

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
CURSO DE MATEMÁTICA APLICADA E COMPUTACIONAL

Pedro da Silva Peixoto

O Uso de modelos econométricos em empresas
Trabalho de conclusão de curso

São Paulo
2005

Pedro da Silva Peixoto

O Uso de modelos econométricos em empresas
Trabalho de conclusão de curso

Trabalho de conclusão de curso apresentado ao Curso de Matemática Aplicada e Computacional da USP, como requisito para a obtenção do grau de BACHAREL em Matemática Aplicada e Computacional com habilitação em Estatística Econômica.

Orientador: Clélia Maria de Castro Tolo

Estatística - IME - USP

Co-orientador: Paulo Picchetti

Economia - FEA - USP

São Paulo

2005

À minha família.

Aos amigos, pelo apoio e companheirismo.

Resumo

Na realidade atual das empresas de grande porte, a competitividade gera uma necessidade por novas ferramentas que otimizem o tempo de tomada de decisão e minimizem o erro de produção. Neste sentido entendemos que modelos estatísticos têm grande potencial na empresa. Assim, este trabalho visa testar um conjunto de modelos que seja aplicável ao mundo empresarial, tendo como base modelos econométricos, em especial regressões sobre séries temporais. Realizamos uma revisão bibliográfica que irá servir de base nas construções dos modelos e propomos um processo prático para a construção de um modelo econométrico. Exemplificamos o processo de modelagem e utilizamos os softwares E-Views e Minitab para as análises estatísticas. Testamos modelos de regressão multivariados e auto-regressivos. Os modelos gerados podem sugerir à empresa melhores ajustes de preços, melhores investimentos e uma previsão de vendas futura mais precisa, resultando portanto em uma otimização das vendas.

Palavras-chaves: Regressões Lineares, Econometria, Previsão, Modelagem, Séries Temporais, Vendas...

Abstract

Now a days on multinacional or large companies, there is a high need of news tools that can optimize time to take decisions and reduce the error of production. So we understand that statistical models have a great potencial in industries. This work aims to test a group of models that are applicable to the business world, using as source econometric models, specially those involving regression of time series. We will review most of the bibliography so that it serves as base to the modeling construction and we propose a practical process for this construction. We exemplify the modeling process and we used the softwares E-Views and Minitab to do the statistical analysis. We try out models based of multi linear regression and auto regressive models. The models generated can suggest to the company better price adjustments, better investments and a more precise forecasting of the sales, resulting in a more optimized sales.

Keywords: Linear Regression, Econometrics, Modeling, Time Series, Forecasting, Sales...

Agradecimentos

À professora Clélia Maria de Castro Toloí pela orientação, ensinamentos e paciência.

Ao professor Paulo Picchetti pelas aulas de econometria e auxílio durante algumas etapas do trabalho.

À professora Sônia Regina Leite Garcia pela orientação pessoal e profissional durante todo o curso e durante a realização deste trabalho.

Aos professores do Departamento de Matemática Aplicada, de Estatística e Economia pelos seus ensinamentos, e aos funcionários do curso, que durante esses anos, contribuíram de algum modo para o nosso enriquecimento pessoal e profissional.

Ao gestor Artur Ferreira Campos Filho, gerente de Suporte a Vendas da Johnson & Johnson Comércio e Distr. Ltda, por me iniciar e ensinar sobre mundo empresarial, econômico e financeiro.

Aos companheiros de trabalho da Johnson & Johnson, profissionais que depositaram em mim conhecimento, incentivo e dedicação.

À todos os meus parentes, pelo encorajamento e apoio. Em especial minha mãe que acompanhou de perto cada etapa do trabalho.

Aos colegas de faculdade, que compartilharam conhecimento e enriqueceram minha formação como profissional e pessoa.

Aos meus amigos pessoais que me deram força e companheirismo.

*“A mente que se abre a uma nova idéia
jamais voltará ao seu tamanho original”.*

Albert Einstein

Sumário

Lista de Figuras	9
Lista de Tabelas	12
1 Introdução	13
1.1 Caracterização do problema	13
1.2 Cenário considerado para o estudo	13
2 Objetivo	15
2.1 Objetivo Geral	15
2.2 Objetivos Específicos	15
3 Revisões Bibliográficas	16
3.1 Metodologia da Econometria	16
3.1.1 Abordagem Clássica	16
3.1.2 Abordagem Inglesa	17
3.2 Definições Estatísticas	19
3.2.1 Diagrama de Dispersão	19
3.2.2 Média amostral	20
3.2.3 Variância Amostral	20
3.2.4 Covariância Amostral	20
3.2.5 Coeficiente de Correlação Amostral	21
3.3 Modelo de Regressão Linear Múltipla	22
3.3.1 O Modelo estatístico de uma regressão linear múltipla	22
3.3.2 Estimador de parâmetros por Mínimos Quadrados Ordinários (MQO)	23

3.3.3	Coeficiente de Determinação R^2	24
3.3.4	Matriz de variância-covariância de $\hat{\beta}$	24
3.3.5	Inferência Estatística sobre os Estimadores	25
3.4	Análise de Autocorrelação	25
3.5	Análise de resíduos	27
3.6	Séries Temporais	29
3.6.1	Estimativa por Mínimos Quadrados	29
3.6.2	Processo Estocástico Estacionário	31
3.6.3	Processo de Média Móvel	31
3.6.4	Processo Autoregressivo	32
3.6.5	Modelo ARMA	32
3.6.6	Correlogramas	32
3.6.7	Metodologia de Box, Jenkins e Reinsel (1994)	36
4	Ferramenta Computacional	37
4.1	E-Views	37
4.1.1	Carga dos Dados	37
4.1.2	Fazendo Regressões	38
4.2	Minitab	41
4.3	Discussão: Minitab vs E-Views	43
5	Caracterização dos Dados	45
5.1	Coleta e Interpretação dos Dados	45
5.2	Dados	47
6	Modelos Propostos	56
6.1	Modelo de Variáveis Explicativas	57
6.2	Modelo Misto	70

6.3	Modelos ARMA	82
6.4	Melhor Modelo	86
7	Discussão e Recomendações	88
7.1	Aplicabilidade	88
7.2	Melhorias	89
8	Conclusões	90
9	Anexo: Dados Brutos	92
	Referências Bibliográficas	94

Lista de Figuras

1.1	Exemplo de Indicadores que podem afetar a Demanda de um produto. . .	14
3.1	Exemplo de Diagrama de Dispersão. Fonte: Microsoft Excel	19
3.2	Exemplos de Diagramas de Dispersão de acordo com o coeficiente de correlação, neste exemplo $\hat{\rho} = r$	21
3.3	Exemplo de Padrão de Autocorrelação Positivo, padrão sistemático nas perturbações. Fonte: Microsoft Excel.	26
3.4	Exemplo de ausência de Autocorrelação, erros aparentemente aleatórios.Fonte: Microsoft Excel.	27
3.5	Exemplo de evolução dos resíduos. Fonte: Minitab.	28
3.6	Exemplo de análise de normalidade dos resíduos. Fonte: Minitab.	29
3.7	Exemplo de Correlograma. Fonte: E-Views	34
4.1	Exemplo de arquivo para carga de dados no E-Views. Fonte: E-Views. . .	38
4.2	Exemplo de carga de dados no E-Views. Fonte: E-Views.	39
4.3	Exemplo referente a função de regressão do E-Views. Fonte: E-Views. . . .	40
4.4	Exemplo de resultados de uma regressão no E-Views. Fonte: E-Views. . . .	41
4.5	Exemplo de tela do Minitab. Fonte: Minitab.	42
4.6	Exemplo de Resultados do Minitab. Fonte: Minitab.	43
4.7	Comparativo entre os vícios por Softwares [5]	44
5.1	Evolução da Demanda. Fonte: Microsoft Excel.	50
5.2	Evolução da variável Índice de Preço e sua correlação com a variável Demanda. Fonte: Microsoft Excel.	50
5.3	Evolução da variável Negociações e Promoções e sua correlação com a variável Demanda. Fonte: Microsoft Excel.	51

5.4	Evolução da variável Inovação e sua correlação com a variável Demanda. Fonte: Microsoft Excel.	51
5.5	Evolução da variável Verbas Táticas e sua correlação com a variável Demanda. Fonte: Microsoft Excel.	52
5.6	Evolução da variável Distribuição Numérica e sua correlação com a variável Demanda. Fonte: Microsoft Excel.	52
5.7	Evolução da variável Preço Líquido e sua correlação com a variável Demanda. Fonte: Microsoft Excel.	53
5.8	Evolução da variável Despesas Publicitárias e sua correlação com a variável Demanda. Fonte: Microsoft Excel.	53
5.9	Análise da variável Log(Despesas Publicitárias) e sua correlação com a variável Demanda. Fonte: Microsoft Excel.	55
6.1	Fluxograma usado na montagem dos modelos.	57
6.2	Detalhes do Modelo 16 somente com variáveis independentes explicativas. Fonte: E-Views.	64
6.3	Análise de resíduos e Ex-Post do modelo 16 somente com variáveis explicativas. Fonte: E-Views.	65
6.4	Teste de Normalidade dos resíduos do Modelo 16 apenas com variáveis explicativas. Fonte: E-Views.	65
6.5	Correlograma dos resíduos do Modelo 16 apenas com variáveis explicativas. Fonte: E-Views.	66
6.6	Modelo de Variáveis Explicativas: Modelo 16. Fonte: Minitab	69
6.7	Modelo de Variáveis Explicativas: Resíduos do Modelo 16. Fonte: Minitab	70
6.8	Detalhes do Modelo Misto 8. Fonte: E-Views.	77
6.9	Detalhes dos resíduos do Modelo Misto 8. Fonte: E-Views.	78
6.10	Correlograma dos resíduos do Modelo Misto 8. Fonte: E-Views.	79
6.11	Teste de normalidade dos resíduos do Modelo Misto 8. Fonte: E-Views. . .	79
6.12	Correlograma da Demanda. Fonte: E-Views.	82

6.13	Modelo AR(6) completo. Fonte: E-Views.	83
6.14	Detalhes do Modelo AR(6) incompleto. Fonte: E-Views.	84
6.15	Correlograma dos resíduos do Modelo AR(6) incompleto. Fonte: E-Views.	85
6.16	Detalhes dos resíduos do Modelo AR(6) incompleto. Fonte: E-Views. . . .	86
6.17	Teste de normalidade dos resíduos do Modelo AR(6) incompleto. Fonte: E-Views.	87

Lista de Tabelas

3.1	Padrões teóricos de FAC e FACP. Fonte: Gujarati 2000 [6].	34
5.1	Detalhes sobre as variáveis.	47
5.2	Descrição das variáveis.	48
5.3	Análises das variáveis.	54
6.1	Modelos Testados somente com variáveis explicativas 1	58
6.2	Modelos Testados somente com variáveis explicativas 2	59
6.3	Modelos Testados somente com variáveis explicativas 3	60
6.4	Modelos Testados somente com variáveis explicativas 4	61
6.5	Modelos Testados somente com variáveis explicativas 5	62
6.6	Modelos Testados somente com variáveis explicativas 6	63
6.7	Modelos Mistos testados 1	71
6.8	Modelos Mistos testados 2	72
6.9	Modelos Mistos testados 3	73
6.10	Modelos Mistos testados 4	74
6.11	Modelos Mistos testados 5	75
6.12	Modelos Mistos testados 6	76
9.1	Dados Brutos: Jan02 a Jun03	92
9.2	Dados Brutos: Jul03 a Jul05	93

1 Introdução

Nesta introdução vamos descrever o problema e situar seu contexto.

1.1 Caracterização do problema

Com o avanço da globalização e aumento na competitividade entre as empresas há uma demanda por inovações a fim de otimizar custos de produção e agilizar os processos empresariais. As empresas de grande porte possuem em geral um grande número de informações armazenadas em enormes bancos de dados. Geralmente tais informações são usadas em análises gerenciais, planejamento de verbas, estabelecimento de metas futuras e previsões de custo e demanda, porém são usadas basicamente análises gráficas simples, tabelas com dados sumarizados, e estatísticas simples como média e tendência. Com isso as análises para realização de planejamento e previsões passam a ser principalmente baseadas na intuição e experiência do gerente do negócio observando os dados que lhe são apresentados, o que pode levar a um alto grau de incerteza no acerto da previsão e horas de reunião. Evidentemente esse planejamento é bem fundamentado em teorias administrativas e usam algumas ferramentas como auxílio, porém, levando em conta que hoje em dia está cada vez mais comum o uso de banco de dados em empresas, observamos que há um grande leque de ferramentas estatísticas que podem tornar as decisões mais rápidas e precisas do ponto de vista de previsão, complementando o planejamento.

1.2 Cenário considerado para o estudo

Olhando para esse cenário observa-se que ferramentas estatísticas têm grande potencial em empresas. Modelos estatísticos podem ser incorporados no dia a dia gerencial e tornar as decisões baseadas em fatos, com uma margem de erro controlável e com maiores possibilidades de análises destes fatos. Neste estudo vamos focar em análises estatísticas ligadas a Regressões Lineares e também realizar algumas análises utilizando a metodologia de Séries Temporais. Neste tipo de modelo é possível tomar como referência um

indicador, foco das análises, e entender seu comportamento frente a outros indicadores. Neste estudo vamos focar em análises sobre a Demanda gerada por um determinado produto, assim é possível chegarmos em uma equação que represente o comportamento da demanda com base em outros fatores ou com base no seu próprio passado. Observe a figura 1.1 que mostra justamente alguns indicadores que podem afetar a demanda de um produto, destacando fatores internos, aqueles que a própria empresa possui controle que estão no sistema interno de dados da empresa, e externos, aqueles que são controlados pelo mercado como um todo e estão em sistemas externos de dados fora da empresa, conforme descrito por Parente (2003)[11].

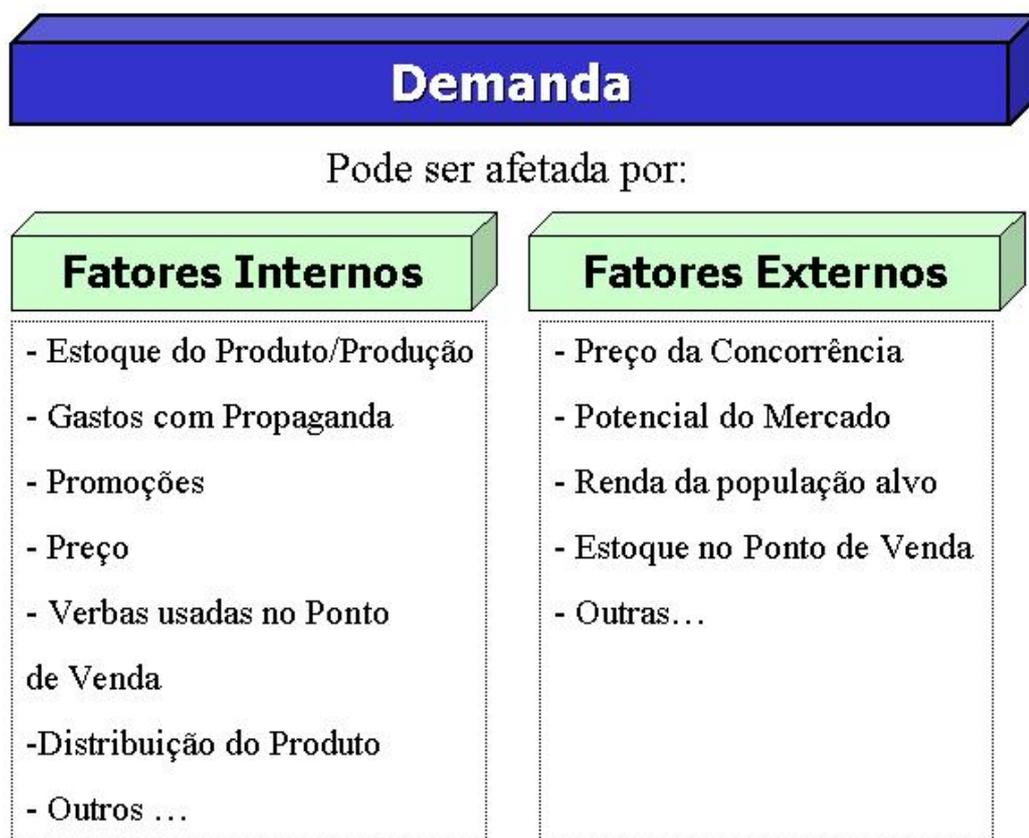


Figura 1.1: Exemplo de Indicadores que podem afetar a Demanda de um produto.

Evidentemente tanto os fatores como o foco a ser analisado podem e irão divergir conforme a empresa e o interesse de seu estudo.

2 Objetivo

2.1 Objetivo Geral

Mostrar a viabilidade prática e teórica de um modelo de previsão de volume de demanda, baseado em técnicas Econométricas, em análises gerenciais de empresas.

2.2 Objetivos Específicos

1. Oferecer uma alternativa inovadora de controle administrativo para planejamento de produção e elaboração de estratégias em empresas através de modelos econométricos.
2. Realizar uma revisão bibliográfica referente a técnicas de modelagem em econometria.
3. Descrever uma metodologia prática para criação de modelos baseados em Regressões Lineares Múltiplas e de Séries Temporais.
4. Aplicar sobre um conjunto de dados coletados, definidos como exemplo para este trabalho, modelos de regressão linear em conjunto com modelos de séries temporais, para ilustrar e assim obter uma melhor compreensão do efeito de variáveis sobre a demanda de um produto e poder realizar previsões desta demanda.
5. Criar um trabalho que vá servir de apoio a estatísticos que desejam implementar modelos econométricos em empresas.

3 Revisões Bibliográficas

Neste capítulo fornecemos o conteúdo teórico necessário para a construção dos modelos.

3.1 Metodologia da Econometria

Conforme descrito por Barossi e Braga (2000) [1], a aplicação de métodos matemáticos e estatísticos à análise de conjuntos de dados econômicos, com o objetivo de prover suporte empírico às teorias econômicas, constitui-se no ponto fundamental da Econometria. Na verdade, podemos definir a Econometria como um método de análise econômica que agrega a Estatística, a Matemática e a Teoria Econômica. Obviamente, não se trata da caracterização de uma ciência adicional, mas de uma forma científica de traduzir o modelo teórico para uma formulação empiricamente testável. Nesse sentido, discutimos, ao longo desta seção, alguns aspectos relacionados à modelagem econométrica. Partimos da abordagem tradicional de livro-texto, cujo método consiste em tomar um modelo econométrico simples, capaz de admitir a introdução de maiores generalizações, desenvolvido principalmente por Koopmans (1957)[9] que procura, como mencionado acima, testar um modelo previamente fornecido pela teoria. Em seguida, discutimos o paradigma alternativo apresentado por Hendry (1987)[7] que se constitui exatamente na atitude inversa à do primeiro método: partindo-se dos dados existentes, tenta-se montar um modelo o mais adequado possível à "história" contada pelos dados.

3.1.1 Abordagem Clássica

Um modelo nada mais é do que uma representação simplificada de um processo do mundo real. De acordo com Koopmans (1957)[9], um modelo representativo deve ser o mais simples possível e obedecer à regra da parcimônia no que se refere ao espectro de variáveis relevantes para a explicação do fenômeno a que se propõe. Dessa forma, um modelo econômico constitui-se em um conjunto de hipóteses que retratam de forma aproximada o comportamento de uma economia. Nos anos 50 e 60, a preocupação latente de um econometrista encerrava-se na construção de modelos econômicos completos que contivessem implicações

econométricas testáveis.

Nesse sentido, a preocupação de uma geração de econométricos esteve centrada na formulação de modelos econômicos cuja representação econométrica produzisse hipóteses testáveis relevantes à ciência econômica, principalmente quanto à utilização dos resultados na previsão e avaliação dos resultados de políticas econômicas.

Um modelo econométrico, dentro desse arcabouço, é descrito por um conjunto de equações comportamentais derivadas do modelo econômico, as quais envolvem variáveis observáveis e um termo aleatório ou errático, que contém todos os fatores que não foram incorporados ao modelo em análise. Além disso, contém afirmações sobre a existência de erros de observação em variáveis do modelo e sobre a especificação da distribuição de probabilidades do termo aleatório. O objetivo dessa formulação é prover uma forma representativa passível de teste empírico, por meio de estimação, teste e checagem do diagnóstico produzido.

Sumarizando, podemos caracterizar a metodologia econométrica até aqui discutida como um método de análise que tem como ponto de partida a teoria econômica. Com base nisso, o pesquisador é capaz de construir um modelo teórico que encerre as principais conclusões teóricas. Em seguida, ao caracterizarmos a distribuição de probabilidades do termo aleatório, produzimos um modelo econométrico que está pronto para ser estimado e testado. A etapa seguinte envolve a utilização de métodos estatísticos de estimação do modelo econométrico, além do teste das implicações produzidas por ele. O conjunto de dados empíricos observáveis, assim como a inferência estatística, é o suporte dessa etapa. Finalmente, caso o modelo econométrico não seja passível de rejeição, pode ser utilizado na previsão do comportamento de variáveis econômicas ou para a elaboração de cenários, quando da formulação de políticas.

3.1.2 Abordagem Inglesa

A metodologia econométrica da escola inglesa é apresentada por Hendry (1987)[7] e Spanos (1989) [14] como uma crítica à modelagem clássica. Entre as várias críticas imputadas a esta, destacamos três que se auto-explicam em relevância. Primeiro, existe uma diferença entre o modelo econométrico proposto e o processo gerador dos dados produzidos pela realidade. Isto é, a associação entre uma variável teórica e a sua representação estatística não é nada mais do que uma forma de ajustamento do modelo aos dados. Segundo,

Hendry (1987)[7] admite que as informações contidas nos bancos de dados podem ir além do previsto por um arcabouço teórico qualquer. Logo, é criticável a posição clássica de tomar como ponto de partida para a modelagem econométrica dado enfoque teórico. Na verdade, é imprescindível deixar espaço para que informações adicionais contidas nos dados aflorem e permitam um espaço científico mais amplo ao pesquisador.

Isso não implica abandonar as técnicas utilizadas pela Econometria tradicional, notadamente aquelas apresentadas na seção de Modelo de regressão linear. A diferença mais significativa é a maneira como esse modelo (de regressão) é inicialmente formulado. Ou este é advindo da teoria (abordagem tradicional), ou é o mais genérico possível, para daí, extraíndo os "excessos" (via análise estatística), chegarmos a um modelo "enxuto" representativo da realidade.

Finalmente, não podemos tomar a tradução imediata de um modelo econômico em modelo econométrico como um fato sem contestações. Quando introduzimos um termo aleatório no modelo econômico e afirmamos ser ele então um modelo econométrico, estamos qualificando um modelo estatístico cuja base de dados conveniente é experimental. Essa afirmação, segundo Hendry (1987) e Spanos (1989), não é tão óbvia assim, porque não há correspondência entre um modelo estatístico e os dados observáveis que escolhemos para representar as variáveis teóricas nele contidas. Além disso, a identificação entre dados experimentais e aqueles gerados pela natureza não é evidente.

De forma resumida, a modelagem econométrica na escola inglesa deve ser construída com base em um modelo estatístico cuja especificação seja o mais generalizada possível, retratando um mecanismo que mimetize as informações contidas no conjunto de dados, tendo em vista os modelos estimáveis. Podemos apresentar um sumário dessa metodologia da seguinte forma: a teoria econômica, como arcabouço lógico de idéias, não pode estar totalmente descolada do processo gerador que caracteriza o comportamento dos dados na natureza. Com base nisso e na interação entre esse processo e o conjunto de dados coletados, construímos um modelo econômico estimável. Esse modelo é produto da construção de um suporte estatístico generalizado, ou seja, o modelo estatístico proposto não admite a simplicidade como ponto de partida, e sim a generalização.

Portanto, existe inter-relação clara entre os vários estágios decisivos do método, ao mesmo tempo em que se manifesta uma preocupação em seguir um processo analítico que parte da generalização e chega à simplificação. As etapas de estimação e previsão são

os últimos estágios do método, como não poderia deixar de acontecer. Não poderíamos deixar de salientar que essa metodologia é também passível de críticas. Com o intuito de ilustrar essa afirmação, podemos considerar o fato de que, partindo de uma generalização em modelos econométricos, é possível que se deixe de lado a simplicidade e a parcimônia que tanto interessa ao pesquisador em ciência econômica. Nesse caso, mesmo sabendo que a construção de um modelo de demanda por um bem qualquer não seja uma tarefa simples, a generalização admitida pela escola inglesa no cumprimento dos requisitos que lhe são relevantes pode ofuscar, por exemplo, o papel do mecanismo de preços nesse sistema. Precisando ainda mais, ao seguirmos as normas do método, podemos contemplar o modelo com uma qualidade estatística esplêndida, em detrimento de um significado econômico, cuja base é a teoria.

3.2 Definições Estatísticas

3.2.1 Diagrama de Dispersão

O diagrama de dispersão é um gráfico onde pontos no espaço cartesiano XY são usados para representar simultaneamente os valores de duas variáveis quantitativas medidas em cada elemento do conjunto de dados. O diagrama de dispersão é usado principalmente para visualizar a relação/associação entre duas variáveis [12].

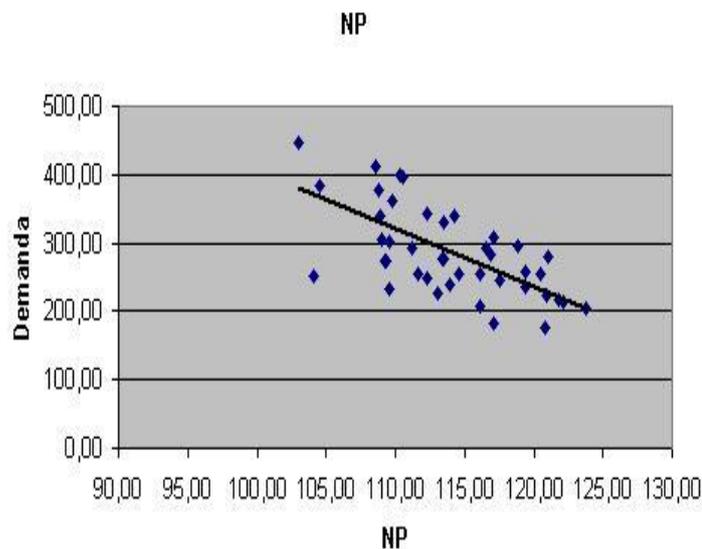


Figura 3.1: Exemplo de Diagrama de Dispersão. Fonte: Microsoft Excel

3.2.2 Média amostral

Seja X uma variável aleatória com média μ ($E(X) = \mu$). Supondo uma amostra de X , assumindo valores (x_1, x_2, \dots, x_n) , definimos a média amostral com a seguinte estatística:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (3.1)$$

Podemos provar que \bar{X} é um estimador não viesado de μ , e que portanto a esperança ou valor esperado de \bar{X} , indicado por $E(\bar{X})$, é igual a sua média populacional, μ .

$$E(\bar{X}) = \mu \quad (3.2)$$

3.2.3 Variância Amostral

Seja X um variável aleatória, com média μ e variância σ^2 , $E(X) = \mu$ e $\text{VAR}(X) = \sigma^2$. Supondo uma amostra de X assumindo valores (x_1, x_2, \dots, x_n) , a distribuição, ou dispersão, dos valores da amostra em torno da média amostral pode ser medida pela variância amostral, que podemos definir usando a estatística a seguir:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1} \quad (3.3)$$

A raiz quadrada da variância amostral é denominada de desvio padrão amostral ($\hat{\sigma}$).

Assim como no caso da média amostral, a variância amostral, assim estimada, é uma estatística não viesada da variância populacional σ^2 .

3.2.4 Covariância Amostral

Sejam X e Y duas variáveis aleatórias com médias μ_x e μ_y , respectivamente. Então a covariância entre as duas variáveis é definida como:

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)] \quad (3.4)$$

Supondo amostras de X e Y definidas como (x_1, x_2, \dots, x_n) e (y_1, y_2, \dots, y_n) e suas respectivas médias amostrais \bar{X} e \bar{Y} , podemos definir a covariância amostral como:

$$\widehat{\text{Cov}}(X, Y) = \frac{\sum_{i=1}^n x_i y_i}{n} - \bar{X} \bar{Y} \quad (3.5)$$

3.2.5 Coeficiente de Correlação Amostral

O Coeficiente de Correlação é um indicador para obtermos o grau de relacionamento entre duas variáveis quaisquer. Dadas X e Y, duas variáveis com n elementos, e sejam \bar{X} e \bar{Y} suas médias ou esperanças amostrais, respectivamente, define-se como coeficiente de correlação amostral pela seguinte fórmula [10]:

$$\widehat{\rho}_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{j=1}^n (x_j - \bar{X})^2 \sum_{j=1}^n (y_j - \bar{Y})^2}} \quad (3.6)$$

Note que $-1 < \hat{\rho} < 1$. Caso $\hat{\rho}$ se aproxime de 1 há indícios de correlação positiva, ou seja, X e Y tendem a variar no mesmo sentido. Caso $\hat{\rho}$ se aproxime de -1 há indícios de correlação negativa, ou seja, X e Y tendem a variar no sentido contrário. Se não existe correlação entre as variáveis ρ se aproxima de 0 e graficamente temos uma "nuvem" de pontos. Observe a figura 3.2.

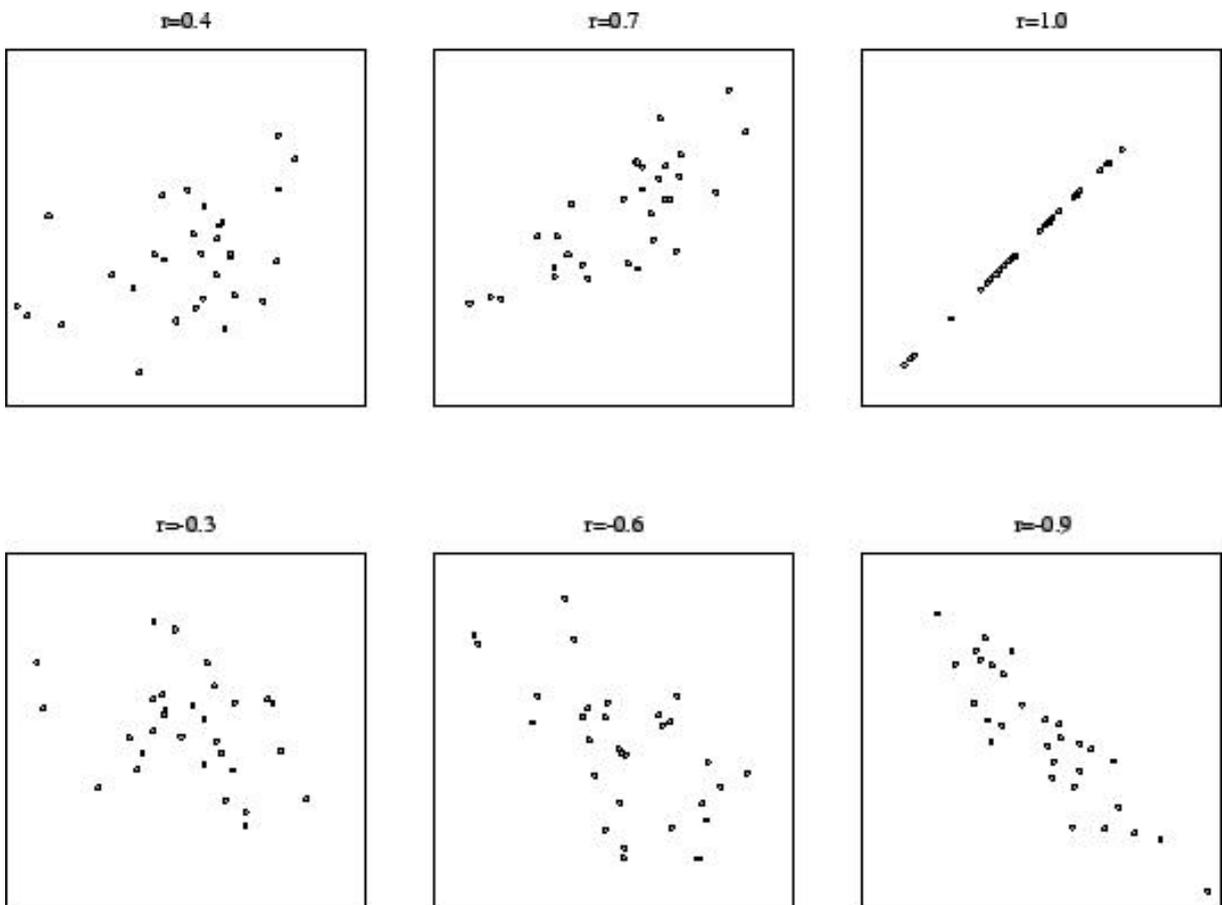


Figura 3.2: Exemplos de Diagramas de Dispersão de acordo com o coeficiente de correlação, neste exemplo $\hat{\rho} = r$.

O coeficiente de correlação pode ser definido em função da covariância e variâncias amostrais entre X e Y [6]:

$$\hat{\rho} = \frac{\widehat{cov}(X, Y)}{\hat{\sigma}_x \hat{\sigma}_y} \quad (3.7)$$

3.3 Modelo de Regressão Linear Múltipla

Vamos nos basear nos conceitos descritos por Gujarati (2000) [6], Hoffman e Vieira (1977)[8] e com algumas referências de Wooldridge (2002)[15].

3.3.1 O Modelo estatístico de uma regressão linear múltipla

Temos uma regressão linear múltipla (MLR - Multi Linear Regression) quando admitimos que o valor da variável dependente é função linear de duas ou mais variáveis independentes. O Modelo estatístico de uma regressão linear múltipla com k variáveis independentes é:

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + u_j, j = 1, \dots, n. \quad (3.8)$$

Ou, matricialmente,

$$y = X\beta + u \quad (3.9)$$

Onde:

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ u_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix}$$

Onde:

- Y_j indica a j-ésima observação da variável dependente.

- X_{kj} indica a j -ésima observação da variável independente k .
- β_k indica o coeficiente referente a variável independente k .
- u_j indica uma variável aleatória com média zero e variância constante.

Pressuposições [8]:

1. A variável dependente Y_j é função linear das variáveis independentes.
2. Os valores das variáveis independentes são fixos.
3. $E(u_i) = 0$.
4. Os erros são homocedásticos, isto é, $E(u_j^2) = \sigma^2$.
5. Os erros são não-correlacionados entre si, isto é, $E(u_j u_h) = 0$ para $j \neq h$.
6. Os erros tem distribuição normal¹.

3.3.2 Estimador de parâmetros por Mínimos Quadrados Ordinários (MQO)

Para obter a estimativa de MQO de β , vamos primeiro escrever a função de regressão da amostra de tamanho n em relação a k variáveis:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + u_i \quad (3.10)$$

que pode ser escrita mais concisamente na notação matricial como

$$y = X\beta + u \quad (3.11)$$

em que β é um vetor coluna de k elementos com estimadores de MQO dos coeficientes de regressão e u é um vetor coluna $n \times 1$ com n resíduos [6]. Os estimadores de MQO são obtidos minimizando:

$$u^t u = \sum_{i=1}^n u_i^2 = (y - X\beta)^t (y - X\beta) = y^t y - 2\beta^t X^t y + \beta^t X^t X \beta \quad (3.12)$$

Usando diferenciação matricial obtemos:

$$\frac{\partial(u^t u)}{\partial \beta} = -2X^t y + X^t X \beta \quad (3.13)$$

¹Entenda como erro o vetor de perturbações u

Igualando a equação 3.13 a zero chegamos no estimador para β :

$$(X^t X)\hat{\beta} = X^t y \implies \hat{\beta} = (X^t X)^{-1} X^t y \quad (3.14)$$

contanto que exista a inversa indicada.

3.3.3 Coeficiente de Determinação R^2

Indica uma medida da proporção da variabilidade em uma variável que é explicada pela variabilidade das outras. No caso de uma regressão simples, ou seja, $k=1$, o quadrado do coeficiente de correlação de Pearson, definido anteriormente, é chamado de coeficiente de determinação ou simplesmente R^2 . No caso geral definimos R como[6]:

$$R^2 = \frac{\hat{\beta}(X^t y) - n\bar{Y}^2}{y^t y - n\bar{Y}^2} \quad (3.15)$$

na representação matricial (lembrando que \bar{Y} é um estimador para a média da variável dependente Y).

Uma importante propriedade do R^2 é que conforme aumentamos o números de variáveis explicativas o R^2 nunca diminui, sempre aumenta, ou fica igual. Como alternativa para isso podemos usar o Coeficiente de Determinação Ajustado (R_{ADJ}^2):

$$R_{ADJ}^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2 / (n - k)}{\sum_{i=1}^n y_i^2 / (n - 1)} \quad (3.16)$$

onde k = número de parâmetros no modelo, incluindo o termo de intercepto (β_0). Esta estatística mostra um coeficiente de determinação ajustado pelo tamanho da amostra e pelo número de parâmetros. O maior valor que R_{ADJ}^2 pode assumir é igual ao do próprio coeficiente de determinação descrito na equação 3.15.

3.3.4 Matriz de variância-covariância de $\hat{\beta}$

Define-se como:

$$Cov(\hat{\beta}) = \begin{bmatrix} var(\hat{\beta}_1) & cov(\hat{\beta}_1, \hat{\beta}_2) & \dots & cov(\hat{\beta}_1, \hat{\beta}_k) \\ cov(\hat{\beta}_2, \hat{\beta}_1) & var(\hat{\beta}_2) & \dots & cov(\hat{\beta}_2, \hat{\beta}_k) \\ \cdot & \cdot & \cdot & \cdot \\ 4. & \cdot & \cdot & \cdot \\ cov(\hat{\beta}_k, \hat{\beta}_1) & cov(\hat{\beta}_k, \hat{\beta}_2) & \dots & var(\hat{\beta}_k) \end{bmatrix} = \sigma^2 (X^t X)^{-1} \quad (3.17)$$

onde σ^2 é estimado por $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{\hat{u}^t \hat{u}}{n - k} \quad (3.18)$$

Esta matriz será usada para a realização de inferência sobre os estimadores.

3.3.5 Inferência Estatística sobre os Estimadores

Para a realização de inferência usaremos a suposição de uniformidade dos erros:

$$u \sim N(0, \sigma^2 I) \quad (3.19)$$

onde u e 0 representam vetores coluna $n \times 1$ e I é uma matriz identidade $n \times n$, sendo 0 o vetor nulo. A partir desta suposição podemos mostrar que:

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^t X)^{-1}) \quad (3.20)$$

Como, na prática, σ^2 (a variância da população) é desconhecido, ele é estimado por $\hat{\sigma}^2$. Assim segue-se que cada elemento de $\hat{\beta}$ tem distribuição t-student com $n-k$ graus de liberdade. Simbolicamente,

$$t = \frac{\hat{\beta}_i - \beta_i}{ep(\hat{\beta}_i)} \quad (3.21)$$

onde $ep(\hat{\beta}_i) = \sqrt{var(\hat{\beta}_i)}$.

Essa estatística nos permitirá testar a hipótese de um coeficiente ser nulo ou não, isto é:

$$H_0 : \beta_i = 0 \quad (3.22)$$

$$H_1 : \beta_i \neq 0 \quad (3.23)$$

Caso a estatística t se encontre na região crítica então não podemos rejeitar H_0 , e portanto a respectiva variável relacionada com o estimador de β_i é representativa para o modelo, e possui coeficiente não nulo. Mais detalhes sobre testes de hipótese em Bussab e Morettin (2003) [3].

3.4 Análise de Autocorrelação

O termo autocorrelação pode ser definido como: correlação entre membros de um série de observações separados por k unidades de tempo. A autocorrelação ocorre na regressão

linear múltipla quando as perturbações (u_i) são autocorrelacionadas, ou seja, dada uma série definida por uma variável aleatória X , e uma respectiva amostra (x_1, x_2, \dots, x_n) , temos autocorrelação se:

$$\rho_X(t, t+k) \neq 0 \quad (3.24)$$

que é estimado por $\widehat{\rho}_X(t, t+k)$ indicando o coeficiente de correlação amostral entre 2 subconjuntos, ou sub-séries, da amostra extraída de X separados por k períodos.

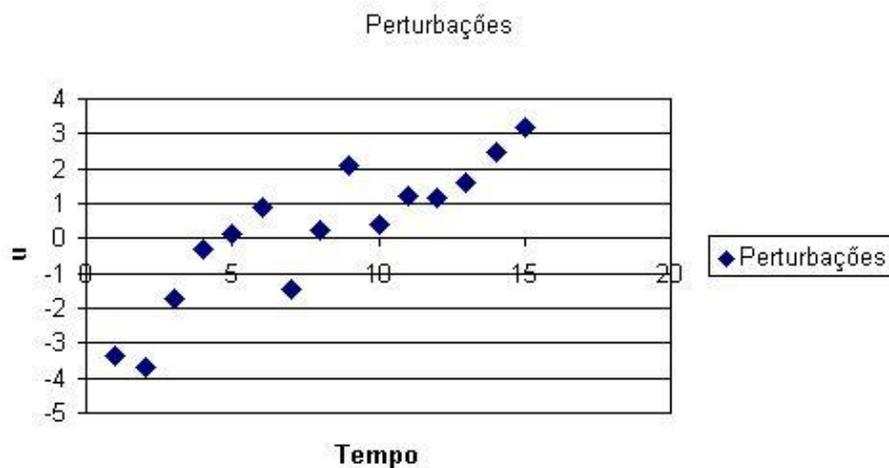


Figura 3.3: Exemplo de Padrão de Autocorrelação Positivo, padrão sistemático nas perturbações. Fonte: Microsoft Excel.

No modelo que definimos anteriormente para estimar Regressão Linear Múltipla supomos que não deveria existir autocorrelação. Portanto a sua estimativa pode causar um viés de especificação do modelo. Assim precisamos testar os modelos por autocorrelação, caso ela exista será necessário reavaliar o conjunto de variáveis independentes e fazer possíveis alterações no modelo ².

Teste d de Durbin-Watson

Podemos testar se existe autocorrelação de primeira ordem nas perturbações usando a estatística definida como:

$$d = \frac{\sum_{i=2}^n (\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=2}^n \hat{u}_i^2} \quad (3.25)$$

²Mais detalhes em [6]

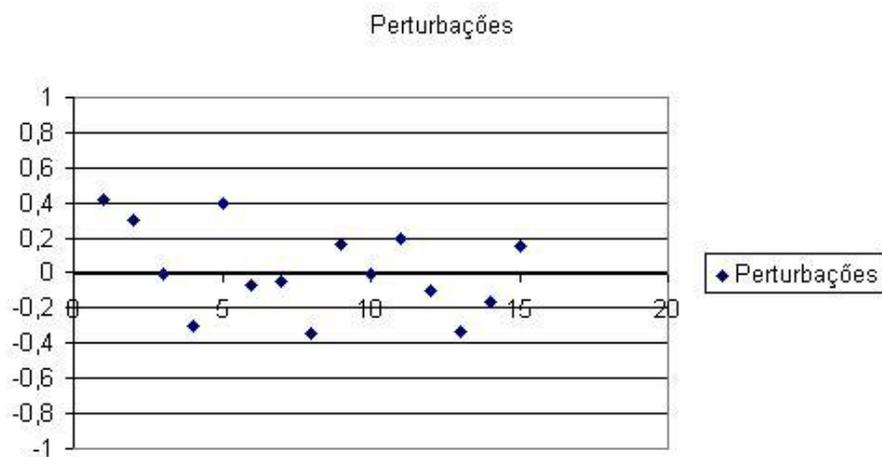


Figura 3.4: Exemplo de ausência de Autocorrelação, erros aparentemente aleatórios. Fonte: Microsoft Excel.

Rejeitamos a hipótese de presença de autocorrelação caso d esteja perto de 2. Para descobrirmos os valores críticos para os quais passa-se a aceitar tal hipótese devemos consultar uma tabela padrão para esta estatística.³

3.5 Análise de resíduos

É sempre prudente checarmos se o modelo estimado está de acordo com as hipóteses definidas, para isso uma análise sobre os resíduos encontrados podem fornecer muitas informações referentes a autocorrelação e normalidade das perturbações.

Graficamente podemos avaliar a existência ou não de autocorrelação olhando para a evolução dos resíduos no tempo, caso nesta evolução perceba-se algum padrão sistemático é indício de existência de autocorrelação, caso haja aparente aleatoriedade dos dados é um indício de ausência de autocorrelação.

Outra análise importante é checar a normalidade dos resíduos, que também pode ser feita graficamente com um histograma ou um QQ-Plot versus uma normal, observe a figura 3.6⁴.

A análise gráfica permite um bom julgamento a respeito da aceitação ou não da

³A tabela completa existe em [6]

⁴Para detalhes sobre QQ-Plot consultar [3]

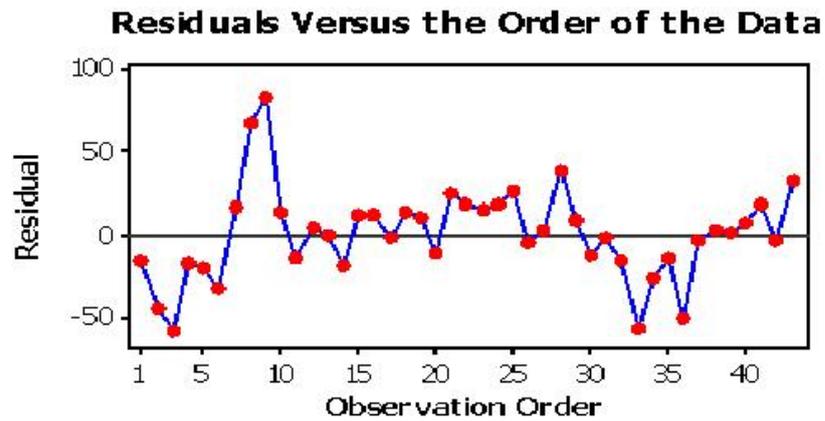


Figura 3.5: Exemplo de evolução dos resíduos. Fonte: Minitab.

hipótese de normalidade, assim, se o histograma se aproximar da distribuição conhecida como Normal e o QQ-Plot de uma reta, há grandes indícios de normalidade dos resíduos.

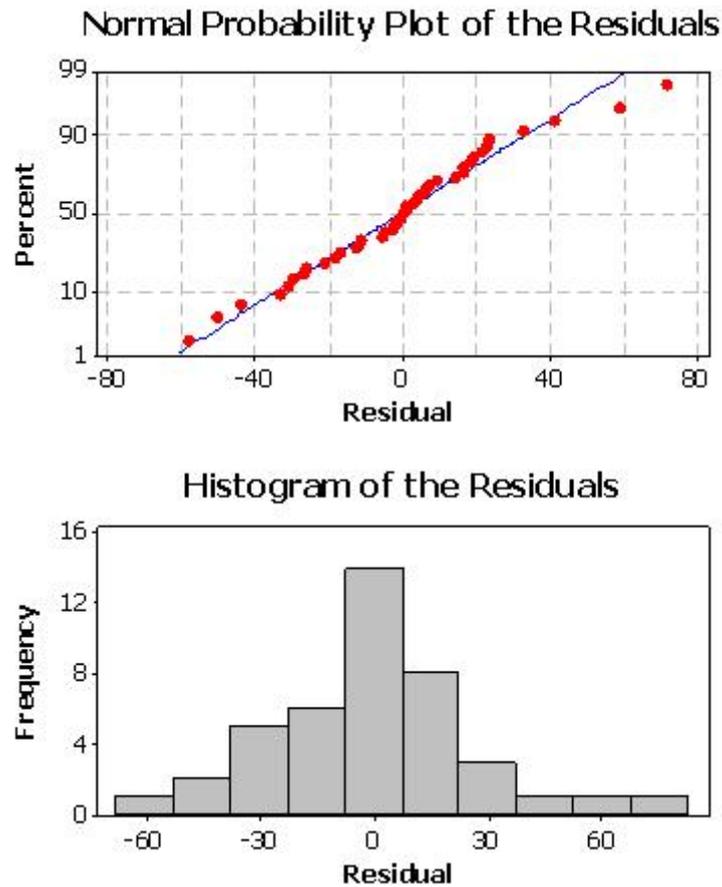


Figura 3.6: Exemplo de análise de normalidade dos resíduos. Fonte: Minitab.

3.6 Séries Temporais

Passaremos agora a descrever a metodologia para estimarmos modelos derivados de um conjunto de dados ordenados cronologicamente, basicamente vamos focar nas revisões realizadas por Wooldridge (2002) [15] e Enders (2003) [4]. Vamos descrever alguns processos de séries temporais mais conhecidos e muito usados.

3.6.1 Estimativa por Mínimos Quadrados

Podemos descrever um modelo de séries temporais como⁵:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t, t = 1, \dots, n. \quad (3.26)$$

Onde:

⁵Aqui descrevemos a forma estendida do modelo, mas perceba que, assim como fizemos com os modelos de MLR, podemos representar os modelos de séries temporais na forma matricial

- $t = 1..n$, onde n indica o número de períodos ou observações.
- $j = 1..k$, onde k indica o número parâmetros.
- Y_t indica a observação da variável dependente no período de tempo t .
- X_{kt} indica a observação da variável independente k no instante de tempo t .
- β_k indica o coeficiente referente a variável independente k .
- u_t indica resíduo relativo ao período de tempo t .

Podemos usar o mesmo método descrito anteriormente de Mínimos Quadrados para a estimativa dos coeficientes do modelo de regressão, porém vamos ter algumas diferenças nas pressuposições para que a estimativa por MQO não tenha viés.

Pressuposições [15]:

1. Linearidade em relação aos parâmetros
2. O valor esperado dos resíduos, dado todo o conjunto de variáveis independentes, deve ser nulo, $E(u_t/X) = 0$.
3. As variáveis independentes não podem ter colinearidade perfeita entre elas.
4. Homoscedasticidade dos resíduos, $Var(u_t/X) = \sigma^2$ para todo t .
5. Ausência de correlação serial, $Corr(u_t, u_s/X) = 0$ para todo $t \neq s$.
6. Os resíduos, u_t , são independentes das variáveis independentes e têm distribuição normal($0, \sigma^2$).

A partir destas pressuposições podemos chegar a:

$$Var(\hat{\beta}_j/X) = \sigma^2/[SST_j(1 - R_j^2)], j = 1, \dots, k. \quad (3.27)$$

onde $SST_j = \sum_{t=1}^n (x_{tj} - \bar{x}_j)^2$, é a variação total da amostra em x_j , e R^2 é o coeficiente de determinação definido da mesma forma que na equação 3.15. A variância σ^2 pode ser estimado sem viés por $\hat{\sigma}^2 = SSR/(n - k - 1)$ onde $SSR = \sum_{t=1}^n \hat{u}_t^2$. Assim podemos realizar o mesmo tipo de inferência já descrito anteriormente na seção a respeito de Inferência estatística sobre os Estimadores e realizar análises de autocorrelação e de resíduos, também já descritos anteriormente.

3.6.2 Processo Estocástico Estacionário

Conforme descreve Gujarati (2000) [6], dados de qualquer série temporal podem ser pensados como sendo gerados por um processo estocástico ou aleatório; e um conjunto concreto de dados pode ser considerado como uma (particular) realização (isto é, uma amostra) do processo estocástico subjacente.

A distinção entre o processo estocástico e sua realização é parecida com a distinção entre população e amostra em dados de corte. Assim como utilizamos dados amostrais para fazer inferências sobre uma população, em séries temporais usamos a realização para fazer inferências sobre o processo estocástico subjacente.

Um tipo de processo estocástico que tem recebido grande atenção e exame por parte de analistas de séries temporais é o chamado processo estocástico estacionário. A grosso modo, diz-se que um processo estocástico é estacionário se suas média e variância forem constantes ao longo do tempo e o valor da covariância entre dois períodos de tempo depender apenas da distância ou defasagem entre os dois períodos, e não do período de tempo efetivo em que a covariância é calculada.

$$E(Y_t) = \mu \quad (3.28)$$

$$Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2 \quad (3.29)$$

$$\gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)] \quad (3.30)$$

A partir de uma série estacionária podemos propor modelos como o Autoregressivo, de Média Móvel ou ARMA.

3.6.3 Processo de Média Móvel

Podemos descrever um processo de média móvel de ordem q (MA(q) - *Moving Average*) da seguinte forma:

$$y_t = u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \dots + \alpha_q u_{t-q} \quad (3.31)$$

onde u_t representa uma seqüência iid⁶ com média zero e variância σ_u^2 . Assim o processo de média móvel representa em y_t uma média ponderada dos termos u_{t-k} para $k=0, \dots, q$.

⁶Independente e igualmente distribuída

3.6.4 Processo Autoregressivo

Podemos descrever um processo de auto regressão de ordem p (AR(q) - *Autoregressive*) da seguinte forma:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + u_t \quad (3.32)$$

onde u_t representa uma seqüência iid com média zero e variância σ_u^2 . Esse processo estima o valor esperado para a variável de estudo y_t em função de seu próprio passado, atribuindo pesos a cada período ocorrido no passado. Por uma questão de estabilidade devemos descrever condições limitantes para os coeficientes α_k , no caso de um AR(1), por exemplo, $|\alpha_1| < 1$.

3.6.5 Modelo ARMA

É possível combinar um processo autoregressivo com um processo de média móvel, gerando o que chamamos de ARMA - Auto Regressive Moving Average, obtendo um modelo com a seguinte forma:

$$ARMA(p, q) : y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \sum_{i=0}^q b_i u_{t-i} \quad (3.33)$$

Para estimarmos um modelo ARMA precisamos que este descreva um processo estável, e portanto haverá restrições quanto aos coeficientes.

3.6.6 Correlogramas

Os correlogramas ajudam-nos a identificar um modelo do tipo ARMA. Vamos descrever duas funções e depois mostramos a importância dos correlogramas.

Função de Autocorrelação (FAC)

Um teste simples de estacionariedade baseia-se na chamada função autocorrelação (FAC), e vai nos ajudar a estimar os componentes de um modelo ARMA. A FAC na defasagem k , indicada por ρ_k , é definida como:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (3.34)$$

onde $\gamma_k e \gamma_0$ são dados por 3.30. Como tanto a covariância quanto a variância são medidas nas mesmas unidades de medida, ρ_k é um número sem unidade, ou puro. Fica entre -1 e + 1, como qualquer coeficiente de correlação. Se representarmos graficamente ρ_k contra k , o gráfico que obtemos é conhecido como correlograma da população ou da amostra em questão se assim quisermos.

Se uma série temporal for puramente aleatória, denominada por **ruído branco**, então os coeficientes de autocorrelação amostral são, aproximadamente, distribuídos normalmente com média zero e variância $\frac{1}{n}$, onde n indica o tamanho da amostra.

Na estimativa de um modelo do tipo ARMA temos que supor, por hipótese, que os resíduos sejam um ruído branco, assim o correlograma pode ajudar a fazermos essa diagnose.

Função de Autocorrelação Parcial (FACP)

O conceito de autocorrelação parcial é análogo ao conceito de coeficiente de regressão parcial. No modelo de regressão múltipla de k variáveis, o k -ésimo coeficiente de regressão β_k mede a taxa de variação no valor médio do regressando para uma mudança de uma unidade no k -ésimo regressor X_k , mantendo constante a influência de todos os outros regressores. Da mesma maneira, a autocorrelação parcial ϕ_{kk} mede a correlação entre observações (séries temporais) que estejam k períodos afastados, depois de controlar às correlações nas defasagens intermediárias (isto é, defasagens menores que k). Em outras palavras, autocorrelação parcial é a correlação entre Y_t e Y_{t-k} depois de se remover o efeito dos Y s intermediários. Assim, se o processo for $AR(p)$, temos que:

$$AR(p) : \begin{cases} \phi_{kk} \neq 0, & k \leq p \\ 0, & k > p \end{cases}$$

A partir dos gráficos dos correlogramas (FAC e FACP amostrais), podemos usar a tabela 3.1 para montarmos possíveis modelos do tipo ARMA, AR ou MA.

A tabela 3.1 mostra situações teóricas, na prática o decaimento exponencial pode não ser tão evidente. Essa tabela e o correlograma podem indicar um ou mais modelos preliminares que podem ser adequados para ajustar uma série temporal.

Na figura 3.7 mostramos um exemplo de correlograma, com as funções de autocorrelação e de autocorrelação parciais. Observe que na figura 3.7 temos um caso que

Padrões teóricos de FAC e FACP

Tipo de modelo	Padrão típico de FAC	Padrão típico de FACP
AR(p)	Decai exponencialmente para zero ou com padrão de onda senoidal amortecida, ou ambos.	Valores significativos, ou seja, não nulos até a defasagem p
MA(q)	Valores significativos, ou seja, não nulos até a defasagem q	Decai exponencialmente para zero
ARMA(p,q)	Decai exponencialmente para zero	Decai exponencialmente para zero

Tabela 3.1: Padrões teóricos de FAC e FACP. Fonte: Gujarati 2000 [6].

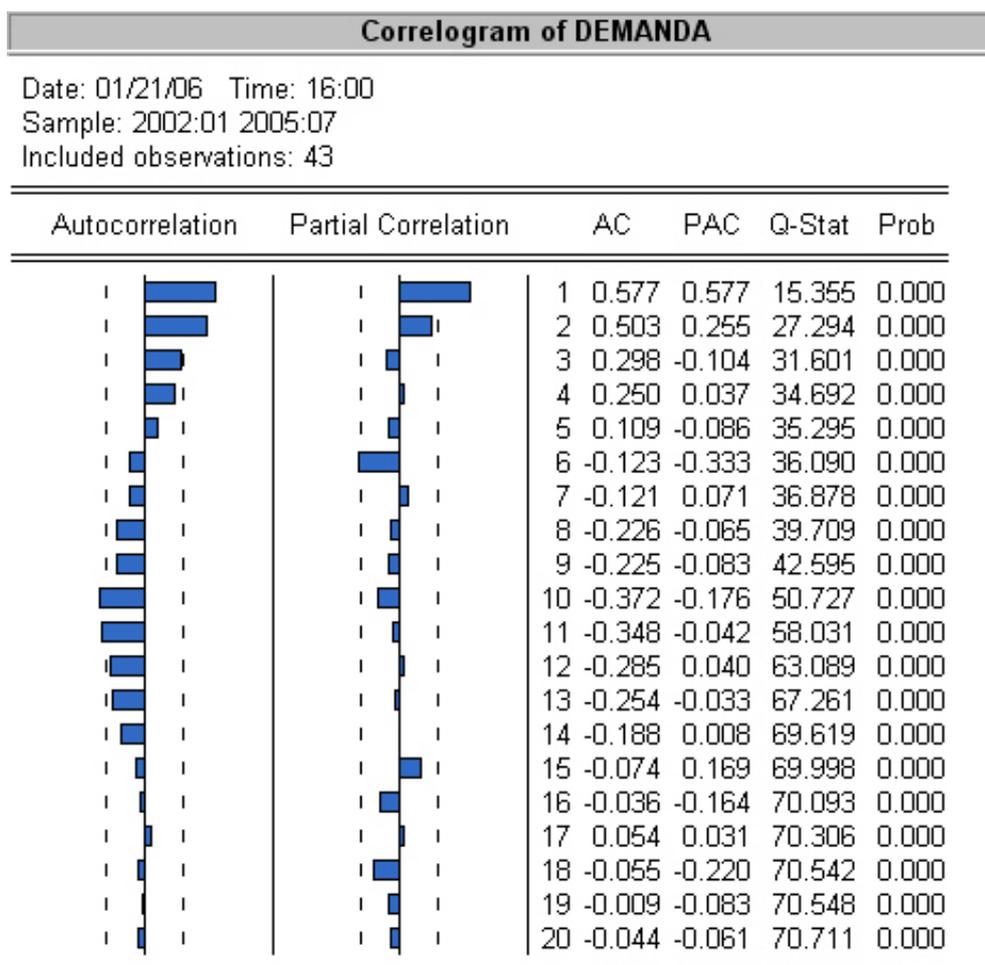


Figura 3.7: Exemplo de Correlograma. Fonte: E-Views

pode possivelmente descrever um $AR(p)$, decaimento exponencial senoidal na FAC e com picos na FACP.

3.6.7 Metodologia de Box, Jenkins e Reinsel (1994)

Obviamente, a grande questão é: olhando para uma série temporal, como sabemos se ela segue um processo puramente AR (e, nesse caso, qual o valor de p) ou um processo puramente MA (e, nesse caso, qual o valor de q) ou um processo ARMA (e, nesse caso, quais são os valores de p e q). A metodologia de BJ vem a calhar para responder à questão anterior. O método consiste de quatro etapas:

1. **Identificação.** Ou seja, descobrir os valores apropriados de p e q . Para isso usamos o correlograma e o correlograma parcial para perceber em que períodos de defasagem existe mais correlação com a variável dependente ou de correlação entre as observações com k períodos de defasagem, ou seja, correlação entre Y_t e Y_{t-k} e dos componentes de média móvel u_t e u_{t-k} .
2. **Estimativa.** Depois de identificar os valores apropriados de p e q , o próximo passo é estimar os parâmetros dos termos auto-regressivo e de média móvel incluídos no modelo. Neste estudo a estimativa é feita utilizando o método dos mínimos quadrados ordinários, conforme já foi descrito anteriormente.
3. **Checagem de diagnóstico.** Depois de escolher um modelo ARMA em particular, e estimar seus parâmetros, vemos em seguida se o modelo escolhido se ajusta aos dados razoavelmente bem, pois é possível que um outro modelo ARMA possa se ajustar melhor. Daí por que a modelagem ARMA de Box-Jenkins é mais uma arte do que uma ciência; é necessária considerável habilidade para escolher o modelo ARMA correto. Um teste simples do modelo escolhido é ver se os resíduos estimados desse modelo são ruídos brancos; se são, podemos aceitar o ajuste específico; se não são, devemos começar tudo de novo. Assim, a metodologia de BJ é um processo iterativo.
4. **Previsão.** Em muitos casos, as previsões obtidas com esse método são mais confiáveis do que as obtidas com a modelagem econométrica tradicional, especialmente para previsões a curto prazo. Naturalmente, é preciso checar cada caso.

4 Ferramenta Computacional

Usamos o software Econometric Views (E-Views) para a maioria das análises de regressões, e para algumas análises também usamos o Minitab, sendo o Excel a base de armazenamento dos dados e análises estatísticas e gráficas simples. Outros softwares como R-Statistic, SAS, SPSS poderiam também realizar essas mesmas análises. Vamos apresentar os detalhes técnicos apenas das análises de regressão referente ao E-Views e o Minitab.

4.1 E-Views

O software Econometric Views, também chamado de E-Views, é um pacote estatístico desenvolvido por economistas e com a maioria de aplicações na Economia, mas pode ser usado em outras áreas. Este é um software que produz regressões e previsões, desenvolve relações estatísticas entre os dados e usa estas relações para prever valores futuros da série. As áreas onde o E-Views pode ser útil são: previsão de vendas, análise de custos, previsões em análises financeiras, simulação e previsão macroeconômica, análise científica e avaliação de dados. A estimação de regressão do E-Views é feita utilizando a técnica de mínimos quadrados residual. Embora seu manuseio seja um pouco mais complexo que o Minitab, fornece resultados bem completos em análises de regressões. Detalharemos apenas as funções essenciais para a compreensão do estudo e deixamos a cargo do leitor seus conhecimentos aprofundar no uso do E-Views, se for de seu interesse.

4.1.1 Carga dos Dados

Os dados devem estar dispostos em uma planilha de Excel, conforme a figura 4.1, para serem carregados, com as variáveis nas colunas e os nomes na primeira linha.¹

No E-Views, para carregar os dados, selecionamos a partir do menu, *File* -- *> Import* -- *> ReadTextExcel*, a função de carga de dados, e assim chegamos na tela capturada na figura 4.2, aonde devemos detalhar os tipos de dados que pretendemos

¹É possível carregar os dados a partir de outros tipos de arquivos, mas consideramos esse processo mais fácil para compreensão.

	A	B	C	D	E	F	
1	Demanda	Despesas Publicitárias (DP)	Net Price (NP)	Dist. Num. (DN)	Verbas Táticas (VT)	Inovação (I)	Negocia
2	227,30	429,40	113,11	0,60	62,37	0,00	
3	206,44	508,51	116,16	0,71	120,59	0,00	
4	176,22	377,93	120,78	0,65	1.495,24	0,00	
5	360,28	463,94	109,80	0,71	1.917,94	0,00	
6	343,97	447,43	112,29	0,72	1.901,27	0,00	
7	281,19	384,00	121,05	0,78	85,21	0,00	
8	295,51	298,44	118,90	0,78	1.547,57	0,00	
9	307,50	310,65	117,09	0,73	2,53	0,00	
10	331,53	234,90	113,52	0,64	143,20	0,00	
11	203,23	155,31	123,79	0,67	134,51	0,00	
12	254,36	6,85	116,14	0,84	149,86	0,00	
13	216,52	0,00	121,85	0,74	123,28	0,00	
14	245,69	0,00	117,54	0,78	107,63	0,00	
15	222,91	0,00	120,88	0,87	26,73	0,00	
16	214,79	0,00	122,12	0,71	20,04	0,00	

Figura 4.1: Exemplo de arquivo para carga de dados no E-Views. Fonte: E-Views.

carregando.

4.1.2 Fazendo Regressões

Usaremos o comando *Quick* --- $>$ *EstimateEquation* para estimar regressões lineares utilizando mínimos quadrados ou modelos do tipo ARMA. Observe na figura 4.3 um exemplo de como entrar com os dados na regressão.

Os resultados gerados são os seguintes:

1. Na parte superior aparecem algumas informações referentes ao tipo de método usado, o tamanho da amostra, data e hora em que foi gerado o modelo e quem é a variável dependente em questão.
2. A primeira tabela mostra informações a respeito dos coeficientes estimados na regressão: Valor do coeficiente, desvio padrão, estatística t e p-valor.
3. Na tabela no final aparecem as estatísticas de ajuste calculadas pelo E-Views.

As estatísticas de ajuste calculadas pelo E-Views são:

- R^2 - É a fração da variabilidade da variável dependente explicada pelas variáveis independentes.

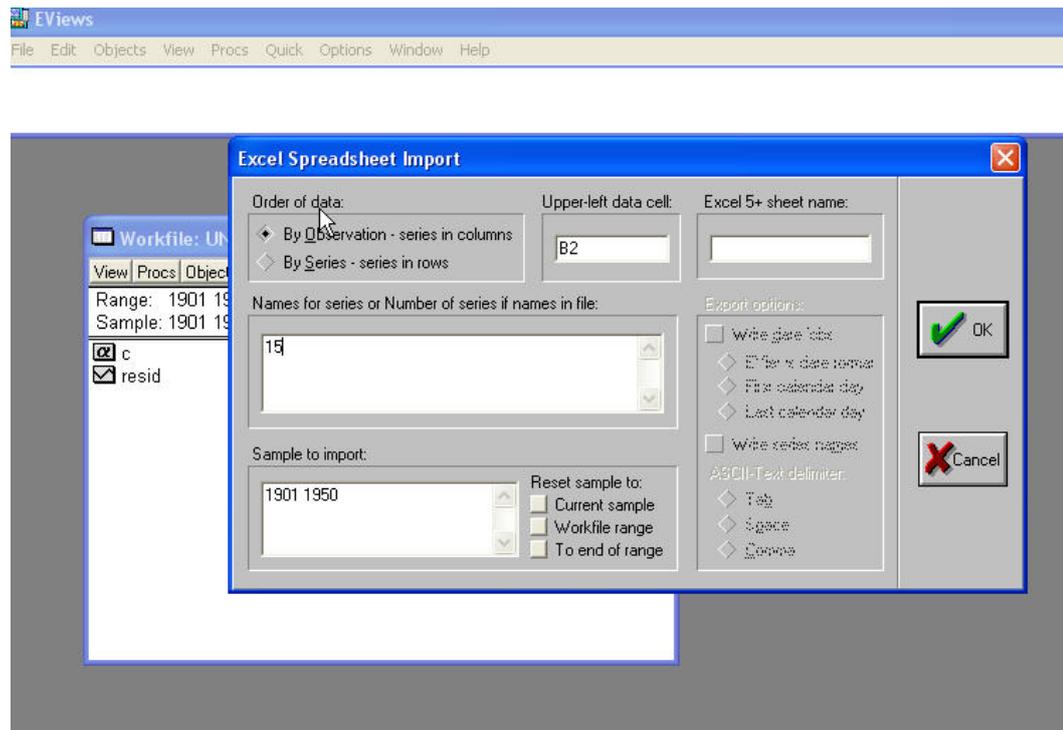


Figura 4.2: Exemplo de carga de dados no E-Views. Fonte: E-Views.

- R^2 Ajustado - É a fração da variabilidade da variável dependente explicada pelas variáveis independentes ajustado de acordo com o tamanho da amostra e número de parâmetros.
- S.E. of regression - O erro padrão da regressão é uma medida resumo baseada na variância estimada dos resíduos.
- Sum of Squares Residual - A soma de quadrados residual pode ser usada em uma variedade de cálculos estatísticos, e é apresentada separadamente por conveniência.
- Log Likelihood - E-Views reporta o valor da função de log-verossimilhança (assumindo erros normalmente distribuídos) avaliada nos valores estimados dos coeficientes.
- Durbin-Watson Statistic - A estatística de Durbin-Watson (DW) mede a correlação serial nos resíduos. Como regra, DW deve estar perto de 2 para indicar ausência de autocorrelação ou ausência de correlação serial.
- A média e o desvio-padrão da variável dependente.
- Akaike Information Criteria - O AIC é freqüentemente usado na seleção de modelos.

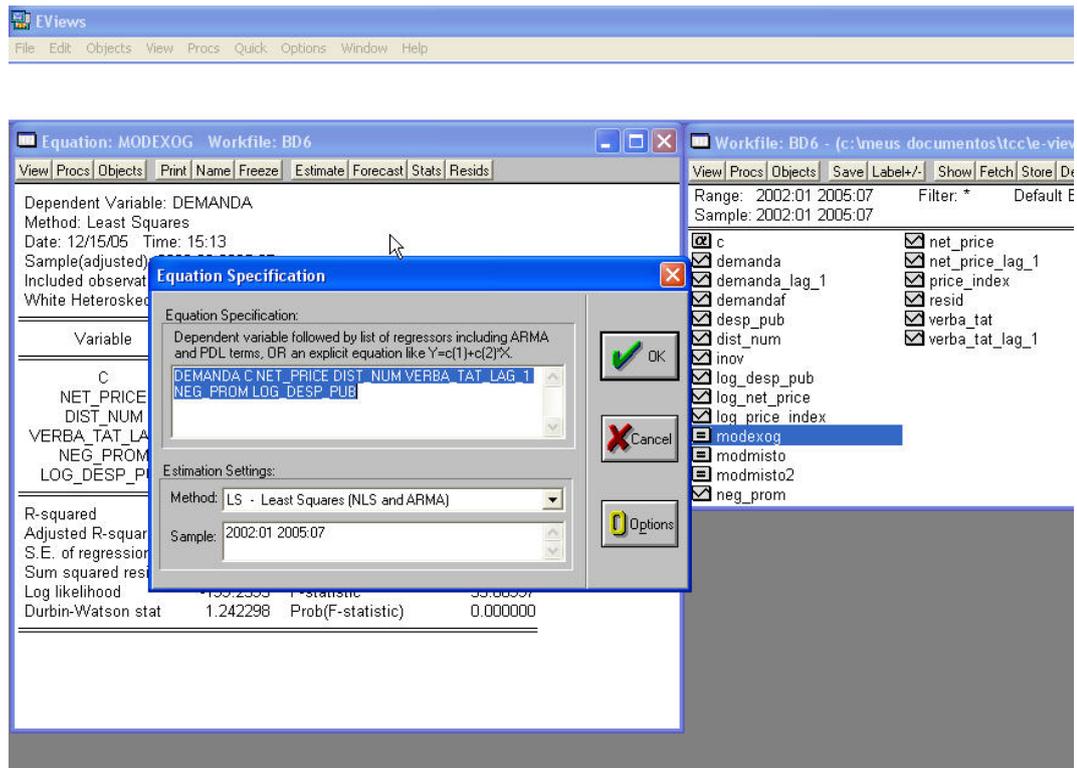


Figura 4.3: Exemplo referente a função de regressão do E-Views. Fonte: E-Views.

Modelos que apresentam os menores valores de AIC são os preferidos.

- Schwarz Criterion (SC) - O Critério de Schwarz é uma alternativa ao AIC que impõe uma penalidade alta para coeficientes adicionais.
- Prob(F-Statistic) e Prob - A estatística-F testa a hipótese de que todos os coeficientes de inclinação (excluindo a constante, ou o intercepto) são nulos. O valor-p denota o nível descritivo do teste-F.

Observe na figura 4.4 os resultados de uma regressão realizada no E-Views.

Dependent Variable: DEMANDA
Method: Least Squares
Date: 12/15/05 Time: 15:13
Sample(adjusted): 2002:02 2005:07
Included observations: 42 after adjusting endpoints
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	836.3274	148.8889	5.617124	0.0000
NET_PRICE	-6.117022	1.312139	-4.661871	0.0000
DIST_NUM	147.9761	39.17049	3.777744	0.0006
VERBA_TAT_LAG_1	0.018691	0.007467	2.503300	0.0170
NEG_PROM	46.13739	12.86909	3.585132	0.0010
LOG_DESP_PUB	5.135142	2.692612	1.907123	0.0645
R-squared	0.823573	Mean dependent var		289.3531
Adjusted R-squared	0.799069	S.D. dependent var		66.97581
S.E. of regression	30.02215	Akaike info criterion		9.773311
Sum squared resid	32447.86	Schwarz criterion		10.02155
Log likelihood	-199.2395	F-statistic		33.60997
Durbin-Watson stat	1.242298	Prob(F-statistic)		0.000000

Figura 4.4: Exemplo de resultados de uma regressão no E-Views. Fonte: E-Views.

4.2 Minitab

O Minitab é um software de interface amigável para análise de dados, oferece opções de regressão e análise de séries temporais, porém não é muito completo para estudos econométricos. Como ele possui uma interface de planilha parecida com a do Excel torna-se mais simples carregar os dados, basta colar as séries nas colunas deixando na primeira linha o nome das variáveis.

No Minitab usamos o comando *Stat* --- *> Regression* para realizar regressões simples ou multivariadas. Para rodar a regressão basta entrar com a variável dependente em *Response* e com as independentes em *Predictors*. A opção *Graphs* fornece alguns gráficos de Resíduos que auxiliam nas análises [13].

Os resultados fornecidos pela regressão apresentam alguns grupos de informações:

- Equação, teste de significância dos coeficientes, desvio padrão dos resíduos (S),

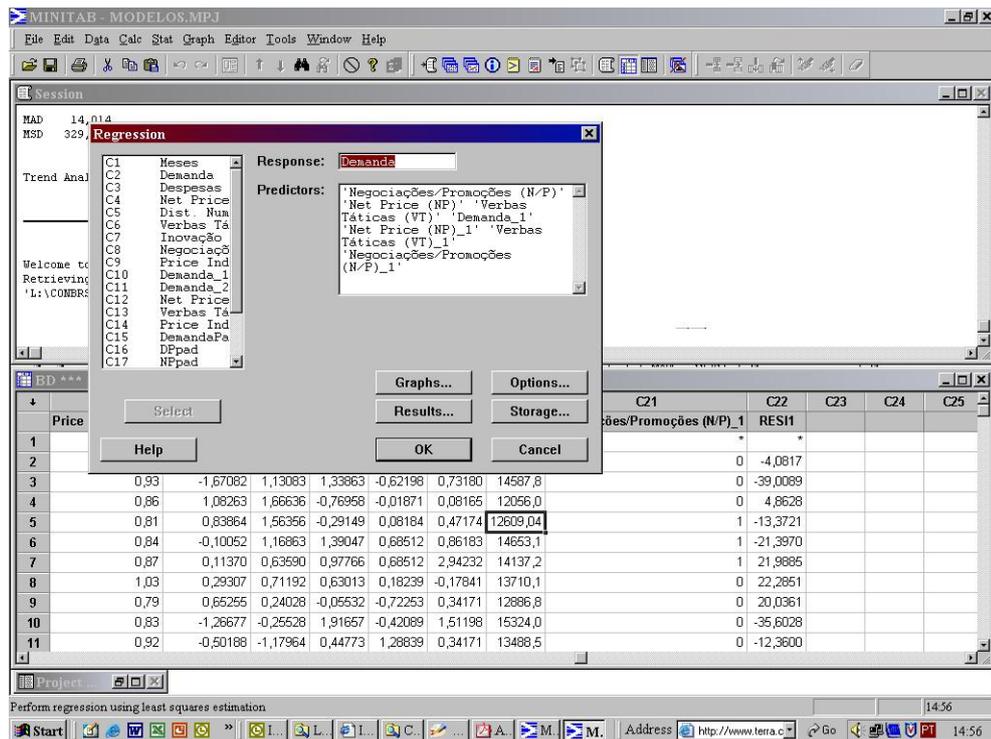


Figura 4.5: Exemplo de tela do Minitab. Fonte: Minitab.

coeficiente de determinação R^2 e coeficiente de determinação ajustado²

- Análise de Variância, testa a significância global dos coeficientes.
- Resíduos mais discrepantes (se existirem).
- Outras estatísticas opcionais, como o Teste Durbin-Watson

²Este coeficiente é equivalente ao coeficiente de determinação, porém é ponderado pelo número de variáveis do modelo, se há muitas variáveis irrelevantes para o modelo este número costuma se distanciar do R^2 , caso contrário fica próximo do R^2 .

Regression Analysis: Demanda versus Net Price (NP); Dist. Num. (DN); ...

The regression equation is
 Demanda = 776 - 5,89 Net Price (NP) + 189 Dist. Num. (DN) + 0,0193 VT Lag -1
 + 48,1 Negociações/Promoções (N/P)
 + 5,16 Log Despesas Publicitárias (LDP)

42 cases used, 1 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	776,5	116,4	6,67	0,000
Net Price (NP)	-5,895	1,007	-5,85	0,000
Dist. Num. (DN)	189,37	50,04	3,78	0,001
VT Lag -1	0,019266	0,006589	2,92	0,006
Negociações/Promoções (N/P)	48,06	13,92	3,45	0,001
Log Despesas Publicitárias (LDP)	5,156	2,543	2,03	0,050

S = 29,0685 R-Sq = 83,5% R-Sq(adj) = 81,2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	153497	30699	36,33	0,000
Residual Error	36	30419	845		
Total	41	183916			

Source	DF	Seq SS
Net Price (NP)	1	82579
Dist. Num. (DN)	1	32728
VT Lag -1	1	21508
Negociações/Promoções (N/P)	1	13208
Log Despesas Publicitárias (LDP)	1	3473

Unusual Observations

Obs	Net Price	Demanda	Fit	SE Fit	Residual	St Resid
9	114	331,53	256,72	8,95	74,81	2,70R
33	110	233,67	293,49	9,45	-59,82	-2,18R
36	104	251,39	305,37	13,04	-53,98	-2,08R

R denotes an observation with a large standardized residual.

Durbin-Watson statistic = 1,34123

Figura 4.6: Exemplo de Resultados do Minitab. Fonte: Minitab.

4.3 Discussão: Minitab vs E-Views

Neste estudo realizamos análises em ambos os softwares de análise de regressão, E-Views e Minitab, e percebemos algumas divergências entre os resultados fornecidos por regressões idênticas, com as mesmas séries, e entendemos que isso possa afetar ligeiramente a escolha do modelo.

Comparações entre estes softwares e ainda com o SPSS e SAS realizadas por Franco et al (2004) [5] em regressões do tipo ARMA mostram que a diferença entre os coeficientes estimados em cada software não é significativa (95% de confiança). Observe na figura 4.7 os vícios, divergências entre estimativas real e realizada pela software, dos estimadores dos coeficientes de um modelo do tipo ARMA(p,q).

Portanto é natural que encontremos algumas pequenas diferenças entre as estimativas do E-Views quando comparadas com o Minitab, já que cada software parece utilizar um algoritmo diferente para fazer a regressão. Temos que lembrar ainda que o

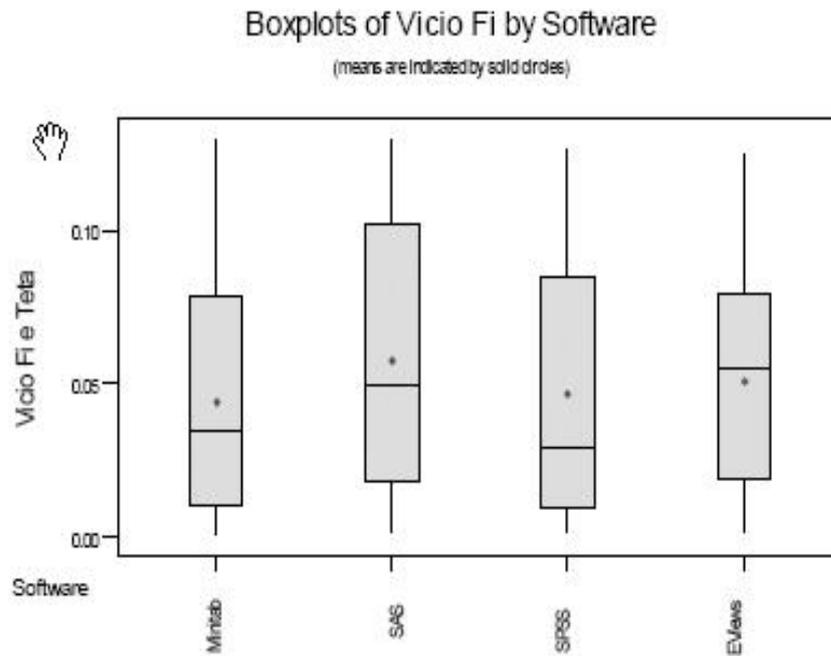


Figura 4.7: Comparativo entre os vícios por Softwares [5]

tamanho ou arredondamento ou truncamento dos números armazenados pelo software também pode influenciar o resultado da regressão. Para este estudo, nós entendemos que essa diferença não é significativa a ponto de alterar bruscamente a escolha de um modelo.

5 Caracterização dos Dados

5.1 Coleta e Interpretação dos Dados

A coleta e interpretação dos dados consistem em uma etapa fundamental para alcançar bons modelos estatísticos. Dado que o objetivo está em estimar a demanda de um determinado produto, e compreender que variáveis têm impacto nesta demanda, o primeiro passo consiste em compreender o cenário em que o produto está inserido, internamente na empresa e externamente no mercado, assim fica possível decidir que tipos de dados serão coletados e suas fontes. No caso deste estudo foram coletados dados em séries temporais de diversas fontes. A partir do cenário em que o produto está inserido pode-se levantar as possíveis variáveis que têm relação com a venda deste produto e suas respectivas fontes. Por exemplo, internamente podem ser levantados o gasto com propagandas, os descontos ou verbas aplicados, o nível de estoque, preço (naturalmente cada empresa armazena tais informações de forma diferente), e externamente podem ser levantados a taxa de distribuição do produto nas lojas, o preço em relação à concorrência, o share de mercado, entre outros, que podem ser encontrados em instituições de pesquisa de mercado, conforme observamos também na figura 1.1. Quanto maior for a abrangência das variáveis maior serão as chances de se chegar em um bom modelo estatístico para este produto. Para cada variável observa-se a fonte, a extensão do histórico, a periodicidade de sua coleta, sua unidade. Neste estudo utilizamos variáveis com periodicidade mensal e com histórico de pelo menos 36 meses. Note que um histórico maior pode trazer mais informações para o modelo, o que pode ser muito importante para entendermos o comportamento das vendas de um produto, principalmente se ele tiver algum tipo de sazonalidade. Para facilitar a compreensão dos resultados do estudo vamos focar a metodologia para dados coletados desta forma e neste formato, mas fica evidente que estes métodos podem e devem ser adaptados para cada realidade.

Destacaremos aqui quatro características fundamentais que as séries temporais devem possuir:

1. Os dados devem aparecer em seqüência cronológica.

2. Os intervalos de tempo entre as observações devem ser iguais, por exemplo, diários semanais ou mensais.
3. Os dados devem ser coletados no mesmo instante dentro de cada período, por exemplo as 15 horas, no caso de coleta diária; às quintas feiras à tarde, para coleta semanal, etc.
4. A seqüência deve ser completa, isto é, não pode haver dados faltantes. Quando há falta de um dado pode-se estimá-lo como a média aritmética entre antecessor e sucessor imediatos.

Conforme descrito por Parente (2003) [11], os dados internos devem ser coletados a partir de seus próprios sistemas de informação e os dados externos podem ser coletados em instituições de pesquisa ou empresas como o Ibope, Instituto Brasileiro de Geografia e Estatística (IBGE), Banco Central do Brasil, ACNielsen, Indicator, associações setoriais como a Associação Brasileira de Supermercados (Abras) ou a Federação Brasileira de Associantes de Bancos (Febraban) e outras instituições ou empresas. Caso o dado relevante para o estudo não se encontre nem internamente nem externamente pronto é crucial a realização de um levantamento destes dados, no caso interno deve-se começar a medir tal dado e no caso externo indica-se uma pesquisa de mercado.

A metodologia usada neste estudo, para coleta de dados, foi a seguinte: inicialmente foram realizadas reuniões com diversos gerentes envolvidos no processo em que o produto está inserido, como profissionais de Marketing, Trade, Finanças, Produção e Vendas, para coletar informações sobre o produto e possíveis fatores que poderiam afetar sua demanda. Daí obteve-se um enorme número de variáveis que somadas a tudo aquilo que existe efetivamente nos bancos de dados da empresa poderiam ser avaliadas se são ou não adequadas para o modelo seguindo alguns critérios, como:

- Existe histórico da variável?
- Qual a periodicidade de sua coleta?
- É possível coletar dados desta variável?
- Qual o grau de relevância da variável para a empresa?
- A variável pode ter uma representação numérica?

- A variável é restrita a esse produto ou a vários produtos?
- A variável é restrita a uma região ou canal de distribuição?

Definindo-se a extensão do histórico que será usado e a periodicidade da coleta de dados, tomando o cuidado de manter como referência as variáveis mais importantes destacadas nas reuniões com os conhecedores do produto, inevitavelmente o número de variáveis se reduzirá muito.

Com isso chegamos então a um conjunto de variáveis que deve, na seqüência, ser coletado nos sistemas de informação respectivos de cada variável. Partimos então para a etapa de entender um pouco de cada variável, conhecer seu conceito, sua evolução e relação com o produto e as demais variáveis.

5.2 Dados

Os dados usados neste trabalho são frutos de um estudo realizado em uma empresa multinacional de grande porte no ramo de Saúde, Higiene e Beleza, porém para manter o sigilo da empresa os dados foram "camuflados". Assim temos um banco de dados fictício gerado a partir de dados reais, e entendemos que isso não surtirá impacto negativo sobre as conclusões do estudo. Para este estudo usamos variáveis com históricos mensais desde Janeiro de 2002 conforme as tabelas 5.1 e 5.2.

Variáveis	Relação no modelo	Tipo	Unidade
Demanda (D)	Variável Dependente	Contínua	Caixas
Despesas Publicitárias (DP)	Variável Independente	Contínua	R\$
Preço Líquido (NP)	Variável Independente	Contínua	R\$
Distribuição Numérica (DN)	Variável Independente	Contínua	%
Verbas Táticas (VT)	Variável Independente	Contínua	R\$
Inovação (I)	Variável Independente	Dummy	0 ou 1
Negociações/Promoções (N/P)	Variável Independente	Dummy	0 ou 1
Índice de Preço (PI)	Variável Independente	Contínua	%

Tabela 5.1: Detalhes sobre as variáveis.

Variáveis	Descrição
D	Volume em milhares de caixas do produto X.
DP	Gastos com publicidade realizado em milhares de R\$ realizados com TV, Rádio, Revistas, Outdoors, TV-Cabo.
NP	Valor líquido em R\$ por caixa do produto X A sigla NP tem origem do inglês, Net Price.
DN	% de Lojas, referente ao mercado em que o produto está inserido, no qual o produto está presente.
VT	Gasto para abatimentos direto no cliente, como descontos pontuais, brindes, propagandas como folders, ilhas, etc.
I	Indica 1 para os períodos após a implementação de um novo fator agregando valor no produto, zero caso contrário. Variável booleana.
N/P	Indica 1 caso no mês em questão tenha ocorrido uma promoção extra e de grande impacto ou uma negociação ou compra especial com algum cliente. Variável booleana.
PI	Índice que mostra a relação do preço do produto X em relação aos concorrentes. Ex: Variável mostrando 115 indica que o produto X está com preços 15% acima da concorrência em média. A sigla PI tem origem do inglês, Price Index.

Tabela 5.2: Descrição das variáveis.

Foi realizada uma seleção preliminar partindo de aproximadamente 100 indicadores de possíveis variáveis, que foram sendo filtrados por diversos motivos, como não haver histórico, histórico curto, histórico não ser mensal, ser pouco relevante para o estudo, ser redundante em relação a outras variáveis; assim chegamos em uma conjunto razoável de variáveis para serem testadas nos modelos conforme já descrevemos na seção anterior. Essas análises partem de uma análise dos gestores do produto: aqueles que realmente conhecem o produto têm melhores condições de indicar possíveis variáveis para um modelo. Além da compreensão destas variáveis é importante a compreensão da história do produto no mercado, assim é importante lembrar se houve alguma mudança no produto, se nesse período há algum mês em que ocorreu algum evento fora do padrão, e outras características que podem e devem ser levadas em conta no modelo.

Após selecionadas as variáveis para serem de fato estudadas no trabalho, iniciamos um conjunto de análises em relação a essas variáveis:

1. Análise e entendimento da evolução da Demanda - Graficamente
2. Análise da evolução da variável no tempo - Graficamente
3. Análise de correlação com a demanda - Diagrama de Dispersão
4. Análise de correlação entre as variáveis - Fórmula de Correlação de Pearson
5. Análise de defasagens sobre a variável - Graficamente por Diagrama de Dispersão
6. Análise de ajustes algébricos sobre a variável - Graficamente por Diagrama de Dispersão

Observando a figura 5.1 percebemos a evolução da série de que queremos compreender o comportamento, a demanda do produto gerada entre Jan/2002 até Jul/2005 mês a mês. Fica claro que existem picos onde ocorre uma venda maior em dois períodos, meados de 2002 e começo de 2004, aonde de fato ocorreram promoções e negociações mais intensamente. Nota-se que nos últimos períodos existe uma certa estabilidade que ocorreu depois de um momento de baixa nas vendas.

Na tabela 5.3 vamos analisar algumas observações a respeito das variáveis com base nos gráficos respectivos de cada variável, prevendo que sinal de coeficiente

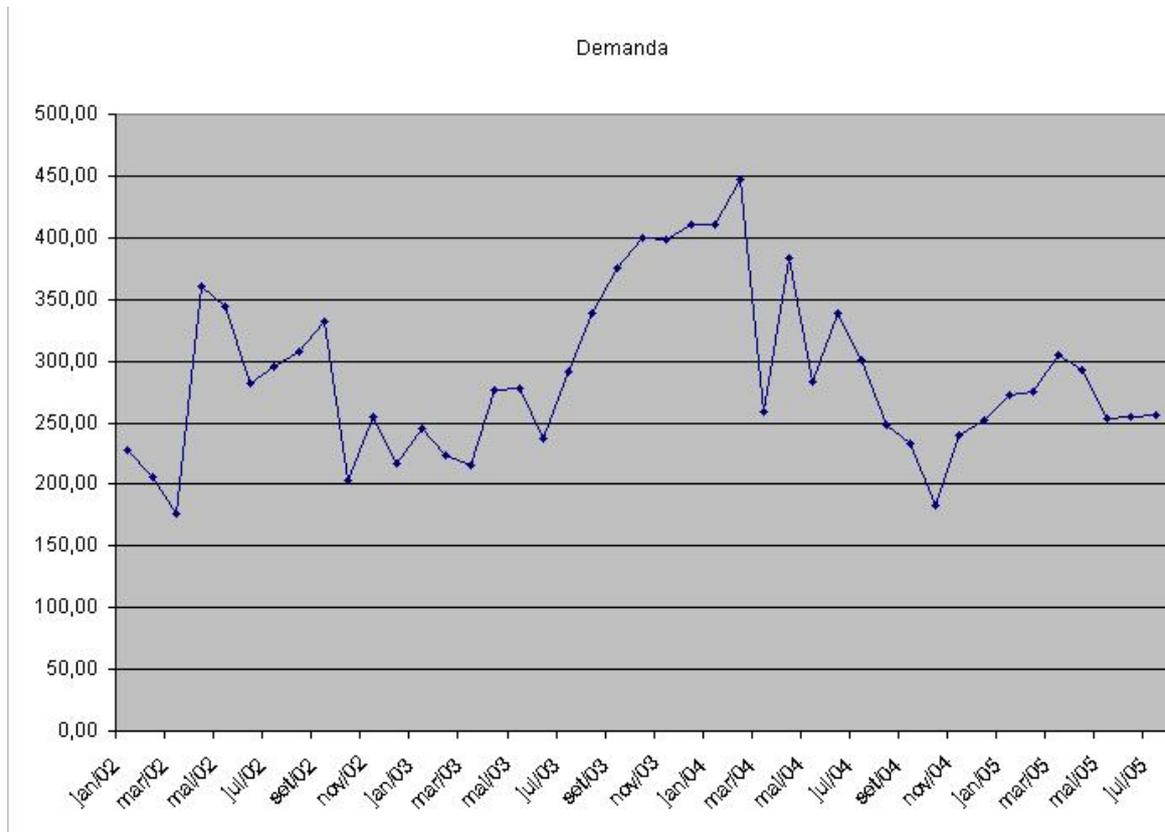


Figura 5.1: Evolução da Demanda. Fonte: Microsoft Excel.

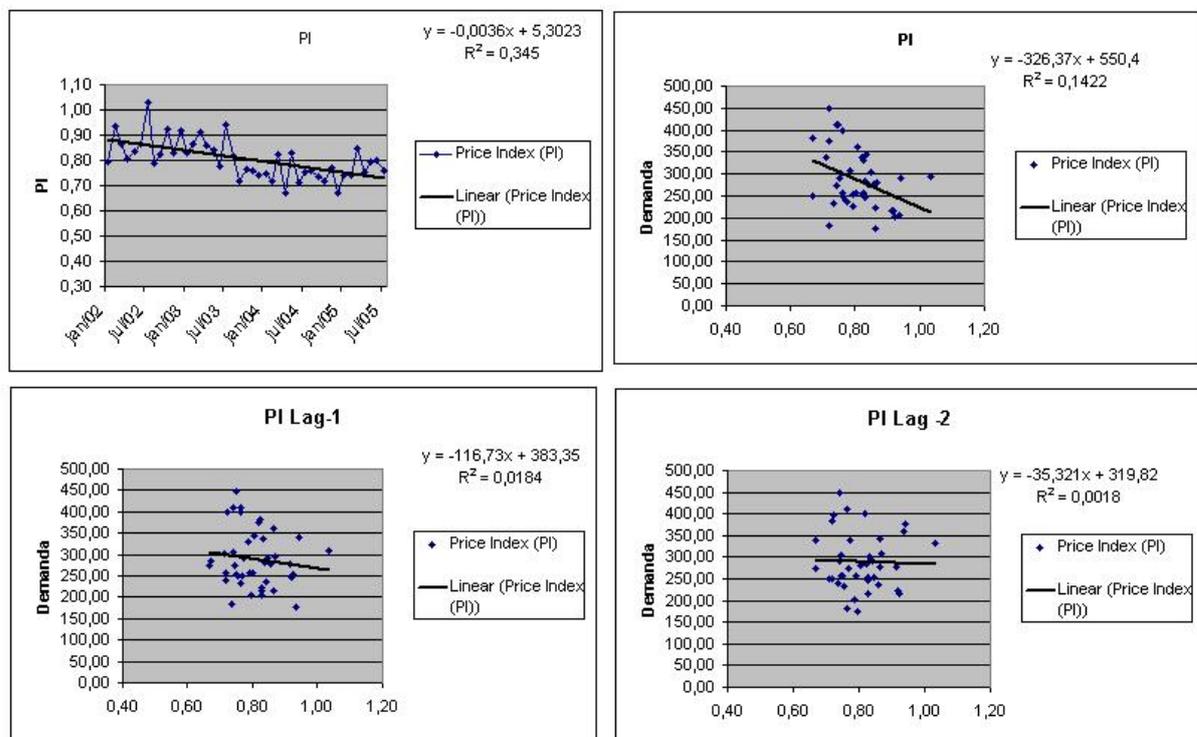


Figura 5.2: Evolução da variável Índice de Preço e sua correlação com a variável Demanda. Fonte: Microsoft Excel.

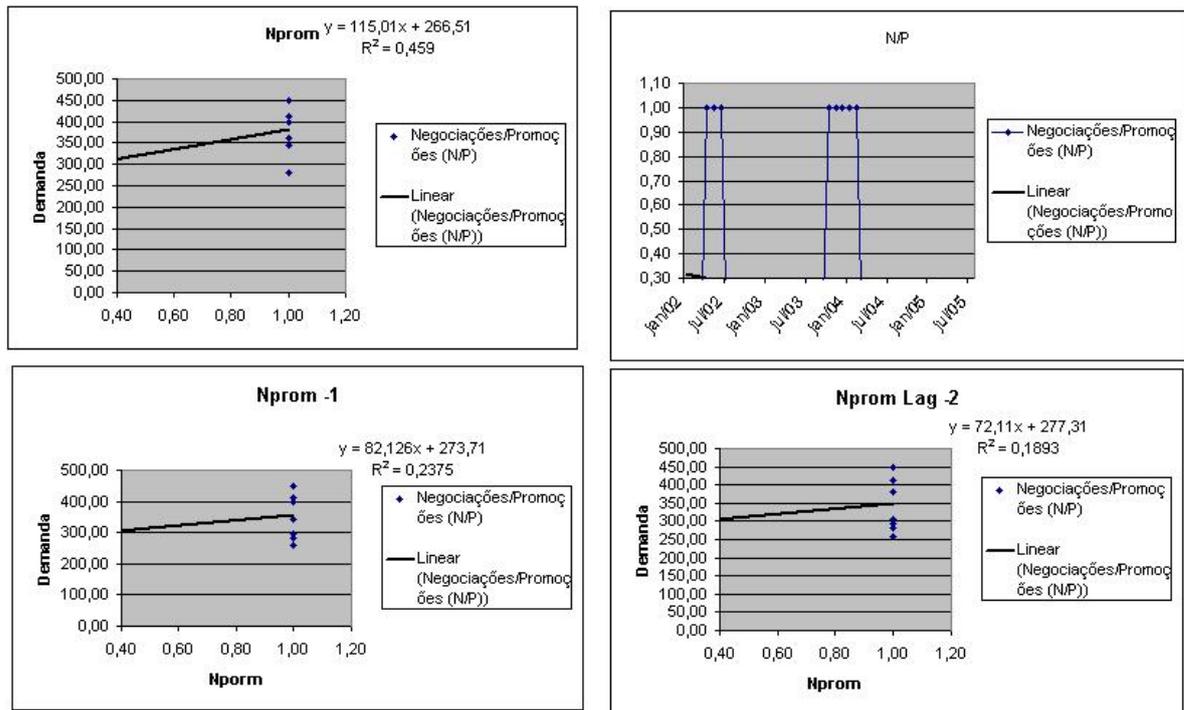


Figura 5.3: Evolução da variável Negociações e Promoções e sua correlação com a variável Demanda. Fonte: Microsoft Excel.

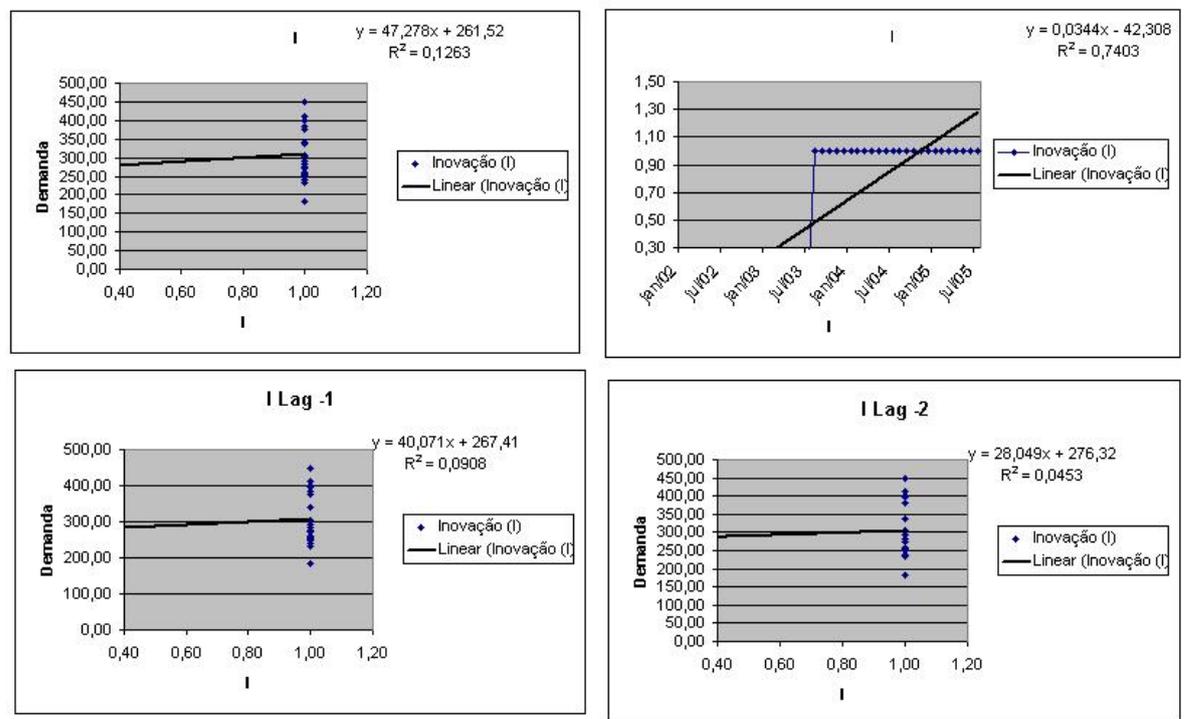


Figura 5.4: Evolução da variável Inovação e sua correlação com a variável Demanda. Fonte: Microsoft Excel.

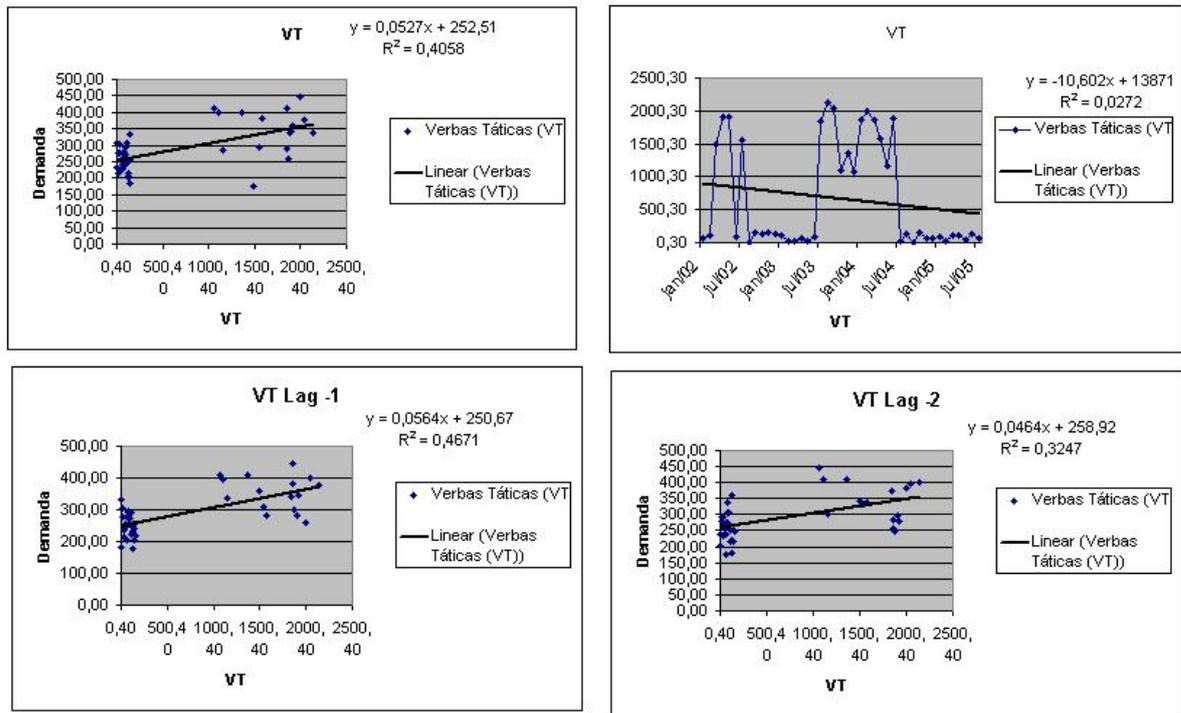


Figura 5.5: Evolução da variável Verbas Táticas e sua correlação com a variável Demanda. Fonte: Microsoft Excel.

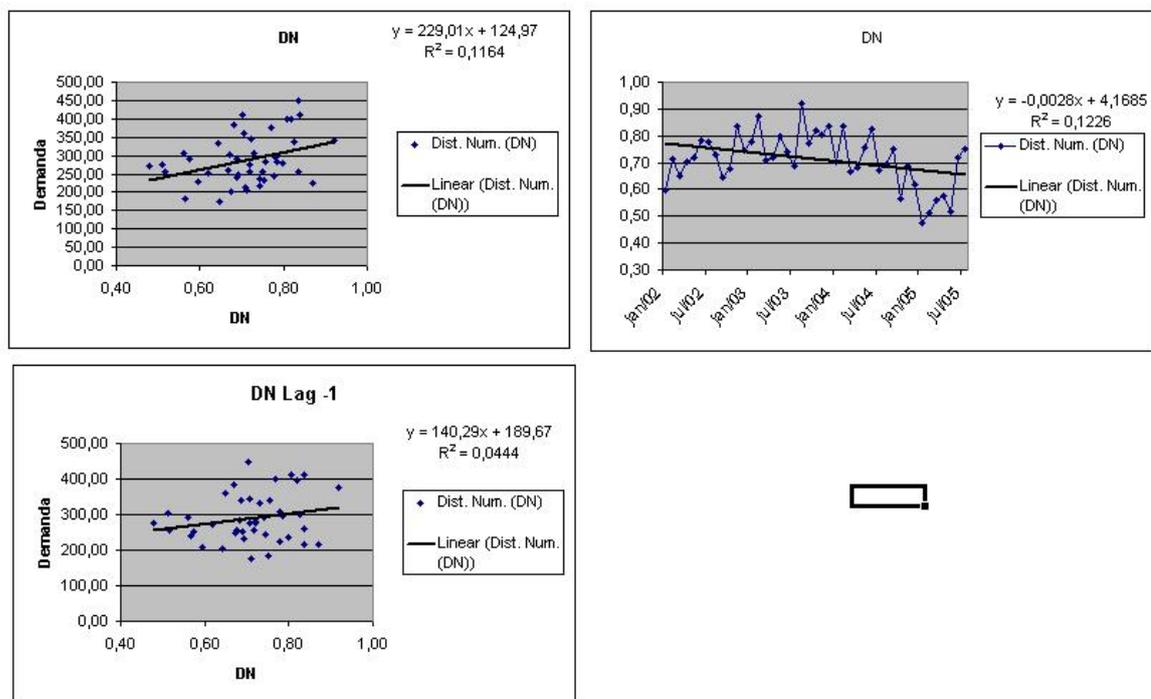


Figura 5.6: Evolução da variável Distribuição Numérica e sua correlação com a variável Demanda. Fonte: Microsoft Excel.

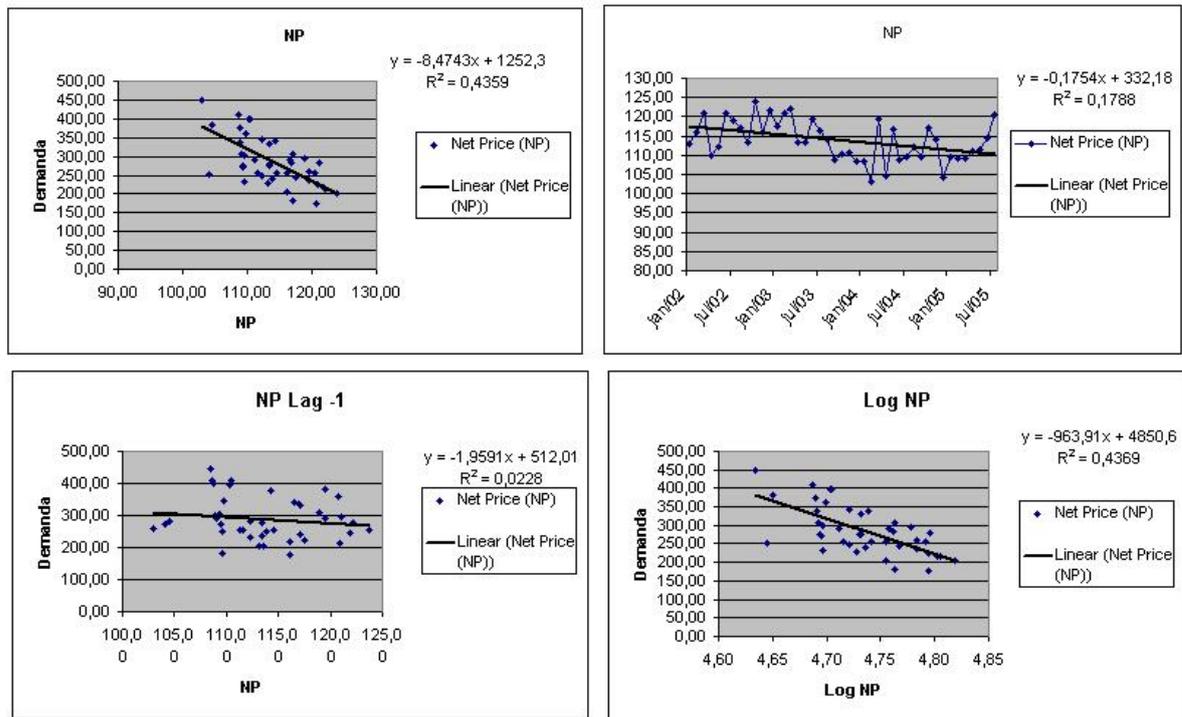


Figura 5.7: Evolução da variável Preço Líquido e sua correlação com a variável Demanda.

Fonte: Microsoft Excel.

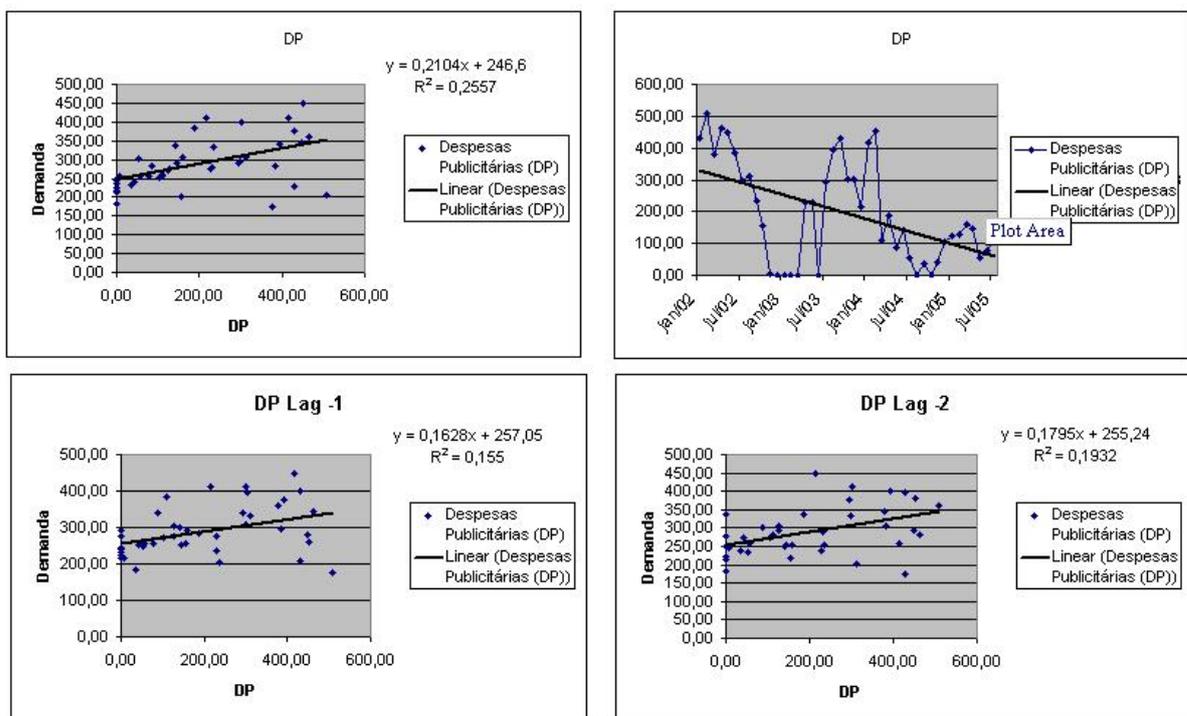


Figura 5.8: Evolução da variável Despesas Publicitárias e sua correlação com a variável Demanda. Fonte: Microsoft Excel.

esperamos para essa variável, seu grau de correlação com a demanda e algumas observações particulares. Denominaremos os graus de correlação de acordo com a seguinte escala¹:

- Correlação Baixa: $R^2 < 0,2$
- Correlação Média: $0,2 < R^2 < 0,4$
- Correlação Alta: $R^2 > 0,4$

Variáveis	Coefficiente (Positivo/Negativo)	Correlação
DP	+	Média
NP	-	Alta
DN	+	Baixa
VT	+	Alta com defasagem de 1 mês
I	+	Baixa
N/P	+	Alta
PI	-	Baixa

Tabela 5.3: Análises das variáveis.

Com base na tabela 5.3, e nos respectivos gráficos cada uma das variáveis, percebemos que as variáveis Preço Líquido (NP), Verbas Táticas (VT) e Negociações/Promoções (N/P) são as variáveis mais relevantes para o modelo. Nota-se também que tanto a distribuição numérica (DN), quanto o preço versus concorrência (PI), parecem impactar pouco na demanda, e que a inovação (I) introduzida no produto surtiu pouco efeito nas vendas. Um fato interessante é o fato da variável verbas táticas ter maior correlação quando defasada em 1 mês, indicando que provavelmente um investimento realizado no ponto de venda demora a surtir efeito sobre as vendas. Já a variável Despesas Publicitárias mostrou um impacto médio sobre as vendas, a melhor correlação ocorre quando aplicamos logaritmo sobre a variável, porém como esta variável possui alguns meses com valores muito baixos e nulos esse artifício pode distorcer as características da variável.

¹Esta escala foi criada pelo autor para ajudar neste estudo e é relativa apenas a esse conjunto de variáveis, e nos ajudará no sentido de comparar a correlação das variáveis com a demanda

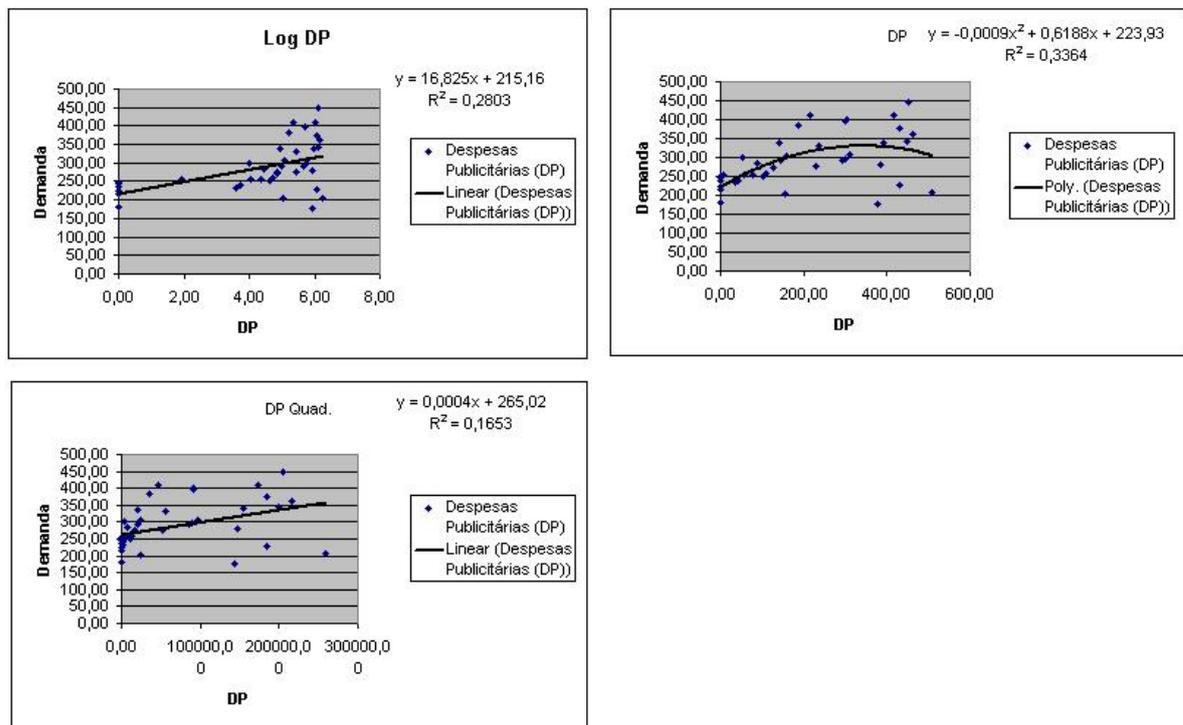


Figura 5.9: Análise da variável Log(Despesas Publicitárias) e sua correlação com a variável Demanda. Fonte: Microsoft Excel.

6 Modelos Propostos

Neste estudo realizamos diversos testes para chegarmos a um modelo adequado as condições desejadas e vamos propor inicialmente 3 tipos de modelos:

1. **Modelo de Variáveis Explicativas** - Modelo de regressão múltipla linear usando somente variáveis independentes explicativas para compreender o comportamento da demanda, ou seja, aquelas que explicam o fator de estudo; demanda do produto em questão.
2. **Modelo Misto** - Usa, em regressões múltiplas de séries temporais, tanto as variáveis independentes explicativas como componentes defasados da própria demanda para explicar o comportamento da demanda.
3. **Modelo ARMA** - Usando somente componentes autoregressivos e de média móvel da própria demanda para explicar seu comportamento.

Para montar os modelos do tipo 1 e 2 utilizamos a metodologia resumida no gráfico 6.1, com base na metodologia descrita por Hendry (1987) [7] e Spanos (1989) [14], onde deixamos as variáveis e o cenário econômico revelarem o modelo. No caso dos modelos do tipo 3, ARMA, usamos a metodologia de Box, Jenkins e Reinsel (1994) [2].

A grande dificuldade é avaliar qual modelo é realmente bom para representar um certo processo econômico, e saber o que faz um modelo ser melhor que o um outro modelo. Evidentemente, primeiro devemos avaliar as hipóteses sob as quais o modelo de regressão por mínimos quadrados está inserido; a não validade de alguma das hipóteses pode levar a um viés de especificação do modelo ou erros em sua estimação, por exemplo, pode ocorrer omissão de variáveis importantes para o modelo, ou inclusão de variáveis supérfluas, ou ainda algum erro na forma funcional do modelo. Porém, junto com isso, é importante salientar a coerência do modelo com o negócio, ou seja, tão importante quanto o modelo estar bem especificado está o fato do modelo realmente representar o que ocorre na economia.

Neste estudo entendemos que estamos razoavelmente limitados a um conjunto de dados previamente escolhidos, supostamente bem escolhidos no sentido teórico para

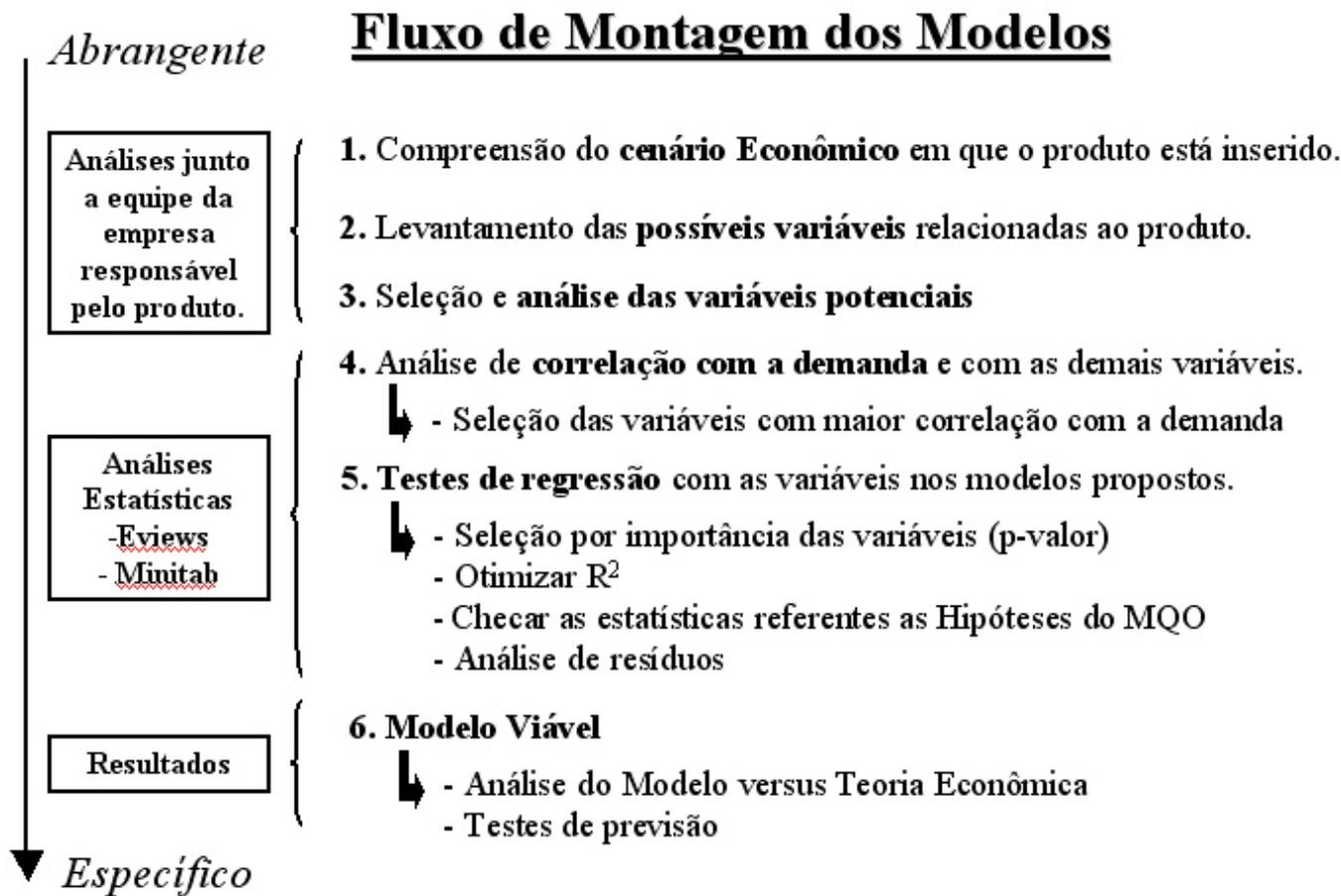


Figura 6.1: Fluxograma usado na montagem dos modelos.

representar o negócio, e que por se tratar de um trabalho de conclusão de curso, não trataremos de métodos de estimação demasiadamente sofisticados. Por esses dois motivos é possível que encontremos nos modelos perspectivas de melhorias tanto no sentido de informação, ou seja, um conjunto de dados maior e mais relevante, quanto no sentido de teoria econométrica, incrementando possivelmente os modelos com metodologias mais sofisticadas de estimação.

6.1 Modelo de Variáveis Explicativas

Usando as variáveis descritas no capítulo dedicado aos dados, levando em conta aquelas que têm maior correlação com a demanda e possíveis ajustes que possam aumentar essa correlação, aplicaremos o Método dos Mínimos Quadrados Ordinário (MQO) a fim de estimar os parâmetros da equação que melhor represente o comportamento da demanda.

Usamos as variáveis de acordo com as análises previamente realizadas individualmente sobre as variáveis, observando todas as suas características quanto a conceito, relevância e estatísticas. O objetivo agora é entender a relação conjunta das variáveis sobre a demanda, determinando qual o conjunto de variáveis que melhor a representa. Testamos uma série de modelos no E-Views e a seguir mostramos os principais testes realizados.

Modelos Testados somente com variáveis explicativas

Modelo:	Variáveis:	Coefficiente:	P-Valor:
1	<i>Constante</i>	926,22	0,0000
R^2	<i>Preço Líquido</i>	-7,25	0,0000
0,816488	<i>Distribuição Numérica</i>	138,67	0,0202
Durbin-Watson	<i>Verbas Táticas</i>	0,02	0,0062
0,956079	<i>Negociações/Promoções</i>	63,14	0,0003
Prob(F-Statistic)	<i>Despesas Publicitárias</i>	0,00	0,3659
0,000000	<i>Inovação</i>	2,65	0,8686
Observação	<i>Índice de Preço</i>	79,46	0,4655
Variáveis	0	0,00	0
com p-valor alto	0	0,00	0
2	<i>Constante</i>	962,71	0,0000
R^2	<i>Preço Líquido</i>	-8,28	0,0000
0,822553	<i>Distribuição Numérica</i>	122,86	0,0404
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0035
1,145,423	<i>Negociações/Promoções</i>	50,76	0,0041
Prob(F-Statistic)	<i>Despesas Publicitárias</i>	0,00	0,5906
0,000000	<i>Inovação</i>	3,72	0,8114
Observação	<i>Índice de Preço</i>	191,88	0,0876
VT melhor	0	0,00	0
com defasagem	0	0,00	0

Tabela 6.1: Modelos Testados somente com variáveis explicativas 1

Pensando na importância das variáveis, otimizando o R^2 e verificando a aceitabilidade das hipóteses necessárias usarmos o MQO, concluímos que o modelo 16, observado primeiramente na tabela 6.6, é adequado para representar o comportamento da demanda, já que este modelo apresenta boa aderência e variáveis estatisticamente significativas im-

Modelo:	Variáveis:	Coefficiente:	P-Valor:
3	<i>Constante</i>	1016,31	0,0000
R^2	<i>Preço Líquido</i>	-7,34	0,0000
0,817012	<i>Distribuição Numérica</i>	138,87	0,0199
Durbin-Watson	<i>Verbas Táticas</i>	0,02	0,0058
0,960653	<i>Negociações/Promoções</i>	62,96	0,0003
Prob(F-Statistic)	<i>Despesas Publicitárias</i>	0,00	0,3575
0,000000	<i>Inovação</i>	3,04	0,8487
Observação	<i>Log(Índice de Preço)</i>	72,32	0,4269
PI melhor com	0	0,00	0
logaritmo	0	0,00	0
4	<i>Constante</i>	1161,43	0,0000
R^2	<i>Preço Líquido</i>	-8,36	0,0000
0,822368	<i>Distribuição Numérica</i>	123,56	0,0394
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0036
1,158	<i>Negociações/Promoções</i>	50,11	0,0045
Prob(F-Statistic)	<i>Despesas Publicitárias</i>	0,00	0,5980
0,000000	<i>Inovação</i>	3,78	0,8091
Observação	<i>Log(Índice de Preço)</i>	159,82	0,0895
Variáveis	0	0,00	0
com p-valor alto	0	0,00	0
5	<i>Constante</i>	4457,21	0,0000
R^2	<i>Log(Preço Líquido)</i>	-937,58	0,0000
0,821011	<i>Distribuição Numérica</i>	122,75	0,0414
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0038
1,162	<i>Negociações/Promoções</i>	50,56	0,0044
Prob(F-Statistic)	<i>Despesas Publicitárias</i>	0,00	0,6085
0,000000	<i>Inovação</i>	4,23	0,7869
Observação	<i>Índice de Preço</i>	192,61	0,0884
Preço Líquido	0	0,00	0
com logaritmo	0	0,00	0

Tabela 6.2: Modelos Testados somente com variáveis explicativas 2

Modelo:	Variáveis:	Coefficiente:	P-Valor:
6	<i>Constante</i>	4700,84	0,0000
R^2	<i>Log(Preço Líquido)</i>	-948,82	0,0000
0,821044	<i>Distribuição Numérica</i>	123,45	0,0402
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0039
1,175	<i>Negociações/Promoções</i>	49,91	0,0048
Prob(F-Statistic)	<i>Despesas Publicitárias</i>	0,00	0,6123
0,000000	<i>Inovação</i>	4,33	0,7820
Observação	<i>Log(Índice de Preço)</i>	161,75	0,0880
Variáveis	0	0,00	0
com p-valor alto	0	0,00	0
7	<i>Constante</i>	975,33	0,0000
R^2	<i>Preço Líquido</i>	-8,09	0,0000
0,820082	<i>Distribuição Numérica</i>	121,07	0,0361
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0017
1,056	<i>Negociações/Promoções</i>	46,51	0,0036
Prob(F-Statistic)	<i>Índice de Preço</i>	151,45	0,0946
0,000000	0	0,00	0
Observação	0	0,00	0
PI com	0	0,00	0
coef. invertido	0	0,00	0
8	<i>Constante</i>	4401,17	0,0000
R^2	<i>Log(Preço Líquido)</i>	-918,27	0,0000
0,818555	<i>Distribuição Numérica</i>	120,63	0,0375
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0018
1,076	<i>Negociações/Promoções</i>	46,39	0,0038
Prob(F-Statistic)	<i>Índice de Preço</i>	151,65	0,0962
0,000000	0	0,00	0
Observação	0	0,00	0
NP com log	0	0,00	0
e PI com coef. invertido	0	0,00	0

Tabela 6.3: Modelos Testados somente com variáveis explicativas 3

Modelo:	Variáveis:	Coefficiente:	P-Valor:
9	<i>Constante</i>	1132,84	0,0000
R^2	<i>Preço Líquido</i>	-8,16	0,0000
0,819956	<i>Distribuição Numérica</i>	121,58	0,0353
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0017
1,066	<i>Negociações/Promoções</i>	46,05	0,0040
Prob(F-Statistic)	<i>Log(Índice de Preço)</i>	126,36	0,0961
0,000000	0	0,00	0
Observação	0	0,00	0
PI com	0	0,00	0
coef. invertido	0	0,00	0
10	<i>Constante</i>	986,23	0,0000
R^2	<i>Preço Líquido</i>	-8,39	0,0000
0,822251	<i>Distribuição Numérica</i>	120,55	0,0385
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0017
1,139	<i>Negociações/Promoções</i>	50,52	0,0037
Prob(F-Statistic)	<i>Índice de Preço</i>	182,93	0,0791
0,000000	Despesas Publicitárias	0,00	0,5177
Observação	0	0,00	0
DP com p-valor	0	0,00	0
alto	0	0,00	0
11	<i>Constante</i>	892,88	0,0000
R^2	<i>Preço Líquido</i>	-7,17	0,0000
0,827945	<i>Distribuição Numérica</i>	137,47	0,0205
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,02	0,0150
1,170	<i>Negociações/Promoções</i>	44,80	0,0048
Prob(F-Statistic)	<i>Índice de Preço</i>	93,24	0,3521
0,000000	Log(Despesas Publicitárias)	3,81	0,2143
Observação	0	0,00	0
Variáveis	0	0,00	0
com p-valor alto	0	0,00	0

Tabela 6.4: Modelos Testados somente com variáveis explicativas 4

Modelo:	Variáveis:	Coefficiente:	P-Valor:
12	<i>Constante</i>	917,78	0,0000
<i>R²</i>	<i>Preço Líquido</i>	-6,57	0,0000
0,805639	<i>Distribuição Numérica</i>	130,55	0,0293
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,02	0,0069
1,101	<i>Negociações/Promoções</i>	48,98	0,0059
Prob(F-Statistic)	<i>Despesas Publicitárias</i>	0,00	0,8199
0,000000	0	0,00	0
Observação	0	0,00	0
DP com p-valor	0	0,00	0
alto	0	0,00	0
13	<i>Constante</i>	993,02	0,0000
<i>R²</i>	<i>Preço Líquido</i>	-7,89	0,0000
0,798769	<i>Despesas Publicitárias</i>	0,00	0,5186
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0002
1,086	<i>Negociações/Promoções</i>	56,15	0,0020
Prob(F-Statistic)	<i>Índice de Preço</i>	204,38	0,0608
0,000000	0	0,00	0
Observação	0	0,00	0
DP com p-valor	0	0,00	0
alto	0	0,00	0
14	<i>Constante</i>	916,43	0,0000
<i>R²</i>	<i>Preço Líquido</i>	-5,79	0,0000
0,777828	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0009
Durbin-Watson	<i>Negociações/Promoções</i>	54,94	0,0032
1,061	<i>Despesas Publicitárias</i>	0,00	0,7670
Prob(F-Statistic)	0	0,00	0
0,000000	0	0,00	0
Observação	0	0,00	0
DP com p-valor	0	0,00	0
alto	0	0,00	0

Tabela 6.5: Modelos Testados somente com variáveis explicativas 5

Modelo:	Variáveis:	Coefficiente:	P-Valor:
15	<i>Constante</i>	890,97	0,0000
R^2	<i>Preço Líquido</i>	-7,96	0,0000
0,822794	<i>Preço Líquido (defasado 1 mês)</i>	0,84	0,4691
Durbin-Watson	<i>Distribuição Numérica</i>	98,82	0,1298
0,997	<i>Verbas Táticas (defasada 1 mês)</i>	0,03	0,0015
Prob(F-Statistic)	<i>Negociações/Promoções</i>	49,10	0,0031
0,000000	<i>Índice de Preço</i>	136,37	0,1436
Observação	0	0,00	0
NP defasado	0	0,00	0
com p-valor alto	0	0,00	0
16	<i>Constante</i>	917,42	0,0000
R^2	<i>Preço Líquido</i>	131,05	0,0000
0,805355	<i>Distribuição Numérica</i>	0,02	0,0266
Durbin-Watson	<i>Verbas Táticas (defasada 1 mês)</i>	50,52	0,0045
1,124	<i>Negociações/Promoções</i>	-6,57	0,0019
Prob(F-Statistic)	0	0,00	0
0,000000	0	0,00	0
Observação	0	0,00	0
Modelo viável	0	0,00	0
DW ruim	0	0,00	0
17	<i>Constante</i>	802,53	0,0000
R^2	<i>Distribuição Numérica</i>	97,08	0,1426
0,811463	<i>Verbas Táticas (Defasada 1 mês)</i>	0,03	0,0027
Durbin-Watson stat	<i>Negociações/Promoções</i>	53,72	0,0013
1,072	<i>Preço Líquido</i>	-6,59	0,0000
Prob(F-Statistic)	<i>Preço Líquido (Defasado 1 mês)</i>	1,23	0,2874
0,000000	0	0,00	0
Observação	0	0,00	0
NP (-1) com	0	0,00	0
p-valor alto	0	0,00	0

Tabela 6.6: Modelos Testados somente com variáveis explicativas 6

portantes para o negócio. Veja nas tabelas de modelos de variáveis explicativas de 1 a 7 os principais testes realizados no e-views e algumas estatísticas dos testes. A equação de regressão do modelo 16 é a seguinte:

$$D_t = 917,4 + 131DN_t + 0,02VT_{t-1} + 50,5N/P_t - 6,56NP_t \quad (6.1)$$

Observe na figura 6.2, 6.3 e 6.4 detalhes sobre o modelo, seus resíduos e sua previsão Ex-Post¹. Esse modelo foi estimado utilizando o Software E-Views.

Dependent Variable: DEMANDA

Method: Least Squares

Date: 01/07/06 Time: 17:31

Sample(adjusted): 2002:02 2005:07

Included observations: 42 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	917.4229	113.6495	8.072389	0.0000
DIST_NUM	131.0479	56.74000	2.309622	0.0266
VERBA_TAT_LAG_1	0.023399	0.007729	3.027247	0.0045
NEG_PROM	50.51944	15.12882	3.339285	0.0019
NET_PRICE	-6.565748	1.042277	-6.299430	0.0000
R-squared	0.805355	Mean dependent var		289.3531
Adjusted R-squared	0.784313	S.D. dependent var		66.97581
S.E. of regression	31.10500	Akaike info criterion		9.823958
Sum squared resid	35798.28	Schwarz criterion		10.03082
Log likelihood	-201.3031	F-statistic		38.27251
Durbin-Watson stat	1.124099	Prob(F-statistic)		0.000000

Figura 6.2: Detalhes do Modelo 16 somente com variáveis independentes explicativas.

Fonte: E-Views.

O modelo apresentou um coeficiente de determinação de 82% e possui variáveis não todas conjuntamente nula, pelo teste F. Observe os gráficos dos resíduos, o correlo-

¹A Previsão Ex-Post é aquela em que usamos o modelo para estimar o próprio passado e assim comparar essa previsão com os dados reais ocorridos, na prática aplicamos os valores do histórico das variáveis independentes sobre a equação gerada no modelo e calculamos uma demanda média para cada período, que é então comparada com a demanda real ocorrida no mesmo período.

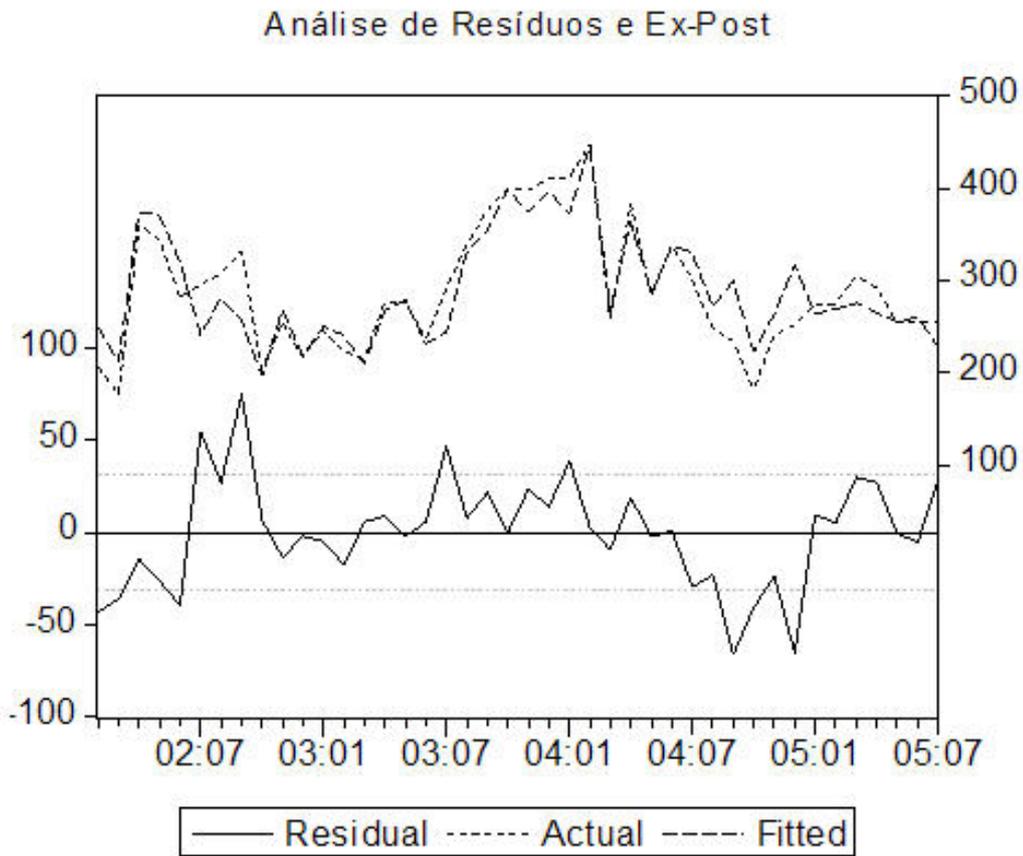


Figura 6.3: Análise de resíduos e Ex-Post do modelo 16 somente com variáveis explicativas.

Fonte: E-Views.

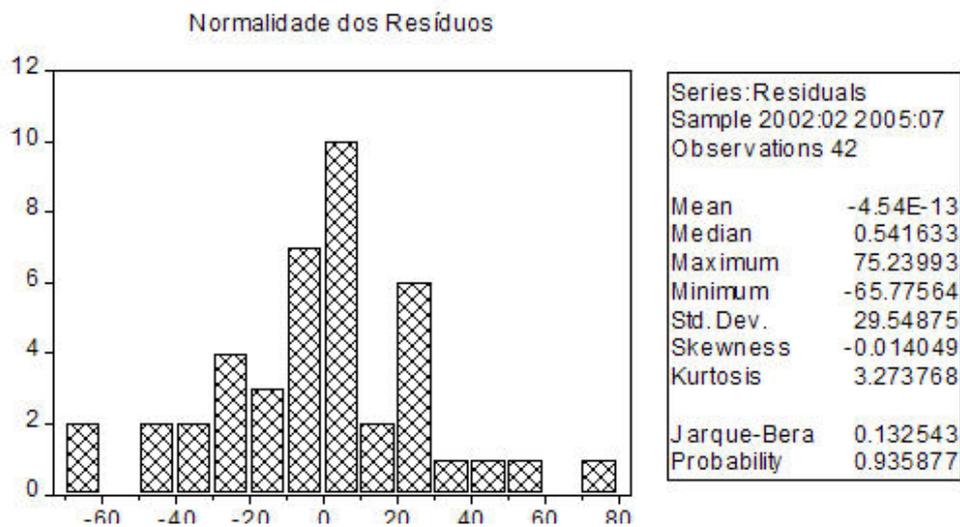


Figura 6.4: Teste de Normalidade dos resíduos do Modelo 16 apenas com variáveis explicativas. Fonte: E-Views.

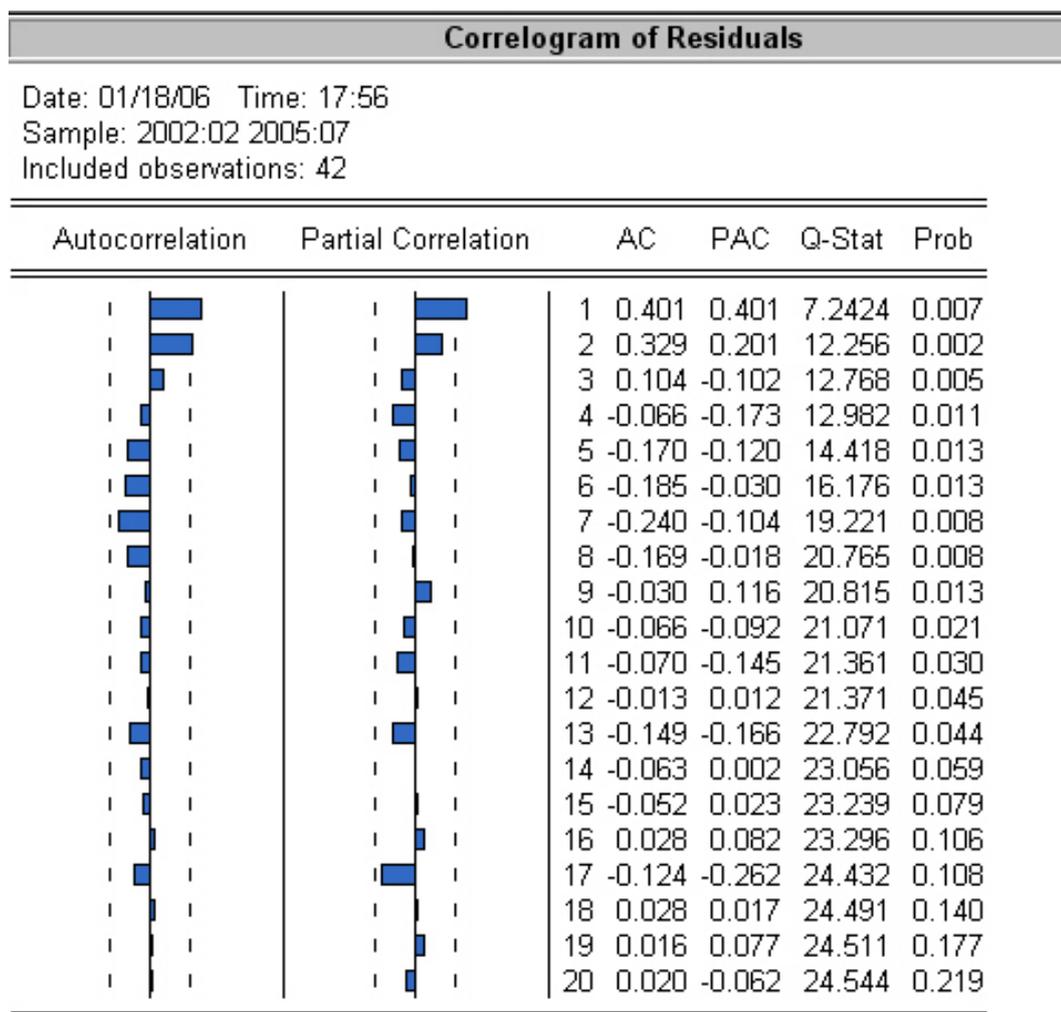


Figura 6.5: Correlograma dos resíduos do Modelo 16 apenas com variáveis explicativas.

Fonte: E-Views.

grama (figura 6.5) e o teste de Durbin Watson, e observe a figura 6.2: neste caso, no teste Durbin-Watson temos um valor de 1,12 e temos 4 variáveis explicativas e 42 períodos, e ao olharmos na tabela de referência para o teste² podemos notar que este valor é ligeiramente menor que a estatística d_l de limite inferior para autocorrelação positiva, o que pode estar indicando uma possível presença de autocorrelação de primeira ordem nos resíduos; este fato também pode ser percebido notando que os resíduos não estão dispostos de maneira aparentemente aleatória, conforme a figura 6.3. Para complementar, o correlograma, da figura 6.5, nos mostra que a função de autocorrelação parcial tem componentes fora do intervalo de confiança, e que de fato há indícios de autocorrelação serial.

Quanto às variáveis, sob um nível de 95% de significância, todas passam no teste t e podemos rejeitar a hipótese de serem nulas. São elas:

- Preço Líquido - Com coeficiente negativo, indicando que um preço mais baixo resulta em aumento de vendas, compreendemos que esta variável é muito importante para o comportamento da demanda.
- Negociações/Promoções - Com coeficiente positivo, indicando o aumento médio que pode causar uma promoção ou negociação especial nas vendas.
- Distribuição Numérica - Com coeficiente positivo, indicando que uma presença maior nas lojas resulta em aumento de vendas.
- Verbas Táticas - Com coeficiente positivo, indicando o aumento médio causado para cada real gasto com verbas no ponto de vendas sobre as vendas do produto, porém repare que esta variável está no modelo com uma defasagem de 1 mês, indicando que esse tipo de atividade promocional demora a surtir efeito sobre as vendas, em média 1 mês.

Quanto às demais variáveis:

- Despesas Publicitárias - a variável se adaptou melhor no modelo após aplicarmos seu logaritmo, porém como esta variável possui alguns meses com valor muito baixo, ou nulo, não há muito sentido em se aplicar o logaritmo, e quando não usamos o logaritmo seu p-valor ficou muito alto. OBS: Nos modelos Lin-Log o coeficiente da

²Encontrada em [6]

variável independente com logaritmo pode ser representado por:

$$\beta = \frac{\Delta Y}{\frac{\Delta X}{X}} \quad (6.2)$$

onde: $\frac{\Delta X}{X}$ representa a variação relativa de X e ΔY a variação absoluta de Y.

- Índice de Preço (Preço versus Concorrência) - Nos modelos em que testamos o Índice de Preço junto com o Preço Líquido, esta variável apresentou coeficiente positivo, o que não faz sentido econômico. Porém, retirando a variável Preço Líquido (NP) do modelo, a variável Índice de Preço (PI) ganha sentido, mas não é tão relevante para explicar o comportamento da demanda quanto o Preço Líquido.
- Inovação - Esta variável não apresentou boa correlação com a demanda, seu p-valor ficou alto em todos os testes, indicando que este evento de melhoria no produto não representa aumento significativo na demanda.

Estimamos este modelo também no Minitab, apenas para efeito de comparação entre os softwares; observe os resultado na figura 6.6 e seus resíduos em 6.7. Há pouca ou nenhuma diferença entre as estimativas realizadas pelo e-views.

Este tipo de modelo é conveniente quando temos poder de ação sobre as variáveis independentes e quando queremos justamente realizar mudanças no comportamento de uma variável independente e entender qual será o impacto da mudança sobre a demanda do produto. Este modelo não é o mais adequado para estimar simples previsões de demanda, já que será necessário também estimar com precisão a previsão de cada uma das variáveis independentes, que pode ser tanto ou mais difícil que estimar a previsão da própria demanda, porém pode sugerir uma demanda média supondo prováveis cenários para as variáveis independentes. Esse modelo ajuda o gerente do negócio a entender o comportamento de suas ações sobre o produto, sugerindo como ele deve regular o preço do produto, suas promoções e verbas, podendo com isso otimizar a venda do produto. Porém, estatisticamente, percebemos que o melhor modelo encontrado nos testes realizados possui problemas nos resíduos, indicando possíveis problemas de especificação. Ainda assim achamos que este modelo pode ser muito útil na compreensão das variáveis econômicas que afetam a demanda.

Regression Analysis: Demanda versus Net Price (NP); Dist. Num. (DN); ...

The regression equation is

$$\text{Demanda} = 917 - 6,55 \text{ Net Price (NP)} + 130 \text{ Dist. Num. (DN)} + 0,0235 \text{ VT Lag}_1 + 50,5 \text{ Negociações/Promoções (N/P)}$$

42 cases used, 1 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	916,5	113,8	8,06	0,000
Net Price (NP)	-6,553	1,042	-6,29	0,000
Dist. Num. (DN)	130,13	56,78	2,29	0,028
VT Lag_1	0,023468	0,007732	3,04	0,004
Negociações/Promoções (N/P)	50,47	15,15	3,33	0,002

S = 31,1370 R-Sq = 80,5% R-Sq(adj) = 78,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	148044	37011	38,17	0,000
Residual Error	37	35872	970		
Total	41	183916			

Source	DF	Seq SS
Net Price (NP)	1	82579
Dist. Num. (DN)	1	32728
VT Lag_1	1	21980
Negociações/Promoções (N/P)	1	10756

Durbin-Watson statistic = 1,12227

Figura 6.6: Modelo de Variáveis Explicativas: Modelo 16. Fonte: Minitab

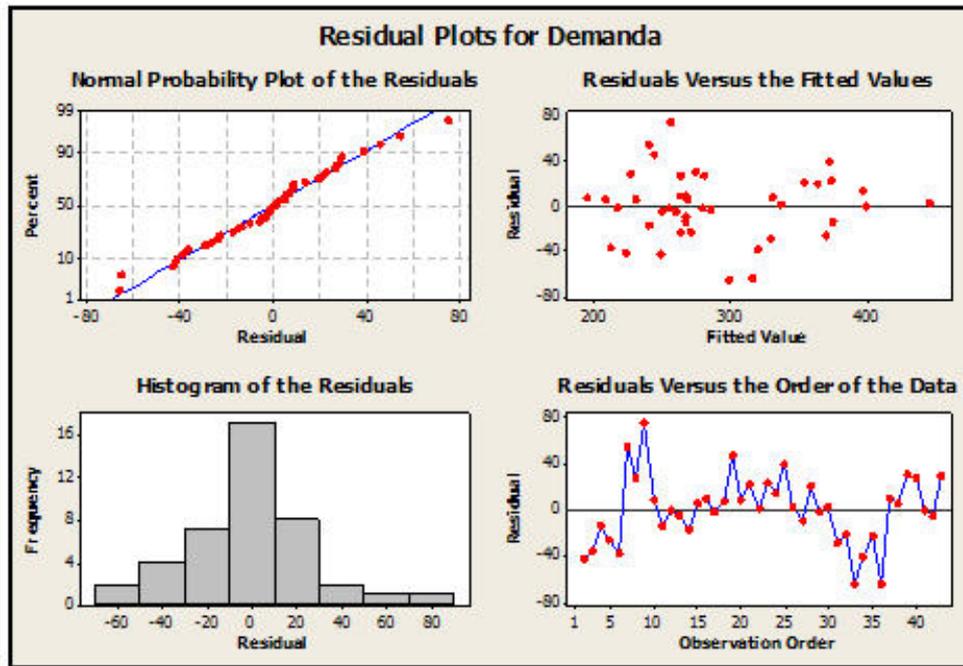


Figura 6.7: Modelo de Variáveis Explicativas: Resíduos do Modelo 16. Fonte: Minitab

6.2 Modelo Misto

Olhando para o modelo descrito na seção anterior notamos que este tem uma boa aplicabilidade na empresa porém lembramos que neste modelo havia uma possível presença de autocorrelação nos resíduos. No modelo que vamos propor agora visamos justamente tentar corrigir este problema. O problema da autocorrelação pode indicar que provavelmente estamos omitindo alguma variável relevante para o modelo ou que esta faltando alguma informação no modelo. Para sanar essa possível falta de informação vamos incluir como uma das variáveis explicativas o passado da própria demanda, já que esta pode trazer informações relevantes sobre o seu comportamento no futuro. Denominamos esse modelo de misto pois este inclui tanto variáveis explicativas quanto a própria demanda.

Seguem nas tabelas abaixo (6.7, 6.8, 6.9, 6.10, 6.11 e 6.12) os principais testes realizados para os Modelos Mistos.

De acordo com os testes realizados podemos perceber que realmente incluir a demanda defasada em 1 mês como variável explicativa melhora a aderência do modelo e reduz ligeiramente o problema de autocorrelação. Dentre os modelos testados um que podemos destacar como forte candidato para representar a demanda é o modelo 8, que aparece na tabela 6.9. Este modelo rejeita a hipótese de autocorrelação entre os resíduos

Modelos Mistos

Modelo:	Variáveis:	Coefficiente:	P-Valor:
1	<i>Constante</i>	875,95	0,0000
R^2	<i>Demanda (Defasada 1 mês)</i>	0,19	0,0529
0,824851	<i>Distribuição Numérica</i>	145,09	0,0123
Durbin-Watson stat	<i>Verbas Táticas (Defasada 1 mês)</i>	0,01	0,0916
1,439	<i>Negociações/Promoções</i>	42,12	0,0086
Prob(F-Statistic)	<i>Preço Líquido</i>	-6,70	0,0000
0,000000			
Observação			
Modelo de Var. Exp. + Demanda defasada			
2	<i>Constante</i>	835,98	0,0000
R^2	<i>Demanda (Defasada 1 mês)</i>	0,11	0,2857
0,834601	<i>Demanda (Defasada 2 meses)</i>	0,11	0,2051
Durbin-Watson stat	<i>Distribuição Numérica</i>	134,18	0,0199
1,331	<i>Verbas Táticas (Defasada 1 mês)</i>	0,01	0,0884
Prob(F-Statistic)	<i>Negociações/Promoções</i>	43,52	0,0061
0,000000	<i>Preço Líquido</i>	-6,37	0,0000
Observação			
Defasagem de 2 meses não é significativa			

Tabela 6.7: Modelos Mistos testados 1

Modelo:	Variáveis:	Coefficiente:	P-Valor:
3	<i>Constante</i>	934,47	0,0000
R^2	<i>Demanda (Defasada 1 mês)</i>	0,19	0,0428
0,840255	<i>Distribuição Numérica</i>	135,12	0,0163
Durbin-Watson stat	<i>Verbas Táticas (Defasada 1 mês)</i>	0,02	0,0449
1,451	<i>Negociações/Promoções</i>	37,87	0,0153
Prob(F-Statistic)	<i>Preço Líquido</i>	-8,26	0,0000
0,000000	<i>Índice de Preço</i>	154,92	0,0747
Observação Inclusão do Índice de Preço mostrou coef. invertido			
4	<i>Constante</i>	872,23	0,0000
R^2	<i>Demanda (Defasada 1 mês)</i>	0,21	0,0389
0,828231	<i>Distribuição Numérica</i>	145,01	0,0128
Durbin-Watson stat	<i>Verbas Táticas (Defasada 1 mês)</i>	0,01	0,1819
1,426	<i>Negociações/Promoções</i>	35,57	0,0453
Prob(F-Statistic)	<i>Preço Líquido</i>	-6,73	0,0000
0,000000	<i>Despesas Publicitárias</i>	0,00	0,4122
Observação Inclusão de DP pouco significativa			
5	<i>Constante</i>	1003,17	0,0000
R^2	<i>Demanda (Defasada 1 mês)</i>	0,25	0,0199
0,834542	<i>Distribuição Numérica</i>	133,56	0,0201
Durbin-Watson stat	<i>Verbas Táticas (Defasada 1 mês)</i>	0,02	0,0782
1,513	<i>Negociações/Promoções</i>	34,05	0,0399
Prob(F-Statistic)	<i>Preço Líquido</i>	-7,79	0,0000
0,000000	<i>Inovação</i>	-19,30	0,1611
Observação Inclusão de Inovação pouco significativa			

Tabela 6.8: Modelos Mistos testados 2

Modelo:	Variáveis:	Coefficiente:	P-Valor:
6	<i>Constante</i>	853,67	0,0000
R^2	<i>Demanda (Defasada 1 mês)</i>	0,22	0,0089
0,842409	<i>Distribuição Numérica</i>	137,08	0,0111
Durbin-Watson stat	<i>Negociações/Promoções</i>	41,08	0,0067
1,325	<i>Preço Líquido</i>	-6,55	0,0000
Prob(F-Statistic)	<i>Verbas Táticas</i>	0,02	0,0102
0,000000			
Observação			
VT sem defasagem mais significativo			
7	<i>Constante</i>	824,82	0,0000
R^2	<i>Demanda (Defasada 1 mês)</i>	0,16	0,0812
0,848558	<i>Distribuição Numérica</i>	131,36	0,0149
Durbin-Watson stat	<i>Negociações/Promoções</i>	42,55	0,0050
1,250	<i>Preço Líquido</i>	-6,32	0,0000
Prob(F-Statistic)	<i>Verbas Táticas</i>	0,02	0,0155
0,000000	<i>Demanda (Defasada 2 meses)</i>	0,09	0,3060
Observação			
Defasagem de 2 meses não é significativa			
8	<i>Constante</i>	269,09	0,0620
R^2	<i>Demanda (Defasada 1 mês)</i>	0,55	0,0000
0,894792	<i>Negociações/Promoções</i>	42,22	0,0007
Durbin-Watson stat	<i>Preço Líquido</i>	-6,91	0,0000
1,916	<i>Preço Líquido (Defasada 1 mês)</i>	5,55	0,0000
Prob(F-Statistic)	<i>Verbas Táticas</i>	0,01	0,0238
0,000000			
Observação			
Modelo Viável			

Tabela 6.9: Modelos Mistos testados 3

Modelo:	Variáveis:	Coefficiente:	P-Valor:
9	<i>Constante</i>	503,95	0,0924
R^2	<i>Demanda (Defasada 1 mês)</i>	0,28	0,1307
0,909042	<i>Negociações/Promoções</i>	59,96	0,0022
Durbin-Watson stat	<i>Preço Líquido</i>	-5,69	0,0000
1,657	<i>Preço Líquido (Defasada 1 mês)</i>	3,01	0,0878
Prob(F-Statistic)	<i>Verbas Táticas</i>	0,02	0,0053
0,000000	<i>Demanda (Defasada 12 meses)</i>	-0,05	0,6108
Observação			
Defasagem de 1 ano pouco significativa			
10	<i>Constante</i>	862,36	0,0001
R^2	<i>Negociações/Promoções</i>	71,50	0,0002
0,899759	<i>Preço Líquido</i>	-5,76	0,0000
Durbin-Watson stat	<i>Preço Líquido (Defasada 1 mês)</i>	0,76	0,4147
1,474	<i>Verbas Táticas</i>	0,03	0,0000
Prob(F-Statistic)	<i>Demanda (Defasada 12 meses)</i>	-0,12	0,1492
0,000000			
Observação			
Variáveis com p-valor alto			
11	<i>Constante</i>	952,74	0,0000
R^2	<i>Negociações/Promoções</i>	67,34	0,0002
0,897001	<i>Preço Líquido</i>	-5,76	0,0000
Durbin-Watson stat	<i>Verbas Táticas</i>	0,03	0,0000
1,500	<i>Demanda (Defasada 12 meses)</i>	-0,14	0,0960
Prob(F-Statistic)			
0,000000			
Observação			
Modelo Aparentemente Viável			

Tabela 6.10: Modelos Mistos testados 4

Modelo:	Variáveis:	Coefficiente:	P-Valor:
12	<i>Constante</i>	949,85	0,0000
R^2	<i>Negociações/Promoções</i>	66,57	0,0010
0,897036	<i>Preço Líquido</i>	-5,75	0,0000
Durbin-Watson stat	<i>Verbas Táticas</i>	0,03	0,0001
1,507	<i>Demanda (Defasada 12 meses)</i>	-0,14	0,1075
Prob(F-Statistic)	<i>Demanda (Defasada 1 mês)</i>	0,01	0,9274
0,000000			
Observação			
Variáveis com			
p-valor alto			
13	<i>Constante</i>	848,52	0,0000
R^2	<i>Negociações/Promoções</i>	74,80	0,0000
0,869624	<i>Preço Líquido</i>	-4,76	0,0000
Durbin-Watson stat	<i>Verbas Táticas</i>	0,04	0,0000
1,253	<i>Demanda (Defasada 6 meses)</i>	-0,17	0,0072
Prob(F-Statistic)			
0,000000			
Observação			
D(-6) melhora o			
modelo			
14	<i>Constante</i>	832,16	0,0000
R^2	<i>Negociações/Promoções</i>	62,07	0,0010
0,876761	<i>Preço Líquido</i>	-4,87	0,0000
Durbin-Watson stat	<i>Verbas Táticas</i>	0,03	0,0000
1,553	<i>Demanda (Defasada 1 mês)</i>	0,12	0,1900
Prob(F-Statistic)	<i>Demanda (Defasada 6 meses)</i>	-0,18	0,0044
0,000000			
Observação			
Variáveis com			
p-valor alto			

Tabela 6.11: Modelos Mistos testados 5

Modelo:	Variáveis:	Coefficiente:	P-Valor:
15	<i>Constante</i>	1078,35	0,0000
<i>R²</i>	<i>Negociações/Promoções</i>	57,01	0,0005
0,919791	<i>Preço Líquido</i>	-6,31	0,0000
Durbin-Watson stat	<i>Verbas Táticas</i>	0,03	0,0000
1,814	<i>Demanda (Defasada 6 meses)</i>	-0,15	0,0133
Prob(F-Statistic)	<i>Demanda (Defasada 12 meses)</i>	-0,19	0,0159
0,000000			
Observação			
Modelo			
Aparentemente Viável			
16	<i>Constante</i>	1035,85	0,0000
<i>R²</i>	<i>Negociações/Promoções</i>	58,95	0,0007
0,908392	<i>Verbas Táticas (Defasada 1 mês)</i>	0,03	0,0000
Durbin-Watson stat	<i>Demanda (Defasada 6 meses)</i>	-0,21	0,0030
2,293	<i>Demanda (Defasada 12 meses)</i>	-0,18	0,0378
Prob(F-Statistic)	<i>Preço Líquido</i>	-5,84	0,0000
0,000000			
Observação			
17	<i>Constante</i>	874,44	0,0000
<i>R²</i>	<i>Negociações/Promoções</i>	75,86	0,0000
0,889928	<i>Preço Líquido</i>	-5,22	0,0000
Durbin-Watson stat	<i>Verbas Táticas</i>	0,03	0,0000
1,336	<i>Demanda (Defasada 13 meses)</i>	-0,09	0,2346
Prob(F-Statistic)			
0,000			
Observação			

Tabela 6.12: Modelos Mistos testados 6

e possui a melhor aderência já testada para estes dados, ele inclui além da defasagem da demanda uma defasagem de um mês sobre a variável Preço Líquido. A equação de regressão do modelo 8 é a seguinte:

$$D_t = 269,1 + 0,55D_{t-1} + 42,2N/P_t - 6,9NP_t + 5,55NP_{t-1} + 0,01VT_t \quad (6.3)$$

Observe nas figuras 6.8, 6.9, 6.10 e 6.11 detalhes sobre o modelo.

Dependent Variable: DEMANDA

Method: Least Squares

Date: 01/08/06 Time: 17:37

Sample(adjusted): 2002:02 2005:07

Included observations: 42 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	269.0894	139.6698	1.926612	0.0620
DEMANDA(-1)	0.553204	0.090230	6.131052	0.0000
NEG_PROM	42.21501	11.41113	3.699459	0.0007
NET_PRICE	-6.914555	0.764971	-9.038975	0.0000
NET_PRICE(-1)	5.548891	1.036276	5.354646	0.0000
VERBA_TAT	0.013052	0.005532	2.359407	0.0238
R-squared	0.894792	Mean dependent var		289.3531
Adjusted R-squared	0.880180	S.D. dependent var		66.97581
S.E. of regression	23.18367	Akaike info criterion		9.256336
Sum squared resid	19349.36	Schwarz criterion		9.504575
Log likelihood	-188.3831	F-statistic		61.23616
Durbin-Watson stat	1.916045	Prob(F-statistic)		0.000000

Figura 6.8: Detalhes do Modelo Misto 8. Fonte: E-Views.

Quanto às variáveis temos:

- Demanda defasada em 1 mês - Esta variável mostra o quanto da demanda será gerada no mês seguinte a partir do que ocorreu no mês anterior. Por exemplo, caso tenhamos realizado uma venda de 100 unidades no mês anterior, o modelo já nos diz que teremos ao menos 55 unidades de vendas no mês seguinte(já que o coeficiente é de aproximadamente 0,55).

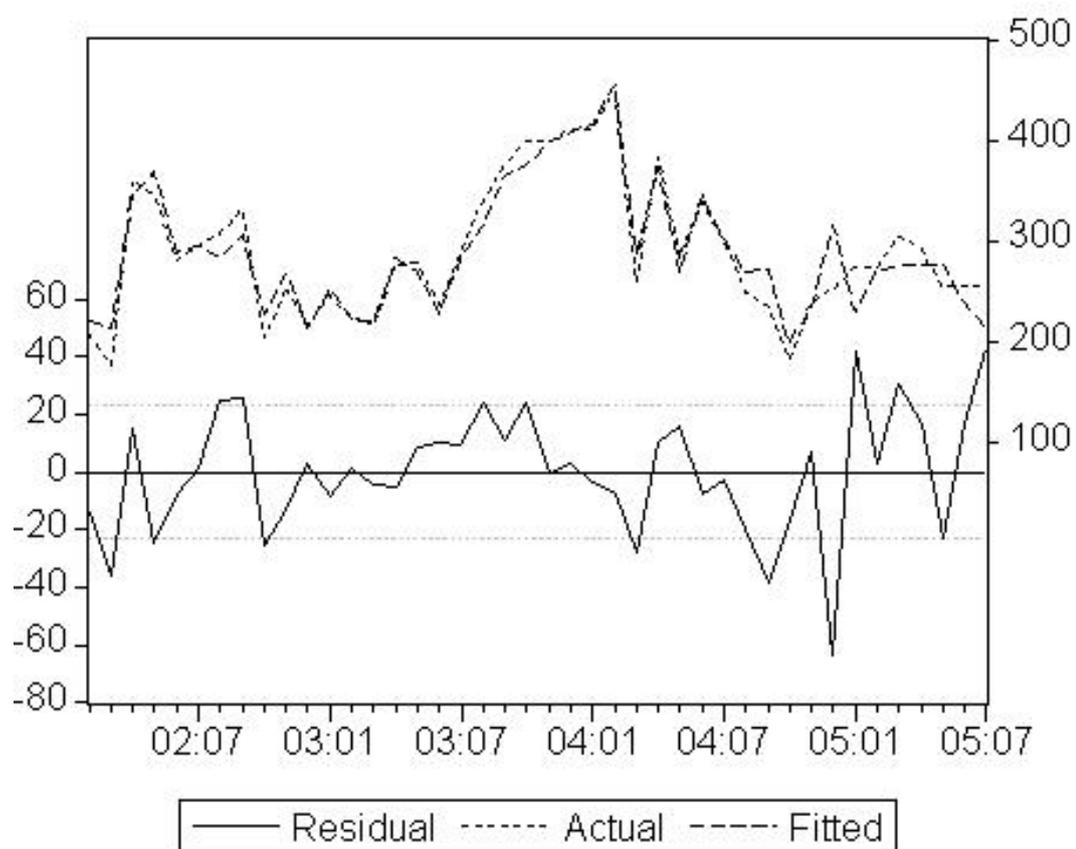


Figura 6.9: Detalhes dos resíduos do Modelo Misto 8. Fonte: E-Views.

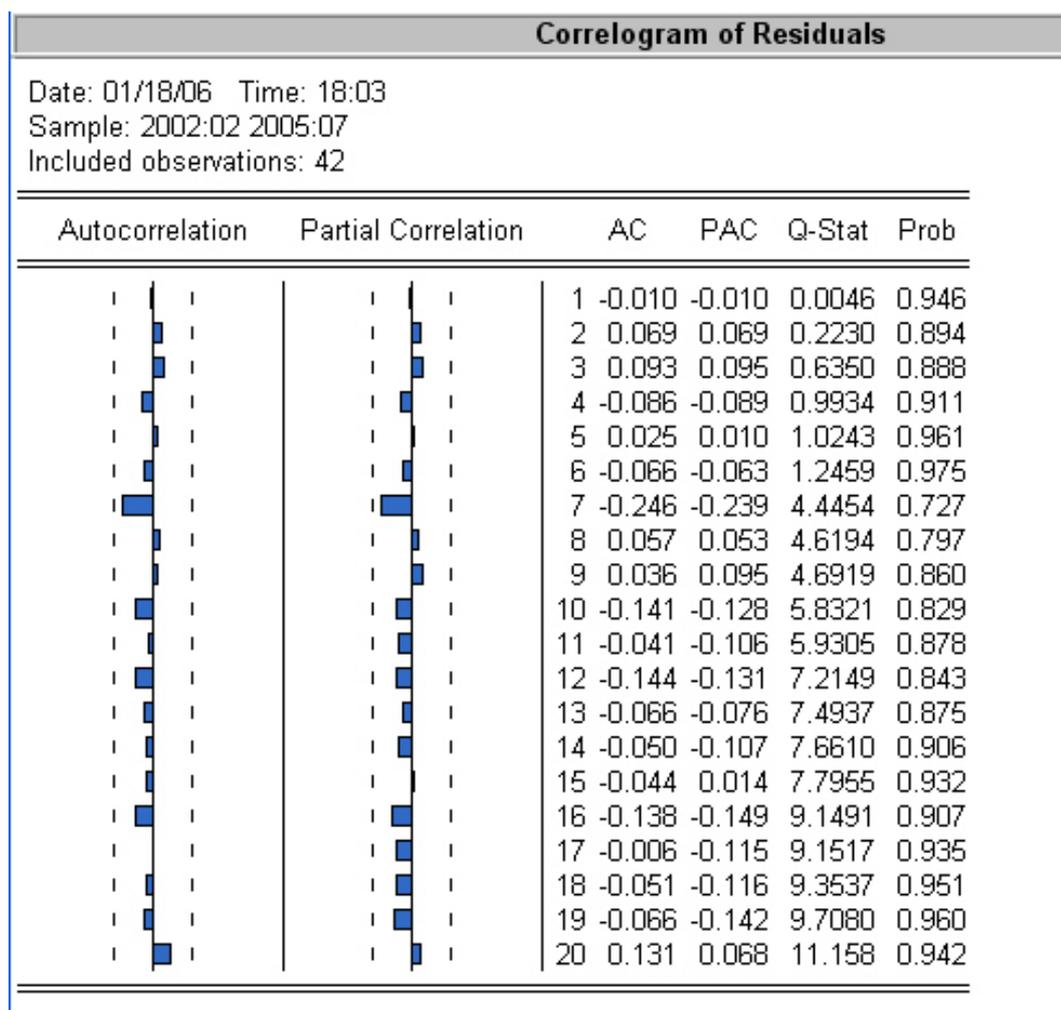


Figura 6.10: Correlograma dos resíduos do Modelo Misto 8. Fonte: E-Views.

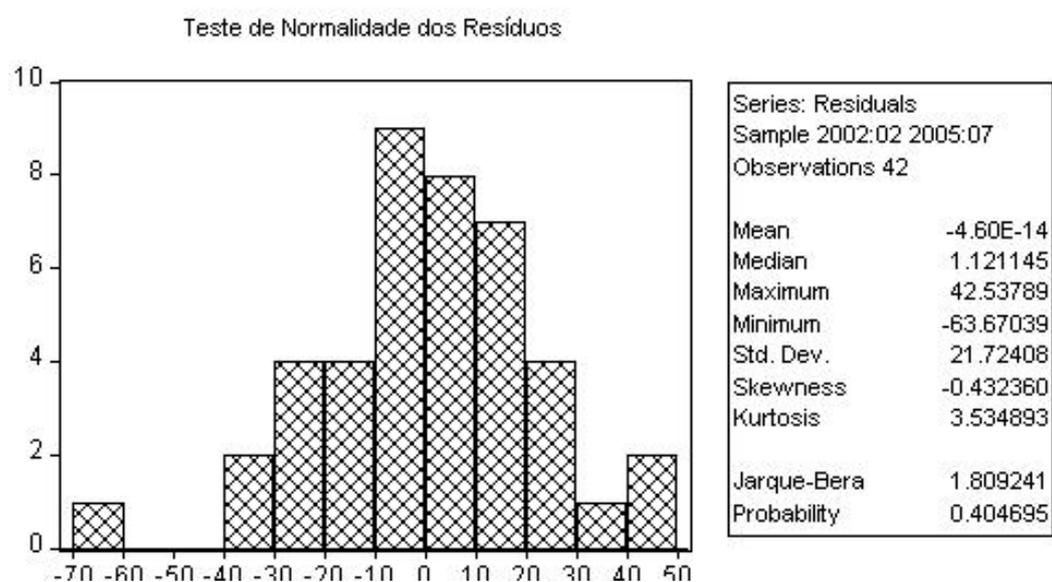


Figura 6.11: Teste de normalidade dos resíduos do Modelo Misto 8. Fonte: E-Views.

- Preço Líquido - Com coeficiente negativo, indicando que um preço mais baixo resulta em aumento de vendas, compreendemos que esta variável é muito importante para o comportamento da demanda. Neste modelo em especial esta variável é complementada pela sua defasagem em 1 mês, o Preço Líquido (-1), conforme podemos observar na figura 6.8, e possui coeficiente positivo. Essa combinação entre as duas variáveis mostra um fenômeno econômico interessante, indicando como ocorre o impacto sobre a demanda de uma possível variação no preço do produto. Por exemplo, neste modelo se ocorrer uma redução de preços seguida de uma nova redução o impacto não será tão grande no aumento na demanda quanto se realizar uma única redução de preço. Possivelmente isso reflete um comportamento do consumidor frente a mudanças de preços do produto.
- Negociações/Promoções - Com coeficiente positivo, indicando o aumento médio que pode causar uma promoção ou negociação especial nas vendas.
- Verbas Táticas - Com coeficiente positivo, indicando o aumento médio causado para cada real gasto com verbas no ponto de vendas sobre as vendas do produto, porém repare que esta variável que estava no modelo de variáveis explicativas com uma defasagem de 1 mês, agora aparece com melhor significância sem a defasagem, possivelmente pois esse efeito de defasagem agora é captado pela defasagem da demanda.

Quanto às outras variáveis:

- Despesas Publicitárias - Mostrou-se pouco significativa nos testes.
- Preço versus Concorrência - Nos modelos em que testamos o Índice de Preço junto com o Preço Líquido esta variável apresentou coeficiente positivo, o que não faz sentido econômico. Porém retirando a variável Preço Líquido do modelo esta variável ganha sentido, mas não é tão relevante para explicar o comportamento da demanda quanto o Preço Líquido.
- Inovação - Esta variável não apresentou boa correlação com a demanda, seu p-valor ficou alto em todos os testes, indicando que este evento de melhoria no produto não representou aumento significativo na demanda.

- Defasagens da Demanda - Ao incluirmos no modelo defasagens de 6 meses e 12 meses percebemos que existe uma certa sazonalidade semestral e anual, porém a inclusão dessas defasagens mostrou-se ainda pior que a defasagem de 1 mês, por isso não foram incluídas no modelo final.

Entendemos que este modelo mostra-se muito útil do ponto de vista de planejamento de produção e ação gerencial sobre as vendas do produto. Este modelo oferece a possibilidade de serem realizadas previsões de vendas para o mês seguinte, com base em um possível cenário das variáveis Preço, Verbas Táticas e Negociações e Promoções.

6.3 Modelos ARMA

Na seção anterior chegamos a um modelo conveniente para análises econômicas, porém dependente de realizações de previsões de outras variáveis independentes da demanda. Agora pretendemos propor um modelo que ajude simplesmente a estimar a média de vendas prevista para o mês seguinte usando como base apenas a própria demanda. Dessa forma esse modelo não será afetado por possíveis erros na estimação das previsões de variáveis como o Preço, Verbas Táticas e outras.

Para chegarmos neste modelo vamos utilizar a metodologia descrita por Box, Jenkins e Reinsel (1994)[2]. Assim, observando a figura 6.12, podemos analisar o correlograma da variável Demanda.

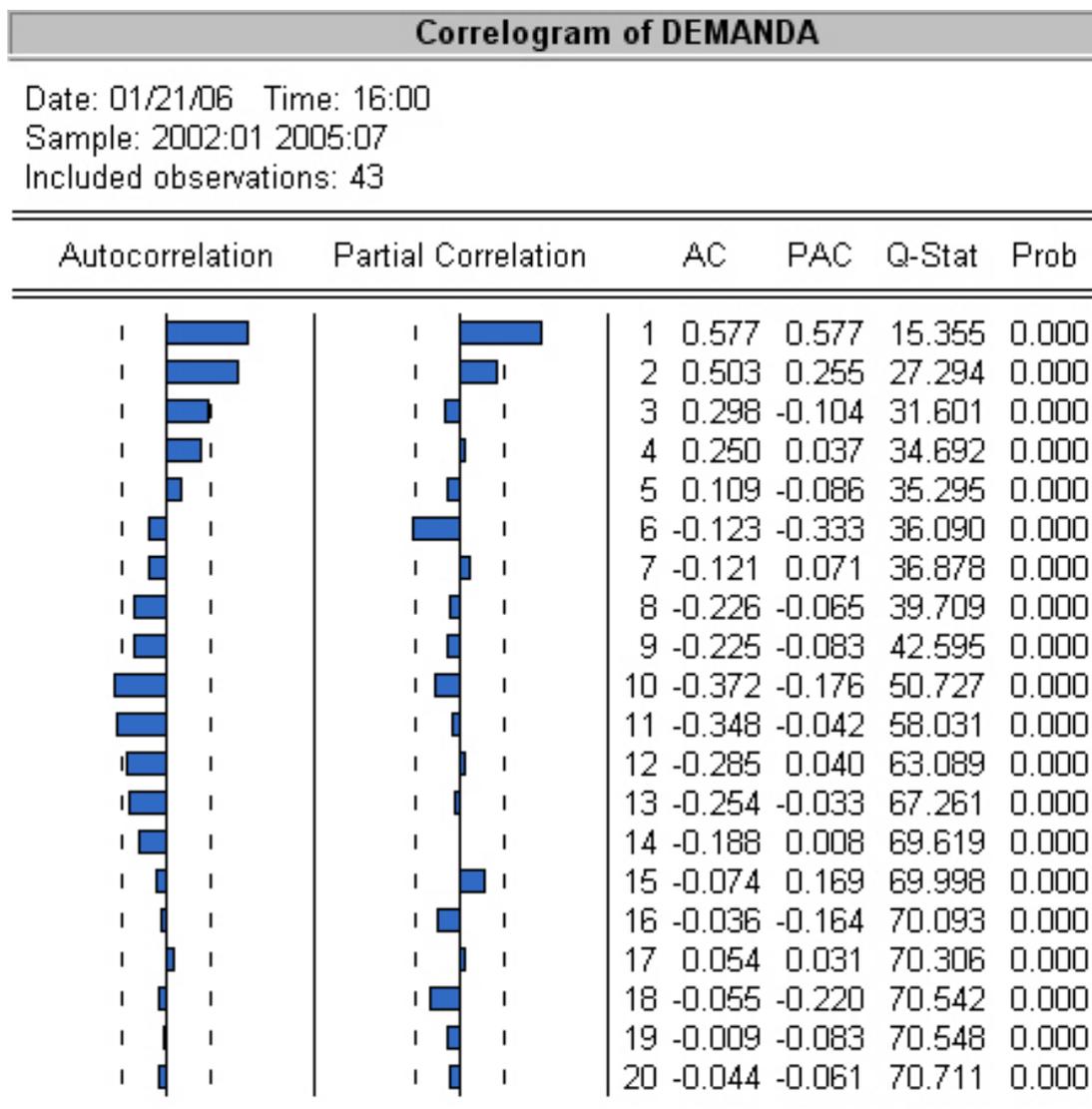


Figura 6.12: Correlograma da Demanda. Fonte: E-Views.

Com base no correlograma, podemos perceber que os coeficientes de autocorrelação amostral da função de autocorrelação parcial (FACP) têm valores significativos apenas até a sexta defasagem, e a função de autocorrelação (FAC) decai para zero com padrão de onda senoidal amortecida, indicando que a série possa ser descrita por um AR(6). Observe na figura 6.13 detalhes em relação a regressão do modelo AR(6) completo.

Dependent Variable: DEMANDA
 Method: Least Squares
 Date: 01/31/06 Time: 21:00
 Sample(adjusted): 2002:07 2005:07
 Included observations: 37 after adjusting endpoints
 Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	287.1282	17.79792	16.13269	0.0000
AR(1)	0.380161	0.162608	2.337895	0.0263
AR(2)	0.495136	0.169498	2.921191	0.0066
AR(3)	-0.092512	0.173976	-0.531754	0.5988
AR(4)	0.093007	0.158877	0.585405	0.5627
AR(5)	0.020516	0.150590	0.136236	0.8925
AR(6)	-0.324534	0.139439	-2.327432	0.0269
R-squared	0.597857	Mean dependent var		291.4791
Adjusted R-squared	0.517429	S.D. dependent var		65.82681
S.E. of regression	45.72814	Akaike info criterion		10.65196
Sum squared resid	62731.89	Schwarz criterion		10.95673
Log likelihood	-190.0613	F-statistic		7.433396
Durbin-Watson stat	1.945920	Prob(F-statistic)		0.000061
Inverted AR Roots	.87+.28i	.87-.28i	.07+.76i	.07-.76i
	-.75-.33i	-.75+.33i		

Figura 6.13: Modelo AR(6) completo. Fonte: E-Views.

Podemos observar na figura 6.13 que os valores-p das defasagens 3, 4 e 5 da demanda estão acima de 5%, e portanto não podemos rejeitar a hipótese nula de seus coeficientes. Assim vamos propor um modelo AR(6) incompleto, que irá incluir as defasagens 1, 2 e 6, conforme descrito na figura 6.14.

01.jpg

Dependent Variable: DEMANDA
 Method: Least Squares
 Date: 01/08/06 Time: 17:45
 Sample(adjusted): 2002:07 2005:07
 Included observations: 37 after adjusting endpoints
 Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	286.5565	16.90448	16.95151	0.0000
AR(1)	0.358010	0.142851	2.506187	0.0173
AR(2)	0.503544	0.147000	3.425471	0.0017
AR(6)	-0.292987	0.107040	-2.737162	0.0099
R-squared	0.590166	Mean dependent var		291.4791
Adjusted R-squared	0.552908	S.D. dependent var		65.82681
S.E. of regression	44.01505	Akaike info criterion		10.50875
Sum squared resid	63931.72	Schwarz criterion		10.68290
Log likelihood	-190.4118	F-statistic		15.84012
Durbin-Watson stat	1.906215	Prob(F-statistic)		0.000001
Inverted AR Roots	.87 -.28i -.73 -.36i	.87+.28i -.73+.36i	.04+.72i	.04 -.72i

Figura 6.14: Detalhes do Modelo AR(6) incompleto. Fonte: E-Views.

Analisando o correlograma dos resíduos, descritos na figura 6.15, notamos que não há valores significativos para os coeficientes de autocorrelação e autocorrelação parcial, e que, portanto, os resíduos são um ruído branco. Pela metodologia de Box, Jenkins e Reinsel (1994) [2] chegamos em um modelo apropriado para descrever a demanda deste produto.

Detalhamos nas figuras 6.16 e 6.17 o gráfico da evolução dos resíduos, a previsão ex-post e um histograma dos resíduos.

Com isso a equação que define o modelo resultante é:

$$D_t = 286,5 + 0,358D_{t-1} + 0,5D_{t-2} - 0,29D_{t-6} \quad (6.4)$$

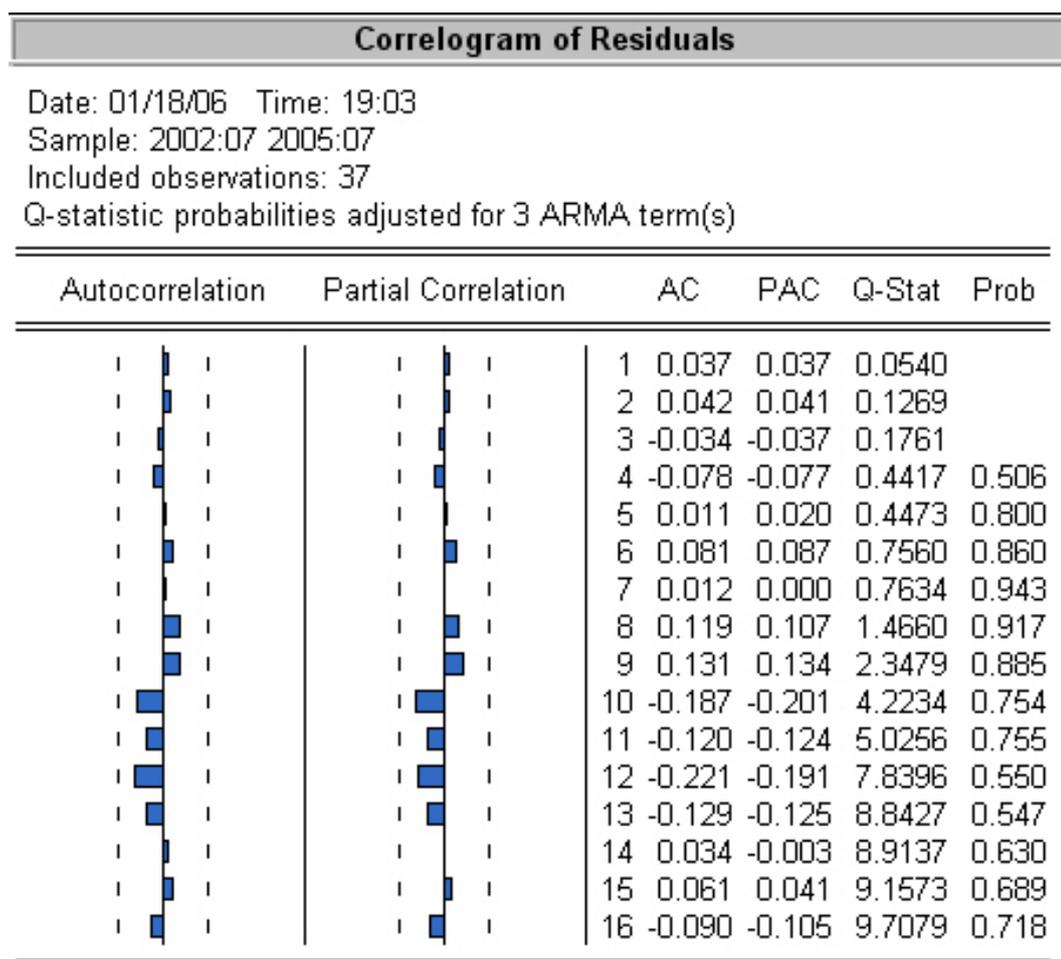


Figura 6.15: Correlograma dos resíduos do Modelo AR(6) incompleto. Fonte: E-Views.

Assim fica claro que há um comportamento com ciclos semestrais, e que a demanda também é fortemente influenciada pela demanda dos dois últimos meses que à antecedem.

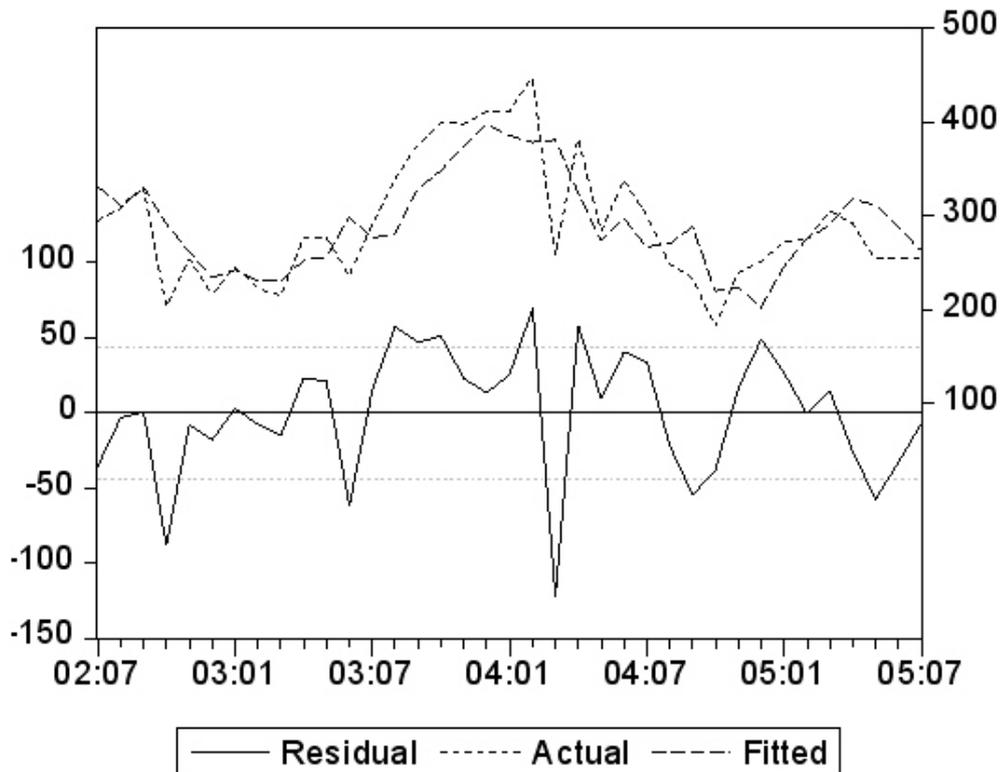


Figura 6.16: Detalhes dos resíduos do Modelo AR(6) incompleto. Fonte: E-Views.

6.4 Melhor Modelo

Destacamos neste capítulo 3 tipos de modelos. O primeiro, somente com variáveis explicativas, mostrou um forte problema de autocorrelação dos resíduos. Apresentamos como solução o modelo que denominamos de misto, e este mostrou ser um bom modelo para este conjunto de dados, resultando em uma boa aderência e é efetivo para otimizar e prever as vendas do produto usado como exemplo. Para complementar o estudo implementamos o modelo do tipo ARMA, que também mostrou ser um modelo adequado para representar a demanda deste produto e também muito útil para entendermos o comportamento e evolução desta demanda, mostrando, por exemplo, ciclos de sazonalidade do produto bimestrais e semestrais.

Para decidirmos qual dos dois modelos, misto ou autoregressivo, é mais adequado para representar a demanda deste produto, podemos usar os critérios de seleção de modelos de Akaike e Schwarz. Tanto usando o critério de seleção de modelos de Akaike (AIC), quanto o critério de Schwarz, o modelo misto mostrou ser melhor modelo para representar a demanda deste produto em relação ao modelo autoregressivo, já que a es-

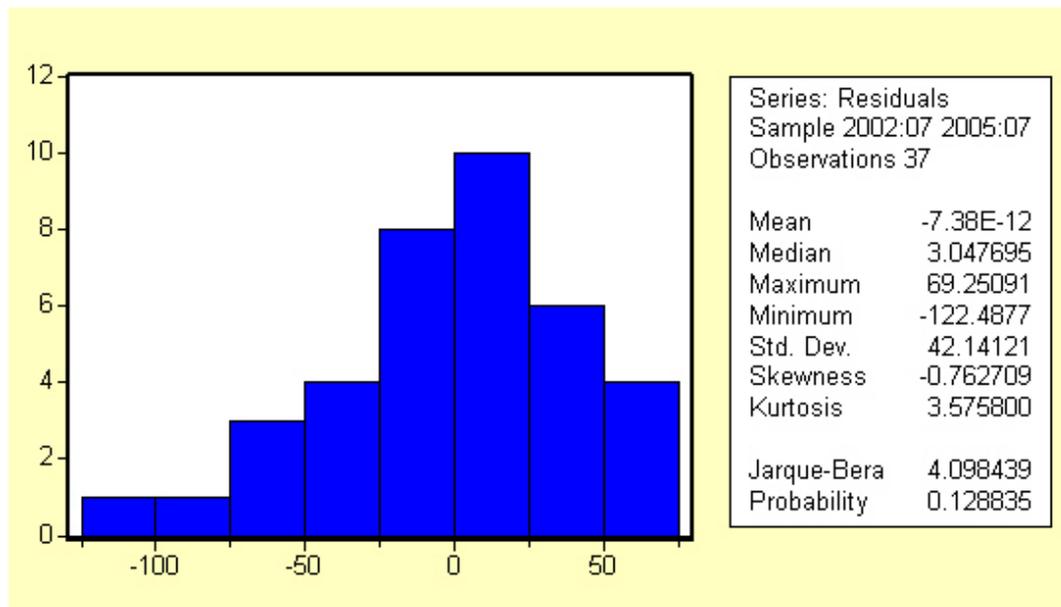


Figura 6.17: Teste de normalidade dos resíduos do Modelo AR(6) incompleto. Fonte: E-Views.

tatística AIC e SIC do modelo misto são menores que do modelo autoregressivo (observe nas figuras 6.8 e 6.14 os critérios de seleção de AIC e SIC).

7 Discussão e Recomendações

Durante o desenvolvimento deste trabalho surgiram questões de aplicabilidade dos modelos e possíveis melhorias que iremos discutir a seguir.

7.1 Aplicabilidade

Quanto ao uso dos modelos

Um problema levantado durante estudo, e ainda não apresentado, é referente ao real uso dos modelos. Entendemos que, na prática, os modelos devam ser utilizados em uma plataforma que permita simulações rápidas e também que já estejam incorporados no dia a dia da empresa. Assim, transportar os modelos para o Excel, por exemplo, pode facilitar seu uso, bem como softwares de análise de risco, como o Crystal Ball, podem ser úteis em simulações. Não detalhamos neste estudo a implementação dos modelos em Excel, e deixamos a cargo do leitor realizar essa parte de modo a adaptar da melhor forma possível um modelo à sua respectiva empresa.

Atualização dos Modelos

Evidentemente o mercado em que um produto está inserido é muito dinâmico, portanto um modelo bom hoje pode ser ruim daqui alguns meses. Porém, pela dificuldade e tempo dispendido na montagem de um modelo, não é viável remontarmos um modelo mês a mês, especialmente se tratarmos de mais de um produto e mais de um modelo. Portanto entendemos que devemos re-estimar os modelos conforme a volatilidade do mercado no qual o produto está inserido, no caso de um produto razoavelmente estável, pode ser trimestralmente ou semestralmente.

7.2 Melhorias

Viés dos Modelos

Neste estudo sempre nos baseamos em estimadores pelo Método dos Mínimos Quadrados Ordinários, porém esse método talvez não seja o mais indicado para estimarmos regressões sobre séries temporais, apesar disso ocorrer freqüentemente na literatura e ser comum no mercado. O fato é que as hipóteses iniciais do MQO, em especial a hipótese de que os resíduos não tem correlação com as variáveis independentes, podem não estar sendo cumpridas, especialmente no caso em que incluímos como variável independente uma componente defasada da variável dependente. Esse problema pode gerar um possível viés e inconsistência dos parâmetros estimados. Porém esse método de estimação é bastante prático e mostrou-se muito adequado para o estudo, mas entendemos que isso nos permite a possibilidade de aprimorar o método de estimação dos modelos.

Com base em algumas referências levantadas deixamos como sugestão de estudo futuro implementar modelos semelhantes aos descritos neste trabalho usando Funções de Transferência e Estimadores por Máxima Verossimilhança, que possivelmente possam vir a corrigir os problemas acima descrito. Alguns detalhes podem ser encontrados no Wooldridge (2002) [15] e no Enders (2003) [4].

8 Conclusões

Análises econométricas ainda são raramente usadas em empresas como ferramenta de auxílio no planejamento de vendas, com isso entendemos que a proposta de modelagem do trabalho é inovadora e aprimora as atuais técnicas existentes de previsão de vendas.

Nosso principal objetivo neste trabalho era mostrar a teoria e exemplificar a construção de um modelo baseado em regressões lineares por Mínimos Quadrados Ordinários. Conseguimos no estudo descrever, nas revisões bibliográficas, detalhes da teoria que servem como base para a construção dos modelos. Mostramos também o processo prático, envolvendo tanto o lado empresarial, de coleta de dados e interpretações dos resultados, quanto a parte técnica, de implementação das regressões nos softwares Minitab e E-Views e o processo de análise estatística dos modelos.

Quanto aos resultados encontrados a partir do banco de dados usado como exemplo para o estudo, destacamos 3 tipos de modelos. O primeiro, somente com variáveis explicativas, mostrou um forte problema de autocorrelação dos resíduos. Apresentamos como solução o modelo que denominamos de misto, e este mostrou ser o modelo mais adequado para este conjunto de dados, resultando em uma boa aderência e efetividade para otimizar e prever as vendas do produto usado como exemplo. Para complementar o estudo implementamos o modelo do tipo ARMA, que no entanto mostrou ser um modelo pior que o misto ao utilizamos critérios de seleção de modelos (AIC e SIC). Porém o modelo autoregressivo mostrou-se muito útil para entendermos o comportamento da demanda, mostrando, por exemplo, ciclos de sazonalidade do produto.

Considerando este conjunto de dados, destacamos, portanto, a equação 6.3 como o melhor modelo para definir o comportamento descrito pela demanda deste produto. Além disso, podemos destacar como algumas conseqüências importantes dos modelos descritos em relação a este produto:

- O preço é realmente um fator determinante nas vendas deste produto, porém esse produto mostrou-se com preço levemente independente do preço concorrente, já que a variável Índice de Preço não se mostrou tão relevante para representar a demanda do produto. Entendemos, portanto, que uma variação na concorrência

pouco afetaria este produto, enquanto que uma variação em seu próprio preço irá mudar o comportamento das vendas.

- As despesas gastas com publicidade não foram efetivas e surtiram pouco ou nenhum impacto sobre as vendas.
- A inovação realizada no produto também não surtiu muito efeito sobre as vendas.
- Negociações, promoções e verbas no ponto de venda podem ser, e de fato mostraram-se, relevantes na alavancagem de vendas.
- As vendas de um mês possuem relação com as vendas ocorridas no mês anterior, no bimestre anterior e no semestre anterior, indicando um certo grau de sazonalidade do produto.

Podemos observar que o estudo e o modelo mostram-se úteis nas análises e projeções de vendas de um produto, e resultam em informações importantes para melhorar a efetividade das vendas. Este trabalho também pode ajudar estatísticos que queiram implementar modelos econométricos nos processos de produção e vendas de grandes empresas.

9 Anexo: Dados Brutos

Meses	Demanda	DP	NP	DN	VT	I	N/P	PI
Jan-02	227,30	429,40	113,11	0,60	62,37	0,00	0,00	0,79
Feb-02	206,44	508,51	116,16	0,71	120,59	0,00	0,00	0,93
Mar-02	176,22	377,93	120,78	0,65	1495,24	0,00	0,00	0,86
Apr-02	360,28	463,94	109,80	0,71	1917,94	0,00	1,00	0,81
May-02	343,97	447,43	112,29	0,72	1901,27	0,00	1,00	0,84
Jun-02	281,19	384,00	121,05	0,78	85,21	0,00	1,00	0,87
Jul-02	295,51	298,44	118,90	0,78	1547,57	0,00	0,00	1,03
Aug-02	307,50	310,65	117,09	0,73	2,53	0,00	0,00	0,79
Sep-02	331,53	234,90	113,52	0,64	143,20	0,00	0,00	0,83
Oct-02	203,23	155,31	123,79	0,67	134,51	0,00	0,00	0,92
Nov-02	254,36	6,85	116,14	0,84	149,86	0,00	0,00	0,83
Dec-02	216,52	0,00	121,85	0,74	123,28	0,00	0,00	0,92
Jan-03	245,69	0,00	117,54	0,78	107,63	0,00	0,00	0,83
Feb-03	222,91	0,00	120,88	0,87	26,73	0,00	0,00	0,86
Mar-03	214,79	0,00	122,12	0,71	20,04	0,00	0,00	0,91
Apr-03	276,30	228,93	113,48	0,72	61,85	0,00	0,00	0,86
May-03	277,26	229,95	113,35	0,80	25,27	0,00	0,00	0,84
Jun-03	236,75	0,00	119,44	0,74	81,38	0,00	0,00	0,77

Tabela 9.1: Dados Brutos: Jan02 a Jun03

Meses	Demanda	DP	NP	DN	VT	I	N/P	PI
Jul-03	291,16	294,01	116,51	0,69	1845,08	0,00	0,00	0,94
Aug-03	339,16	392,49	114,23	0,92	2137,89	1,00	0,00	0,82
Sep-03	375,82	429,72	108,79	0,77	2042,04	1,00	0,00	0,72
Oct-03	399,13	303,11	110,27	0,82	1102,38	1,00	1,00	0,76
Nov-03	397,71	301,62	110,50	0,81	1366,97	1,00	1,00	0,76
Dec-03	411,01	214,96	108,57	0,84	1068,23	1,00	1,00	0,74
Jan-04	411,23	415,20	108,53	0,70	1858,00	1,00	1,00	0,75
Feb-04	447,69	452,15	102,96	0,84	1997,26	1,00	1,00	0,72
Mar-04	258,38	110,97	119,44	0,67	1860,91	1,00	0,00	0,82
Apr-04	382,81	186,65	104,59	0,68	1580,11	1,00	0,00	0,67
May-04	283,72	86,58	116,91	0,76	1153,76	1,00	0,00	0,83
Jun-04	338,21	141,61	108,94	0,82	1878,95	1,00	0,00	0,71
Jul-04	300,65	53,60	109,56	0,67	28,77	1,00	0,00	0,75
Aug-04	248,51	1,02	112,32	0,69	124,18	1,00	0,00	0,76
Sep-04	233,67	36,00	109,57	0,75	5,16	1,00	0,00	0,73
Oct-04	182,31	0,00	117,12	0,57	143,80	1,00	0,00	0,72
Nov-04	239,93	42,24	113,90	0,69	58,47	1,00	0,00	0,77
Dec-04	251,39	103,92	104,09	0,62	71,79	1,00	0,00	0,67
Jan-05	272,94	125,71	109,40	0,48	87,68	1,00	0,00	0,74
Feb-05	274,50	127,30	109,27	0,51	13,92	1,00	0,00	0,74
Mar-05	305,13	158,18	109,05	0,56	114,18	1,00	0,00	0,85
Apr-05	292,11	145,00	111,25	0,57	99,34	1,00	0,00	0,75
May-05	253,98	56,59	111,62	0,52	42,37	1,00	0,00	0,79
Jun-05	255,23	77,82	114,56	0,72	123,59	1,00	0,00	0,80
Jul-05	255,97	108,63	120,51	0,75	56,35	1,00	0,00	0,76

Tabela 9.2: Dados Brutos: Jul03 a Jul05

Referências Bibliográficas

- [1] Barossi Filho, M. e Braga, M. Bobik, *Metodologia da Econometria* IN: Vasconcellos et al, *Manual de Econometria*, Editora Atlas, São Paulo, 2000.
- [2] Box, George, Gwilym Jenkins e Gregory Reisel. *Time Series Analysis, Forecasting and Control*, San Francisco, Holden Day, 1994.
- [3] Bussab, Wilton e Morettin, Pedro A. *Estatística Básica*, SARAIVA, 5a. Edição, 2003.
- [4] Enders, Walter *Applied Econometric Time-Series*, 2nd Edition, 2003.
- [5] GC Franco, JP Soares e JA Ribeiro, *Comparação dos Softwares MINITAB, SAS, SPSS e EVIEWS na Estimação de Modelos ARIMA(p,d,q)* UFMG, 2004.
- [6] Gujarati, D. *Econometria Básica*, terceira edição, Makron Books, 2000.
- [7] Hendry D. F., Econometric methodology: a personal perspective. In: BEWLEY, T. F. *Advances in econometrics*. Cambridge: Cambridge University Press, 1987. Cap. 10.
- [8] Hoffman, Rodolfo e Vieira, Sônia, *Análise de Regressão: Uma introdução a Econometria*, São Paulo, HUCITEC, 1977.
- [9] Koopmans, T. C. *Three essays on the state of economics science*. New York : McGraw-Hill, 1957.
- [10] Magalhães, Marcos N. e Lima, Antonio C. P. *Noções de Probabilidade e Estatística*, EDUSP, 2002.
- [11] Parente, Juracy G., *O Sistema de Informação de Marketing e a Pesquisa de Marketing in: Dias, Sérgio R. et al, Gestão de Marketing*, Saraiva, 2003.
- [12] Shimakura, Silvia <http://www.est.ufpr.br/~silvia/CE055/node15.html>.
- [13] Siqueira Campos, Marcos *Desvendando o MinitabTM*, Qualitymark, 2003.

-
- [14] Spanos, A. *Statistical foundations of econometric modelling*. Cambridge : Cambridge University Press, 1989. .
- [15] Wooldridge, J. *Introductory Econometrics - A Modern Approach*, Thomson-SW, 2002.