

UNIVERSIDADE FEDERAL DE MINAS GERAIS – UFMG

INSTITUTO DE CIENCIAS EXATAS – ICEX

DEPARTAMENTO DE ESTATÍSTICA

**ANÁLISE DO INDICE DE MOBILIDADE DE PASSAGEIROS EM
TRANSPORTE COLETIVO EM BH**

ROGÉRIO FARIA D'ÁVILA

ORIENTADOR: ELA MERCEDES MEDRANO DE TOSCANO

BELO HORIZONTE/MG: Novembro/2010

UNIVERSIDADE FEDERAL DE MINAS GERAIS – UFMG

INSTITUTO DE CIENCIAS EXATAS – ICEX

DEPARTAMENTO DE ESTATÍSTICA

**ANÁLISE DO INDICE DE MOBILIDADE DE PASSAGEIROS EM
TRANSPORTE COLETIVO EM BH**

**MONOGRAFIA SUBMETIDA AO DEPARTAMENTO DE ESTATÍSTICA
DA UNIVERSIDADE FEDERAL DE MINAS GERAIS, COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
ESPECIALISTA.**

AGRADECIMENTOS

Agradeço primeiramente a Deus pela oportunidade e graça de mais uma conquista. Aos meus pais pelo apoio incondicional à realização de minhas metas e à minha esposa pela paciência neste período que, por muitas vezes, tive que ficar trancado no quarto para conseguir cumprir os deveres de casa propostos pelos professores. Um agradecimento especial à Prof. Ela Mercedes que com muita paciência e dedicação me orientou neste trabalho que acredito ser um início satisfatório para objetivos acadêmicos muitos maiores. Enfim, quero agradecer a todos que de uma forma direta ou indireta contribuíram para a conclusão de mais uma etapa.

“Quanto mais um homem se aproxima de suas metas,
tanto mais crescem as dificuldades.”

Johann Goethe

RESUMO

ANÁLISE DO ÍNDICE DE MOBILIDADE DE PASSAGEIROS EM TRANSPORTE COLETIVO EM BH

O crescimento desordenado nas regiões metropolitanas decorrentes da urbanização trouxe sérios problemas para o setor de transporte. Um dos maiores problemas deste desenvolvimento é que a estrutura geoeconômica ultrapassa seus próprios limites geográficos fazendo com que os indivíduos tenham que se deslocar para outros setores geográficos. Estes deslocamentos são realizados por um mercado de trabalho através do que chamamos de transporte coletivo e alguns outros de transporte rodoviário de passageiros. A caracterização do sistema de transporte coletivo de passageiros divide-se em características da oferta e da demanda. Sob esta perspectiva, este trabalho tem como objetivo *identificar as características de um determinado comportamento de passageiros ao longo de um período para que dessa forma as empresas deste ramo possam obter um planejamento estratégico que possibilite a melhora no serviço de transporte coletivo que vem sendo tão criticado por seus usuários. O cenário atual deste setor é preocupante devido uma parcela significativa da população utilizar este serviço e a quantidade de frota que é disponibilizada, tendo em vista que hoje, os usuários do transporte coletivo gastam um tempo significativo em espera de ônibus nos pontos designados, além de excesso de passageiros e uma qualidade ruim dos veículos.* De acordo com a característica mensal e o panorama geral dos passageiros acreditamos prever uma modelo capaz de ser utilizado para que as empresas possam se planejar melhor nos meses de maior movimentação de passageiros para que o serviço de transporte coletivo não perca a credibilidade.

TABELAS

Tabela 1: Estatística Descritiva	22
Tabela 2: Comparação de Modelos – AIC e Schwarz	28
Tabela 3: Comparação de Modelos – MAPE, MAE, MSE	29
Tabela 4: Comparação de Modelos – AIC e Schwarz	32
Tabela 5: Comparação de Modelos – MAPE, MAE, MSE	32

FIGURAS

Figura 1: Metodologia de Box e Jenkins.....	20
---	----

GRÁFICO

Gráfico 1 – Previsão de Séries Temporais.....	12
Gráfico 2- Histograma – IMPU	22
Gráfico 3 – Box Plot	23
Gráfico 4 – Time Series Plot IMPU	24
Gráfico 5 – Crescimento do PIB – Fonte IBGE	24
Gráfico 6 – Scatterplot Série IMPU	26
Gráfico 7 – FAC e FACP da série IMPU (Jan/92 a Dez/2003).....	29
Gráfico 8 – Gráfico Winter’s	30
Gráfico 7 – FAC Modelo 2.....	36
Gráfico 8 – FACP do Modelo 2	36
Gráfico 9 – Gráfico de Resíduos do Modelo 2.....	37
Gráfico 10 – Teste de Normalidade	37

SUMÁRIO

1	INTRODUÇÃO.....	7
1.1	APRESENTAÇÃO.....	7
1.2	JUSTIFICATIVA.....	9
1.3	OBJETIVO.....	9
1.3.1	OBJETIVOS ESPECÍFICOS:.....	9
2	METODOLOGIA.....	11
2.1	OBJETIVO.....	11
3	PREVISÃO DE SÉRIES TEMPORAIS.....	12
4	MODELO DE PREVISÃO DE SÉRIES TEMPORAIS.....	13
4.1	MÉTODOS DE DECOMPOSIÇÃO DE SÉRIES TEMPORAIS.....	13
4.1.1	MÉTODOS SIMPLES DE PREVISÃO DE SÉRIES TEMPORAIS.....	14
4.2	MEDIDAS DE ACUIDADE DOS METODOS DE PREVISÃO.....	21
5	ESTUDO DE CASO.....	22
6	MODELOS DE ESTUDO.....	27
6.1	ANÁLISE COM BANCO DE DADOS REDUZIDO - JAN/92 A DEZ/2003.....	27
6.1.1	Modelos ajustados a Série IMPU no Período amostral.....	28
6.1.2	Estatísticas dos Erros de Previsão.....	29
6.2	ANÁLISE COM BANCO DE DADOS COMPLETO - JAN/92 A DEZ/2004.....	31
6.2.1	ESTATÍSTICAS DOS ERROS DE PREVISÃO.....	32
7	ESCOLHA DO MELHOR MODELO.....	33
8	CONCLUSÃO.....	38
9	BIBLIOGRAFIA:.....	39
10	ANEXO.....	40
10.1	BASE DE DADOS.....	40
10.2	ANÁLISE PRELIMINAR DO BANCO DE DADOS.....	42
10.2.1	Diferenciando a série impu duas vezes.....	42
10.3	RESULTADOS MODELO ARIMA.....	43

1 INTRODUÇÃO

1.1 APRESENTAÇÃO

O transporte coletivo interestadual e internacional de passageiros, no Brasil, é um serviço público essencial, responsável por uma movimentação superior a 140 milhões de usuários/ano. A Agência Nacional de Transportes Terrestres é o órgão competente pela outorga de permissão e de autorização, para a operação desses serviços, por meio de Sociedades Empresariais legalmente constituídas para tal fim.

O grau de importância desse serviço pode ser medido quando se observa que o transporte por ônibus é a principal modalidade na movimentação coletiva de usuários. O serviço interestadual, em especial, é responsável por quase 95% do total dos deslocamentos realizados no País. Sua participação na economia brasileira é expressiva, assumindo um faturamento anual superior a R\$ 2,5 bilhões na prestação dos serviços regulares prestados pelas empresas permissionárias e Autorizatórias em Regime Especial (Res. 2868 e 2869/2008), onde são utilizados 13.400 ônibus.

Para efeitos de regulamentação e fiscalização, o transporte de passageiros é tratado nas três esferas de governo:

- As prefeituras municipais cuidam do transporte urbano (dentro da cidade);
- Os governos estaduais respondem pelas linhas intermunicipais dentro de cada Estado (ligando municípios de um mesmo Estado);
- O Governo Federal zela pelo transporte interestadual e internacional de passageiros (transporte de um Estado para outro ou que transpõe fronteiras terrestres com outros países).

Atualmente, a exploração de serviços de transporte rodoviário interestadual e internacional de passageiros encontra-se sob a égide da Lei nº 10.233, de 5 de junho de 2001, no que couber da Lei nº 8.987, de 13 de fevereiro de 1995, e da Lei nº 9.074, de 7 de julho de 1995, regulamentadas pelo Decreto nº 2.521, de 20 de março de 1998, e pelas normas aprovadas em Resolução, pela Diretoria Colegiada da ANTT.

As ações de regulação e fiscalização do setor têm caráter permanente e objetivam a adequação das rotinas e procedimentos para a efetiva operacionalização da Lei nº 10.233/2001, que criou a ANTT, buscando a melhoria dos serviços e a redução dos custos aos usuários do transporte rodoviário de passageiros, quer no transporte regular, quer no de fretamento contínuo, eventual ou turístico.

O transporte público urbano de passageiros é uma atividade que interage diretamente com o desenvolvimento e organização das cidades, recebendo impactos diretos da estrutura urbana, demográfica e de uso e ocupação do solo (SEDU, 2001).

Os deslocamentos populacionais em diferentes períodos da História, envolvendo escalas espaciais diferenciadas (entre campo e cidade, entre cidades na mesma região, entre regiões num mesmo país e entre países), vêm mostrando que expressam organizações sociais complexas (guerras, refugiados, expatriados políticos) e problemas econômicos conjunturais, por isso sugere-se um planejamento feito de forma a limitar as possíveis falhas devido ao crescimento populacional.

O transporte cumpre uma importante função social permitindo a mobilidade das pessoas e estimulando a disseminação de informações e idéias. Além do mais, relacionado com a distribuição dos recursos, topografia e desenvolvimento do comércio, o sistema termina sendo um dos condicionantes da distribuição da população (Martins e Filho, 1998).

O acelerado processo de urbanização está gerando conseqüências indesejáveis, tais como enormes congestionamentos, a perda de qualidade nos transportes coletivos e principalmente uma deteriorização no meio ambiente e na vida daqueles que utilizam tal serviço. Um conjunto de pesquisas feitas no Rio de Janeiro desde 2004 mostra que o tempo perdido nos ônibus durante um determinado trajeto gera um prejuízo de aproximadamente R\$ 8 bilhões ao ano para a cidade, considerando ser um cenário conservador. (Balassiano, 2009)

1.2 JUSTIFICATIVA

Este trabalho se justifica à medida que atualmente, apesar de termos várias empresas que atuam no segmento, não existe uma abordagem estatística para identificar, de forma descritiva e sistemática modelos de previsão para entender o universo do transporte coletivo. Esta abordagem mostra-se significativa à medida que as características populacionais mudam e a necessidade de um transporte coletivo satisfatório com qualidade de eficiência torna-se cada vez mais solicitado pela população. O transporte coletivo de passageiros, pode-se dizer que configura situação de crise, com sintomas e razões diferentes, constatados até mesmo pelo PNLT/CENTRAN, aliás, pelo simples exame dos anuários divulgados pela ANTT.

Em um caso a demanda está estagnada desde o início da década, com tendência até ligeiramente declinante em ligações importantes; no outro modal, o mercado tem crescido a taxas muito elevadas, causando problemas ligados à capacidade da infraestrutura, ao mesmo tempo em que a oferta foi comprometida pela quebra de operadoras e pelas dificuldades atuais para redistribuição plena de suas concessões.

1.3 OBJETIVO

O objetivo do presente estudo é analisar as características do índice de mobilidade de transporte coletivo urbano em BH.

1.3.1 OBJETIVOS ESPECÍFICOS:

- Descrição das principais características do índice de mobilidade urbana, para identificar as tendências mensais e anuais;
- Identificar os meses de maior movimentação;
- Encontrar um melhor modelo que explique a variabilidade do índice de mobilidade urbana;
- Encontrar o melhor modelo para previsão do índice de mobilidade urbana.

No seguinte capítulo será apresentado a metodologia de forma resumida de forma a explicar melhor a teoria por trás do processo de previsão. No capítulo III, será apresentada uma descrição dos dados, com os resultados obtidos da aplicação da metodologia. No capítulo IV será apresentada a conclusão e considerações finais.

2 METODOLOGIA

Uma série temporal é um conjunto de observações ordenado no tempo. Matematicamente uma série temporal é definida pelos valores Y_1, Y_2 de uma variável Y nos tempos t_1, t_2, t_n , ou seja, $Y = F(t)$. [ANDERSON1971].

Caso o levantamento das observações da série possa ser feito a qualquer momento do tempo, a série temporal é dita contínua, sendo denotada por $x(t)$ [GRANGER1977]. Entretanto, de acordo com, na maioria das séries, as observações são tomadas em intervalos de tempo discretos e eqüidistantes.

2.1 OBJETIVO

[GRANGER1977] coloca como objetivo inicial da análise de séries temporais a realização de inferências sobre as propriedades ou características básicas do mecanismo gerador do processo estocástico das observações da série.

Após a formulação do modelo matemático, obtido pela seleção entre as alternativas de classes de modelos identificadas como apropriadas para essa representação e subsequente estimação de seus parâmetros, é possível utilizá-lo para testar alguma hipótese ou teoria a respeito do mecanismo gerador do processo estocástico e realizar a previsão de valores futuros da série temporal.

3 PREVISÃO DE SÉRIES TEMPORAIS

Segundo [BARBANCHO1970], uma previsão é uma manifestação relativa a sucessos desconhecidos em um futuro determinado. A previsão não constitui um fim em si, mas um meio de fornecer informações e subsídios para uma conseqüente tomada de decisão, visando atingir determinados objetivos.

Considerando um conjunto de observações de uma série temporal coletadas até o instante t e de um modelo que represente esses fenômenos, a previsão do valor da série no tempo $t+h$ (gráfico 1) pode ser obtida.

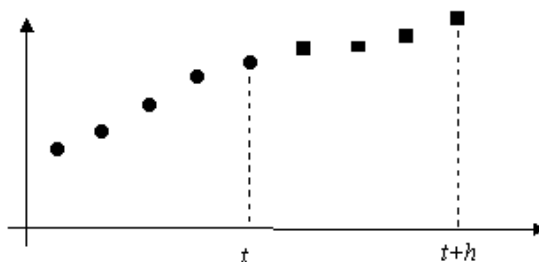


Gráfico 1 – Previsão de Séries Temporais

Dependendo do valor assumido pelo horizonte de previsão, classificam as previsões dos valores futuros de uma série temporal como de curto, médio ou longo prazo. Desse modo, diante da possibilidade de existência de diferentes horizontes de previsão.

Para [SOUZA1989], a garantia da ótima das previsões de uma série temporal somente é alcançada adotando como horizonte de previsão o instante de tempo imediatamente subsequente à origem t . Naturalmente, a investigação do poder preditivo do modelo especificado como o mais adequado para explicar o mecanismo gerador das observações de uma série temporal é um processo empírico de verificação, visto que são feitas comparações entre as previsões e as observações, a fim de confirmar a habilidade do modelo matemático em descrever a estrutura definida pelos dados da série temporal analisada. É importante salientar que o caráter previsional de um determinado modelo, segundo [BARBANCHO1970], não pode ser sustentado quando forem identificadas mudanças estruturais nas observações obtidas a partir da origem, em relação àquelas utilizadas para a elaboração do modelo.

4 MODELO DE PREVISÃO DE SÉRIES TEMPORAIS

Existem inúmeros métodos para se fazer previsão de séries temporais, desde os mais complexos que envolvem diferentes parâmetros aos métodos mais simples e de fácil entendimento. Métodos simples podem apresentar resultados satisfatórios sobre certas condições, além de permitir uma total compreensão de suas limitações facilitando assim a interpretação dos resultados. Sendo assim deve-se primeiro avaliar os benefícios de se utilizar um método simples ou um mais complexo antes de se iniciar a previsão em uma determinada aplicação.

Os métodos de previsão de séries temporais, classificados como métodos quantitativos, baseiam suas previsões na extrapolação de características de observações passadas e no inter-relacionamento entre essas observações, fornecendo previsões acuradas se o futuro apresentar comportamento similar ao passado [WHEELWRIGHT1985].

4.1 MÉTODOS DE DECOMPOSIÇÃO DE SÉRIES TEMPORAIS

Métodos de decomposição geralmente tentam identificar duas componentes distintas do padrão básico subjacente, que tendem a caracterizar a economia e uma série de negócios. Estas são as tendências de ciclo e os fatores sazonais. O fator sazonal refere-se a flutuações periódicas de comprimento constante como, por exemplo, a temperatura, a precipitação, o mês do ano, época de férias. A tendência representa o ciclo de mudanças a longo prazo no nível da série. A tendência de ciclos é por vezes separada em componentes de tendência e cíclicos, mas a distinção é um pouco artificial e muitos procedimentos da decomposição abordam a tendência e ciclo, como um único componente conhecido como a evolução do ciclo.

A decomposição pressupõe que os dados são compostos da seguinte forma:

Eliminar tendências e algumas flutuações sazonais pela média dos preços durante vários anos. [WHEELWRIGHT1985] expressa o relacionamento entre as componentes da série temporal através da equação.

$$x_t = f(S_t, T_t, C_t, E_t)$$

Onde S_t corresponde à componente sazonal para o período t;

T_t é a componente de tendência no período t;

C_t é a componente de ciclo no período t e

E_t é a componente aleatória no período t.

Segundo [MORETTIN1981], a componente sazonal representa as flutuações da série de acordo com algum fator de sazonalidade. O ciclo apresenta um comportamento similar à componente sazonal, embora tenha normalmente comprimento maior que aquela. Justamente pelo fato de não apresentar duração uniforme, a identificação da componente ciclo é mais problemática. A tendência representa o aumento ou declínio gradual nos valores das observações de uma série temporal. Com a remoção das componentes de sazonalidade, ciclo e tendência, a componente aleatória fica determinada.

O objetivo desses procedimentos consiste em remover cada uma das componentes, permitindo que o comportamento da série temporal seja mais bem compreendido e, conseqüentemente, prognosticar valores futuros mais apropriados.

4.1.1 MÉTODOS SIMPLES DE PREVISÃO DE SÉRIES TEMPORAIS

Os métodos assim classificados efetuam a previsão do valor futuro da série temporal pelo alisamento das observações passadas da série de interesse. O propósito desses métodos consiste em identificar o padrão básico presente nos dados históricos e, então, usar esse padrão para prever valores futuros.

Entre os métodos simples de previsão destacam-se o Alisamento Exponencial Simples, o Alisamento Exponencial Linear e o Alisamento Exponencial Sazonal e Linear de Winter, os quais são apresentados sucintamente na seqüência.

4.1.1.1 ALISAMENTO EXPONENCIAL SIMPLES

A princípio, o método conhecido como Alisamento Exponencial Simples se assemelha ao da Média Móvel por extrair das observações da série temporal o comportamento aleatório pelo alisamento dos dados históricos. Entretanto, a inovação introduzida pelo Alisamento Exponencial Simples advém do fato de este método atribuir pesos diferentes a cada observação da série. Enquanto que na Média Móvel as observações usadas para encontrar a previsão do valor futuro contribuem em igual proporção para o cálculo dessa previsão, no Alisamento Exponencial Simples as informações mais recentes são evidenciadas pela aplicação de um fator que determina essa importância.

$$Z_t = aZ_t + (1-a)Z_{t-1}, Z_0 = Z_1, t = 1, \dots, N, 0 \leq a \leq 1$$

a (alfa) é a constante de alisamento, e Z_t (Z_t barra) é o valor exponencialmente alisado. Se usarmos a equação recursivamente, nós podemos observar que valores antigos de Z_t exercem menor influência sobre a previsão Z_t :

$$Z_t = Z_t + a(1-a)Z_{t-1} + a(1-a)^2Z_{t-2} + a(1-a)^3Z_{t-3} + \dots,$$

Significando que o Alisamento Exponencial Simples é uma média ponderada, que dá maior peso às observações mais recentes, e peso menor para observações mais antigas, eliminando uma das desvantagens do método MMS. A previsão de valores futuros da série, a partir da origem t, é dada pela última média calculada:

$\hat{Z}_t(h) = Z_t$, $h = 1, 2, 3, \dots$ h é o horizonte de previsão. Ou ainda:

$$\hat{Z}_t(h) = aZ_t + (1-a)\hat{Z}_t - 1(h + 1), h = 1, 2, 3, \dots$$

A equação consiste de uma atualização da previsão, passo-a-passo, sem a necessidade de levar em consideração toda a série para o cálculo de previsão.

Vantagens:

- flexibilidade obtida da variação de a
- necessidade de armazenar somente Z_t , Z_t e a

A desvantagem é a dificuldade de se obter o valor de α ótimo para a previsão.

De acordo com [MORETTIN1981] e [GRANGER1977], o valor assumido por α determina o ajuste aplicado aos dados. Quanto menor o valor da constante, mais estáveis serão as previsões, visto que a utilização de baixo valor de α implica na atribuição de peso maior às observações passadas e, conseqüentemente, qualquer flutuação aleatória no presente contribui com menor importância para a obtenção da previsão, contudo, não há metodologia que oriente quanto à seleção de um valor apropriado para α , sendo normalmente encontrado por tentativa e erro.

4.1.1.2 ALISAMENTO EXPONENCIAL LINEAR

Quando o método Alisamento Exponencial Simples é aplicado na previsão de séries temporais que apresentam tendência entre as observações passadas, os valores prognosticados superestimam (ou subestimam) os valores reais. Desta forma, a acuidade das previsões fica prejudicada.

Para evitar esse erro sistemático, o método Alisamento Exponencial Linear foi desenvolvido procurando reconhecer a presença de tendência na série de dados. O valor da previsão obtido através deste método é alcançado pela aplicação da equação [WHEELWRIGTH1985].

$$F_{t+m} = S_t + mT_t$$

Onde:

S_t corresponde à previsão no tempo t , conforme equação;

T_t representa a componente de tendência, obtida pela equação e m é o horizonte de previsão.

T_t

$$S_t = \alpha x_t + (1 - \alpha)(S_{t-1} + T_{t-1})$$

onde α é o peso atribuído à observação x_t , $0 < \alpha < 1$.

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1}$$

onde β é o coeficiente de alisamento, análogo a α .

4.1.1.3 ALISAMENTO EXPONENCIAL SAZONAL E LINEAR DE WINTER

Este método produz resultados similares ao Alisamento Exponencial Linear, sendo, no entanto, capaz de manipular séries temporais que além de apresentarem tendência nos dados, apresentam também sazonalidade.

As equações a seguir definem o referido método de previsão.

$$S_t = \alpha \frac{x_t}{I_{t-l}} + (1 - \alpha)(S_{t-1} + T_{t-1})$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1}$$

$$F_{t+m} = (S_t + mT_t)I_{t-l+m} \quad I_t = \gamma \frac{x_t}{S_t} + (1 - \gamma)I_{t-l}$$

Onde:

- l corresponde ao alisamento do fator de sazonalidade $\frac{x_t}{S_t}$;
- l é o intervalo da sazonalidade
- γ corresponde ao peso atribuído ao fator de sazonalidade.

4.1.1.4 MODELOS AUTOREGRESSIVOS E DE MEDIAS MÓVEIS – ARIMA

Os modelos ARIMA são modelos estatísticos lineares para análise de séries temporais. A abreviação em língua inglesa refere-se a “Auto-Regressive Integrated Moving Average model”, ou seja, um modelo auto-regressivo integrado de médias móveis. Os termos auto-regressivos correspondem a defasagens da série transformada (isto é, série estacionária obtida por diferenciação) e as médias móveis a defasagens dos erros aleatórios. O termo “integrado” refere-se ao processo de diferenciação da série original para torná-la

estacionária. O modelo tem como premissa básica que a série temporal é gerada por um processo estocástico cuja natureza pode ser representada através de um modelo. A notação empregada para designação do modelo é normalmente ARIMA (p, d, q) onde p representa o número de parâmetros auto-regressivos, o número de diferenciações para que a série torne-se estacionária e q o número de parâmetros de médias móveis. Casos particulares são o modelo ARMA (p, q), o modelo auto-regressivo AR(p) e o modelo de médias móveis MA(q), todos para séries temporais estacionárias (d=0). Os modelos auto-regressivo AR(p) é definido por, onde a estimativa da variável S para um tempo t depende de uma combinação linear de p termos da série observada, incluindo o termo aleatório $\varepsilon(t)$ de ruído branco (erros de estimação com distribuição normal, média zero, variância constante e não-correlacionados). O modelo AR [p] é limitado, pois assume a existência de uma relação linear entre os elementos da seqüência e baseia-se na hipótese de que a série é estacionária, isto é, a média e o desvio padrão das observações medidas não variam com o tempo. A especificação de um modelo Autoregressivo (AR) é dada pela equação:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t$$

Onde

- x_t corresponde à observação da série temporal no tempo t;
- ϕ_p corresponde ao parâmetro do modelo AR de ordem p;
- ε_t representa o erro de eventos aleatórios que não podem ser explicados pelo modelo.
- ε_t

Caso as observações da série temporal possam ser representadas pela equação, a ordem do modelo puder ser determinada e os parâmetros estimados, são possíveis prever o valor futuro da série em análise.

Um modelo de Médias Móveis (MA) fica definido conforme equação

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

Onde e_t representa o erro de eventos aleatórios que não podem ser explicados pelo modelo e θ^q corresponde ao parâmetro do modelo MA de ordem q.

O modelo misto Autoregressivo e de Médias Móveis (ARMA) através da equação a seguir, como sendo a combinação dos modelos AR e MA.

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Analisando a equação é possível verificar que os modelos ARMA relacionam os valores futuros com as observações passadas, assim como também com os erros passados apurados entre os valores reais e os previstos.

O método de Box e Jenkins consiste na busca de um modelo ARIMA (AutoRegressive Integrate Moving Average) que represente o processo estocástico gerador da série temporal, a partir de um modelo ARMA aplicável na descrição de séries temporais estacionárias, estendendo esse conceito para séries temporais não-estacionárias [NELSON1973].

Genericamente, um processo ARIMA (p, d, q) pode ser representado pela equação.

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

Sendo $w_t = x_t - x_{t-d}$,

Onde ϕ_p e θ_q são os parâmetros dos processos Autoregressivo e de Média Móvel de ordem p e q (ARMA (p, q));

e_t corresponde ao erro de eventos aleatórios que não podem ser explicados pelo modelo;

d equivale ao grau de homogeneidade não-estacionária.

4.1.1.5 METODOLOGIA DE BOX E JENKINS

O objetivo da identificação é determinar os valores de p , d e q do modelo ARIMA (p , d , q) [WHEELWRIGTH1985]. Inicialmente, a série temporal é diferenciada para se obter uma série estacionária. Com isso, o processo fica reduzido a um modelo ARMA (p , q). Em seguida, a ordem do processo ARMA é identificada pela análise dos coeficientes de autocorrelação e autocorrelação parcial. Ainda nesta etapa são efetuadas estimativas preliminares dos parâmetros do modelo identificado.

Após a identificação do modelo que seja uma representação adequada do mecanismo gerador da série, a estimação dos parâmetros desse modelo é efetuada. [JUDGE1988] coloca que os parâmetros do processo AR são estimados através de métodos de regressão; caso o processo MA esteja envolvido, a estimação dos parâmetros deste modelo é obtida pela aplicação de algum algoritmo de otimização não-linear.

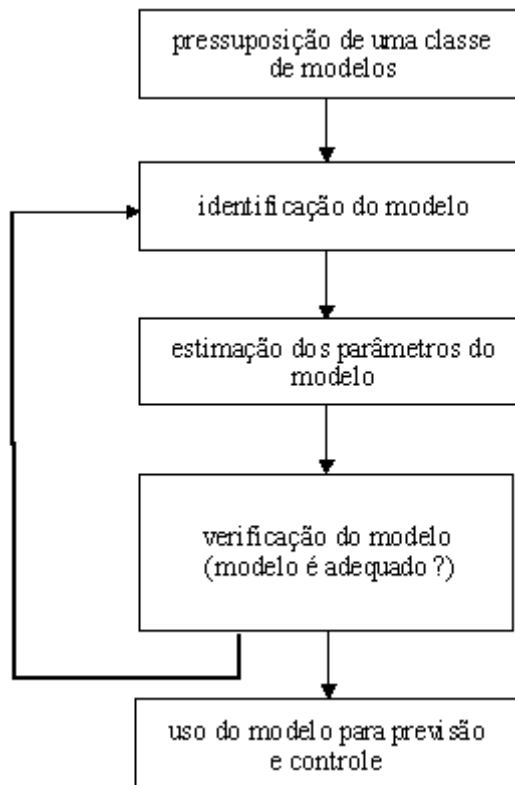


Figura 1: Metodologia de Box e Jenkins

Estimado o modelo, a verificação de sua habilidade em representar os fenômenos observáveis da série temporal é confirmada pela análise dos erros do modelo proposto. Caso a inadequação fique evidenciada, o ciclo de identificação, estimação e verificação é novamente aplicado, até que a representação apropriada seja encontrada.

Após a validação do modelo, a previsão dos valores futuros da série temporal modelada pode, enfim, ser obtida.

4.2 MEDIDAS DE ACUIDADE DOS METODOS DE PREVISÃO

De acordo com Makridakis [WHEELWRIGTH1985], a suposição básica de qualquer técnica de previsão de séries temporais é que o valor observado na série fica determinado por um padrão que se repete no tempo e por alguma influência aleatória. Isto significa dizer que mesmo quando o padrão exato que caracteriza o comportamento da série temporal tenha sido isolado, algum desvio ainda existirá entre os valores da previsão e os valores realmente observados. Essa aleatoriedade não pode ser prevista; entretanto, se isolada, sua magnitude pode ser estimada e usada para determinar a variação ou erro entre as observações e previsões realizadas.

$$\text{Erro médio} = \frac{\sum_{i=1}^n (x_i - \hat{x}_i)}{n}$$

Onde

x_i é o valor observado no instante i ;

\hat{x}_i é o valor previsto no instante i e

n corresponde ao número de previsões efetuadas.

$$\text{erro quadrado médio} = \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}$$

5 ESTUDO DE CASO

A série a ser estudada será denominada IMPU e possui como característica um banco de dados cujo período prolonga de jan/92 a dez/2004 com seus respectivos índices, valores estes obtidos através de uma análise econômica do Banco Central do Brasil. Para obter uma idéia geral deste banco de dados trabalhou-se primeiramente a parte descritiva dela que tem como objetivo principal comprovar a significância dos dados. Assim, foram calculadas frequência, curtose, media, mediana dentre outras metodologias estatísticas como mostra a Tabela 2 a seguir.

Estatística	Valor
Amostra (N)	156
Percent	100
Media	79,65
Desvio Padrão	387,81
Variance	12425,07
Minimo	46,53
Primeiro Quartil (Q1)	60,65
Mediana	80,96
Terceiro Quartil (Q3)	100,35
Máximo	112,05
Intervalo Interquartilico (IQR)	39,71
Curtose	-1,61

Tabela 1: Estatística Descritiva

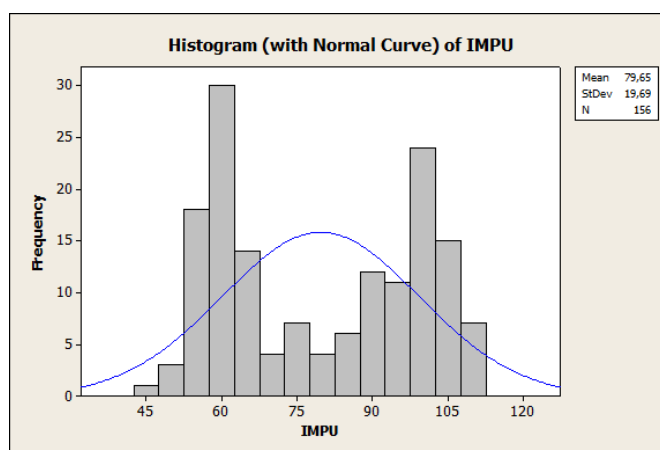


Gráfico 2- Histograma – IMPU

Após a análise descritiva dos dados será feito um gráfico de box plot por ano e por mês. O boxplot é um gráfico que possibilita representar a distribuição de um conjunto de dados com base em alguns de seus parâmetros descritivos, quais sejam: a mediana (q2), o quartil inferior (q1), o quartil superior (q3) e do intervalo interquartil (IQR = q3 - q1). Pelo Gráfico 3 percebe-se visualmente que nos meses de jan e dez há uma sazonalidade forte devido ser meses atípicos de férias. Este comportamento já está previsto nos estudos de transporte e levada em consideração tal sazonalidade.

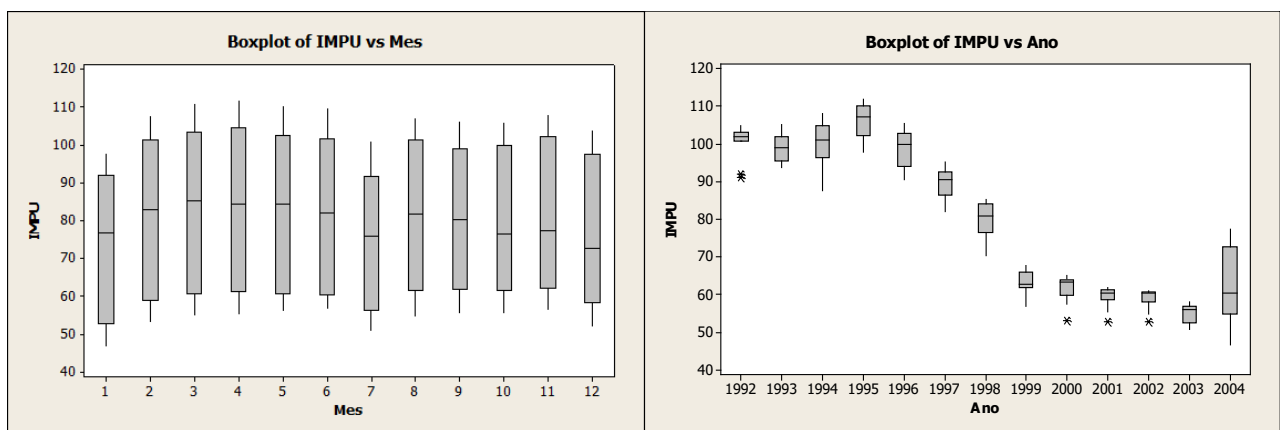


Gráfico 3 – Box Plot

Pelo Gráfico 4 da série completa abaixo pode-se ter uma visão anual do comportamento do índice de passageiros do transporte coletivo e assim ter subsídios suficientes para começar a análise, de fato, da série estudada. Em uma análise visual verifica-se uma variabilidade muito alta entre os anos e pela reta de regressão é aparente um ajuste na série para a normalidade.

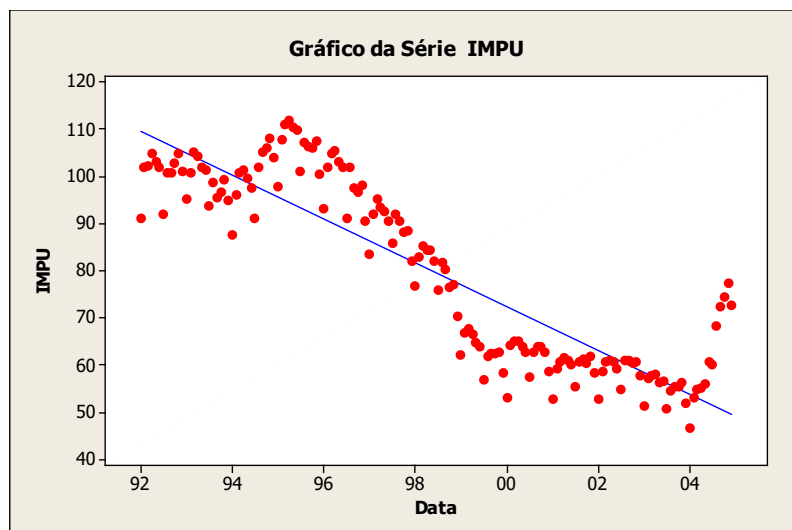


Gráfico 4 – Time Series Plot IMPU

Pelo gráfico observamos que no período de jan/92 a jan/95 há indício de estacionaridade na média e variância. Já no período de jan/95 a jan/00 visualmente há uma tendência aditiva negativa não apresentando estacionaridade na média e variância. Continuando a observação no período de jan/00 a jan/04 há indício de estacionaridade na média e variância e por fim uma tendência aditiva não apresentando estacionaridade na média e variância.

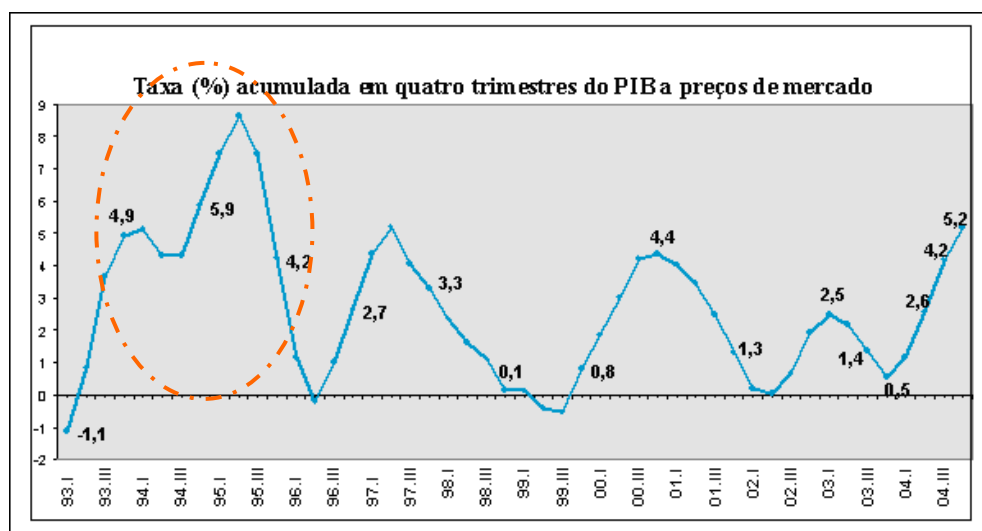


Gráfico 5 – Crescimento do PIB – Fonte IBGE

Para entender um pouco mais esse decréscimo no fluxo de passageiros de jan95 a jan/00 temos que em 1996, quando começou uma fase de instabilidade financeira internacional, com fugas de capitais que atingiram diversos países, foi o começo de um caos

que afetou duramente o setor de transportes internacionais. O Brasil perdia a metade de suas reservas cambiais, que atingiam U\$ 70 bilhões, forçando o governo pensar na desvalorização do Real.

Em outubro de 1997 a crise atingiu Hong Kong e chegou ao Brasil que para impedir a fuga de capitais, os juros foram elevados e para impedir a desvalorização do real, o Banco Central vendeu 9 bilhões de dólares de suas reservas cambiais. Os juros elevados, se por um lado atraem o capital estrangeiro, por outro paralisam os investimentos na economia. As contas dos governos ficam ainda mais vulneráveis, pois se amplia o desequilíbrio orçamentário, já que as dívidas do Estado precisam ser pagas com juros altos. O resultado inevitável é a recessão.

Em 1998 a desconfiança atingiu novamente os investimentos feitos no Brasil, que começou a perder divisas em ritmo acelerado. Em janeiro de 1999, o governo brasileiro foi obrigado a liberar câmbio e assistir á desvalorização do real, com enormes conseqüências em toda a vida toda a vida econômica e política do país. A crise asiática fez crescer a consciência de que a conseqüência de que é preciso conter o efeito destruído do capital especulativo.

Além disso, o forte decréscimo no índice de passageiros de 1995 a 1999 se dá principalmente pela estrutura produtiva, devido principalmente à liberação econômica. Grandes instabilidades afetaram a economia brasileira provocando uma pressão para baixo, principalmente, para as atividades produtivas de São Paulo. Devido às dificuldades, o crescimento foi modesto, tanto em nível nacional, 2,23%, quanto regional, 1,88%.

Voltando ao estudo da série, pelo Gráfico 6 de scatterplot, apresentam-se as variações mensais e nele apresenta visualmente uma grande variabilidade entre os dados e uma tendência alta em alguns meses.

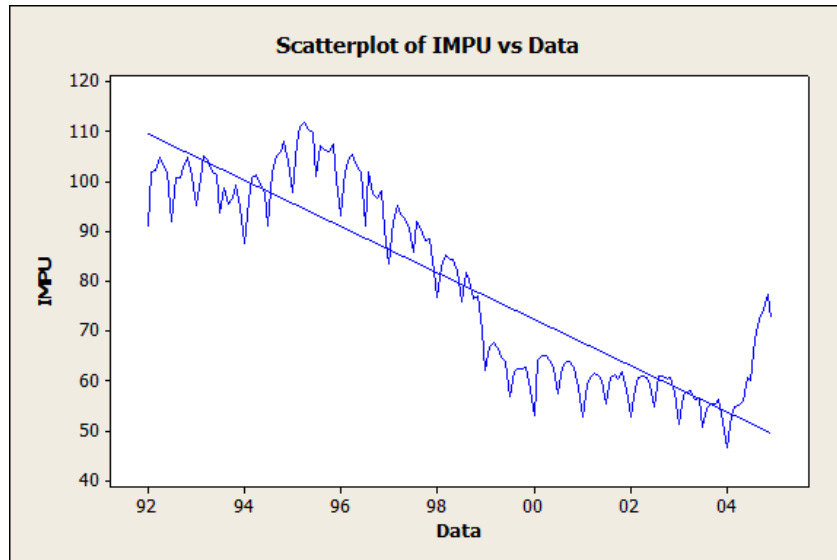


Gráfico 6 – Scatterplot Série IMPU

Por se tratar de não ser estacionária na média inicialmente vamos diferenciá-la para torná-la estabilizar a variância com os procedimentos adequados de diferenciar duas vezes a série e analisar posteriormente os correlogramas de FAC e FACP que identificará o ruído branco para validação do modelo, contudo a primeira diferenciação apresentam os correlogramas FAC com decrescimento exponencial lento e o FACP com dois fatores de auto correlação não significativos nos lags 6 e 12. Dessa forma sugerimos um modelo Auto Regressivo de ordem 2 (AR2).

Diferenciando segunda vez a série obtém-se que o correlograma FAC apresentou um decrescimento exponencial rápido e quatro fatores de autocorrelação significativo nos lags 2, 4, 12 e 14 no correlograma FACP.

Aplicando o modelo ARIMA percebe-se que neste modelo há alguns coeficientes não significativos o que indica que a auto regressiva é incompleta (AR4), dessa forma, para este estudo utiliza-se também o método de análise de regressão para identificar as variáveis mais significativas para o modelo e ajustar o modelo afim de encontrar um ruído branco.

6 MODELOS DE ESTUDO

A seguir propomos alguns modelos que explicam a variável regressora levando em consideração as variáveis propostas no modelo incompleto. Neste estudo propomos alguns modelos que tentarão explicar o comportamento do índice de passageiros de transporte urbano.

Válido ressaltar que os primeiros estudos têm como metodologia a utilização de um banco de dado com menos observações sendo retratado até dez/2003 por verificar pelos gráficos da análise descritiva que há um comportamento tendencioso e que poderá interferir na análise da série.

6.1 ANÁLISE COM BANCO DE DADOS REDUZIDO - JAN/92 A DEZ/2003

MODELO 1

$$D12IMPU = - 0,107 + 0,271 D12IMPU-4 - 0,534 RESI1-12$$

MODELO 1 no período amostral de 01/1992 a 12/2003

$$IMPU_t = IMPU_{t-12} + 0,271IMPU_{t-4} - 0,271IMPU_{t-16} - 0,107 + a_t - 0,534a_{t-12}$$

MODELO 2

$$D12IMPU = - 0,214 + 0,206 D12IMPU-4 - 0,443 RESI1-12 + 3,87 INTERVENÇÃO AGO/96 \\ - 3,77 INTERVENÇÃO SET/96 + 5,09 INTERVENÇÃO JUL/97 + 4,80 INTERVENÇÃO FEV/00$$

MODELO 2 no período amostral de 01/1992 a 12/2003

$$IMPU_t = IMPU_{t-12} + 0,206IMPU_{t-4} - 0,206IMPU_{t-16} + 3,87I_{1t} - 3,77I_{2t} + 5,09I_{3t} \\ + 4,80I_{4t} - 0,214 + a_t - 0,443a_{t-12}$$

Onde:

$$I_{1t} = \begin{cases} 1 & \text{se } t = \text{Ago}/1996 \\ 0 & \text{caso contrário} \end{cases} \quad I_{2t} = \begin{cases} 1 & \text{se } t = \text{Set}/1996 \\ 0 & \text{caso contrário} \end{cases}$$

$$I_{3t} = \begin{cases} 1 & \text{se } t = \text{Jul}1997 \\ 0 & \text{caso contrário} \end{cases} \quad I_{4t} = \begin{cases} 1 & \text{se } t = \text{Fev}2000 \\ 0 & \text{caso contrário} \end{cases}$$

MODELO 3

$$D12IMPU = -0,156 + 0,260 D12IMPU-4 + 3,74 \text{ INTERVENÇÃO AGO/96} \\ - 3,84 \text{ INTERVENÇÃO SET/96} + 4,91 \text{ INTERVENÇÃO JUL/97} - 0,479 \text{ RES11-12}$$

MODELO 3 no período amostral de 01/1992 a 12/2003

$$IMPU_t = IMPU_{t-12} + 0,260IMPU_{t-4} - 0,260IMPU_{t-16} + 3,74I_{1t} - 3,84I_{2t} + 4,91I_{3t} \\ - 0,156 + a_t - 0,479a_{t-12}$$

Onde:

$$I_{1t} = \begin{cases} 1 & \text{se } t = \text{Ago}/1996 \\ 0 & \text{caso contrário} \end{cases} \quad I_{2t} = \begin{cases} 1 & \text{se } t = \text{Set}/1996 \\ 0 & \text{caso contrário} \end{cases}$$

$$I_{3t} = \begin{cases} 1 & \text{se } t = \text{Jul}1997 \\ 0 & \text{caso contrário} \end{cases}$$

6.1.1 MODELOS AJUSTADOS A SÉRIE IMPU NO PERÍODO AMOSTRAL

Tabela 2: Comparação de Modelos – AIC e Schwarz

Parâmetros	Modelo 1 Modelo SARMA sem intervenção			Modelo 2 Modelo SARMA com intervenção			M3 Modelo SARMA com intervenção		
	Coef.	T	P-valor	Coef.	T	P-valor	Coef.	T	P-valor
ϕ	0,271	3,35	0,001	0,206	2,92	0,004	0,206	3,58	0,001
Θ_{12}	-0,534	-6,21	0,000	0,443	-5,86	0,000	0,479	-6,05	0,000
Θ_0	-0,107	-0,72	0,474	-0,214	-1,66	0,100	-0,156	-1,15	0,253
I_{1t} (Ago/96)				3,87	2,95	0,004	3,74	2,70	0,008
I_{2t} (Set/96)				-3,77	-2,89	0,005	-3,84	-2,79	0,006
I_{3t} (Jul/97)				5,09	3,87	0,000	4,91	3,53	0,001
I_{4t} (Fev/00)				4,80	3,57	0,001			

Estatísticas			
S	1,52087	1,29201	1,36598
R ²	32,7%	53,3%	47,3%
Radj ²	31,4%	50,5%	44,7%
AIC	0,8569	0,7949	0,8132
Schwarz	0,9600	0,8774	0,8964

6.1.2 ESTATÍSTICAS DOS ERROS DE PREVISÃO

Tabela 3: Comparação de Modelos – MAPE, MAE, MSE

	Modelo 1 Modelo SARMA sem intervenção	Modelo 2 Modelo SARMA com intervenção	Modelo 3 Alisamento Exponencial
Período Amostral			
MAPE	15874,5	15483,3	15875,5
MAE	76,5	76,5	76,5
MSE	6224,8	6226,2	6224,9
Período de Validação			
MAPE	3432,13	3432,13	3432,13
MAE	64,3830	64,3830	64,3830
MSE	4216,20	4216,20	4216,20

O Gráfico 7 e Gráfico 8 mostram o resultado comparativo da série estudada levando em consideração os modelos com e sem intervenção e o alisamento exponencial. Verificamos pelo gráfico que a série estudada pelo método de Alisamento Exponencial Winters acompanhou bem a série IMPU com medidas de MAPE, MAD e MSD bem baixas além das medidas de variabilidade, sazonalidade e tendência apresentarem os menores valores dentre as possibilidades estudadas.

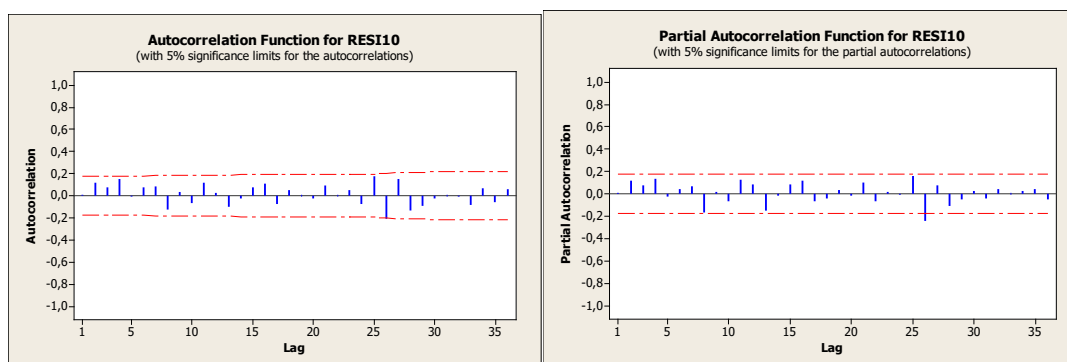


Gráfico 7 – FAC e FACP da série IMPU (Jan/92 a Dez/2003)

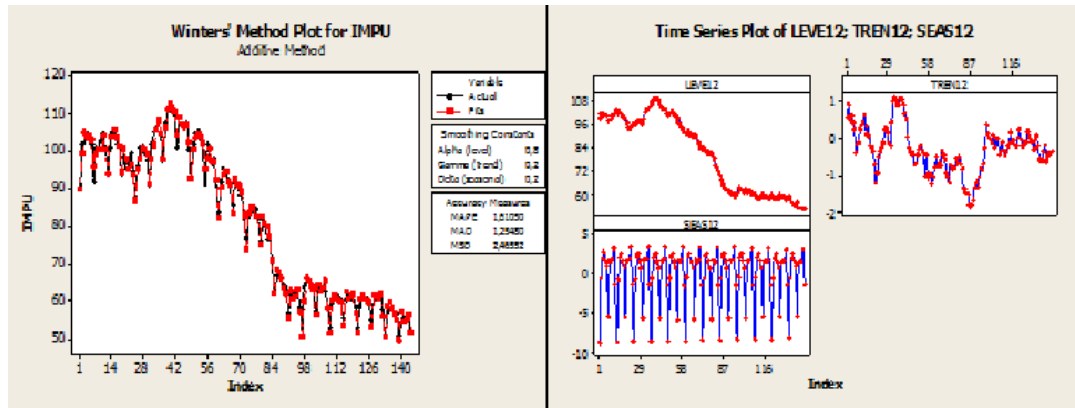


Gráfico 8 – Gráfico Winter's

Para equação do modelo do Alisamento Exponencial Winter temos:

$$L_t = 0,8(Y_t - S_{t-12}) + 0,2(L_{t-1} + T_{t-1})$$

$$T_t = 0,2(L_t - L_{t-1}) + 0,8T_{t-1}$$

$$S_t = 0,2(Y_t - L_t) + 0,8S_{t-12}$$

Equação da Previsão

$$Y_t = L_{t-1} + T_{t-1} + S_{t-12}$$

$$Y_t = 0,8(Y_t - S_{t-12} + T_{t-1} + S_{t-12}) + 0,2(L_{t-1} + T_{t-1} + L_t - L_{t-1} + Y_t - L_t)$$

$$Y_t = 0,8(Y_t + T_{t-1}) + 0,2(Y_t + T_{t-1})$$

6.2 ANÁLISE COM BANCO DE DADOS COMPLETO - JAN/92 A DEZ/2004

MODELO 1

$$D12IMPU = 0,124 + 0,214 D12IMPU-4 - 0,430 D12IMPU-12 + 0,165 RES1-2$$

$$IMPU_t = 0,214IMPU_{t-4} - 0,570IMPU_{t-12} + 0,214IMPU_{t-16} + 0,430IMPU_{t-24} + 0,124 + a_t + 0,165a_{t-2} S$$

MODELO 2

$$D12IMPU = - 0,025 + 0,183 D12IMPU-4 - 0,395 D12IMPU-12 + 4,27 \text{ Intervenção Ago/94} + 5,16 \text{ Intervenção Set/94} \\ - 5,02 \text{ Intervenção Set/96} + 5,40 \text{ Intervenção Jul/97} + 5,08 \text{ Intervenção Fev/00} + 5,07 \text{ Intervenção Jun/04} \\ + 4,47 \text{ Intervenção Jul/04}$$

$$IMPU_t = 0,183IMPU_{t-4} - 0,605IMPU_{t-12} - 0,183IMPU_{t-16} - 0,395IMPU_{t-24} + 4,27I_{1t} + 5,16I_{2t} \\ - 5,02I_{3t} + 5,40I_{4t} + 5,08I_{5t} + 5,07I_{6t} + 4,47I_{7t} - 0,025 + a_t$$

Onde:

$$I_{2t} = \begin{cases} 1 & \text{se } t = \text{Set1994} \\ 0 & \text{caso contrário} \end{cases} \quad I_{3t} = \begin{cases} 1 & \text{se } t = \text{Set1996} \\ 0 & \text{caso contrário} \end{cases}$$

$$I_{4t} = \begin{cases} 1 & \text{se } t = \text{Jul1997} \\ 0 & \text{caso contrário} \end{cases} \quad I_{5t} = \begin{cases} 1 & \text{se } t = \text{Fev2000} \\ 0 & \text{caso contrário} \end{cases} \quad I_{6t} = \begin{cases} 1 & \text{se } t = \text{Jun2004} \\ 0 & \text{caso contrário} \end{cases}$$

MODELO 3

$$D12IMPU = 0,062 + 0,176 D12IMPU-4 - 0,471 D12IMPU-12 + 0,185 RES12-2 - 4,87 \text{ Intervenção} \\ \text{Set/96} + 5,02 \text{ Intervenção Jul/97} + 5,02 \text{ Intervenção Jun/04} + 4,52 \text{ Intervenção Fev/00}$$

$$IMPU_t = 0,176IMPU_{t-4} - 0,529IMPU_{t-12} + 0,176IMPU_{t-16} - 0,471IMPU_{t-24} - 4,87I_{3t} + 5,02I_{4t} \\ + 4,52I_{5t} + 5,02I_{6t} + a_t + 0,185a_{t-2}$$

Onde:

$$I_{3t} = \begin{cases} 1 & \text{se } t = \text{Set1996} \\ 0 & \text{caso contrário} \end{cases} \quad I_{4t} = \begin{cases} 1 & \text{se } t = \text{Jul1997} \\ 0 & \text{caso contrário} \end{cases}$$

$$I_{5t} = \begin{cases} 1 & \text{se } t = \text{Fev2000} \\ 0 & \text{caso contrário} \end{cases} \quad I_{6t} = \begin{cases} 1 & \text{se } t = \text{Jun2004} \\ 0 & \text{caso contrário} \end{cases}$$

6.1.1 MODELOS AJUSTADOS A SÉRIE IMPU NO PERÍODO AMOSTRAL

Tabela 4: Comparação de Modelos – AIC e Schwarz

Parâmetros	Modelo 1 Modelo SARMA sem intervenção			Modelo 2 Modelo SARMA com intervenção			M3 Modelo SARMA com intervenção		
	Coef.	T	P-valor	Coef.	T	P-valor	Coef.	T	P-valor
ϕ	0,214	2,76	0,007	0,183	3,00	0,003	0,176	2,60	0,001
Θ_{12}	-0,430	-5,78	0,000	-0,395	-6,34	0,000	-0,471	-6,79	0,000
Θ_0	0,124	0,82	0,416	-0,025	-0,20	0,844	0,062	0,46	0,649
θ_2	0,165	2,14	0,034				0,185	2,33	0,021
I_{1t} (Ago/94)				4,27	3,02	0,003			
I_{2t} (Set/94)				5,16	3,67	0,000			
I_{3t} (Set/96)				5,02	-3,55	0,001	-4,87	-3,12	0,002
I_{4t} (Jul/97)				5,40	3,87	0,000	5,02	3,30	0,001
I_{5t} (Fev/00)				5,08	3,60	0,000	4,52	2,93	0,004
I_{6t} (Jun/04)				5,07	3,63	0,000	5,02	3,30	0,00
I_{7t} (Jul/04)				4,47	3,320	0,002			
Estatísticas									
S	1,73170			1,38578			1,50958		
R ²	29,7%			57,1%			48,4%		
Radj ²	28,1%			53,9%			45,4		
AIC	0,9379			0,5266			0,7190		
Schwarz	0,9966			0,7222			0,8558		

6.2.1 ESTATÍSTICAS DOS ERROS DE PREVISÃO

Tabela 5: Comparação de Modelos – MAPE, MAE, MSE

	Modelo 1 Modelo SARMA sem intervenção	Modelo 2 Modelo SARMA com intervenção	Modelo 3 Alisamento Exponencial
MAPE	200,3	1,242E+16	1,2731E+16
MAE	0,8526	0,9038	0,9045
MSE	1,2578	2,4116	1,9510

7 ESCOLHA DO MELHOR MODELO

Para a escolha do melhor modelo dentre todos os estudados procurou-se verificar as estatísticas apresentadas pelas macros EP.MAC e CI.MAC dadas pelas estatísticas MAPE, MAE, MSE e pelas estatísticas AIC (Critério de Akaike) e Schwarz respectivamente.

Vários indicadores têm sido desenvolvidos para avaliar e comparar a *performance* de modelos. Tais ferramentas estatísticas têm valor ao permitir o julgamento isento das vantagens dos modelos. Alguns indicadores são de precisão, outros de adequação relativa e, finalmente, outros tentam capturar a capacidade de acerto de tendência dos modelos em questão. Para uma revisão, consulte [DIEBOLD e LOPEZ 1996].

Os critérios usuais de avaliação incluem o valor médio quadrático dos erros de previsão (MSE) ou a sua raiz quadrada (RMSE), o valor médio dos erros absolutos (MAE), o valor médio dos erros absolutos percentuais (MAPE), e o coeficiente de U-Theil.

O RMSE é dependente da escala e, em geral, é utilizado para comparação de preditores:

$$RMSE = \sqrt{\left[\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \right]}$$

onde: n é o número de previsões avaliadas. Aqui cabe uma observação: as comparações são feitas entre previsões i -passos à frente, para o mesmo i e modelos distintos, então, esses índices, rigorosamente, são indexados em i .

Uma variante possível é a generalização da potência de maneira a penalizar a maior dispersão, que deteriora fortemente a previsão da primeira diferença da série. A formulação seria, no caso testado, o caso de d -potência:

$$MdAE = \left[\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|^d \right]$$

O MAE tem propriedades semelhantes e define-se como:

$$MAE = \left[\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \right]$$

O MAPE tenta capturar a importância do erro relativo e define-se como:

$$MAPE = \left[\frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \right]$$

Critério de Informação de Akaike é uma estatística freqüentemente utilizada para a escolha da especificação ótima de uma equação de regressão no caso de alternativas não aninhadas.

Dois modelos são ditos não aninhados quando não existem variáveis independentes comuns aos dois, portanto, quando se quer decidir entre dois modelos não aninhados, o melhor é o que produz o menor valor do critério de Akaike

Por exemplo, o número de defasagens a serem incluídas numa equação com defasagens distribuídas pode ser indicado pela seleção que produz o menor valor do critério de Akaike.

O critério de Akaike (AIC) é definido como:

$$AIC = 2 * (k-L) / N$$

Onde:

L é a estatística log verossimilhança;

N o número de observações;

k o número de coeficientes estimados (incluindo a constante).

Critério de Schwarz é uma estatística semelhante ao critério de Akaike com a característica de impor uma penalidade maior pela inclusão de coeficientes adicionais a serem estimados.

O critério de Schwarz (SIC) é definido como:

$$SIC = (k \cdot \log(N) - 2 \cdot L) / N$$

Aplicando, então, esta teoria de performance de modelos pode-se chegar a uma conclusão que, de todos os modelos estudados, o que melhor se encaixa dentro dos parâmetros é o modelo 2 com o banco de dados completo, primeiro por ter o menor valor do critério de Akaike e depois pelos valores de MAPE, MAE e MSD também serem os menores dentre todos os seis modelos estudados.

Análise de Regressão

$$D12IMPU = - 0,025 + 0,183 D12IMPU-4 - 0,395 D12IMPU-12 + 4,27 \text{Intervenção Ago/94} \\ + 5,16 \text{Interven set/94} - 5,02 \text{interven set/96} + 5,40 \text{int jul97} \\ + 5,08 \text{inte fev/00} + 5,07 \text{int jun/04} + 4,47 \text{int jul/04}$$

$$IMPU_t = 0,183IMPU_{t-4} - 0,605IMPU_{t-12} - 0,183IMPU_{t-16} - 0,395IMPU_{t-24} + 4,27I_{1t} + 5,16I_{2t} \\ - 5,02I_{3t} + 5,40I_{4t} + 5,08I_{5t} + 5,07I_{6t} + 4,47I_{7t} - 0,025 + a_t$$

Onde:

$$I_{2t} = \begin{cases} 1 & \text{se } t = \text{Set1994} \\ 0 & \text{caso contrário} \end{cases} \quad I_{3t} = \begin{cases} 1 & \text{se } t = \text{Set1996} \\ 0 & \text{caso contrário} \end{cases}$$

$$I_{4t} = \begin{cases} 1 & \text{se } t = \text{Jul1997} \\ 0 & \text{caso contrário} \end{cases} \quad I_{5t} = \begin{cases} 1 & \text{se } t = \text{Fev2000} \\ 0 & \text{caso contrário} \end{cases} \quad I_{6t} = \begin{cases} 1 & \text{se } t = \text{Jun2004} \\ 0 & \text{caso contrário} \end{cases}$$

Análise FAC e FACP para estudo do ruído branco

Pela análise dos correlogramas FAC e FACP temos que o resíduo encontrado no modelo apresenta um ruído branco, desta forma, há um forte indício de que o modelo é válido.

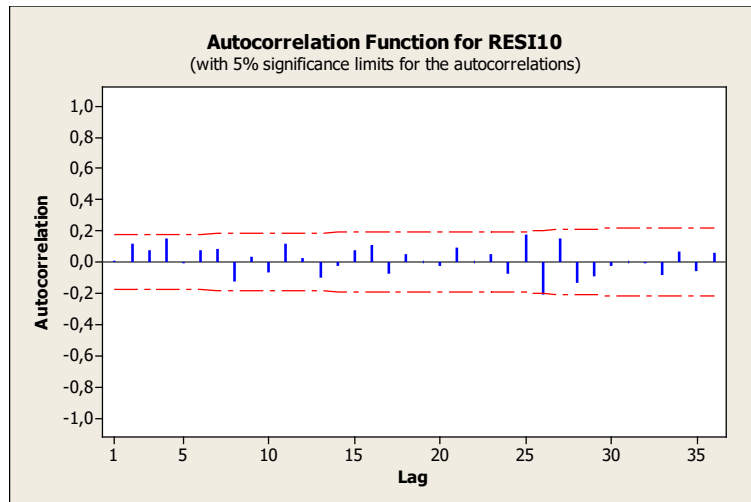


Gráfico 9 – FAC Modelo 2

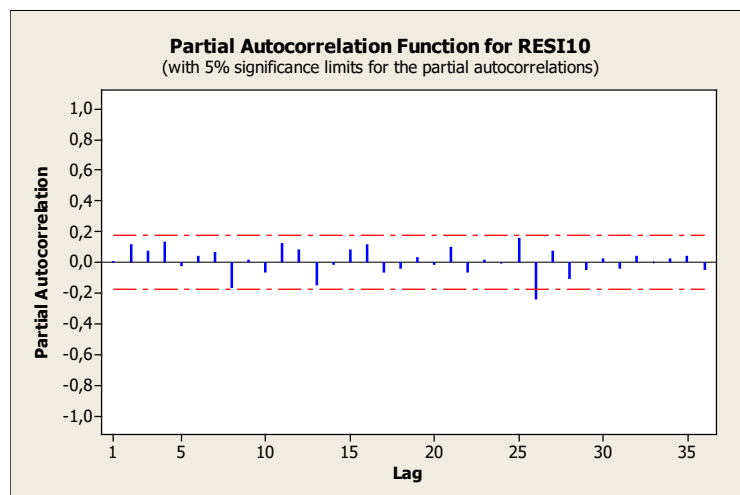


Gráfico 10 – FACP do Modelo 2

Análise Residual

Pelo gráfico de resíduos percebe-se visualmente que os dados estão bem concentrados em uma determinada região, contudo, verifica-se um estacionaridade em torno da média e um forte indicio de normalidade pelo histograma.

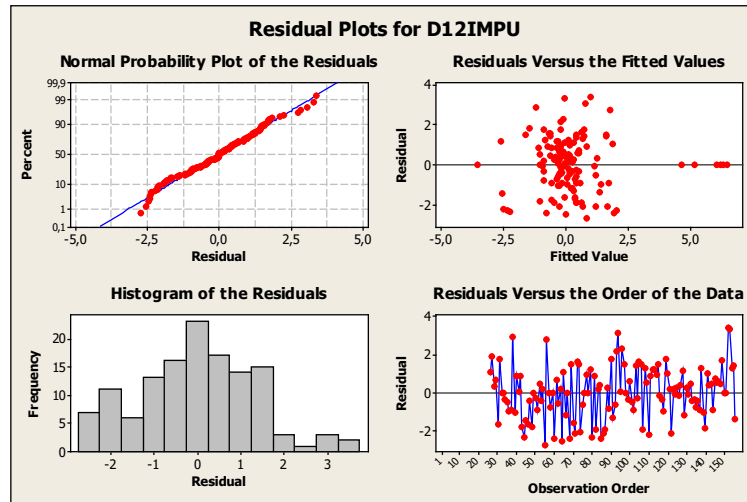


Gráfico 11 – Gráfico de Resíduos do Modelo 2

Teste de Normalidade

Pelo teste de normalidade observa-se que o p-valor está dentro dos parâmetros para a validação do modelo com um desvio padrão relativamente baixo 1,337, portanto o modelo é válido.

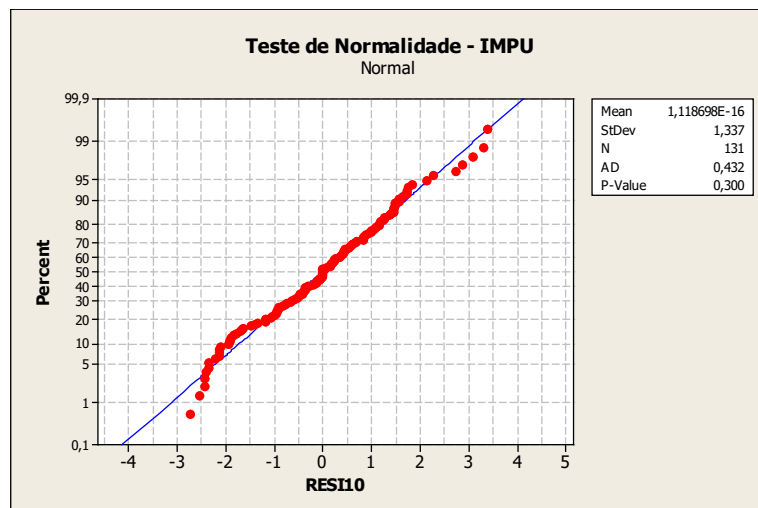


Gráfico 12 – Teste de Normalidade

8 CONCLUSÃO

De acordo com o objetivo do trabalho que era encontrar um modelo capaz de prever a movimentação de passageiros de transporte coletivo pode-se chegar à conclusão que o resultado foi satisfatório, contudo o banco de dados apresentado tem sua série no ano de 1992 a 2004, isso significa que este deverá apresentar uma atualização primeiro pelo crescimento populacional e segundo pela projeção do PIB onde sofrera alterações bruscas por conta da crise econômica financeira mundial em 2009.

As principais características da série estudada foi que nos meses de férias a sazonalidade é alta, por ser mês de férias onde há um deslocamento grande da população das áreas urbanas para as rodovias e neste sentido há uma alteração forte no índice de passageiros do transporte coletivo. Com relação aos anos da série temos uma forte queda dada pelo início da crise financeira no exterior a criação do plano real no período entre 1995 e 2000, período este, por sinal, que causou uma grande mudança na metodologia de previsão da série IMPU.

A escolha do melhor modelo apresenta, de fato, todas as características que eram esperadas, ou seja, intervenções em meses de maior variabilidade, média móvel sazonal por conta dos meses atípicos, estacionaridade em torno da média após as diferenciações, ruído branco dentre outras teorias que validam o modelo.

Válido acrescentar que mesmo o modelo sendo válido a explicação dele é bem baixa com 57,1%, portanto, em uma situação prática de trabalho haveria uma necessidade grande de mais argumentos para a validação do modelo perante um cliente, por exemplo, de concessão onde há milhões de reais em investimentos e o valor de explicação do modelo é um dos indicadores mais fortes para que o projeto seja aceito pela concessionária.

9 BIBLIOGRAFIA:

KRUGMAN, P. (1996) Confronting the Mystery of Urban hierarchy Journal of the Japanese and International Economies 10, pp. 339-418.

IBGE. Censos Demográficos 1991 e 2000. Rio de Janeiro. MTE. Relatório Anual das Informações Sociais (1997-2005)

PNUD, FJP, IPEA, Atlas do Desenvolvimento Humano no Brasil 1991-2000.

CORREIA (2005). R. L. Trajetórias Geográficas. Rio de Janeiro (RJ). Bertrand Brasil. 304 p.

CHRISTALLER W (1966). Central Places in Southern Germany. New Jersey (NY). Prentice Hall. 1966.

KING J. L. (1985) Central Place Theory. Beverly Hills (CA). Sage Publications. (1985)

GRIFFITH, D. A.; AMRHEIN, C. G. (1997) "Multivariate Statistical Analysis for Geographers". Prentice Hall. New Jersey

IBGE. Produto Interno Bruto dos Municípios. Série relatórios metodológicos. Rio de Janeiro: IBGE, 2004.

Ministério do Trabalho e Emprego (MTE/Governo Federal). Bases estatísticas CAGED 1999-2010. www.mte.gov.br

Instituto de Pesquisa Econômica Aplicada (IPEA). Presidência da República. IPEADATA. www.ipeadata.gov.br

F. Lukermann, P. W. Porter (1960) Gravity and Potential Models in Economic Geography Annals of the Association of American Geographers, Vol. 50, No. 4 (Dec., 1960), pp. 493-504

10 ANEXO

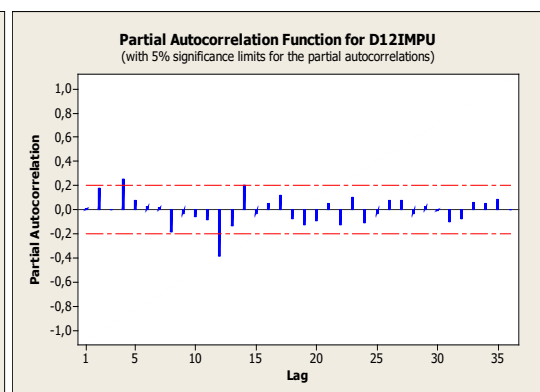
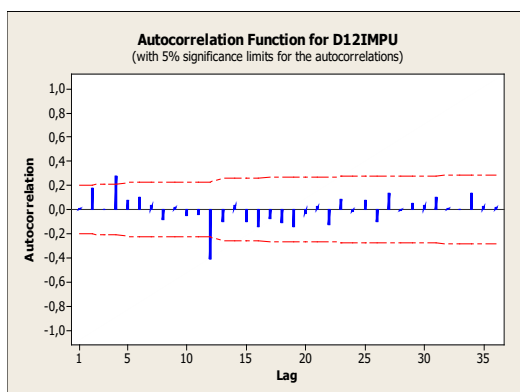
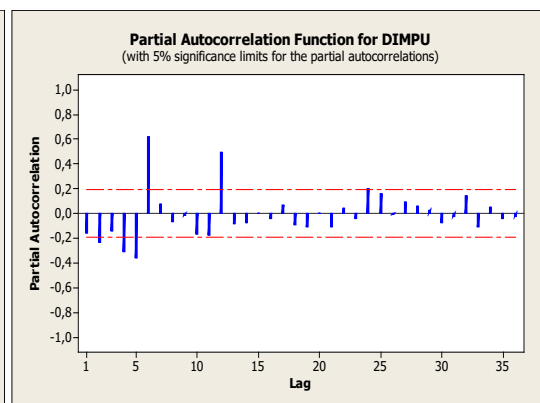
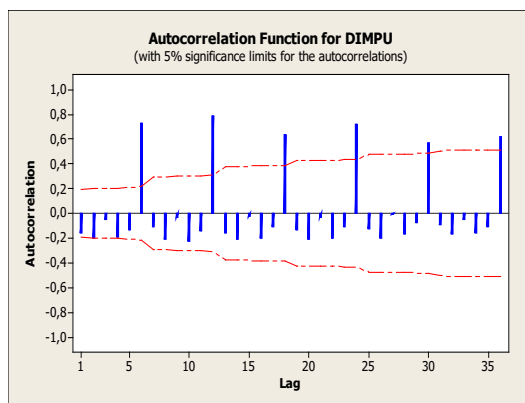
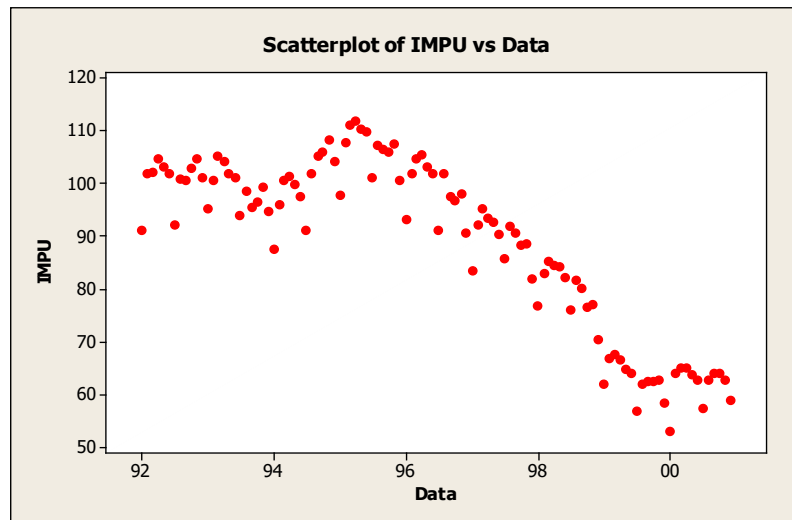
10.1 BASE DE DADOS

Data	Índice de Movimentação	Data	Índice de Movimentação	Data	Índice de Movimentação
jan/92	91.15	jan/95	97.75	jan/98	76.81
fev/92	101.9	fev/95	107.74	fev/98	83.03
mar/92	102.22	mar/95	111.18	mar/98	85.33
abr/92	104.77	abr/95	112.05	abr/98	84.52
mai/92	103.1	mai/95	110.41	mai/98	84.31
jun/92	101.86	jun/95	109.95	jun/98	82.08
jul/92	92.13	jul/95	101.07	jul/98	76.01
ago/92	100.82	ago/95	107.22	ago/98	81.71
set/92	100.73	set/95	106.44	set/98	80.21
out/92	102.96	out/95	106.07	out/98	76.48
nov/92	104.88	nov/95	107.55	nov/98	77.09
dez/92	101.1	dez/95	100.55	dez/98	70.43
jan/93	95.36	jan/96	93.1	jan/99	62.03
fev/93	100.76	fev/96	101.86	fev/99	66.8
mar/93	105.15	mar/96	104.76	mar/99	67.57
abr/93	104.18	abr/96	105.41	abr/99	66.55
mai/93	102.05	mai/96	103.15	mai/99	64.64
jun/93	101.25	jun/96	101.96	jun/99	63.9
jul/93	93.88	jul/96	91.17	jul/99	56.76
ago/93	98.66	ago/96	101.87	ago/99	61.87
set/93	95.47	set/96	97.53	set/99	62.3
out/93	96.58	out/96	96.77	out/99	62.43
nov/93	99.36	nov/96	98.05	nov/99	62.79
dez/93	94.82	dez/96	90.56	dez/99	58.29
jan/94	87.66	jan/97	83.53	jan/00	52.95
fev/94	95.99	fev/97	92.11	fev/00	64.06
mar/94	100.76	mar/97	95.37	mar/00	65.02
abr/94	101.36	abr/97	93.46	abr/00	65.02
mai/94	99.76	mai/97	92.58	mai/00	63.8
jun/94	97.61	jun/97	90.46	jun/00	62.7
jul/94	91.11	jul/97	85.87	jul/00	57.37
ago/94	101.97	ago/97	91.88	ago/00	62.6
set/94	105.24	set/97	90.67	set/00	63.94
out/94	106.17	out/97	88.24	out/00	63.95
nov/94	108.25	nov/97	88.6	nov/00	62.8
dez/94	104.12	dez/97	81.99	dez/00	58.71

Data	Índice de Movimentação	Data	Índice de Movimentação
jan/01	52.8	jan/04	46.53
fev/01	59.15	fev/04	52.96
mar/01	60.58	mar/04	54.76
abr/01	61.4	abr/04	55.23
mai/01	60.96	mai/04	55.93
jun/01	60.04	jun/04	60.66
jul/01	55.52	jul/04	60.1
ago/01	60.64	ago/04	68.26
set/01	61.35	set/04	72.41
out/01	60.45	out/04	74.37
nov/01	61.75	nov/04	77.3
dez/01	58.45	dez/04	72.79
jan/02	52.78		
fev/02	58.7		
mar/02	60.52		
abr/02	61.02		
mai/02	60.53		
jun/02	59.3		
jul/02	54.91		
ago/02	61.09		
set/02	60.83		
out/02	60.46		
nov/02	60.69		
dez/02	57.71		
jan/03	51.28		
fev/03	57.09		
mar/03	57.78		
abr/03	57.97		
mai/03	56.39		
jun/03	56.49		
jul/03	50.75		
ago/03	54.55		
set/03	55.44		
out/03	55.44		
nov/03	56.25		
dez/03	51.78		

10.2 ANÁLISE PRELIMINAR DO BANCO DE DADOS

10.2.1 DIFERENCIANDO A SÉRIE IMPU DUAS VEZES



10.3 RESULTADOS MODELO ARIMA

ARIMA Model: IMPU

Type		Coef	SE Coef	T	P
AR	1	-0,0417	0,1062	-0,39	0,695
AR	2	0,1917	0,1068	1,80	0,076
AR	3	-0,0114	0,1070	-0,11	0,915
AR	4	0,1459	0,1086	1,34	0,183
SMA	12	0,8563	0,0910	9,41	0,000
Constant		-0,07783	0,03762	-2,07	0,041

Differencing: 1 regular, 1 seasonal of order 12

Number of observations: Original series 108, after differencing 95

Residuals: SS = 252,829 (backforecasts excluded)
MS = 2,841 DF = 89

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	4,3	15,4	21,8	34,8
DF	6	18	30	42
P-Value	0,630	0,635	0,862	0,776

Regression Analysis: D12IMPU versus D12IMPU-4; D12IMPU-12

The regression equation is

$$D12IMPU = - 0,005 + 0,210 D12IMPU-4 - 0,443 D12IMPU-12$$

83 cases used, 25 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	-0,0046	0,2095	-0,02	0,983
D12IMPU-4	0,21016	0,09541	2,20	0,031
D12IMPU-12	-0,44279	0,09191	-4,82	0,000

S = 1,90078 R-Sq = 27,6% R-Sq(adj) = 25,8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	110,270	55,135	15,26	0,000
Residual Error	80	289,037	3,613		
Total	82	399,307			

Source	DF	Seq SS
D12IMPU-4	1	26,418
D12IMPU-12	1	83,852

