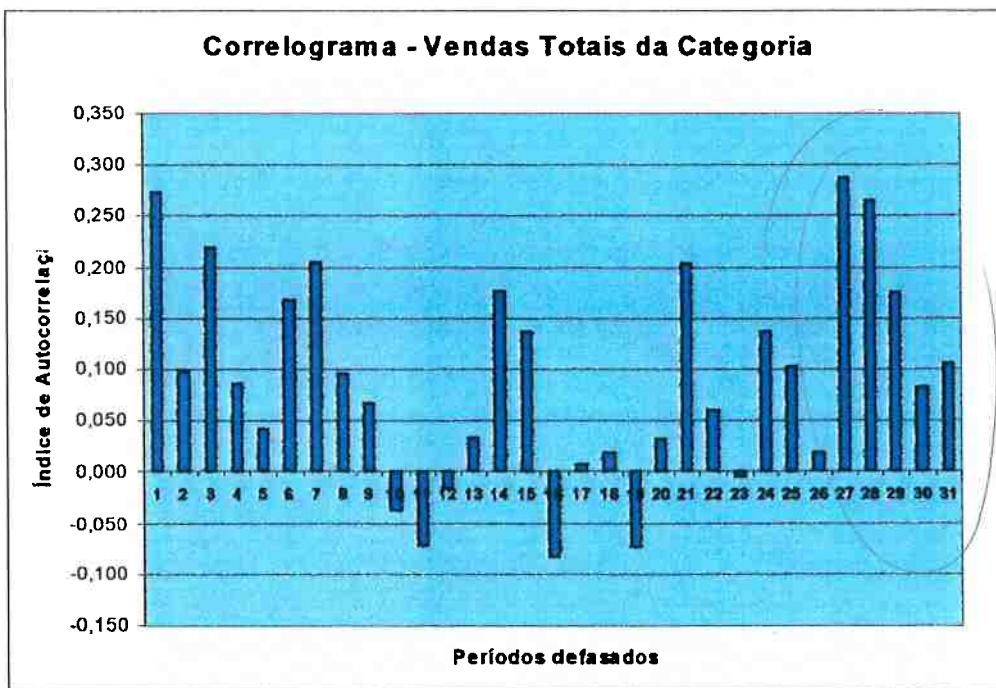


Figura 4.4 – Correlograma das Vendas da Categoria

Do gráfico, percebe-se que os “picos” de r estão em $k=1$, $k=3$, $k=7$, $k=14$, $k=21$, $k=27$ e $k=28$, sugerindo uma sazonalidade semanal, onde os valores se “repetem” a cada 7 dias. Em outras palavras, esse gráfico mostra que, se for observado um dia da série temporal ao acaso, por exemplo uma terça-feira, o valor da sua observação terá as maiores correlações com os dados das terças-feiras anteriores, evidenciando assim uma sazonalidade.

4.4.2. Vendas do Produto 1

Após ter uma visão geral da situação que ocorreu com as vendas da categoria como um todo, agora as análises serão concentradas nas marcas da empresa. Como explicado anteriormente, para se obter os dados totais de vendas de uma marca, deve-se primeiro descobrir todos os tipos de embalagens que foram vendidos ao longo do período estudado para então obter essas várias séries temporais e assim consolidá-las em uma única série. A análise a ser feita a seguir é similar a feita anteriormente com os dados da categoria.

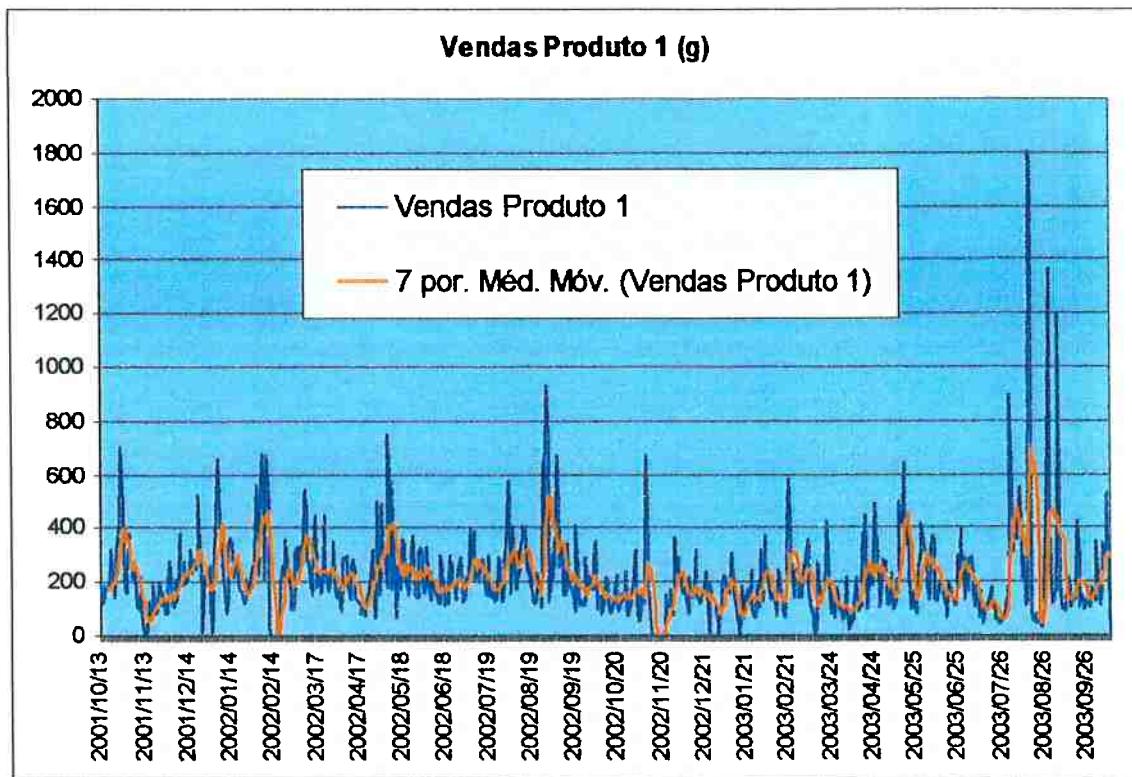


Figura 4.5 – Vendas diárias do Produto 1 com Média Móvel (k=7)

Percebe-se nos dados uma forte sazonalidade diária, assim como ocorria nos dados da categoria. Mais uma vez, os picos no gráfico se devem a eventos que ocorrem para alavancar as vendas dentro da loja, como a presença de demonstradora ou a distribuição de brindes. Porém ao se traçar a média móvel de 7 dias a sazonalidade “remanescente” – que diz respeito à ordem da semana no mês: primeira, segunda, etc. – já não fica tão evidente. Fechando um pouco o foco e analisando os dados mais de perto, temos:

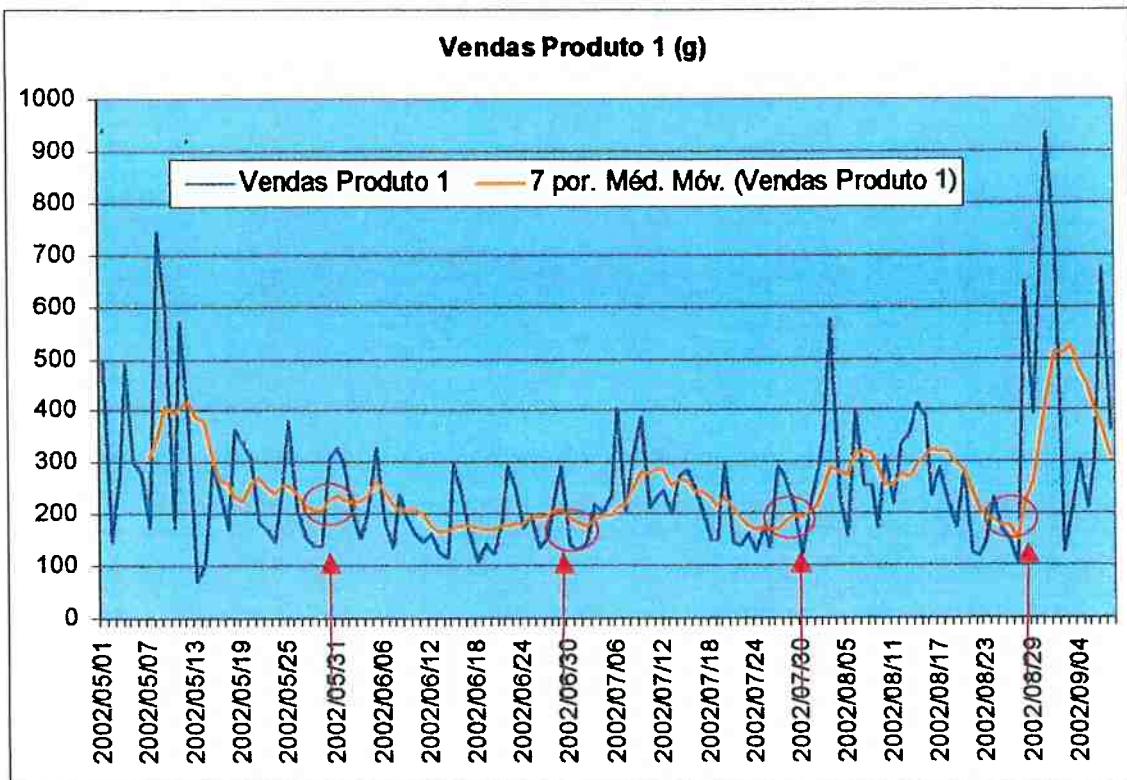


Figura 4.6 – Detalhe do Gráfico do Produto 1 com Média Móvel (k=7)

Nesses dados, o início da elevação que a média móvel apresentava todo fim de mês já não é tão intensa. Isso se deve ao fato de que, sendo os dados acima de uma marca somente, eles são mais sensíveis a pequenas variações por eventos pontuais. Por exemplo, se no último fim de semana de um determinado mês acontece uma forte promoção dessa marca, suas vendas subirão consideravelmente, o que comprometerá a média móvel, camuflando assim essa “segunda” sazonalidade dentro do mês. Isso não ocorre para os dados da categoria pois, tendo eles um volume muito maior, eventos pontuais como esse exemplificado não chegam a alterar o comportamento usual dos dados, eles possuem uma maior “inércia”.

A distribuição das vendas ao longo dos dias da semana e ao longo das semanas no mês para o produto 1 é:

Semana no Mês	Dia da Semana							Total
	DOM	SEG	TER	QUA	QUI	SEX	SAB	
1aSEM	3,9%	2,7%	3,3%	3,8%	3,2%	4,1%	6,3%	27,4%
2aSEM	3,8%	2,5%	2,5%	3,1%	2,3%	2,7%	5,0%	22,0%
3aSEM	4,0%	2,3%	2,5%	2,3%	2,6%	3,0%	5,0%	21,7%
4aSEM	3,1%	2,2%	2,1%	2,7%	2,5%	2,6%	4,6%	19,8%
5aSEM	1,1%	0,2%	0,8%	0,9%	1,2%	1,7%	2,6%	9,1%
Total	15,9%	10,6%	11,2%	12,7%	11,8%	14,1%	23,5%	100,0%

Tabela 4.3 – Participação das Vendas do Produto 1 por dia

Os resultados são muito similares aos obtidos para a categoria. A análise de autocorrelação é mostrada a seguir:

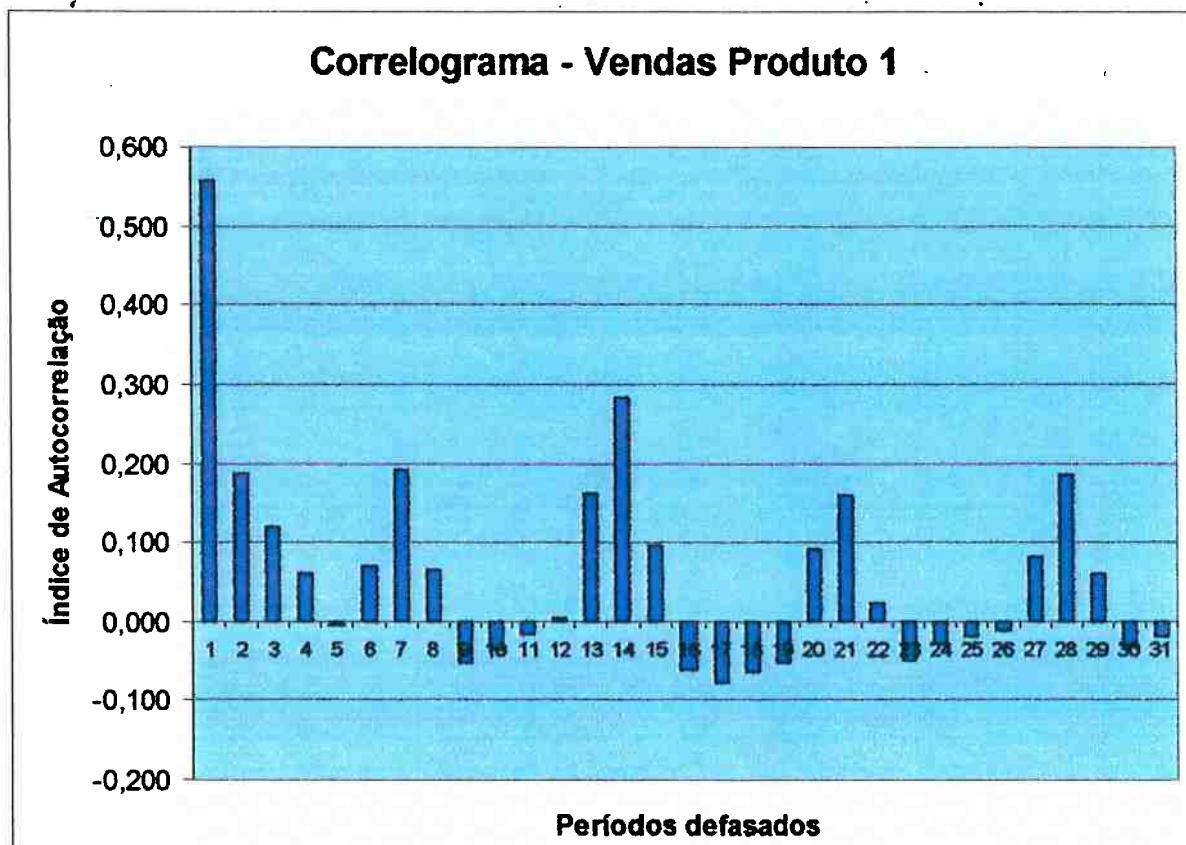


Figura 4.7 – Correlograma das Vendas do Produto 1

No caso das vendas do produto 1, a sazonalidade que existe a cada 7 dias fica mais evidente que nos dados da categoria, estando os picos em $k=1$, $k=7$, $k=14$, $k=21$ e $k=28$. Como explicado anteriormente, geralmente r_1 possui um valor alto (quando esse valor é bem próximo de 1, pode existir uma tendência na série).

4.4.3. Vendas do Produto 2

Sendo similares as análises realizadas para cada uma das quatro marcas, serão colocados aqui somente os resultados numéricos e algumas observações que forem pertinentes.

O gráfico de vendas do produto 2 durante o período estudado está abaixo:

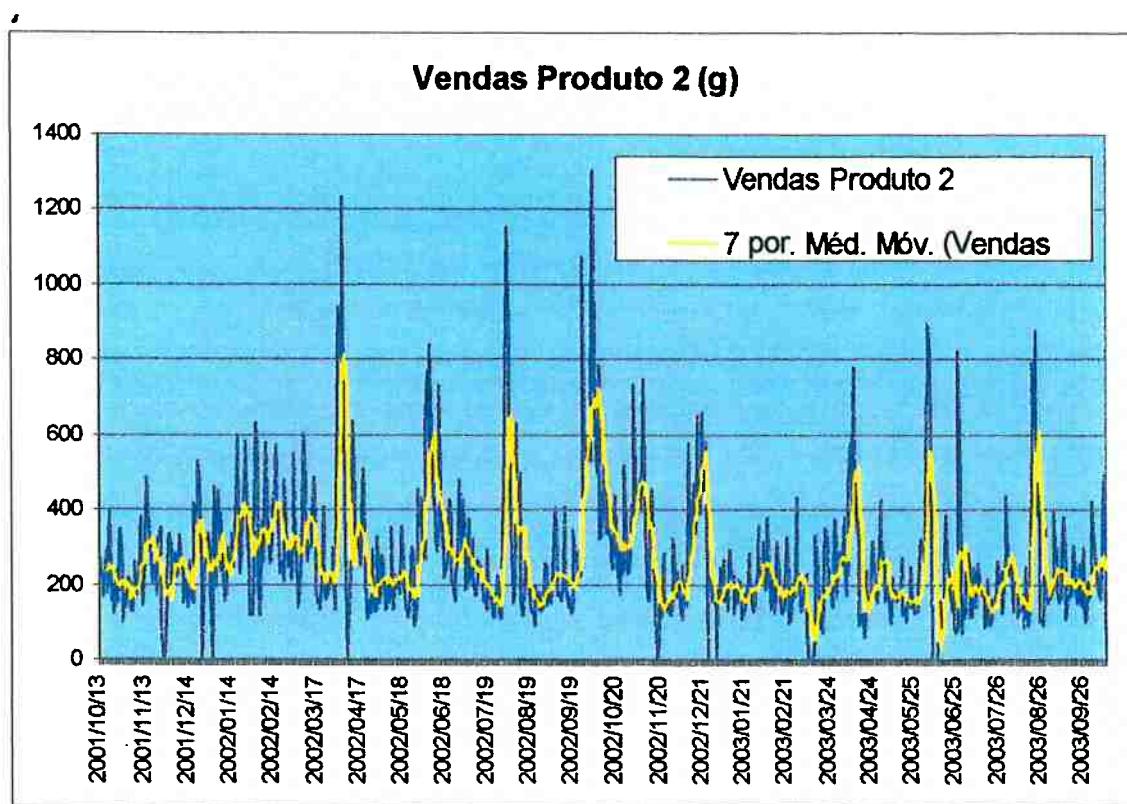


Figura 4.8 – Vendas diárias do Produto 2 com Média Móvel (k=7)

Pode-se perceber que existe uma sazonalidade diária, como era de se esperar. Porém, assim como ocorreu com a análise do produto 1, nos dados de vendas do produto 2 não está tão evidente a (segunda) sazonalidade entre as semanas de um mês, que pode ser “enxergada” quando a média móvel de 7 dias é plotada no gráfico, assim como ocorreu com os dados de vendas da categoria. A tabela de participação de vendas nos dias da semana está a seguir:

Semana no Mês	Dia da Semana							Total
	DOM	SEG	TER	QUA	QUI	SEX	SAB	
1aSEM	4,0%	3,3%	3,1%	3,9%	3,8%	4,2%	5,4%	22,6%
2aSEM	3,9%	2,7%	2,5%	2,9%	2,6%	2,7%	5,4%	22,7%
3aSEM	3,2%	2,4%	2,3%	3,0%	3,1%	3,0%	4,5%	21,4%
4aSEM	3,6%	2,4%	2,3%	2,1%	2,2%	2,7%	5,1%	20,3%
5aSEM	1,3%	0,8%	0,7%	0,8%	1,0%	0,8%	1,4%	6,9%
Total	16,0%	11,6%	11,0%	12,6%	12,6%	13,4%	22,8%	100,0%

Tabela 4.4 – Participação das Vendas do Produto 2 por dia

Para o produto 2, o dia da semana com vendas mais fracas é a terça-feira. O correlograma com os índices de autocorrelação confirma a existência da sazonalidade entre os dias da semana:

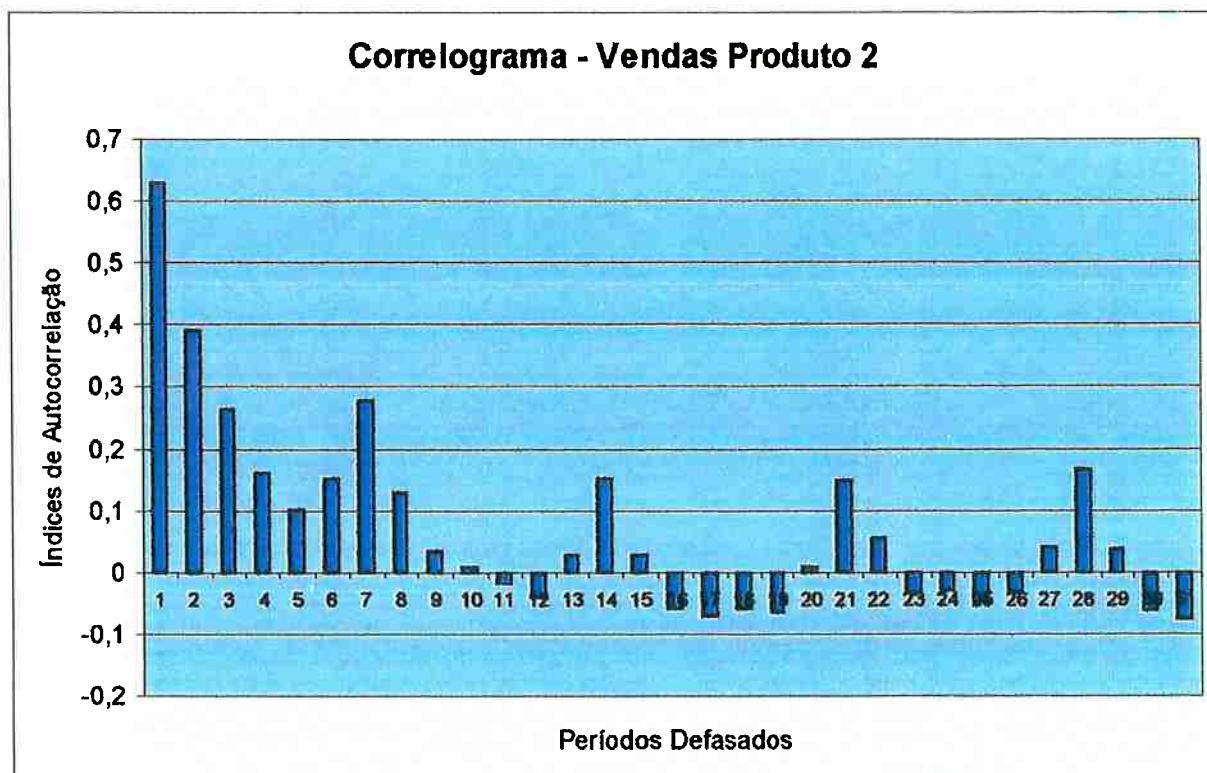


Figura 4.9 - Correlograma das Vendas do Produto 2

4.4.4. Vendas do Produto 3

O produto 3 é um produto com menor giro, como se pode notar no gráfico abaixo:

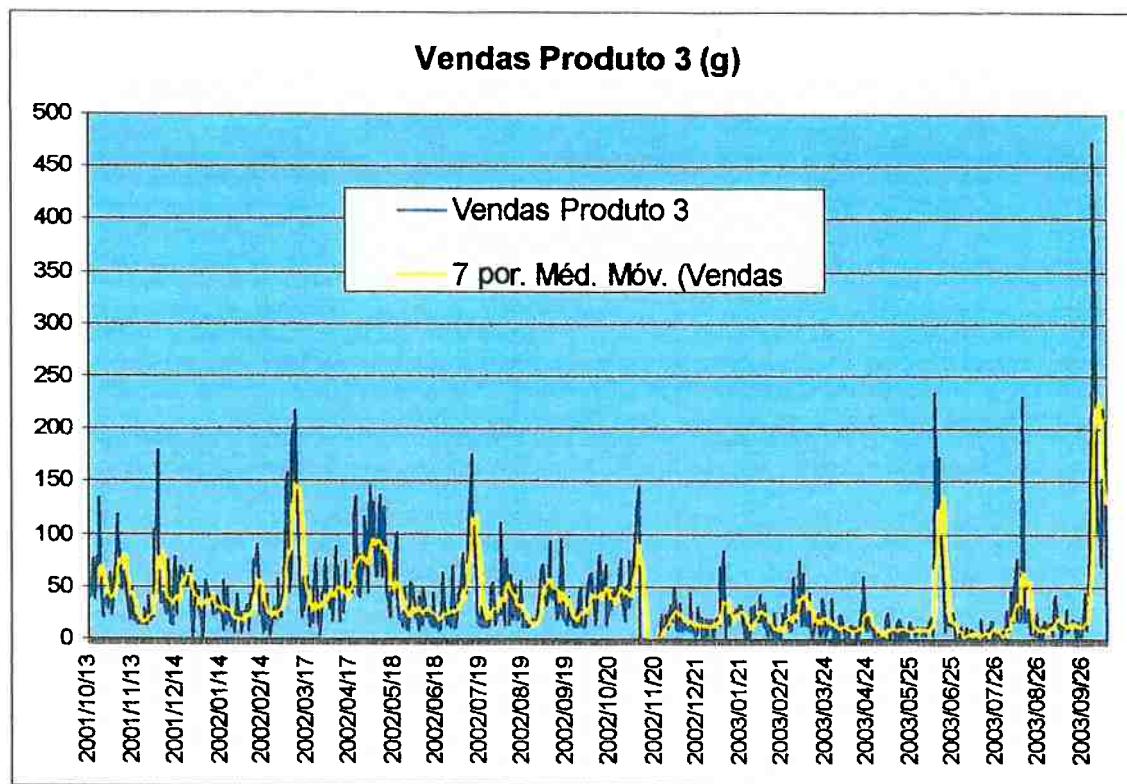


Figura 4.10 – Vendas diárias do Produto 3 com Média Móvel (k=7)

Semana no Mês	Dia da Semana							Total
	DOM	SEG	TER	QUA	QUI	SEX	SAB	
1aSEM	4,8%	3,1%	3,3%	4,4%	4,7%	4,0%	5,9%	31,1%
2aSEM	4,2%	2,2%	2,1%	4,0%	4,2%	4,9%	6,5%	28,2%
3aSEM	3,7%	2,9%	2,3%	2,0%	1,9%	2,1%	3,6%	18,5%
4aSEM	2,8%	1,6%	1,6%	2,2%	2,1%	1,7%	3,4%	15,4%
5aSEM	0,6%	0,7%	0,9%	0,8%	1,1%	1,4%	1,4%	6,8%
Total	16,1%	10,5%	10,2%	13,4%	14,0%	14,0%	21,8%	100,0%

Tabela 4.5 - Participação das Vendas do Produto 3 por dia

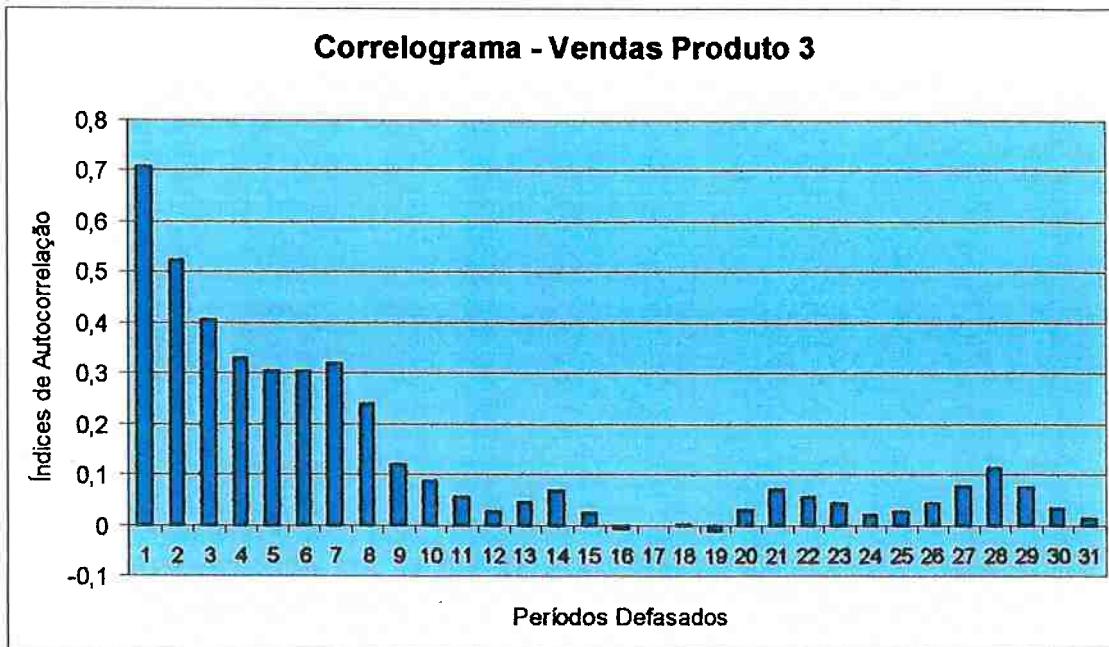


Figura 4.11 - Correlograma das Vendas do Produto 3

Assim como o produto 2, o produto 3 também tem a terça-feira como o dia de menor vendas. A sazonalidade também pode ser notada através do correlograma acima.

4.4.5. Vendas do Produto 4

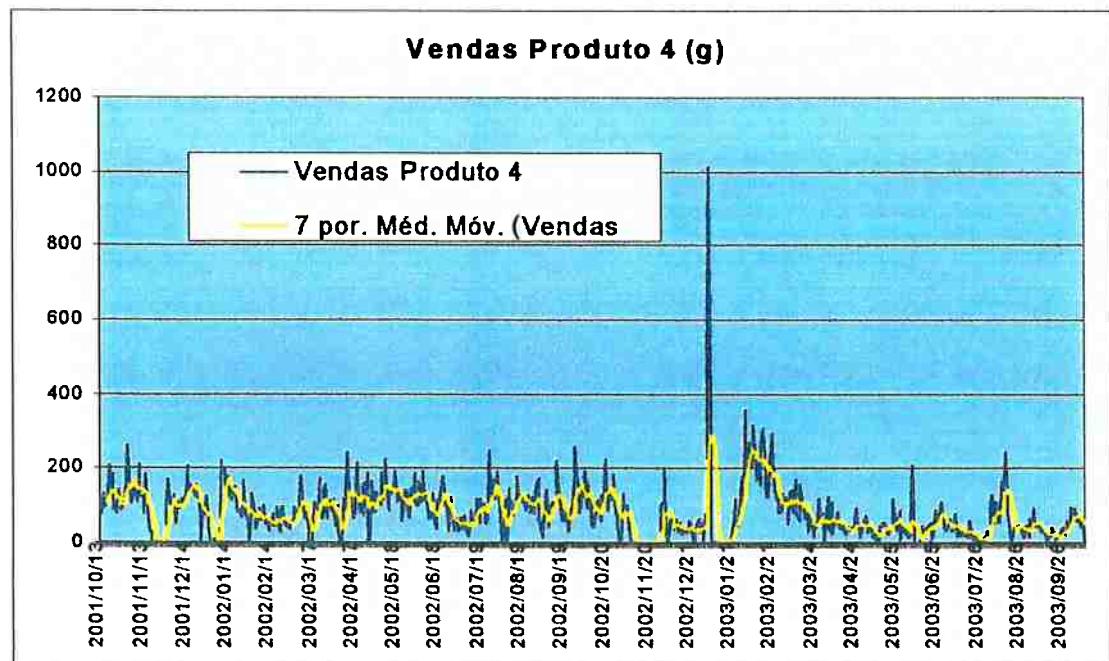


Figura 4.12 - Vendas diárias do Produto 4 com Média Móvel (k=7)

O produto 4 também apresenta um giro baixo, mas ainda um pouco acima do produto 3:

A participação dos dias da semana nas vendas pode ser conferida na tabela seguinte:

Semana no Mês	Dia da Semana							Total
	DOM	SEG	TER	QUA	QUI	SEX	SAB	
1a SEM	4,8%	3,4%	3,4%	3,1%	2,6%	3,8%	5,3%	26,4%
2a SEM	3,4%	2,7%	2,6%	3,4%	2,7%	2,9%	4,8%	22,5%
3a SEM	3,7%	2,8%	2,9%	2,8%	3,3%	3,1%	4,7%	23,2%
4a SEM	2,9%	2,5%	2,3%	2,7%	2,3%	2,9%	5,0%	20,5%
5a SEM	1,1%	1,0%	1,1%	1,1%	0,9%	0,9%	1,3%	7,3%
Total	15,8%	12,3%	12,3%	13,0%	11,9%	13,0%	21,2%	100,0%

Tabela 4.6 - Participação das Vendas do Produto 4 por dia

O dia mais fraco de vendas para esse produto é a quinta-feira. E a terceira semana do mês obteve uma participação das vendas maior do que a segunda semana, contrariando o que era esperado.

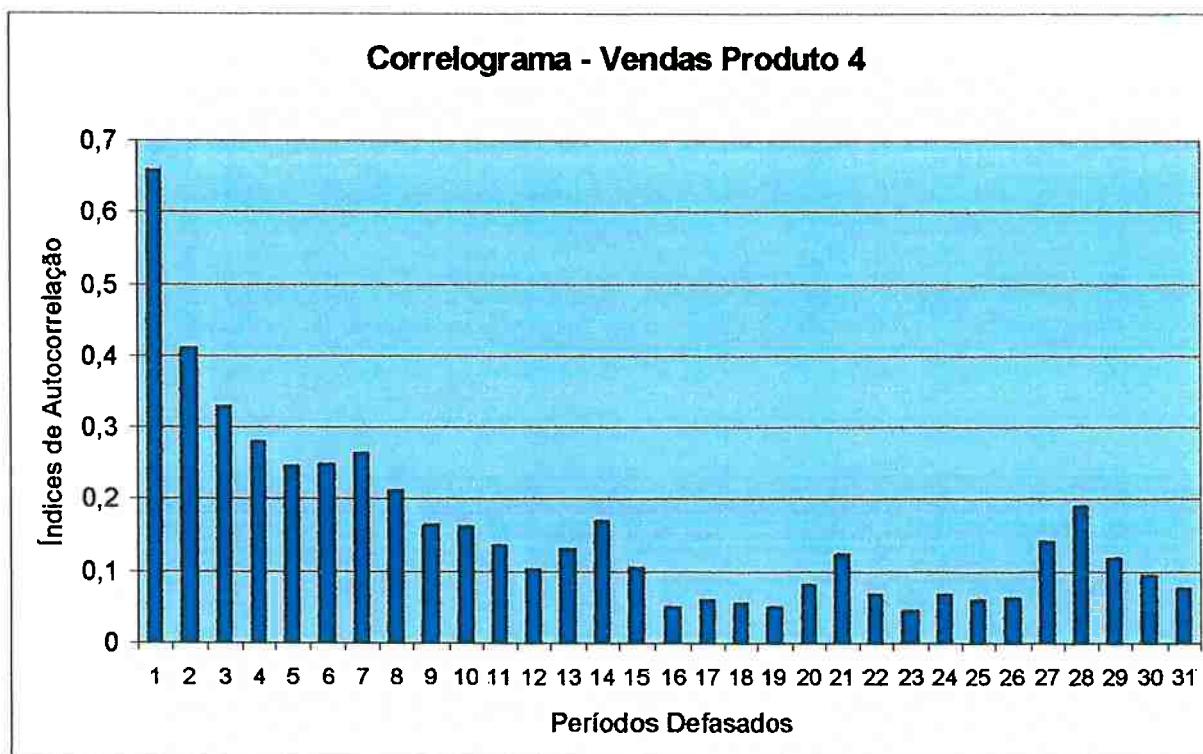


Figura 4.13 - Correlograma das Vendas do Produto 4

O correlograma do produto 4 também apresenta os maiores índices de autocorrelação para os primeiros valores de k e também para $k=14$, $k=21$ e $k=28$.

4.5. Seleção do Modelo de Previsão

Os dados que estão disponíveis listados anteriormente certamente influenciam nas vendas dos produtos em questão. Em outras palavras, as vendas são dependentes dessas variáveis. Em função dessa disponibilidade ampla desses dados, foi decidido desenvolver um sistema de Predição de Vendas. A idéia é desenvolver dois modelos paralelamente, alimentando um deles com informações geradas do outro. Os dois modelos seriam:

- Um modelo temporal de projeção das Vendas, que se baseará estritamente nos dados históricos de vendas, e que fará as previsões.
- Um modelo causal, que tentará explicar a relação entre as vendas e as variáveis que as influenciam. Não serão geradas previsões a partir desse modelo.

Assim, as informações do modelo causal serão consideradas no modelo de projeção para ao fim se obter uma predição das vendas, de forma a melhorar as previsões geradas pelo modelo temporal.

Logo, o modelo causal deverá ser o de regressão múltipla, por possuir mais de uma variável explicativa.

E para selecionar então o modelo de projeção mais adequado à necessidade, deve-se levar em conta os fatores explicados a seguir, de acordo com O'Donovan (1983).

4.5.1. Horizonte de Previsão

Refere-se ao período de tempo no futuro no qual a previsão de vendas é necessária. Este período pode ser classificado da seguinte forma:

- Imediato: menor que um mês
- Curto-prazo: de um a três meses
- Médio-prazo: de três meses a dois anos
- Longo-prazo: mais que dois anos

De acordo com O'Donovan (1983), os modelos mais utilizados para as previsões imediatas são os modelos de Suavização Exponencial Simples, de Winters, de Decomposição e de Box-Jenkins. Em se tratando de curto-prazo, o autor não impõe restrições quanto aos métodos. E para os de médio e longo-prazo, os métodos de regressão e os modelos econométricos são os mais indicados.

4.5.2. Disponibilidade de Dados

Determinados modelos de previsão necessitam de muitos dados para poder ser implantados, como por exemplo o Método de Box-Jenkins. Por isso este fator possui um caráter restritivo.

4.5.3. Comportamento dos Dados

O comportamento dos dados de vendas de uma empresa pode assumir os seguintes padrões, podendo estar combinados ou não:

- Sazonal: apresenta um comportamento periódico regular
- Tendência: de crescimento ou decrescimento
- Permanência: distribuição aleatória, sem tendência de crescimento ou decrescimento aparente
- Ciclico: apresenta flutuações ao redor da tendência, geralmente causada pelas condições econômicas presentes

Esse conhecimento a respeito do comportamento dos dados torna-se muito importante à medida que diferentes métodos variam em suas capacidades de se ajustar a este comportamento encontrado.

4.5.4. Complexidade

Este fator talvez seja um dos mais importantes a ser considerado. Para O'Donovan (1983), a principal razão pela qual os modelos mais matematicamente demandantes, complexos e sofisticados não sejam utilizados é pelo fato de que, se os usuários da previsão não entenderem como ela foi gerada, provavelmente não terão confiança e irão abandoná-la. Hanke; Reitsch (1998) ressaltam que não é essencial que o método apresente um processo matemático elaborado ou sofisticado, mas que ele seja oportuno, preciso e compreendido pelo gestor.

4.5.5. Custos

Na escolha do modelo de previsão alguns custos podem ser relevantes, tais como os custos de coleta e de armazenamento dos dados (por exemplo para se obter informações externas em modelos causais), e os custos de monitoramento e de desenvolvimento dos modelos de previsão (alguns softwares que auxiliam nessa etapa do processo ainda possuem um custo elevado).

4.5.6. Precisão

Muitos fatores influenciam a precisão da previsão, dentre eles dados insuficientes, o uso de um modelo não-adequado com a realidade dos dados ou uma descontinuidade brusca nas variáveis determinantes. O aspecto precisão pode ser visto como um critério de desempate entre métodos que forem aprovados nos aspectos anteriores.

Como explicado anteriormente, o sistema de previsão desenvolvido será composto de um método de projeção e de um modelo causal. A respeito deste último, foi definido que será utilizada a regressão múltipla (devido à quantidade de variáveis explicativas). Para o método de projeção, um dos aspectos colocados acima já restringe os métodos a poucas opções: o

comportamento dos dados. Nas análises anteriores foi demonstrada a sua sazonalidade. Dessa forma, as possibilidades ficam restritas ao Método de Winters, Decomposição e Box-Jenkins. Este último, por ser muito complexo e trabalhoso e não necessariamente melhor, foi descartado. Quanto à Decomposição, de acordo com Makridakis; Wheelwright; Hyndman (1998), é preferível usá-la mais com uma ferramenta de auxílio à compreensão da série temporal do que como um método de previsão propriamente dito. Dessa forma, o método selecionado para se projetar os dados foi o Método de Winters.

4.6. Desenvolvimento do Modelo – Parâmetros

Tendo todos os dados necessários para se fazer as previsões, e após analisá-los e tratá-los, é possível iniciar o desenvolvimento do modelo de previsão selecionado na etapa anterior.

Na primeira etapa será desenvolvido o modelo de projeção da série histórica de vendas dos quatro produtos em estudo e dos números totais da categoria, tendo como base somente os dados de vendas em volume. Na segunda etapa serão utilizados todos os dados de demanda e as variáveis levantadas como possíveis explicativas das vendas dos produtos.

Para auxiliar nos cálculos das previsões e realizar as análises desejadas, foi utilizado o software *E-Views*. Para problemas e cálculos mais simples foi utilizado o *Excel*.

4.6.1. Modelo Temporal

Como explicado no capítulo 3, os métodos de previsão temporais utilizam somente os dados de vendas históricos para adequar um modelo ao seu comportamento e então poder fazer a projeção das vendas futuras. Retomando o que havia sido explicado a respeito do método selecionado, o Método da Suavização Exponencial Linear com Três Parâmetros (Método de Winters), a previsão de vendas F para o período $t+m$ é calculada da seguinte forma:

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (3.26)$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \quad (3.27)$$

$$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s} \quad (3.28)$$

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \quad (3.29)$$

onde:

L_t é estimativa do nível

S_{t-s} é o fator de sazonalidade, sendo s o número de períodos da sazonalidade

b_t é a estimativa da tendência linear

Os coeficientes de suavização, α , β e γ devem ter valor entre 0 e 1. Eles funcionam como calibradores do modelo, dando maior ou menor importância aos dados mais recentes ou antigos.

De acordo com as sazonalidades dos dados analisadas anteriormente, foi definido que o período s da sazonalidade seria de 7 dias. Ou seja, no modelo vão existir 7 fatores de sazonalidade, um para cada dia da semana.

Para que o modelo possa ser aplicado, é necessário a definição de um período de inicialização e outro de testes, como explicado no capítulo 3. Sendo a série histórica de dados composta por aproximadamente 720 pontos, foi decidido que o “hoje adotado” seria na metade da série. O primeiro valor que consta é o do dia 13/out/2001, e o último 12/out/2003. O “hoje adotado” ficou sendo então o dia 13/out/2002, e os dias que o sucedem são tratados como “futuro adotado”, cujos dados serão utilizados para se calcular as medidas de erro de previsão. Retomando o esquema teórico apresentado anteriormente e inserindo as datas respectivas, tem-se:

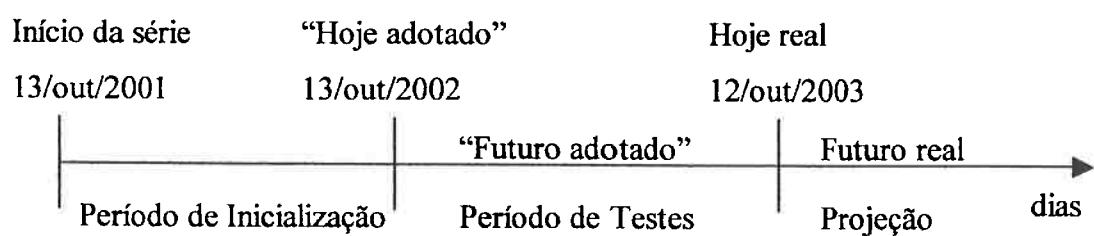


Figura 4.14 - Definição de Períodos Aplicada aos Dados

Esquema adaptado de Santoro (2000)

Para o método de Winters, os cálculos e estimativas são feitas da seguinte forma, baseadas no esquema anterior:

- Durante o período de Inicialização:
 - Estimativa dos valores iniciais de L , b e das sazonalidade S .
 - Estimativa dos coeficientes α , β e γ
- Durante o período de Testes:
 - Cálculo das previsões F
 - Cálculo das medidas de erro de previsão

Para fazer as estimativas baseadas no período de inicialização foi utilizado o software *E-Views*, e para os cálculos do período de testes foi utilizado o *Excel*.

Mesmo tendo sido selecionado o Método de Winters como o mais apropriado, foram testados no *E-Views* outros possíveis métodos para se descobrir o quanto esse modelos se adequariam aos dados do período de inicialização. Cada teste de modelo teve como saída uma medida de erro, que foram utilizados nessa comparação. A série escolhida para esse teste foi a dos dados de vendas do produto 1. Os resultados estão a seguir:

Raiz do Erro Quadrático Médio	
Suavização Exponencial Simples	120,598
Método de Brown	127,135
Método de Holt	120,557
Método de Winters	101,843

Tabela 4.7 – Erro para Diferentes Métodos

Como era de se esperar, o Método de Winters foi o que apresentou a menor medida de erro, confirmando a seleção apropriada do modelo.

O *E-Views* tem como saída, para os cálculos do Método de Winters durante o período de inicialização, os seguintes parâmetros e estimativas:

- Parâmetros α , β e γ (estimados de forma a minimizar a soma dos erros quadráticos das previsões em relação as observações reais, durante o período de inicialização)
- Fatores da equação no último período (13/out/02): nível L_t , tendência b_t e os 7 fatores de sazonalidade S

Com esses dados em mãos, é possível projetar as previsões do “futuro adotado” (a partir de 14/out/02). Isso foi feito com o auxílio do *Excel*. Os resultados são apresentados nos itens a seguir.

Primeiramente é feita a previsão de vendas da categoria como um todo, a fim de prover uma noção geral do comportamento dos dados. Assim como foi feito na pré-análise dos dados, todos os conceitos e análises serão explicados nessa primeira abordagem. Nas subsequentes previsões específicas de cada produto essas explicações não serão repetidas, sendo a abordagem mais direta e objetiva com foco nos resultados obtidos.

4.6.1.1.Previsão de Vendas da Categoria

As saídas do *E-Views* para essa série de dados são as seguintes:

Sample:	10/13/2001 10/13/2002
Included observations:	366
Method:	Holt-Winters Additive Seasonal
Original Series:	VCAT
Forecast Series:	VCATSM
Parameters:	
Alpha	0,500
Beta	0
Gamma	0
Sum of Squared Residuals	325080262,561
Root Mean Squared Error	942,442
End of Period Levels:	
Mean	4427,656
Trend	4,389
Seasonals:	
10/07/2002	-827,327
10/08/2002	-893,182
10/09/2002	-434,423
10/10/2002	-532,320
10/11/2002	-64,308
10/12/2002	2530,809
10/13/2002	220,751

Figura 4.15 – Saída do *E-Views* para a Projeção de Vendas da Categoria

Onde:

Mean é o nível L_t

Trend é tendência b_t

Seasonals são os últimos 7 fatores de sazonalidade (pois $s=7$)

Todos esses valores obtidos são para $t=13/10/02$

O *E-Views* atribui aos parâmetros β e γ valores zero. Isso significa que os valores iniciais da tendência b_t (*trend*) e as últimas sete sazonalidades S não serão atualizados a cada período, permanecendo os mesmos ao longo da projeção. O valor do fator de sazonalidade S do dia 12/10/02 é muito mais alto do que os outros – corresponde a um sábado, dia forte em vendas como comprovado anteriormente. Juntamente com os parâmetros β e γ constam duas medidas de erro, calculadas no período de inicialização (a *Sum of Squared Residuals* foi a medida que foi minimizada a fim de encontrar os valores dos 3 parâmetros). Com todos os parâmetros e valores obtidos, a segunda etapa é iniciada, onde serão estimadas as previsões do “futuro adotado” e serão comparadas com as observações reais. Para isso foi construída uma tabela como a que está a seguir, onde constam todos os parâmetros e dados necessários, e foram calculadas as previsões, de acordo com as equações teóricas apresentadas no capítulo 3. O exemplo abaixo é parte da tabela inteira, e os dados são os relativos às vendas totais da categoria:

PERÍODO	Real	NÍVEL L_t	TENDÊNCIA	SAZONALID.	Previsão	Desvio	Desvio	Desvio	Desvio	Desvio
		CIA bt	ADE St	F		relativo	Absoluto	Quadrático	Relativo %	Absoluto %
2002/10/06	4937,4									
2002/10/07	4979,8				-827,327					
2002/10/08	4008,7				-893,182					
2002/10/09	4210,2				-434,423					
2002/10/10	4017,8				-532,320					
2002/10/11	4440,5				-64,308					
2002/10/12	6805,4				2.530,809					
2002/10/13	4616,2	4.427,656	4,389		220,751					
2002/10/14	3942,0	4.600,701	4,389	-827,327	3.604,718	-337,312	337,312	113779,445	-8,6%	8,6%
2002/10/15	4326,8	4.912,556	4,389	-893,182	3.711,909	-614,931	614,931	378140,5557	-14,2%	14,2%
2002/10/16	3689,0	4.520,197	4,389	-434,423	4.482,521	793,496	793,496	629636,6681	21,5%	21,5%
2002/10/17	2928,5	3.992,700	4,389	-532,320	3.992,266	1.063,771	1.063,771	1131608,088	36,3%	36,3%
2002/10/18	3132,7	3.597,057	4,389	-64,308	3.932,781	800,064	800,064	640101,851	25,5%	25,5%
2002/10/19	5115,2	3.092,941	4,389	2.530,809	6.132,256	1.017,010	1.017,010	1034308,417	19,9%	19,9%
2002/10/20	4290,8	3.583,681	4,389	220,751	3.318,081	-972,702	972,702	946149,1546	-22,7%	22,7%
2002/10/21	2678,8	3.547,110	4,389	-827,327	2.760,743	81,920	81,920	6710,833942	3,1%	3,1%
2002/10/22	2708,5	3.576,599	4,389	-893,182	2.658,318	-50,199	50,199	2519,9856	-1,9%	1,9%
2002/10/23	2531,0	3.273,185	4,389	-434,423	3.146,564	615,605	615,605	378970,039	24,3%	24,3%

Tabela 4.8 – Exemplo de Tabela Utilizada para Projeção das Vendas

Em vermelho está o período de inicialização. No dia 13/10/02 constam os últimos valores de L , b e S , relativos a este período. Os outros seis valores de S estão antes dele. São calculadas então as previsões no período de testes e seus erros, sendo os fatores L , b e S atualizados a

cada período (dependendo do valor dos seus respectivos coeficientes α , β e γ) até o último dado (12/10/03). Os resultados obtidos estão no gráfico a seguir (a escala do eixo y do gráfico foi reduzida para que os dados fiquem mais visíveis; em função disso dois pontos extremos não são mostrados):

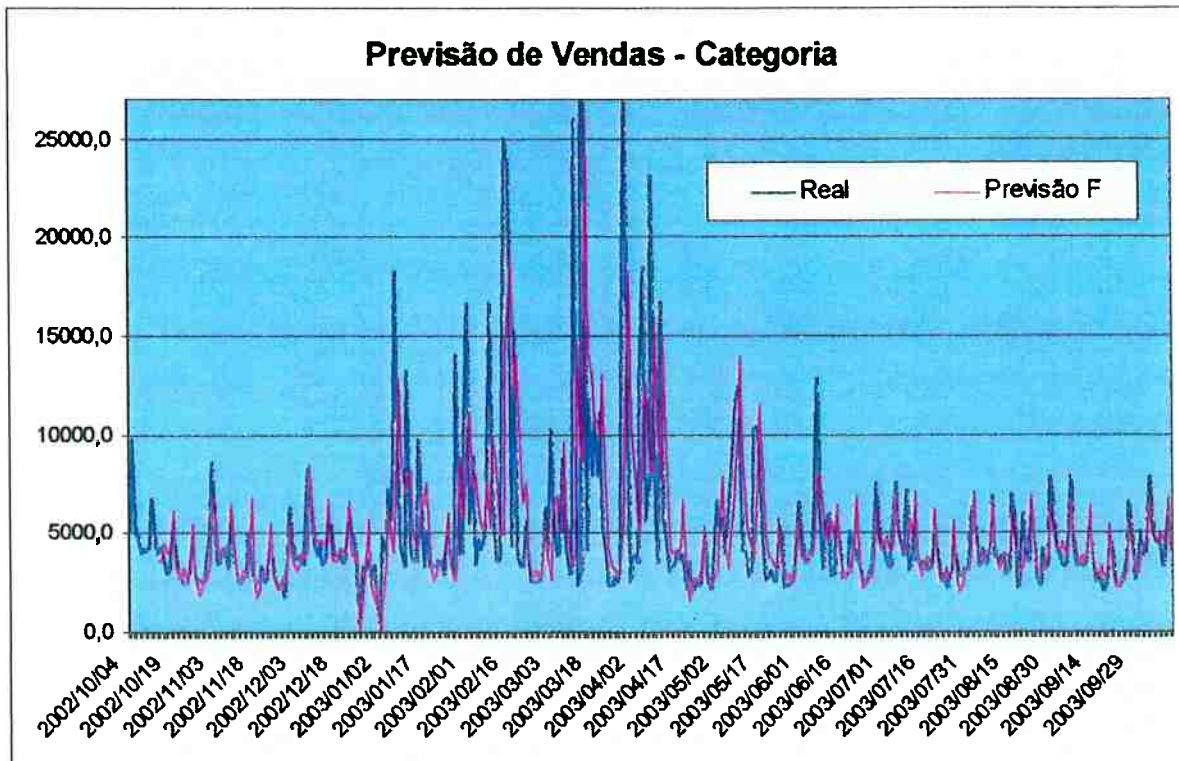


Figura 4.16 – Previsão de Vendas da Categoria

Os valores totais dos erros de previsão computados são:

ME	MAE	MSE	MPE	MAPE
-8.498	1987,420	17.847.086,09	16,6%	36,1%

Tabela 4.9 – Erros de Previsão de Vendas da Categoria

Onde:

ME = Erro Relativo Médio (*Mean Error*)

MAE = Erro Absoluto Médio (*Mean Absolute Error*)

MSE = Erro Quadrático Médio (*Mean Squared Error*)

MPE = Erro Relativo Percentual Médio (*Mean Percentage Error*)

MAPE = Erro Absoluto Percentual Médio (*Mean Absolute Percentage Error*)

Convém frisar que todas as medidas de erro calculadas são relativas às previsões de 1 período à frente, tanto para essa série de dados como para as que serão analisadas a seguir.

Os altos valores encontrados para os erros, o MAPE por exemplo, se devem principalmente a algumas poucas observações que estão muito acima dos valores médios, e dessa forma a elevação ao quadrado desses erros pune excessivamente a medida final. Esses picos, como explicado anteriormente na análise dos dados, se devem a eventos que ocorrem na loja, que fazem as vendas ficarem muito acima dos valores esperados. Esses eventos são previsíveis qualitativamente, mas imprevisíveis pelo modelo adotado aqui. Isso pode ser observado pelo gráfico a seguir:

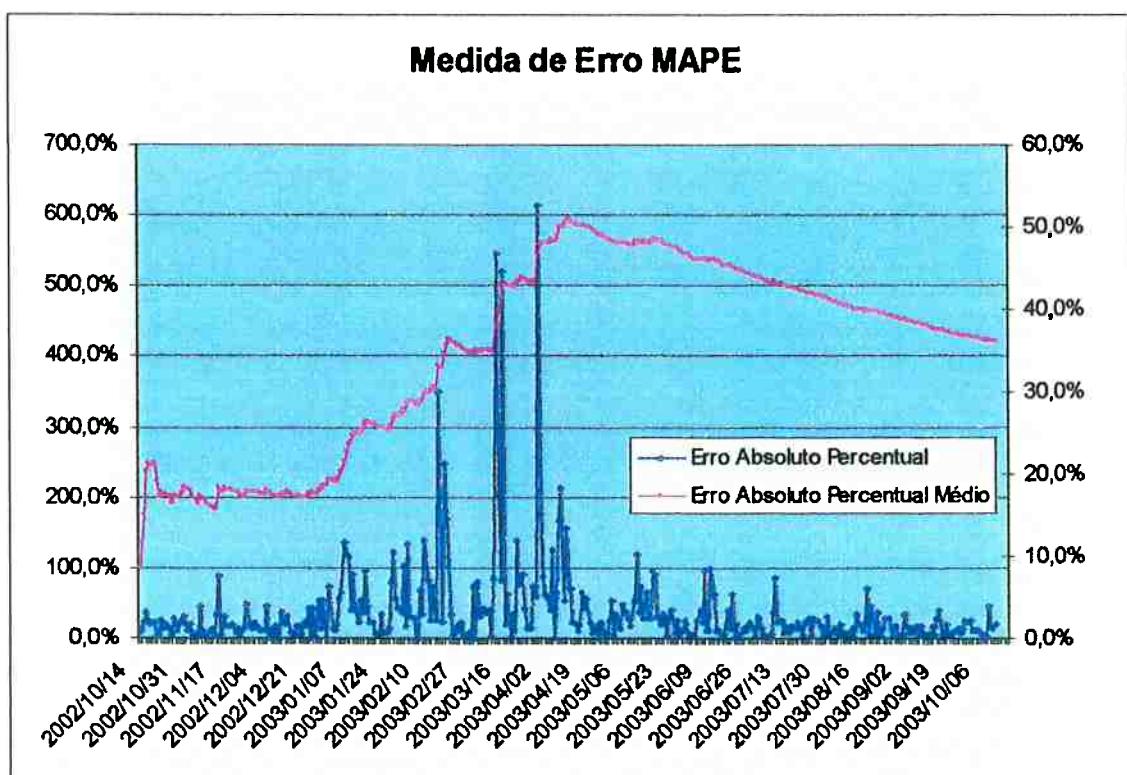


Figura 4.17 – Evolução do MAPE para as Previsões da Categoria

4.6.1.2.Previsão de Vendas do Produto 1

Tendo uma visão geral da categoria, o próximo passo é a previsão de vendas de cada um dos produtos. A saída do *E-Views* para a série de dados do produto 1 é a seguinte:

Sample:	10/13/2001 10/13/2002
Included observations:	366
Method:	Holt-Winters Additive Seasonal
Original Series:	Vendas1
Forecast Series:	Vendas1SM
Parameters:	
Alpha	0,590
Beta	0
Gamma	0
Sum of Squared Residuals	3.796.174,151
Root Mean Squared Error	101,843
End of Period Levels:	
Mean	129,341
Trend	-0,018
Seasonals:	
10/07/2002	-47,470
10/08/2002	-38,828
10/09/2002	-0,604
10/10/2002	-34,788
10/11/2002	-17,258
10/12/2002	118,596
10/13/2002	20,354

Figura 4.18 - Saída do *E-Views* para a Projeção de Vendas do Produto 1

Aqui o software também atribuiu valor zero a β e γ , fazendo com que a medida de tendência (*trend*) e os 7 fatores de sazonalidade não sofram atualizações a cada período. O valor de b praticamente nulo, o que demonstra uma ausência de tendência no comportamento dos dados. Os resultados obtidos para as previsões são os seguintes:

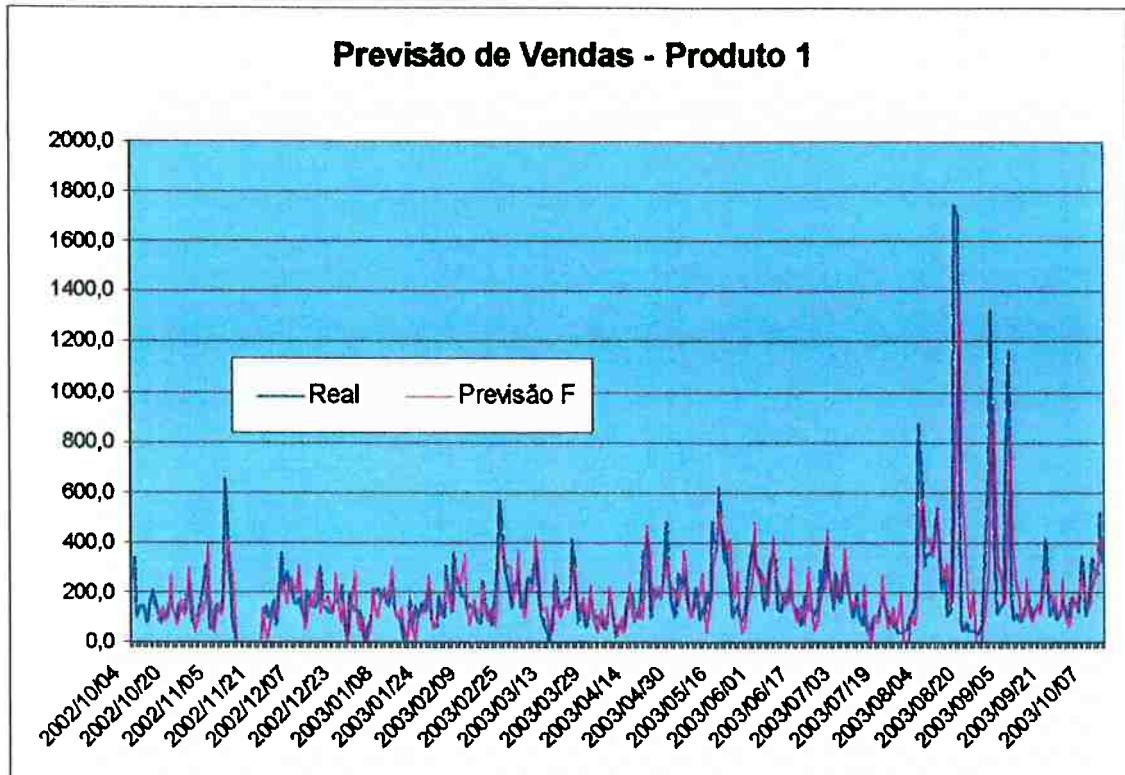


Figura 4.19 - Previsão de Vendas do Produto 1

Os valores das medidas dos erros de previsão calculados são:

ME	MAE	MSE	MPE	MAPE
-2,701	85,218	25407,384	28,0%	63,2%

Tabela 4.10 - Erros de Previsão de Vendas do Produto 1

Assim como ocorreu com os dados da categoria, aqui também foram encontrados valores altos para as medidas acima. Parte desses altos desvios pode ser explicada da mesma forma que foi feito anteriormente: em função de alguns eventos imprevisíveis pelo modelo que ocorrem na loja e que aumentam consideravelmente as vendas, causando enormes erros de previsão que quando elevados ao quadrado tem uma influência significativa no resultado final, como pode ser visto pelo gráfico a seguir:

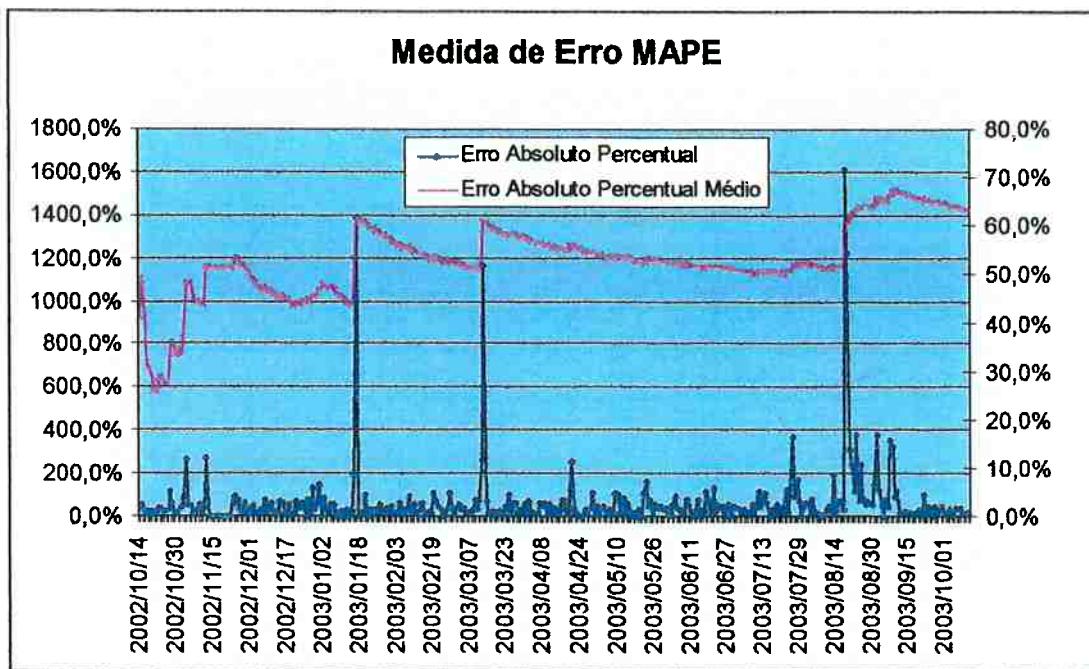


Figura 4.20 - Evolução do MAPE para as Previsões do Produto 1

4.6.1.3. Previsão de Vendas do Produto 2

O E-Views fornece os seguintes parâmetros para o produto 2, estimativas baseadas no período de inicialização definido anteriormente:

Sample:	10/13/2001 10/13/2002
Included observations:	366
Method:	Holt-Winters Additive Seasonal
Original Series:	Vendas2
Forecast Series:	Vendas2SM
Parameters:	
Alpha	0,760
Beta	0
Gamma	0
Sum of Squared Residuals	5.274.717,652
Root Mean Squared Error	120,049
End of Period Levels:	
Mean	371,788
Trend	1,098
Seasonals:	
10/07/2002	-50,347
10/08/2002	-67,710
10/09/2002	-33,728
10/10/2002	-42,029
10/11/2002	-13,884
10/12/2002	175,642
10/13/2002	32,057

Figura 4.21 - Saída do E-Views para a Projeção de Vendas do Produto 2

Nesse caso pode ser notado um aumento na tendência em t ($b_t=0,484$) em relação ao obtido para o produto 1. O único coeficiente com valor não-nulo ainda é o α , responsável por suavizar as medidas do nível L. Os resultados finais para as previsões são:

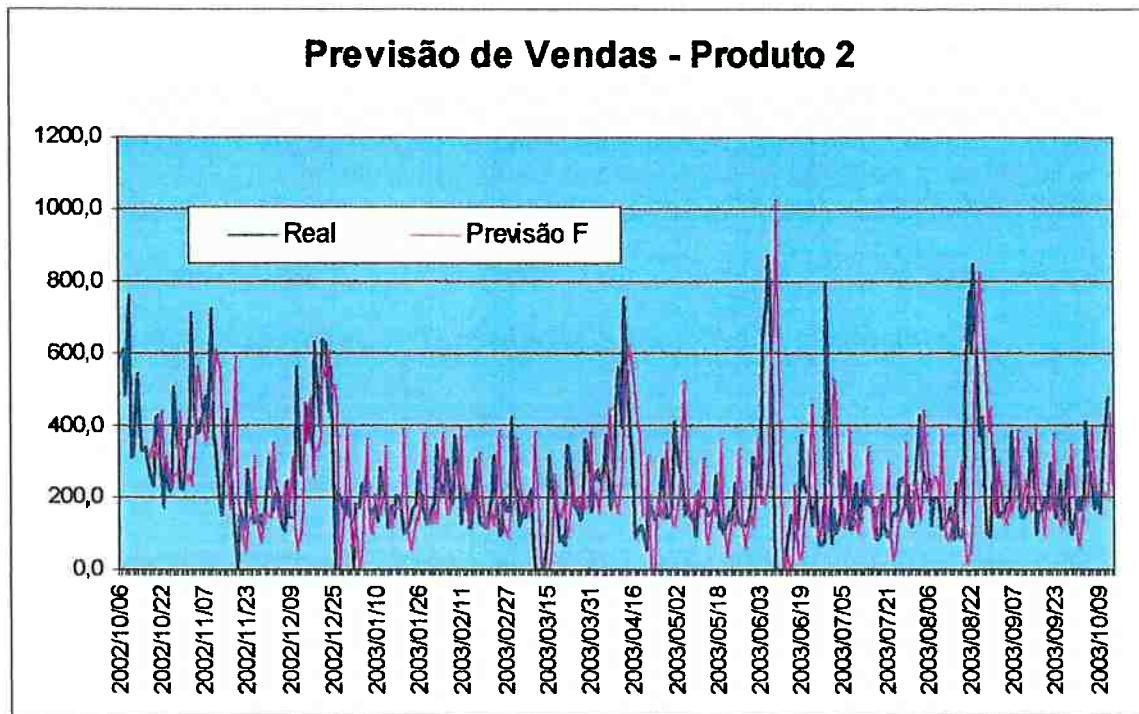


Figura 4.22 - Previsão de Vendas do Produto 2

Os erros de previsão estão na tabela a seguir:

ME	MAE	MSE	MPE	MAPE
-1,000	71,681	11.464,67	35,9%	64,2%

Tabela 4.11 - Erros de Previsão de Vendas do Produto 2

As medidas de erro acima demonstradas também estão altas, pelo mesmo motivo explicado anteriormente (alguns poucos dados extremos), mas nesse caso com mais intensidade. O gráfico do MAPE mais uma vez comprova isso:

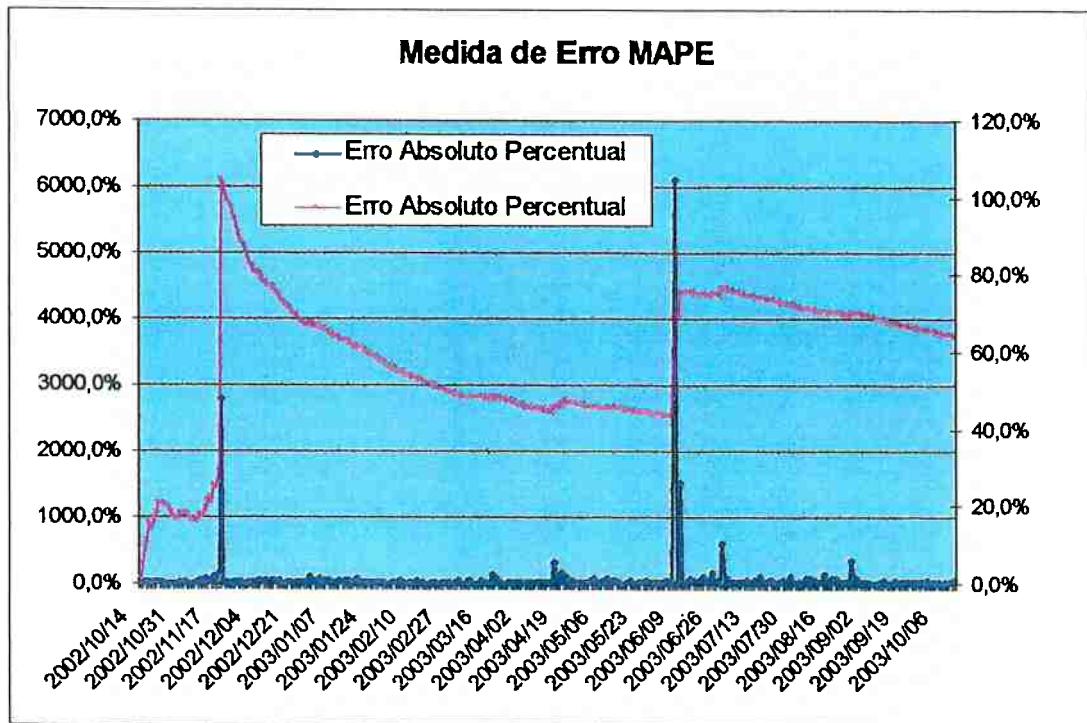


Figura 4.23 - Evolução do MAPE para as Previsões do Produto 2

4.6.1.4. Previsão de Vendas do Produto 3

O gráfico das previsões está a seguir:

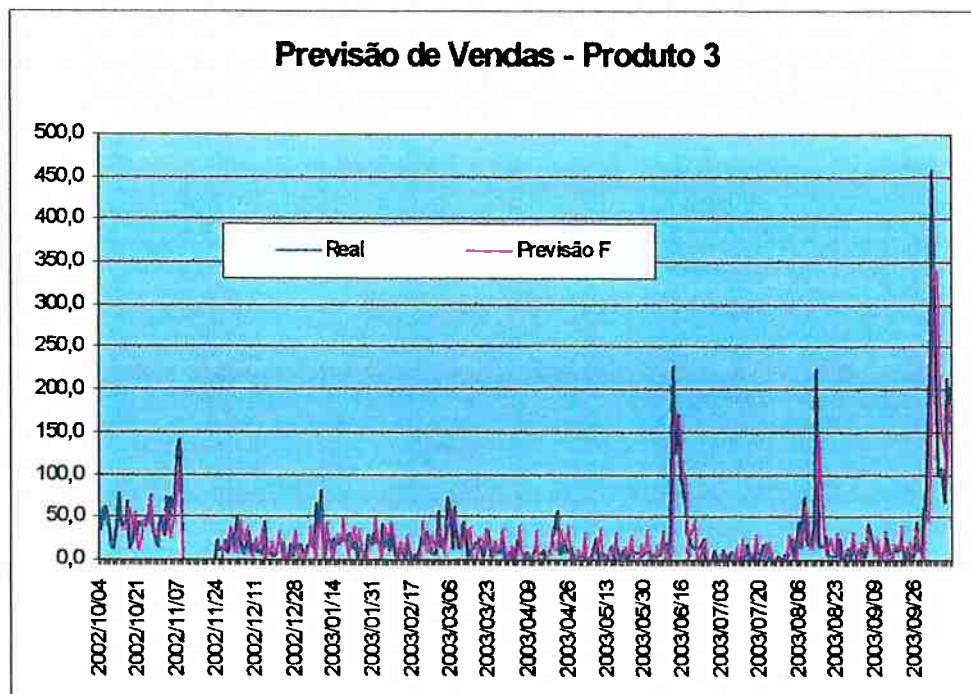


Figura 4.24 - Previsão de Vendas do Produto 3

E as medidas de erro calculadas são:

ME	MAE	MSE	MPE	MAPE
-2,665	18,003	1091,261	30,0%	123,9%

Tabela 4.12 - Erros de Previsão de Vendas do Produto 3

4.6.1.5.Previsão de Vendas do Produto 4

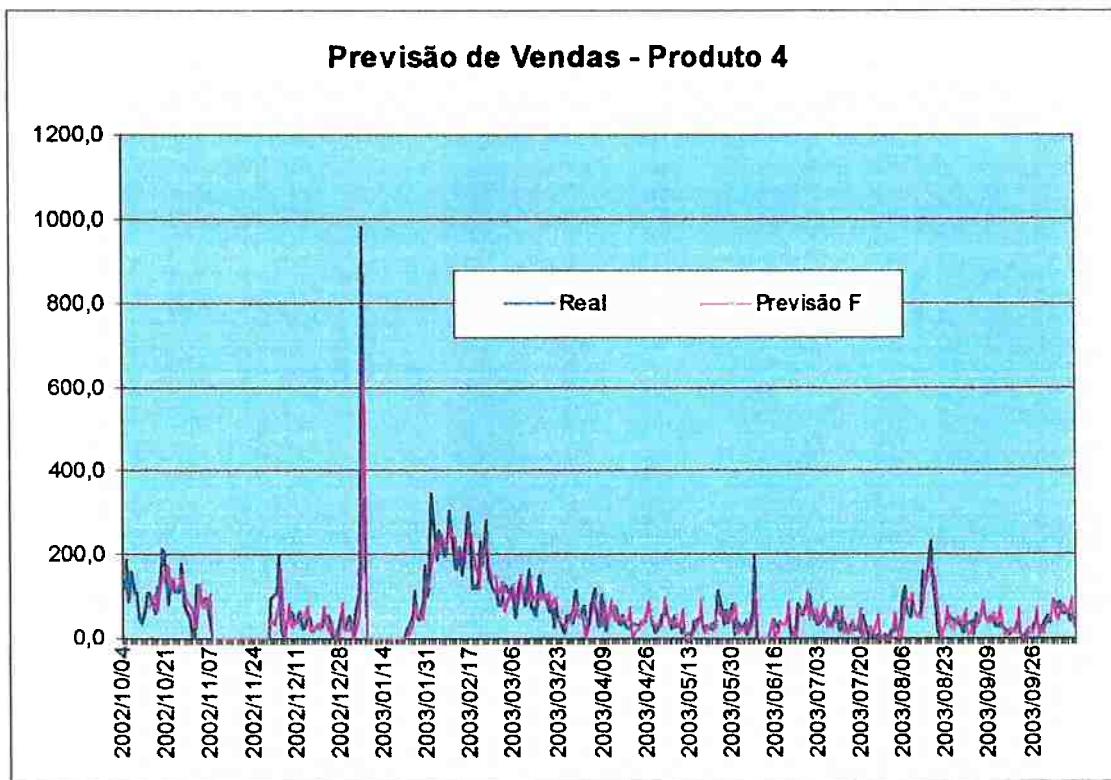


Figura 4.25 - Previsão de Vendas do Produto 4

ME	MAE	MSE	MPE	MAPE
-3,979	30,231	3656,164	18,6%	61,7%

Tabela 4.13 - Erros de Previsão de Vendas do Produto 4

4.6.2. Modelo Causal

Após projetar as vendas dos produtos e da categoria através de um modelo temporal, onde os dados utilizados para tal são estritamente os de vendas passadas, o objetivo agora é o de utilizar um modelo causal para tentar explicar as vendas. Nesses modelos são utilizados como fonte outras variáveis além das vendas passadas. Admite-se que a relação que essas variáveis e as vendas apresentaram no passado continue no futuro. A série de dados de vendas é a variável dependente, enquanto que os outros dados (por exemplo o preço) são as variáveis independentes ou explicativas. Uma equação de regressão múltipla tem a seguinte forma:

$$Y_i = b_0 + b_1 X_{1,i} + b_2 X_{2,i} + \dots + b_k X_{k,i} + e_i \quad (3.43)$$

onde:

Y é a variável dependente – as vendas

X_i são as variáveis independentes ou explicativas

b_i são os coeficientes das variáveis independentes

e_i é o erro residual

As variáveis independentes levantadas foram explicadas anteriormente. A definição de quais seriam essas variáveis potenciais foi feita com o auxílio de especialistas no assunto da empresa. Elas são as seguintes:

- Preço dos produtos da P&G e concorrentes
- Presença em encartes ou não, dos produtos da P&G e concorrentes
- Ocorrência de pacotes especiais ou não, dos produtos da P&G e concorrentes
- Dados de ruptura de gôndola, dos produtos da P&G
- Ocorrência de eventos na loja (cuja finalidade é alavancar as vendas), de produtos da P&G
- Dados relativos ao dia da semana e a semana no mês

No total foram criadas aproximadamente 50 séries de dados, que estão nos anexos do trabalho.

4.6.2.1. Seleção das Variáveis

Para selecionar as variáveis que farão parte do modelo, o ideal seria utilizar algum método como o *best subset regression*, como explicado no capítulo 3. Porém o software que foi utilizado para os cálculos, o *E-Views*, não possui esse recurso na versão utilizada. Em função disso a seleção das variáveis da equação de regressão foi feita manualmente, da seguinte forma:

1. Com a variável dependente (as vendas), eram inseridas as variáveis responsáveis pela sazonalidade (que indicam o dia da semana e a semana no mês)
2. Em seguida eram adicionadas as variáveis relacionadas ao próprio produto, como o seu preço, sua ruptura, presença em encarte, ocorrência de pacotes especiais ou de eventos na loja.
3. Depois eram inseridas as variáveis relativas aos produtos concorrentes.
4. Uma vez que um valor alto para o R^2 era alcançado, eram removidas as variáveis que apresentavam multicolinearidade ou um valor do t-estatístico baixo (geralmente menor do que 2).

Com um certo conhecimento do mercado dos produtos analisados, pode-se direcionar um pouco melhor a escolha das variáveis independentes. Sabendo-se os principais concorrentes do produto 1, por exemplo, a seleção das variáveis corretas é feita de forma mais rápida. O objetivo em todo esse processo é de sempre aumentar o coeficiente de determinação R^2 . Algumas variações das variáveis existentes podem ser criadas a fim de aperfeiçoar o modelo, como por exemplo um índice de preços “(preço do concorrente) / (preço do produto)”. Uma outra opção comumente utilizada em modelos é defasar uma série em x períodos. Assim uma variável “vendas(-1)” seria a própria série de dados de vendas porém defasada de um período. O intuito é o de explicar parte das vendas de hoje devido às vendas de ontem.

Um recurso muito útil em se tratando de regressão múltipla é a matriz de correlação. Selecionando-se algumas variáveis, é construída uma matriz onde constam todos os índices de correlação entre todas essas variáveis. Esses números podem fornecer “dicas” para o desenvolvimento do modelo: por exemplo, se determinada variável explicativa tiver um alto índice de correlação com as vendas, muito provavelmente ela fará parte do modelo. Outra utilidade dessa matriz é a de revelar a existência de multicolinearidade entre as variáveis selecionadas de um modelo. Se duas supostas variáveis explicativas possuírem um alto índice

de correlação ocorrerá um problema de multicolinearidade. Uma delas terá que ser removida ou deverá ser criada uma terceira variável, que será uma combinação das duas.

4.6.2.2. Desenvolvimento

Como a metodologia da análise de regressão múltipla é a mesma para qualquer produto escolhido (selecionar variáveis, verificar multicolinearidade, etc), foi decidido fazê-la para somente um dos produtos. Posteriormente o mesmo processo poderá ser aplicado para os outros três produtos da empresa. Para essa análise foi escolhido o produto 2.

Após serem selecionadas as variáveis do modelo e feita a verificação da existência de multicolinearidade através da matriz de correlação e consequente alteração no modelo em função disso, chegou-se à equação de regressão final para esse produto. Assim como foi feito com o modelo de projeção, neste modelo causal o período de 2 anos foi dividido em “período de inicialização” e “período de testes”. As variáveis do modelo final e seus respectivos coeficientes estimados pelo *E-Views*, todos baseados no período de inicialização, foram:

VARIÁVEL DEPENDENTE: VENDAS PRODUTO 2	
VARIÁVEIS INDEPENDENTES	COEFICIENTES
Constante C	-381,90
Vendas (-1)	0,32
Vendas (-7)	0,08
Quarta	31,48
Sexta	42,65
Sábado	189,75
Semana 1	41,47
Encarte Prod.2	82,64
Eventos Prod.2	235,31
(Preço Prod.5) / (Preço Prod.2)	262,48
Preço Prod.1	43,11
Ruptura Prod.1	66,06
Encarte Prod.8	-42,11

Tabela 4.14 – Coeficientes da Equação de Regressão do Produto 2

O modelo apresentou um $R^2 = 0,727$. O R^2 -ajustado, que leva em consideração os graus de liberdade “punindo” a medida do R^2 de acordo com a quantidade de variáveis explicativas presentes, foi de 0,718. O R^2 -ajustado será sempre menor ou igual ao R^2 . Como se pode observar, no modelo foi adicionado, como variáveis explicativas, as séries de dados de vendas defasadas de um e de sete períodos. Ou seja, é possível explicar parte das vendas de hoje

devido às vendas de ontem e de uma semana atrás. Constam também algumas variáveis que explicam a sazonalidade. Das variáveis diretamente relacionadas ao produto 2, o preço e a ruptura não apresentaram um t-estatístico suficiente para fazerem parte do modelo. Já as variáveis “Encarte” e “Eventos” podem explicar em parte as vendas desse produto, por isso estão presentes.

Analizando-se os coeficientes encontrados verifica-se a coerência entre eles. Por exemplo, para as duas variáveis relativas aos encartes: a “Encarte Prod.2” apresentou um coeficiente positivo (ou seja, quanto mais o produto 2 aparece no encarte, maior é a venda do produto 2), enquanto que a “Encarte Prod.8” apresentou um coeficiente negativo (ou seja, quanto mais o produto 8 – um concorrente – aparece no encarte, menor é a venda do produto 2). Outro ponto interessante é o fato de aparecer a variável “Ruptura Prod.1” entre as significativas. Isso significa que quando ocorre uma falta do produto 1 na prateleira, parte dos consumidores opta pela compra do produto 2, apesar deste não ser – teoricamente – um concorrente direto do primeiro.

O modelo apresentou para a estatística Durbin-Watson um valor de **D-W=1,84**. Isso indica que não há viés significativo nos erros (pois está próximo de 2). A matriz de correlação entre as variáveis do modelo está a seguir:

	Vendas Prod.2	Vendas (-1)	Vendas (-7)	Quarta	Sexta	Sábado	Semana 1
Vendas Prod.2	1	0,62	0,33	-0,08	-0,02	0,40	0,27
Vendas (-1)	0,62	1	0,15	-0,15	-0,09	-0,03	0,24
Vendas (-7)	0,33	0,15	1	-0,09	-0,02	0,41	-0,15
Quarta	-0,08	-0,15	-0,09	1	-0,17	-0,17	0,00
Sexta	-0,02	-0,09	-0,02	-0,17	1	-0,17	0,00
Sábado	0,40	-0,03	0,41	-0,17	-0,17	1	0,00
Semana 1	0,27	0,24	-0,15	0,00	0,00	0,00	1
Encarte Prod.2	0,56	0,52	0,17	0,00	0,00	-0,01	0,15
Encarte Prod.8	-0,07	-0,07	0,05	0,00	0,00	0,00	-0,14
(Preço Prod.5) / (Preço Prod.2)	0,20	0,12	0,10	0,05	-0,01	0,01	-0,01
Preço Prod.1	0,16	0,15	0,12	-0,02	0,02	0,00	-0,07
Ruptura Prod.1	0,05	0,05	0,00	-0,04	-0,04	0,00	-0,08
Eventos Prod.2	0,65	0,58	0,11	-0,03	0,00	0,10	0,28

	Encarte Prod.2	Encarte Prod.8	(Preço Prod.5) / (Preço Prod.2)	Preço Prod.1	Ruptura Prod.1	Eventos Prod.2
Vendas Prod.2	0,56	-0,07	0,20	0,16	0,05	0,65
Vendas (-1)	0,52	-0,07	0,12	0,15	0,05	0,58
Vendas (-7)	0,17	0,05	0,10	0,12	0,00	0,11
Quarta	0,00	0,00	0,05	-0,02	-0,04	-0,03
Sexta	0,00	0,00	-0,01	0,02	-0,04	0,00
Sábado	-0,01	0,00	0,01	0,00	0,00	0,10
Semana 1	0,15	-0,14	-0,01	-0,07	-0,08	0,28
Encarte Prod.2	1	0,10	0,27	0,15	0,13	0,47
Encarte Prod.8	0,10	1	0,09	0,13	-0,05	-0,06
(Preço Prod.5) / (Preço Prod.2)	0,27	0,09	1	0,00	-0,20	0,01
Preço Prod.1	0,15	0,13	0,00	1	-0,18	0,14
Ruptura Prod.1	0,13	-0,05	-0,20	-0,18	1	-0,05
Eventos Prod.2	0,47	-0,06	0,01	0,14	-0,05	1

Tabela 4.15 – Matriz de Correlação

O segundo quadro é continuação do primeiro. Como o índice de correlação entre uma variável x e y é o mesmo que de y e x , somente metade da matriz interessa, pois a outra metade é exatamente igual à primeira. A linha que separa as metades é a diagonal que possui os números 1. Ao analisar os números, verifica-se que as vendas do produto 2 possuem uma correlação mais forte com as vendas do dia anterior, com a presença em encartes e com a ocorrência de eventos na loja. Pelo fato das variáveis que apresentavam multicolinearidade terem sido removidas do modelo, não se observam mais indícios desse problema, ou seja, os índices de correlação entre elas são baixos.

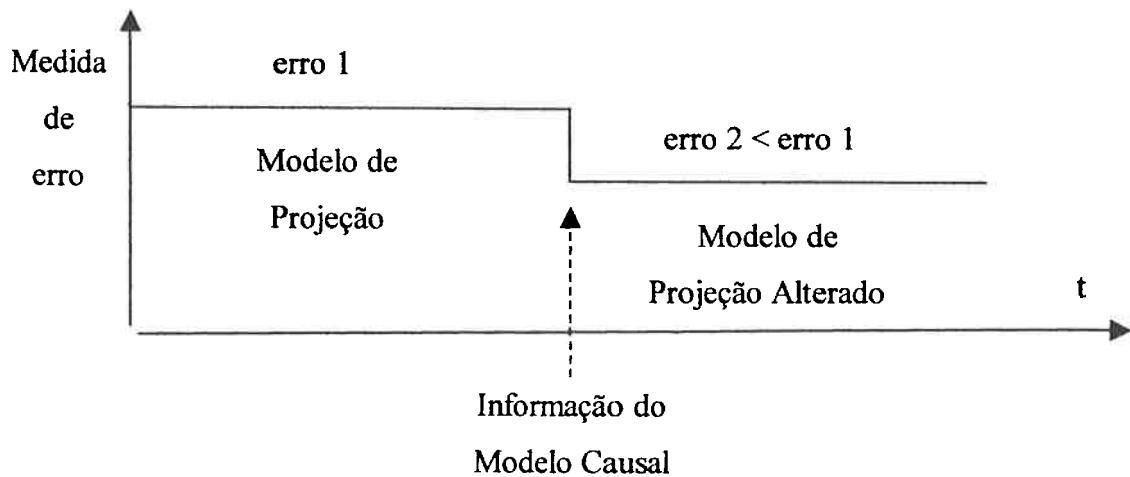
4.6.2.3. Análise de Mercado

Uma outra grande utilidade da análise de regressão múltipla, além da previsão de demanda, é a análise de mercado. Trabalhando-se com todas as variáveis explicativas, é possível tirar conclusões acerca do mercado do produto em questão. No exemplo estudado, foi possível descobrir quais eram os verdadeiros concorrentes do produto 2, e de que forma eles o afetavam.

Os produtos 2 e 8, que são de alta qualidade, aparecem no modelo com as variáveis explicativas “encarte”. Já os produtos 1 e 5, cuja qualidade é um pouco inferior à primeira dupla, aparecem no modelo com as variáveis explicativas “preço”. O variável “preço” do próprio produto 2 não obteve um t-estatístico suficiente para constar no modelo. A impressão que fica é a de que o consumidor realmente percebe essa diferença na qualidade entre esses produtos. Assim, entre dois produtos top de linha, o que pode fazer a diferença na hora da compra é a presença no encarte na loja. Já na comparação entre o produto 2 e os de qualidade um pouco inferior (1 e 5), um ponto decisivo está no preço: o consumidor se dispõe a levar o produto 2 caso a diferença no preço entre eles diminua, pois ele percebe a diferença na qualidade.

4.7. Testes e Resultados do Sistema de Predição de Vendas

Com a equação de regressão e os coeficientes estimados, o passo seguinte é o de tentar melhorar a previsão de vendas no modelo de projeção utilizando informações do modelo causal. A decisão é simples: se com uma informação do modelo causal for possível diminuir o erro na previsão do modelo de projeção, está aprovado. A idéia é representada no esquema a seguir:

**Figura 4.26 – Sistema de Previsão de Vendas**

O processo utilizado para se medir essa possível diminuição do erro de previsão é similar ao utilizado no modelo de projeção anteriormente estudado. O período de 2 anos foi dividido ao meio, sendo o primeiro ano denominado o “período de inicialização” e o segundo ano o “período de testes”, ou o “futuro adotado”, período no qual será utilizado como base para calcular os erros de previsão.

A cada previsão diária F foi adicionado ou subtraído um ajuste. Esse ajuste é calculado a partir das informações e dos coeficientes do modelo causal. Para o caso do produto 2 estudado, foi realizado o teste com as informações a respeito dos encartes que aparecem no modelo, ou seja, a presença em encartes do produto 2 e do produto 8 (concorrente). Os coeficientes estimados para essas variáveis são +82,64 e -42,11 respectivamente. O cálculo do ajuste é simples: se em um determinado dia o produto 2 sai no encarte mas o produto 8 não, o ajuste é dado por $[(1)x(82,64) + (0)x(-42,11)]$, sendo adicionado então o valor de 82,64 à previsão F daquele dia. Fazendo isso para todo o período de testes e recalculando a medida de erro MAPE, obtém-se:

MAPE original	64,2%
MAPE alterado	62,2%

Tabela 4.16 – Resultado do Sistema de Previsão de Vendas

Como se pode observar, apesar da diminuição na medida de erro MAPE não ter sido significativa, a proposta inicial de diminuição do erro de previsão através de informações do método causal aplicada ao método de projeção foi aprovada.

5. CONCLUSÕES

A previsão de demanda no varejo tem sido muito valorizada ultimamente devido a duas preocupações principais das empresas que querem se tornar competitivas: diminuir a quantidade de produtos em estoque e ao mesmo tempo evitar ao máximo perder vendas pela falta destes produtos. Uma má administração da primeira significa dinheiro parado; e a segunda é um “mal-duplo”: quando um consumidor vai até a prateleira e não encontra o produto desejado, muito provavelmente ele comprará um produto similar do concorrente. Somente um consumidor muito leal à marca não fará isso. Dessa forma, além de perder uma venda propriamente dita, foi dada a oportunidade do concorrente conseguir mais uma venda, perdendo assim duas vezes.

A previsão de demanda trata-se de um vasto campo de estudo onde são encontrados desde métodos simples como uma previsão pela média até métodos complexos que envolvem redes neurais ou sistemas econométricos. Em bibliotecas apropriadas, podem ser encontrados livros inteiramente dedicados a um método somente, como por exemplo o método de Box-Jenkins. Como bem frisado pelos autores dos livros estudados nesse trabalho, não é somente a precisão que deve ser levada em conta na seleção de um método. Mas sim deve ser analisado o fator custo/benefício, sendo o termo “custo” levado ao seu mais amplo escopo, que não somente o monetário. Um método que demande um enorme trabalho para gerar uma previsão, desde o levantamento dos dados até sua análise, dificilmente será utilizado continuamente. Outro ponto importante é a sua complexidade: deve ser comprehensível pelo seu gestor, caso contrário não inspirará confiança e cairá em desuso.

Em relação aos métodos utilizados nesse trabalho, foi verificado que:

- Os modelos de projeção são relativamente simples de serem compreendidos e usados, possuem uma grande praticidade e podem ser relativamente precisos. Porém trata-se de uma análise estritamente temporal, não importando as causas da demanda.
- Nos modelos explicativos ou causais, mais especificamente na regressão múltipla, ficou evidente a complexidade em todo o seu processo. Desde o levantamento de dados, o tratamento deles, o desenvolvimento e seleção das variáveis da equação até chegar na análise dos resultados, cada passo significa horas de estudo até se obter um bom resultado. Essa dificuldade se intensificou nesse trabalho pelo fato do mercado do produto estudado ter muitos produtos e ser influenciado por muitas variáveis, que não somente o preço, sendo assim difícil de medir a influência de uma delas especificamente. Caso a regressão múltipla seja utilizada para fazer previsões,

adiciona-se a tudo isso outra dificuldade: cada uma das variáveis independentes que farão parte do modelo deverá ser prevista a fim de se obter a previsão final da variável dependente. Isso poderia frustrar e inibir um gestor que devesse fazer previsões baseadas nesse modelo. Por outro lado, pelo fato da regressão múltipla trabalhar com as causas da demanda, esse método torna-se muito útil se utilizado como uma ferramenta de análise de mercado. Por exemplo, poderão ser levantadas hipóteses a respeito da concorrência dos produtos da empresa, dos dias e períodos mais favoráveis às vendas, das formas mais eficientes de vender, seja através de preços, encartes ou eventos na loja, etc. e através dos resultados obtidos por esses modelos poderão ser tiradas conclusões para melhorar as vendas futuras. Esse foi em parte o conceito utilizado para se fazer o sistema de predição do presente trabalho: foi utilizada uma informação extraída do modelo causal com a finalidade de melhorar um processo formal, que era o caso da previsão de demanda pelo método de projeção. Outro ponto destacável das regressões é a possibilidade de agregar informações qualitativas a um modelo numérico.

Em relação aos resultados finais, eles não foram tão significativos talvez devido à falta de tempo, que certamente dificultou uma maior exploração das análises dos dados. Ainda quanto aos resultados, convém frisar que, apesar dos dados terem sido modificados por meio de um coeficiente, os resultados obtidos não mudaram em relação àqueles que tiveram como base os dados reais.

Um ponto que certamente merece ser destacado foi a oportunidade de aprendizado que o trabalho proporcionou, por dois lados. Primeiro, pelo conhecimento adquirido através do estudo dos diversos métodos de previsão e das etapas de um processo de previsão, todas vivenciadas na prática. Como dito anteriormente, atualmente essa é uma competência na qual toda empresa que almeje ser líder de mercado deve ser bem desenvolvida. E em segundo lugar, o aprendizado relativo ao conflito teoria-prática inerente a um trabalho dessa característica. Um dos primeiros pensamentos que vêm à cabeça de um estudante quando se inicia um estudo em uma empresa é a de que os problemas da realidade poderão ser facilmente resolvidos utilizando-se os métodos matemáticos aprendidos em sala de aula. Porém, logo que se inicia o trabalho prático, percebe-se que a situação real nem sempre está próxima da imaginada, o que dificulta e pode causar uma certa frustração inicial. Certamente,

oportunidades de melhora nos processos utilizados sempre existem quando bem procuradas. Uma questão-chave nesse ponto é a do custo/benefício: em um ambiente empresarial, a carga de trabalho e os estudos demandados para aplicar ou melhorar determinado processo valerá a pena? Cabe ao indivíduo utilizar o bom senso para tal decisão.

6. BIBLIOGRAFIA

- MAKRIDAKIS, Spyros; WHEELWRIGHT, Steven; HYNDMAN, Rob. **Forecasting: Methods and Applications**. 3rd edition. New York, John Wiley & Sons, 1998.
- O'DONOVAN, Thomas M. **Short Term Forecasting – An Introduction to the Box-Jenkins Approach**. 1st edition. New York, John Wiley & Sons, 1983.
- JARRETT, Jeffrey. **Business Forecasting Methods**. 1st edition. Oxford, Basil Blackwell Ltd, 1987.
- FERRARI, Adriana. **Previsão de Vendas para um Bem de Consumo**. Trabalho de Formatura. São Paulo, EPUSP, 1996.
- HANKE, John E.; REITSCH, Arthur G. **Business Forecasting**. 6th edition. Upper Saddle River, Prentice-Hall, 1998.
- GRADIA, Fernando Luiz de Freitas. **Previsão de Demanda para uma Indústria Alimentícia**. Trabalho de Formatura. São Paulo, EPUSP, 1991.
- SANTORO, Miguel Cezar. **Planejamento, Programação e Controle da Produção**. Apostila do curso de PPCP da Escola Politécnica da USP. São Paulo, 2003.
- PROTO, Luiz Otavio Zavalloni. **Desenvolvimento de um Modelo de Previsão de Demanda de Médio e Longo Prazo para Empresa do Setor Cimenteiro**. Trabalho de Formatura. São Paulo, EPUSP, 2002.

7. ANEXOS

PRODUTO 2	VENDAS2	EVENTOS2	RUPTURA2	PRECO2	ENCARTE2	PRODUTO 3	VENDAS3	EVENTOS3	RUPTURA3	PRECO3	ENCARTE3	PRODUTO 4	VENDAS4	EVENTOS4	RUPTURA4
PAC.ESP2	PAC.ESP3	PAC.ESP4													
113,5	0,0	0,0	10,62	0,0	0,0	6,8	0,0	0,0	10,10	0,0	10,10	0,0	29,5	0,0	0,0
120,3	0,0	0,0	10,62	0,0	0,0	11,4	0,0	0,0	10,10	0,0	10,10	0,0	22,7	0,0	0,0
10,1	0,0	0,0	10,61	0,0	0,0	25,0	0,0	0,0	8,15	0,0	8,15	0,0	34,1	0,0	0,0
55,6	0,0	0,0	10,62	0,0	0,0	36,3	0,0	0,0	8,15	0,0	8,15	0,0	38,6	0,0	0,0
228,1	0,0	0,0	10,62	0,0	0,0	56,8	0,0	0,0	8,15	0,0	8,15	0,0	31,8	0,0	0,0
172,5	0,0	0,0	10,61	0,0	0,0	13,6	0,0	0,0	9,78	0,0	9,78	0,0	40,9	0,0	0,0
137,3	0,0	0,0	10,62	0,0	0,0	6,8	0,0	0,0	8,75	0,0	8,75	0,0	38,6	0,0	0,0
151,0	0,0	0,0	10,62	0,0	0,0	25,0	0,0	0,0	7,53	0,0	7,53	0,0	43,1	0,0	0,0
151,0	0,0	0,0	10,62	0,0	0,0	9,1	0,0	0,0	8,59	0,0	8,59	0,0	88,5	0,0	0,0
306,5	0,0	0,0	10,62	0,0	0,0	15,9	0,0	0,0	9,09	0,0	9,09	0,0	45,4	0,0	0,0
213,4	0,0	0,0	10,55	0,0	0,0	6,8	0,0	0,0	8,66	0,0	8,66	0,0	34,1	0,0	0,0
145,3	0,0	0,0	10,62	0,0	0,0	9,43	0,0	0,0	9,43	0,0	9,43	0,0	13,6	0,0	0,0
153,2	0,0	0,0	10,44	0,0	0,0	1,0	0,0	0,0	9,43	0,0	9,43	0,0	36,3	0,0	0,0
234,9	0,0	0,0	10,47	1	0,0	4,5	0,0	0,0	10,10	0,0	10,10	0,0	56,8	0,0	0,0
413,1	0,0	0,0	10,42	1	0,0	4,5	0,0	0,0	10,10	0,0	10,10	0,0	54,5	0,0	0,0
311,0	0,0	0,0	10,39	1	0,0	6,8	0,0	0,0	10,10	0,0	10,10	0,0	72,6	0,0	0,0
280,3	0,0	0,0	10,06	1	0,0	4,5	0,0	0,0	10,10	0,0	10,10	0,0	47,7	0,0	0,0
240,6	0,0	0,0	10,49	1	0,0	2,3	0,0	0,0	10,10	0,0	10,10	0,0	25,0	0,0	0,0
149,8	0,0	0,0	10,52	1	0,0	0,0	0,0	0,0	1,0	0,0	10,10	0,0	52,2	0,0	0,0
186,1	0,0	0,0	10,55	1	0,0	6,8	0,0	0,0	10,10	0,0	10,10	0,0	27,2	0,0	0,0
144,1	0,0	0,0	10,58	0,0	0,0	15,9	0,0	0,0	9,52	0,0	9,52	0,0	36,3	0,0	0,0
160,0	0,0	0,0	10,62	0,0	0,0	25,0	0,0	0,0	9,00	0,0	9,00	0,0	13,6	0,0	0,0
93,1	0,0	0,0	10,59	0,0	0,0	2,3	0,0	0,0	10,10	0,0	10,10	0,0	29,5	0,0	0,0
214,5	0,0	0,0	10,60	0,0	0,0	6,8	0,0	0,0	10,10	0,0	10,10	0,0	9,1	0,0	0,0
173,7	0,0	0,0	10,26	0,0	0,0	6,8	0,0	0,0	10,10	0,0	10,10	0,0	13,6	0,0	0,0
181,6	0,0	0,0	10,62	0,0	0,0	6,8	0,0	0,0	10,10	0,0	10,10	0,0	9,1	0,0	0,0
164,6	0,0	0,0	10,62	0,0	0,0	2,3	0,0	0,0	10,10	0,0	10,10	0,0	40,9	0,0	0,0
148,7	0,0	0,0	10,62	0,0	0,0	9,1	0,0	0,0	10,10	0,0	10,10	0,0	43,1	0,0	0,0
161,2	0,0	0,0	10,40	0,0	0,0	15,9	0,0	0,0	10,10	0,0	10,10	0,0	52,2	0,0	0,0
166,8	0,0	0,0	10,28	0,0	0,0	2,3	0,0	0,0	10,10	0,0	10,10	0,0	49,9	0,0	0,0
263,3	0,0	0,0	10,28	0,0	0,0	11,4	0,0	0,0	9,74	0,0	9,74	0,0	25,0	0,0	0,0
115,8	0,0	0,0	10,28	0,0	0,0	0,0	0,0	0,0	1,0	0,0	9,74	0,0	31,8	0,0	0,0
136,2	0,0	0,0	10,28	0,0	0,0	11,4	0,0	0,0	9,74	0,0	9,74	0,0	22,7	0,0	0,0
97,6	0,0	0,0	10,28	0,0	0,0	15,9	0,0	0,0	9,74	0,0	9,74	0,0	25,0	0,0	0,0
131,7	0,0	0,0	9,85	0,0	0,0	2,3	0,0	0,0	9,74	0,0	9,74	0,0	36,3	0,0	0,0
161,2	0,0	0,0	10,28	0,0	0,0	6,8	0,0	0,0	9,74	0,0	9,74	0,0	29,5	0,0	0,0
133,9	0,0	0,0	10,28	0,0	0,0	18,2	0,0	0,0	9,74	0,0	9,74	0,0	115,8	0,0	0,0
238,4	0,0	0,0	10,28	0,0	0,0	11,4	0,0	0,0	9,74	0,0	9,74	0,0	70,4	0,0	0,0
149,8	0,0	0,0	9,95	0,0	0,0	9,1	0,0	0,0	9,74	0,0	9,74	0,0	31,8	0,0	0,0
131,7	0,0	0,0	9,94	0,0	0,0	9,1	0,0	0,0	9,99	0,0	9,99	0,0	68,1	0,0	0,0
120,3	0,0	0,0	10,28	0,0	0,0	9,1	0,0	0,0	8,24	0,0	8,24	0,0	36,3	0,0	0,0
122,6	0,0	0,0	10,28	0,0	0,0	4,5	0,0	0,0	9,24	0,0	9,24	0,0	84,0	0,0	0,0
124,9	0,0	0,0	10,28	0,0	0,0	6,8	0,0	0,0	9,74	0,0	9,74	0,0	9,1	0,0	0,0
186,1	0,0	0,0	10,28	0,0	0,0	11,4	0,0	0,0	9,74	0,0	9,74	0,0	47,7	0,0	0,0
313,3	0,0	0,0	10,28	0,0	0,0	9,1	0,0	0,0	11,4	0,0	11,4	0,0	38,6	0,0	0,0
292,8	0,0	0,0	10,28	0,0	0,0	4,5	0,0	0,0	9,60	0,0	9,60	0,0	22,7	0,0	0,0
242,9	0,0	0,0	9,77	0,0	0,0	6,8	0,0	0,0	9,60	0,0	9,60	0,0	38,6	0,0	0,0
220,2	0,0	0,0	10,28	0,0	0,0	4,5	0,0	0,0	11,4	0,0	11,4	0,0	20,4	0,0	0,0
635,6	1,0	0,0	9,90	1	0,0	4,5	0,0	0,0	9,40	1	11,4	0,0	0,0	0,0	0,0
730,9	1,0	0,0	10,28	1	0,0	11,4	1	1	0,0	0,0	0,0	0,0	0,0	0,0	0,0

PRODUTO 4		PRODUTO 5		PRODUTO 6		PRODUTO 7		PRODUTO 8		PRODUTO 9	
PRECO4	ENCARTE4	PRECO5	ENCARTE5	PRECO6	ENCARTE6	PRECO7	ENCARTE7	PRECO8	ENCARTE8	PRECO9	ENCARTE9
6.74	0	8.67	0	9.22	0	8.85	0	11.25	0	1	10.62
6.74	0	8.67	0	9.11	0	8.84	0	11.35	0	0	10.99
6.74	0	8.67	0	9.22	0	8.85	0	11.25	0	0	10.99
6.74	0	8.67	0	9.22	0	8.85	0	11.35	0	0	10.99
6.54	0	8.67	0	9.17	0	8.85	0	11.05	0	0	10.62
6.74	0	8.67	0	9.22	0	8.85	0	10.53	0	0	10.62
6.74	0	8.67	0	9.22	0	8.85	0	10.61	1	0	10.62
6.73	0	8.67	0	9.22	0	8.85	0	10.06	1	0	10.99
6.49	0	8.51	0	9.22	0	10.15	0	10.54	1	0	10.62
6.74	0	8.63	0	9.22	0	9.89	0	10.61	1	0	10.62
6.74	0	8.61	0	9.22	0	10.62	0	10.61	1	0	10.62
6.74	0	8.57	0	9.22	0	10.62	0	10.45	1	0	10.62
6.74	0	8.65	0	9.22	0	10.62	0	10.62	1	0	10.62
6.74	0	8.67	0	8.09	0	10.62	0	10.77	0	0	10.62
6.74	0	8.66	0	7.97	1	10.41	0	10.04	0	0	10.62
6.74	0	8.42	0	7.97	1	10.09	0	11.35	0	0	10.62
6.74	0	8.65	0	7.97	1	9.78	0	11.27	0	0	10.62
6.74	0	8.50	0	7.97	1	9.72	0	11.35	0	0	10.62
6.74	0	8.60	0	7.97	1	9.99	0	11.35	0	0	10.62
6.54	0	8.45	0	7.97	1	9.64	0	11.35	0	0	10.62
6.74	0	7.96	0	7.97	1	8.26	0	11.31	0	0	10.62
6.02	0	7.67	1	8.90	0	9.79	0	11.16	0	0	10.62
6.02	0	7.66	1	8.11	0	8.95	0	11.31	0	0	10.62
6.02	0	7.66	1	7.92	0	8.73	0	11.35	0	0	10.62
6.74	0	7.50	1	8.79	0	8.84	0	11.32	0	0	10.62
6.74	0	7.44	1	8.79	0	8.82	0	11.03	0	0	10.52
6.74	0	7.50	1	8.85	0	9.11	0	11.35	0	0	8.69
6.74	0	8.33	1	8.85	0	10.62	0	11.35	0	0	11.35
6.74	0	8.33	1	8.85	0	10.56	0	11.35	0	0	11.35
6.74	0	8.40	1	8.85	0	8.96	0	11.35	0	0	10.62
6.24	0	7.53	1	8.85	0	8.92	0	11.35	0	0	11.35
5.68	0	7.71	1	8.85	0	9.01	0	11.35	0	0	10.62
5.68	0	8.28	1	8.85	0	10.09	0	11.35	0	0	11.35
5.68	0	8.33	1	8.85	0	10.04	0	11.35	0	0	10.62
5.68	0	8.33	1	8.85	0	10.05	0	11.35	0	0	11.35
5.68	0	8.33	1	8.15	1	9.98	0	11.18	0	0	11.35
5.68	0	8.33	1	8.15	1	9.95	0	11.35	0	0	10.62
5.68	0	8.33	1	8.15	1	9.47	0	11.35	0	0	10.62
5.40	0	7.85	1	7.99	1	8.85	0	11.35	0	0	10.62
5.68	0	7.74	1	7.92	1	8.78	0	11.35	0	0	11.35
5.68	0	8.33	1	7.92	1	9.44	0	11.35	0	0	11.35
5.68	0	8.33	1	7.67	1	9.44	0	11.35	0	0	11.35
5.68	0	8.40	1	8.33	1	9.44	0	11.35	0	0	11.35
5.68	0	8.33	1	8.15	1	9.44	0	11.35	0	0	11.35
6.22	0	7.60	0	7.44	1	9.40	0	11.15	0	0	11.35
6.74	0	7.60	0	7.42	1	9.38	0	11.26	0	0	11.35
6.74	0	7.60	0	7.42	1	8.85	0	11.14	0	0	11.35
6.74	0	7.91	0	7.42	1	9.34	0	11.30	0	0	11.35
6.74	0	8.02	0	7.42	1	8.92	0	11.04	0	0	10.62
6.74	0	7.88	0	8.61	1	8.78	0	11.35	0	0	11.35
6.74	0	7.89	0	8.61	1	8.78	0	11.12	0	0	10.62
6.50	0	7.89	0	8.85	0	0	0	0	0	0	0