

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA E
MODELAGEM QUANTITATIVA**

**MODELOS BOX & JENKINS APLICADOS A
PREVISÃO DE DEMANDA DE LEITOS
HOSPITALARES**

MONOGRAFIA DE ESPECIALIZAÇÃO

FRANCISCA MENDONÇA SOUZA

Santa Maria, RS, Brasil

2006

MODELOS BOX & JENKINS APLICADOS A DEMANDA DE LEITOS HOSPITALARES

por

Francisca Mendonça Souza

Monografia apresentada ao curso de Especialização em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Especialista em Estatística e Modelagem Quantitativa.**

Orientador: Prof. Dr. Luis Felipe Dias Lopes

Santa Maria, RS, Brasil

2006

**Universidade Federal de Santa Maria
Centro de Ciências Naturais e Exatas
Departamento de Estatística
Programa de Pós-Graduação em Estatística e Modelagem
Quantitativos**

A Comissão Examinadora, abaixo assinada, aprova a Monografia de
Especialização

**MODELOS BOX & JENKINS APLICADOS A DEMANDA DE LEITOS
HOSPITALARES**

Elaborada por
Francisca Mendonça Souza

Como requisito parcial para a obtenção do grau de
Especialista em Estatística e Modelagem Quantitativa

COMISSÃO EXAMINADORA:

Luis Felipe Dias Lopes / UFSM
(Presidente/Orientador)

Roselaine Ruviano Zanin / UFSM

Anaelena Bragança de Moraes / UFSM

Santa Maria, 15 de janeiro de 2006.

DEDICATÓRIA

Dedico este trabalho, aos meus pais, a quem devo toda esta caminhada, a eles o meu reconhecimento e carinho.

Ao Adriano, pelo estímulo, experiências e ensinamentos.

Ao Leonardo, pela compreensão e força nos momentos mais necessários.

AGRADECIMENTO

Ao Prof. Dr. Luis Felipe Dias Lopes, pela orientação segura e disponível dispensada a este trabalho.

Aos professores do Departamento de Estatística – UFSM e aos colegas de aula, pessoas com quem sempre pude contar.

Aos familiares pelo incentivo e apoio, que se fizeram sempre presentes.

Ao CNPq, entidade governamental brasileira promotora do desenvolvimento científico e tecnológico, pelo auxílio financeiro (Processo 476508/2004-5 – Universal2004/Edital CNPq 19/2004 – Universal).

Ao setor de estatística do Hospital Universitário de Santa Maria na pessoa das colaboradoras Mareli e Miriam pela cedência dos dados e esclarecimentos necessários para a realização deste trabalho, muito obrigada.

RESUMO

Monografia de Especialização
Programa de Pós-Graduação em Estatística e Modelagem Quantitativa
Universidade Federal de Santa Maria

MODELOS BOX & JENKINS APLICADOS A PREVISÃO DE DEMANDA DE LEITOS HOSPITALARES

Autora: Francisca Mendonça Souza
Orientador: Luis Felipe Dias Lopes

Data e Local da Defesa: Santa Maria, 15 de janeiro de 2006.

Desde a sua fundação em 1970 o Hospital Universitário de Santa Maria – HUSM é uma referência em saúde pública para a região central do estado do Rio Grande do Sul. Sendo parte da Universidade Federal de Santa Maria, a instituição atua como Hospital Escola preocupando-se com o ensino e o desenvolvimento de pesquisas e assistência a saúde. O objetivo desta pesquisa é prever a taxa de ocupação hospitalar nos setores denominados Hospital Geral e Pronto Atendimento do HUSM, utilizando a metodologia Box & Jenkins e a análise de intervenção que proverão o conhecimento futuro destas variáveis a curto prazo. Os dados foram coletados no setor de estatística do HUSM, com observações mensais de janeiro de 2000 a dezembro de 2004. O modelo que melhor explica a taxa de ocupação no pronto atendimento foi um modelo autoregressivo integrado de médias móveis com uma diferença, ARIMA(1,1,1) com uma intervenção no período 36 do tipo abrupta temporária, o qual forneceu melhores explicações sobre a variável analisada. O hospital geral foi explicado por um modelo autoregressivo integrado de médias móveis com uma diferença, ARIMA(1,1,2). Os valores previstos servirão como suporte para pesquisas futuras nesta área, assim como as previsões encontradas servirão como uma forma de avaliação da demanda dos leitos no HUSM para os próximos meses como base para uma melhor organização do hospital.

Palavras Chave: Previsão de demanda, Modelos Box & Jenkins, Análise de intervenção, Séries Temporais.

ABSTRACT

Monograph of Specialization
Program of Masters Degree in Statistics and Quantitative Modeling
Federal University of Santa Maria

BOX & JENKINS MODEL APPLIED TO FORECAST THE HOSPITAL BEDS DEMAND

Author: Francisca Mendonça Souza
Advisor: Luis Felipe Dias Lopes

Date and place of defense: Santa Maria, 15th January of 2006.

Since its foundation in 1970 the University Hospital of Santa Maria – HUSM is a reference in public health for the central region of Rio Grande do Sul. Being part of the Federal University of Santa Maria, the institution operates as a school hospital aiming at the teaching development research and health assistance. The main purpose of this research is to forecast the hospital occupation rate in sectors named: General Hospital, Emergency Room of HUSM, using the Box & Jenkins methodology and Intervention analysis, that will provide to future known of these variables at short-term. The data were collected at the HUSM statistical sector, with monthly observations from January of 2000 to December of 2004. The model that better explain the occupation rate in Emergency Room was an autoregressive integrated moving average model with one difference, $ARIMA(1,1,1)$ with one intervention of lag 36 of type abrupt temporary. Which give more explanation about the variable analyzed. To the General Hospital a autoregressive integrated moving average model with one difference, the variable was explained by $ARIMA(1,1,2)$. Obtained the forecasted values will serve as support for future researches in this area, as well as the reached foresees will mainly become a way to evaluate the demand for beds at HUSM in the following months serving as basis for a better hospital organization.

Key Words: Demand Forecasting, Box & Jenkins models, intervention analysis, Times Series Analysis.

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Tema da Pesquisa	14
1.2 Justificativa e Importância	14
1.3 Objetivos	15
1.3.1 Objetivo Geral	15
1.3.2 Objetivos Específicos	15
1.4 Delimitação da Pesquisa	15
2 METODOLOGIA DE SÉRIES TEMPORAIS	16
2.1 Conceitos Gerais e Definições	16
2.2 Análise das Componentes de uma Série Temporal	19
2.3 Modelos Univariados	20
2.3.1 Modelos Estacionários	20
2.3.2 Modelos Não-Estacionários	27
2.4 Metodologia Box&Jenkins	30
2.4.1 Identificação	32
2.4.2 Estimação dos Parâmetros	35
2.4.3 Adequação do Modelo	40
2.4.4 Previsão	43
2.5 Análise de Intervenção	47
3 METODOLOGIA	63
3.1 Os Dados	63
3.2 Análise Descritiva	63
3.3 Modelagem das Séries por Meio da Metodologia Box & Jenkins....	64

4 RESULTADOS E DISCUSSÕES	65
4.1 Hospital Universitário de Santa Maria – HUSM	65
4.2 Variáveis e Períodos Analisados	65
4.3 Análise da Série de Taxa de Ocupação do Pronto Atendimento – PA – HUSM	66
4.4 Análise da Série de Taxa de Ocupação do Pronto Atendimento – PA – HUSM	73
5 CONCLUSÃO	79
6 BIBLIOGRAFIA	81

LISTA DE FIGURAS

FIGURA 1 – Processo estocástico interpretado como uma família de variantes aleatórias	17
FIGURA 2 – Esquema ilustrativo dos modelos univariados	20
FIGURA 3 – Comportamento das funções de autocorrelação e autocorrelação parcial de um modelo AR (1)	22
FIGURA 4 – Comportamento das funções de autocorrelação e autocorrelação parcial de um modelo MA(1)	23
FIGURA 5 – FAC e FACP amostrais da série ARIMA (1, 1, 0), com $\phi = 0,8$; ARIMA (0, 1, 1), com $\theta = 0,3$; ARIMA(1, 1, 1), com $\phi = 0,8$ e $\theta = 0,3$	29
FIGURA 6 – Fluxograma do ciclo iterativo de Box & Jenkins	31
FIGURA 7 – Série Y, que representa um processo integrado de ordem 2 com suas respectivas funções de autocorrelações	34
FIGURA 8 – Representação de um modelo dinâmico	49
FIGURA 9 – Representações de um sistema dinâmico com intervenção	51
FIGURA 10 – Efeitos de Intervenção	54
FIGURA 11 – Função Transferência	58
FIGURA 12 – Taxa de ocupação do PA	67
FIGURA 13 – Função de autocorrelação e autocorrelação parcial, respectivamente da série do PA	67
FIGURA 14 – Taxa de ocupação – PA – HUSM diferenciada	68
FIGURA 15 – Função de autocorrelação e autocorrelação parcial, respectivamente da série do PA diferenciada	68
FIGURA 16 – Gráfico da Distribuição Normal dos resíduos do Modelo ARIMA (1, 1, 1) com intervenção	70
FIGURA 17 – FAC e FACP dos resíduos do modelo ARIMA (1, 1, 1) com intervenção	70

FIGURA 18 – Taxa de Ocupação PA – Intervenção	71
FIGURA 19 – Previsão da Demanda PA - HUSM	72
FIGURA 20 – Taxa de Ocupação do Hospital Geral	73
FIGURA 21 – Função de autocorrelação e auto correlação parcial, respectivamente	74
FIGURA 22 – Taxa de ocupação HG – HUSM diferenciada	74
FIGURA 23 – Gráfico da Distribuição Normal do modelo ARIMA (1, 1, 2) do HG – HUSM	75
FIGURA 24 – FAC – FACP dos resíduos, respectivamente, do modelo delo ARIMA (1, 1, 2) HG – HUSM	75
FIGURA 25 – Previsão da Demanda HG – HUSM	77

LISTA DE TABELAS

TABELA 1 – Comportamento da FAC e da FACP de um Processo ARIMA(p,d,q)	30
TABELA 2 – Medidas descritivas das taxas de ocupação percentual do PA, por ano	66
TABELA 3 – Modelos propostos para o Pronto Atendimento	69
TABELA 4 – Previsões para PA – HUSM	71
TABELA 5 – Medidas descritivas das taxas de ocupação percentual do HG, por ano	73
TABELA 6 – Modelos concorrentes para o pronto atendimento do HG....	75
TABELA 7 – Previsões para HG – HUSM, de janeiro a dezembro de 2005	77

LISTA DE QUADROS

QUADRO 1 – Comportamento das funções FAC e FACP para modelos estacionários.....	26
---	----

1 INTRODUÇÃO

O aprimoramento tecnológico, nos últimos 30, anos facilitou e, ao mesmo tempo, incentivou a aplicação de técnicas estatísticas, com o objetivo de prever e melhor gerenciar a demanda por produtos e/ou serviços prestados. Segundo Makridakis et al. (1998), realizar previsões de demanda é importante para auxiliar na determinação dos recursos necessários, para o bom desempenho de uma empresa, e, em tempos de abertura de mercado, essa atividade torna-se fundamental.

Etimologicamente (*prae e videre*), a palavra previsão sugere que se quer ver alguma coisa, antes que ela exista. Alguns autores preferem a palavra predição, para indicar algo que deverá existir no futuro. Ainda, outros, utilizam o termo projeção.

É importante salientar que a previsão não constitui um fim em si, mas apenas um meio de fornecer informações para uma conseqüente tomada de decisões. As previsões de demanda são elaboradas utilizando-se técnicas quantitativas e qualitativas ou, ainda, uma mistura de ambas. Segundo Pellegrini & Fogliatto (2000), métodos quantitativos utilizam dados históricos para prever a demanda em períodos futuros. Uma das técnicas quantitativas mais difundidas é a metodologia de Box & Jenkins, descrita por esses autores na década de 70. Os modelos de Box & Jenkins partem da idéia de que cada valor de uma serie temporal pode ser explicado por valores prévios, a partir do uso da estrutura de correlação temporal, que, geralmente, há entre valores da série. Segundo Abdel-Aal & Al-Garni (1997), os modelos Box-Jenkins têm sido largamente utilizados para modelagem de previsão em aplicações médicas, ambientais, financeiras e de engenharia. Os próprios autores aplicaram essa metodologia para prever o consumo mensal de energia elétrica, no leste da Arábia Saudita.

Os modelos Box & Jenkins, genericamente conhecidos por ARIMA (*Auto Regressive Integrated Moving Averages*), e na literatura, em português, por Auto-regressivos Integrados de Médias Móveis, são modelos matemáticos que visam captar o comportamento da correlação seriada, ou autocorrelação entre os valores da série temporal, e, com base neste comportamento, realizar previsões futuras.

1.1 Tema da Pesquisa

O tema desta pesquisa está relacionado com a previsão de séries temporais em particular com a aplicação da metodologia de Box & Jenkins, nos dados de demanda do setor de Pronto Atendimento – PA, (Pronto Atendimento é um serviço que se caracteriza pelo atendimento a pacientes com necessidades sentidas como urgentes), Hospital Geral – HG, (Hospital Geral é o local de atenção à saúde de referência para alta complexidade, formação de profissionais de saúde e o desenvolvimento tecnológico numa perspectiva de inserção em rede aos serviços de saúde.), do Hospital Universitário de Santa Maria - HUSM.

1.2 Justificativa e Importância

O HUSM atrai pessoas dos mais diversos lugares, inclusive de outros estados, gerando uma gama enorme de dados, à respeito de cada paciente atendido. Como vários pacientes são internados faz-se necessário uma previsão, sobre a demanda dos leitos do hospital, analisando-se as várias formas de ingresso no hospital, PA, HG, buscando por meio da análise de séries temporais ajustar um modelo para obter uma previsão de demanda no processo de internação, espera e vacância de leitos. Até o momento, estatísticas descritivas que não fornecem um suporte criterioso para uma tomada de decisão e planejamento.

A saúde, sem precedência, é um dos assuntos de maior importância na sociedade, principalmente quando se trata de um hospital regional, como o HUSM, além de ser também um hospital escola. Também sabe-se da grande demanda que o hospital enfrenta. Por isso, um estudo que contemple a previsão da taxa de ocupação traz um subsídio muito importante para o hospital. Assim, a compreensão deste fenômeno, ao longo do tempo, será útil para que medidas de gerenciamento sejam tomadas.

1.3 Objetivos

1.3.1 Objetivo geral

- Realizar a previsão da demanda da taxa de ocupação do PA, HG do HUSM, utilizando-se a técnica de Box-Jenkins de análise de intervenção.

1.3.2 Objetivos Específicos

- Mostrar uma aplicação da análise de séries temporais, no setor de saúde, para que possa ser estendidas aos demais, locais que necessitem dessa técnica;

- Detalhar a aplicação de análise de intervenção;

- Verificar a adequação dos modelos ajustados às variáveis a serem estudadas;

- Encontrar previsões, a curto prazo, para a demanda de leitos do HUSM, de forma a auxiliar na geração de subsídios para a tomada de decisão.

1.4 Delimitação da Pesquisa

Essa pesquisa delimita-se ao HUSM, suas formas de ingresso, taxa de internação do PA e HG, não se estendendo a pesquisa aos demais hospitais da cidade e nem fazer um estudo comparativo com outras técnicas de previsão.

2 METODOLOGIA DE SÉRIES TEMPORAIS

Neste capítulo, trata-se alguns conceitos básicos em séries temporais, da classe geral dos modelos ARIMA e de análise de intervenção, pois esses serão os tópicos necessários para o desenvolvimento desta pesquisa.

2.1 Conceitos Gerais e Definições

Neste tópico serão definidos os conceitos e definições necessárias para o bom entendimento da pesquisa.

- Série Temporal

Uma série temporal é qualquer conjunto de observações ordenadas no tempo. Os dados contidos na referida série devem estar equidistantes, apresentando uma forte dependência serial. Representa-se uma série temporal da seguinte forma: Z_t , $t = 1, 2, \dots, n$, onde n representa o tamanho da série.

Se Z_t é um conjunto de observações correlacionadas entre si, e seguindo-se uma ordem cronológica de tempo tem-se uma série temporal, o caso mais freqüente é aquele em que a série só poderá ser descrita com uma distribuição de probabilidade. Neste caso, tem-se uma série não-determinística.

Determinadas as relações de dependência de Z_t , criar-se-á um modelo matemático para que forneça previsões dos valores futuros da série.

A série temporal Z_t pode ser discreta, ou contínua, univariada (com uma variável), ou multivariada (com mais de duas variáveis) e o tempo "T" pode ser unidimensional (uma dimensão) ou multidimensional (com mais de uma dimensão).

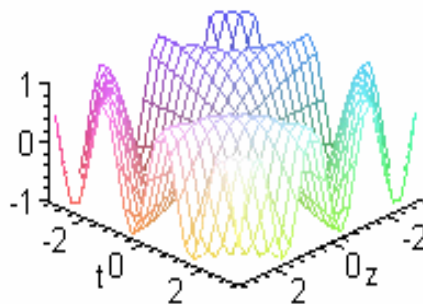
O pesquisador que se utiliza de uma série temporal tem por objetivo estudar o processo gerador da série fazer previsões em função dos seus valores passados e

descrever o comportamento da série. Sendo que este estudo pode ser feito no domínio do tempo, por meio das funções de autocorrelações, ou no domínio da frequência, por meio de análise espectral.

- Processo Estocástico

Os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos controlados por leis probabilísticas. Definição: seja T um conjunto arbitrário, um processo estocástico é uma família $Z = \{ Z(t), t \in T \}$, tal que para cada $t \in T$, $Z(t)$ é uma variável aleatória.

O conjunto T é, normalmente, considerado como o conjunto dos inteiros Z , ou conjunto dos R , e o conjunto de variáveis aleatórias, distribuídas equiespaçadamente no tempo, definidas num mesmo espaço de probabilidade (Ω, A, P) , (Morettim & Tolo, 1987).



Fonte: Resultado da pesquisa, *Maple*

Figura 1 - Um processo estocástico interpretado como uma família de variáveis aleatórias

- Processo Estacionário

O conjunto de variáveis aleatórias $Z = \{ Z_t, t \in T \}$ é dito estacionário se as estatísticas do conjunto de dados não variarem nos instantes t e $t+k$, $k = 0, 1, 2, \dots, n$, ou seja:

$$E[Z_t] = E[Z_{t+k}]$$

$$\text{Var}[Z_t] = \text{Var}[Z_{t+k}]$$

Tem-se outros tipos de estacionariedade, tais como: processo estritamente estacionário (ou forte) e estacionariedade de segunda ordem (ou fraca).

O processo será estritamente estacionário se todas as informações finito dimensionais permanecerem as mesmas sob translação do tempo, ou seja, $F(Z_1, \dots,$

$Z_n; t_1+\tau, \dots, t_n+\tau) = F(Z_1, \dots, Z_n; t_1, \dots, t_n)$; para qualquer t_1, \dots, t_n , de τ de T . Isso significa, em particular, que todas as distribuições unidimensionais são invariantes sob translação do tempo, logo a média $\mu(t)$ e a variância $V(t)$ são constante, isto é, $\mu(t) = \mu$, $V(t) = \sigma^2$, para todo $t \in T$.

Um processo estocástico $Z=\{Z(t), t \in T\}$ será fracamente estacionário (ou estacionário de segunda ordem) se, e somente se:

- i) $E\{Z(t)\} = \mu(t) = \mu$; constante, para todo $t \in T$;
- ii) $E\{Z^2(t)\} < \infty$, para todo $t \in T$;
- iii) $V(t_1, t_2) = \text{cov}\{Z(t_1), Z(t_2)\}$, é uma função t_1-t_2 . Um processo Z tal que (ii) esteja satisfeito diz-se um processo de segunda ordem.

- Processos Homogêneos e Diferenças

Existem séries temporais que não são estacionárias, tais como as séries econômicas, por exemplo, e esta não-estacionariedade pode ser transformada em estacionariedade através de diferenças sucessivas, utilizando-se o operador diferença ΔZ_t ; logo, ter-se-á um processo estacionário homogêneo.

$$\Delta Z_t = Z_t - Z_{t-1}$$

$$\Delta^2 Z_t = \Delta Z_t - \Delta Z_{t-1}$$

- Ruído Branco

É a seqüência de variáveis aleatórias, não correlacionadas, e identicamente distribuídas, com média zero e variância constante, com distribuição normal.

$$a_t \sim N(0, \sigma_a^2)$$

- Processo Ergódico

Na prática, é uma única série temporal, a qual representa todo o processo. O processo ergódico descreve todas as características do processo estocástico, trabalhando através de amostras.

- Modelo

É um artifício matemático, que tem por fim representar a realidade na lei de seus efeitos, espelhando a representação formal de uma realidade empírica. O

modelo deve ser parcimonioso, ou seja, que forneça a racionalidade de todos os fatos empíricos considerados (Morettin, 1987).

2.2 Análise das Componentes de uma Série Temporal

Conforme Morettin e Tolo (1987) uma série temporal é qualquer conjunto de observações ordenadas no tempo, geralmente, compostas por quatro elementos:

1. Tendência (T_t): verifica o sentido de deslocamento da série, ao longo de vários anos, podendo aumentar, diminuir ou permanecer constante.
2. Ciclo (C_t): movimento ondulatório que, ao longo de vários anos, tende à periodicidade, ou seja, é o movimento da série que se repete ao longo dos períodos de tempo, tendo uma longa duração, a qual varia de ciclo para ciclo ou se repete, em períodos muito longos.
3. Sazonalidade ou Fator Sazonal (S_t): relata as flutuações periódicas de comprimento constante, repetindo em períodos fixos, o comprimento do período é denotado por “S”, associado, na maioria dos casos, a mudanças climáticas.
4. Ruído aleatório ou erro (a_t): é tudo aquilo que não é explicado pelas outras componentes da série ou seja, é o que o modelo estimado não consegue captar. Um erro é dito ruído branco quando possui distribuição normal, a média de seus componentes é zero e a variância constante, a esses, são não-correlacionados indicando, assim, que o modelo elaborado conseguiu explicar o máximo de série de dados.

$$a_t \sim N(0, \sigma_a^2)$$

Sendo representada da seguinte maneira:

Dados = Modelo + erro;

Dados = $f(T_t, C_t, S_t)$ + erro.

2.3 Modelos Univariados

São aqueles que se baseiam somente na informação referente à série temporal em estudo. Citam-se dois tipos de modelos, aqueles que levam em consideração somente a informação contida na série histórica Z_t e, aqueles que, além das informações contidas em Z_t , permitem a inclusão de outras informações relevantes não contidas na série histórica. Esses modelos são baseados em estatística clássica e estatística Bayesiana, Hamilton(1994).

Na Figura 2, apresenta-se um esquema dos modelos univariados.

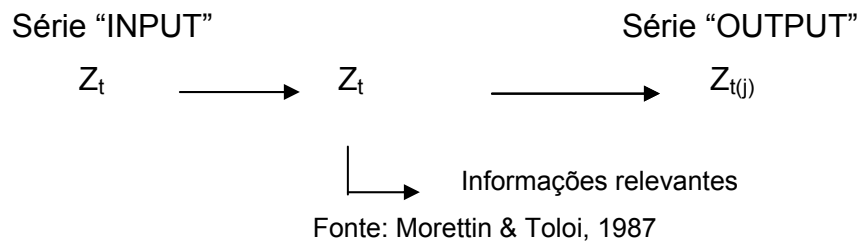


Figura 2 – Esquema ilustrativo dos modelos univariados

Os modelos Box & Jenkins podem ser classificados em estacionários e não estacionários, pois dependendo desta classificação, tem-se a classe geral ARMA(p, q), ou ARIMA(p, d, q), respectivamente.

A seguir, serão apresentados os modelos estacionários e não estacionários.

2.3.1 Modelos Estacionários

Se a série temporal desenvolve-se no tempo em torno de um valor constante, não necessitando de nenhuma transformação matemática para estacionarizá-la, tem-se uma série estacionária.

Modelos estacionários são aqueles que assumem que o processo está em equilíbrio. Um processo é considerado fracamente estacionário se suas médias e variância se mantêm constantes, ao longo do tempo, e a função de autocovariância depende, apenas, da defasagem entre os instantes de tempo. Um processo é fortemente estacionário se todos os momentos conjuntos são invariantes à translação no tempo.

A seguir, descreve-se os modelos estacionários AR, MA e ARMA.

- Modelo Autorregressivo de ordem p (AR(p))

O nome autorregressivo é pelo fato de que Z_t , no instante t é função dos Z's nos instantes anteriores a t e de um erro no instante t. O modelo que contém $\phi(B)=1$ é chamado modelo autorregressivo, sendo denotado por AR (p), onde "p" indica a ordem do modelo, isto é, o número de defasagens.

$$\tilde{Z}_t = \varphi_1 \tilde{Z}_{t-1} + \varphi_2 \tilde{Z}_{t-2} + \dots + \varphi_p \tilde{Z}_{t-p} + a_t \quad 2.1$$

$$\varphi(B) \tilde{Z}_t = a_t. \quad 2.2$$

O modelo autoregressivo de ordem 1 e 2 pode ser representado pelas expressões (3) e (4), respectivamente

$$AR(1) \rightarrow \tilde{Z}_t = \varphi_1 \tilde{Z}_{t-1} + a_t \text{ e} \quad 2.3$$

$$AR(2) \rightarrow \tilde{Z}_t = \varphi_1 \tilde{Z}_{t-1} + \varphi_2 \tilde{Z}_{t-2} + a_t \quad 2.4$$

onde, a notação \tilde{Z}_t corresponde a valores subtraídos da média do processo.

$$\tilde{Z}_t = Z_t - \mu. \quad 2.5$$

O modelo será considerado estacionário se $\phi(B)$ convergir para $|B| \leq 1$ como $\varphi(B)=0$, chamada condição de estacionaridade. A condição de estacionaridade do AR(p) estabelece que todas as raízes da equação devem cair fora do círculo unitário. E, sendo o processo de ordem finita, será sempre inversível.

Vê-se, portanto, que as autocovariâncias não dependem do t, e sim da ordem p. Como $|\varphi| < 1$, pela condição de estacionaridade, quanto maior o valor de p, ou seja, quanto maior a distância entre as observações, menor a autocovariância.

A identificação do modelo será feita através das funções de autocorrelação e autocorrelação parcial, que indicam a ordem do modelo, conforme a Figura 3.

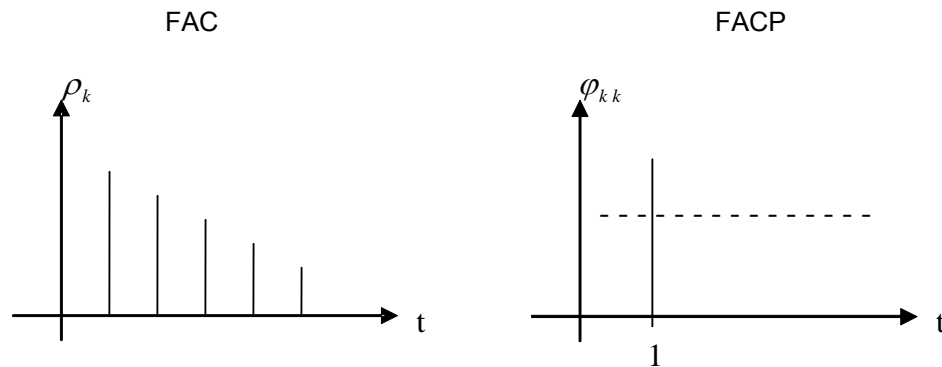


Figura 3 - Comportamento das funções de autocorrelação e autocorrelação parcial de um modelo AR (1)

A função de autocorrelação (FAC) decai exponencialmente, alternando ou não de sinal, em geral é uma mistura de exponenciais e ondas senóides amortecidas e a autocorrelação parcial (FACP) apresentará um corte rápido no *lag* significativo, indicando a ordem p do modelo.

- Modelo de Médias Móveis de Ordem q (MA(q))

O nome do modelo vem do fato de que Z_t é uma função soma algébrica ponderada dos a_t , que se movem no tempo. O nome, em si não é exato, pois Z_t não é média, visto que os parâmetros θ 's não somam, obrigatoriamente, a unidade.

Notação: MA (q) onde “ q ” indica a ordem do modelo e θ é um parâmetro.

$$\tilde{Z}_t = a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}; \quad 2.6$$

$$\tilde{Z}_t = \theta(B) a_t. \quad 2.7$$

A notação \tilde{Z}_t corresponde a valores do processo subtraídos de sua média. Pode-se observar os modelos, MA(1), e MA(2);

$$MA(1) \rightarrow \tilde{Z}_t = a_t + \theta_1 a_{t-1} \quad 2.8$$

$$MA(2) \rightarrow \tilde{Z}_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}. \quad 2.9$$

- Estacionariedade

O processo é de ordem finita. Logo, será sempre estacionário.

- Inversibilidade

A série $\pi(B)$ deve convergir para $|B| < 1$ com $\pi(B) = \theta^{-1}B$. As raízes da equação devem cair fora do círculo unitário.

A identificação do modelo é feita através das autocorrelações parciais, sendo que a função de autocorrelação fornece a ordem do modelo:

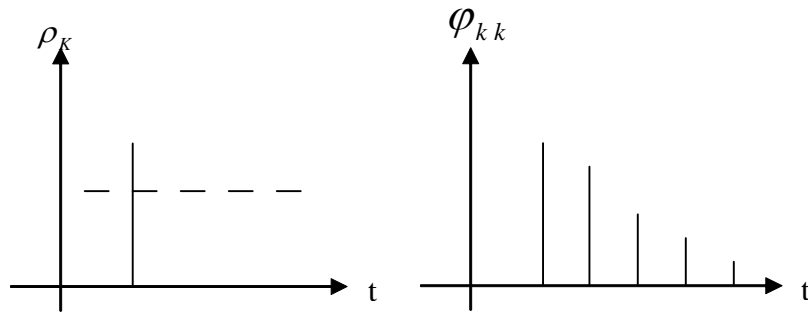


Figura 4 - Comportamento das funções de autocorrelação parcial e autocorrelação de um modelo MA (1)

A autocorrelação apresentará um corte rápido no 1ag significativo, indicando a ordem do modelo, enquanto que a autocorrelação parcial decairá exponencialmente, se todos os parâmetros forem positivos. Caso contrário, formará uma senóide amortecida (Souza, 1991).

- Modelo ARMA de Ordem p e q (Auto-regressivo de médias móveis) (ARMA(p, q))

No modelo ARMA ajusta-se somente às séries temporais estacionárias na média e na variância, e é definido pela equação que segue

$$Z_t = c + \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad 2.10$$

ou

$$z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p} = c + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}. \quad 2.11$$

Há restrições específicas aos valores que possam assumir os parâmetros das equações (2.10) e (2.11). Para $p=1$, $-1 < \phi_1 < 1$. Para $p = 2$, $-1 < \phi_2 < 1$, $\phi_2 + \phi_1 < 1$ e $\phi_2 - \phi_1 < 1$. Para $p > 3$, condições mais complicadas prevalecem. Similarmente, para $q=1$,

$-1 < \theta_1 < 1$. Para $q = 2$, $-1 < \theta_2 < 1$, $\theta_2 + \theta_1 < 1$ e $\theta_2 - \theta_1 < 1$. Para $q > 3$, também prevalece condições mais complicadas (Makridakis et al., 1998, p. 334).

Um dispositivo de notação útil é o operador de deslocamento retroativo B , cujo uso é o seguinte (Makridakis et al., 1998, p.334):

$$B X_t = X_{t-1}, \quad 2.12$$

ou, em outras palavras, B operando sobre Z_t tem o efeito de deslocar os dados para trás em período. Assim,

$$B(B X_t) = B^2 X_t = X_{t-2} \quad 2.13$$

e, eventualmente.

$$B^{12} X_t = X_{t-12}. \quad 2.14$$

Utilizando-se o operador B , a equação (2.14) pode ser escrita como segue:

$$(1 - \phi_1 B - \dots - \phi_p B^p) X_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) e_t. \quad 2.15$$

Pois é da expressão (2.14) que o modelo em questão recebe sua denominação específica de ARMA (p , q), isto é, modelo auto-regressivo de média móvel, de ordem p e q , conforme assinalado na expressão (2.15).

$$\underbrace{(1 - \phi_1 B - \dots - \phi_p B^p)}_{\text{AR}(p)} X_t = c + \underbrace{(1 - \theta_1 B - \dots - \theta_q B^q)}_{\text{MA}(q)} e_t. \quad 2.16$$

Os coeficientes $\phi_1, \phi_2, \dots, \phi_p$, e $\theta_1, \theta_2, \dots, \theta_q$ são estimados para que se ajustem à série temporal que está sendo modelada através dos métodos de mínimos quadrados, ou da máxima verossimilhança. A aplicação do método dos mínimos quadrados é semelhante à da regressão linear, entretanto, sem as formas simples desta aplicação, sendo as estimativas feitas interativamente, por *softwares*, até que a soma dos erros quadráticos seja minimizada. No segundo método, a verossimilhança de um conjunto de observações, denominada L , é proporcional à probabilidade se de obter as observações, dado o modelo, sendo, portanto, uma medida da plausibilidade o atual conjunto de observações, dado um particular conjunto de valores dos parâmetros. O método da máxima verossimilhança estima, interativamente, através de *softwares*, os valores dos parâmetros que maximizem a verossimilhança (Makridakis et al., 1998 p. 359).

O modelo ARMA, como qualquer modelo de previsão adequadamente ajustado a uma série temporal, deve ter seus erros, e_t , distribuídos aleatoriamente, isto é, comportando-se como uma série de ruído branco. Teoricamente, todos os coeficientes de autocorrelação de uma série de números aleatórios devem ser zero, pois os mesmos não guardam qualquer relação entre si. Entretanto, a distribuição amostral dos coeficientes de autocorrelação de um conjunto de n amostras, retiradas de uma população de números aleatórios, pode ser aproximada por uma curva normal de média zero e desvio padrão de $1/\sqrt{n}$. Portanto, espera-se que 95% das n amostras de coeficientes de autocorrelação de uma série aleatória estejam entre $\pm 1,96 / \sqrt{n}$. Se não for este o caso, a série provavelmente não é de ruído branco (Makridakis et al., 1998, p. 318).

Uma abordagem alternativa, à verificação dos coeficientes de autocorrelação dos resíduos da aplicação do modelo ARMA, é a utilização dos testes de Portmanteau, que considera o conjunto dos k primeiros coeficientes de autocorrelação, ao contrário de considerar cada um separadamente. É correto concluir que os resíduos não são ruído branco e a estatística Q , sugerida por Box-Pierce, ou Q^* , sugerida por Ljung-Box, posicionarem no extremo de 5% da cauda direita da distribuição do χ^2 . As expressões, para o cálculo das estatísticas Q , de Box-Pierce, ou Q^* , de Ljung-Box, são as seguintes:

$$Q = n \sum_{k=1}^h r_k^2 \quad 2.17$$

e

$$Q^* = n + (n + 2) \sum_{k=1}^h \frac{r_k^2}{(n - k)} \quad 2.18$$

Nas equações (2.17) e (2.18), n é o número de observações da série e h é a máxima defasagem dos coeficientes de correlação dos resíduos. Alguns autores aconselham que seja tomado $h = 20$ (Makridakis et al., 1998, p.319), outros, $h = n/4$ (Box et al., 1994, p. 32). O valor de h deve ser utilizado como número de graus de liberdade da distribuição χ^2 .

Quando nem a função de autocorrelação nem a função de autocorrelação parcial caem bruscamente a zero, a recomendação de uso de um modelo ARIMA é

reforçada. Alguns autores, entretanto, apresentam procedimentos para identificação das ordens p e q do modelo ARMA, a partir dos gráficos FAC e FACP, como no caso apresentado no Quadro 1, ou é sugerida uma abordagem por tentativas, na qual modelos cada vez mais complexos são, sucessivamente, ajustados à série, até que os resíduos apresentem um comportamento de ruído branco, ou seja, não apresentem correlação entre si. Portanto, segundo a recomendação de Johnson e Montgomery (1976, p.469), a melhor tentativa é o ajuste de um modelo ARMA (1,0) ou simplesmente AR(1).

Modelo	FAC	FACP
ARMA ($p, 0$)	Decaimento gradativo.	Decaimento brusco, após defasagem p .
ARMA ($0, q$)	Decaimento brusco, após defasagem q .	Decaimento gradativo.
ARMA (p, q)	Decaimento gradativo, com onda senoidal amortecida, após a defasagem ($q-p$).	Decaimento gradativo, com onda senoidal amortecida, após a defasagem ($p-q$).

Fonte: Adaptado de Johnson; Montgomery. (1974, p. 469)

Quadro 1 - Comportamento das Funções FAC e FACP para modelos estacionários

Como vimos anteriormente, os processos AR(p), MA(q) e ARMA(p, q) apresentam a função de autocorrelação FAC com características especiais. Assim:

- (i) um processo AR(p) tem FAC que decai de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão;
- (ii) um processo MA(q) tem FAC finita, no sentido que ela apresenta um corte após o “lag” q ;
- (iii) um processo ARMA (p, q) tem FAC infinita em extensão, a qual decai de acordo com exponenciais e/ou senóides amortecidas após o “lag” $q-p$.

Essas observações serão úteis no procedimento de identificação do modelo aos dados observados; calculando-se as estimativas das FAC, que se acredita reproduzir, adequadamente, as verdadeiras FAC`s desconhecidas, e comparando seu comportamento com o descrito acima. Para cada modelo, deve-se escolher um modelo que descreva o processo estocástico.

2.3.2 Modelos Não Estacionários

Como os processos encontrados na vida prática são raramente estacionários, tem-se que encontrar algum tipo de operador que, por transformações matemáticas, transforme o conjunto de dados em estacionários.

- Modelos não estacionários (ARIMA (p, d, q))

Até agora, obteve-se, a partir de um processo de ruído branco a_t , um processo Z_t estacionário, porém com relação de dependência. Como os processos encontrados são raramente estacionários, tem-se que encontrar algum tipo de operador que produza, a partir de Z_t , um processo estacionário.

Considera-se, nesta modelagem, somente os processos chamados não estacionários homogêneos.

Dada uma série Z_t , a primeira diferença de Z_t é definida por:

$$\Delta Z_t = Z_t - Z_{t-1} \quad 2.19$$

e a segunda diferença é dada por:

$$\Delta^2 Z_t = \Delta(Z_t - Z_{t-1}); \quad 2.20$$

$$\Delta^2 Z_t = Z_t - 2 Z_{t-1} + Z_{t-2}. \quad 2.21$$

genericamente, a d-ésima diferença de Z_t é definida por:

$$\Delta^d Z_t = \Delta(\Delta^{d-1} Z_t). \quad 2.22$$

Essas séries são não estacionárias homogêneas, mas, com um número finito de diferenças, consegue-se estacionarizá-las.

Dada Z_t , não estacionária, se $W_t = \Delta^d Z_t$, for estacionária pode-se representar W_t por um modelo ARMA (p,q), isto é:

$$\phi(B)W_t = \theta(B)a_t. \quad 2.23$$

Diz-se que Z_t segue um modelo autorregressivo integrado de médias móveis ARIMA (p, d, q)

$$\phi(B)\Delta^d Z_t = \theta(B)a_t. \quad 2.24$$

como $\Delta Z_t = (1-B) Z_t$, vê-se que $\Delta = 1-B$ e (2.24) pode ser escrito como:

$$\phi(B)(1-B)^d Z_t = \theta(B) Z_t. \quad 2.25$$

Na maioria dos casos é suficiente tomar $d = 1$ ou 2 diferenças para que $\Delta^d Z_t$ seja estacionária. Séries que apresentam periodicidade podem ser modeladas através de um modelo ARIMA sazonal, da forma:

$$\phi(B) \Phi(B^S) \Delta^d \Delta_S^D Z_t = \theta(B) \Theta(B^S) a_t \quad 2.26$$

onde, S é o período da sazonalidade, que pode ser trimestral, semestral ou anual, conforme o caso a seguir:

$$\Delta_S = 1 - B^S. \quad 2.27$$

A equação (2.27) representa o operador diferença sazonal, D da equação (2.26) é o número de diferenças sazonais, $\phi(B^S) = 1 + \phi_1 B^S + \dots + \phi_p^{SP}$ é o operador autorregressivo sazonal de ordem P e $\Theta(B^S) = 1 + \Theta_1 B^S + \dots + \Theta_Q B^{SQ}$ é o operador de médias móveis sazonal de ordem Q .

A equação (2.26) implica que devemos tomar D diferenças sazonais da série Z_t , de modo que a série

$$Y_t = \Delta^d \Delta_S^D Z_t = (1-B)^d (1-B^S)^D Z_t, \quad 2.28$$

seja estacionária. O modelo (2.28) é denominado ARIMA sazonal de ordem (p, d, q) $(P, D, Q)_s$. Apresenta-se alguns casos particulares do modelo:

- (i) ARIMA (0, 1, 1): $\Delta Z_t = (1 - \theta B) a_t$;
- (ii) ARIMA(1, 1, 1): $(1 - \phi B) \Delta Z_t = (1 - \theta B) a_t$;
- (iii) ARIMA (p, 0, 0) = AR(p); ARIMA (0, 0, q) = MA(q);
- (iv) ARIMA (p, 0, q) = ARMA (p, q).

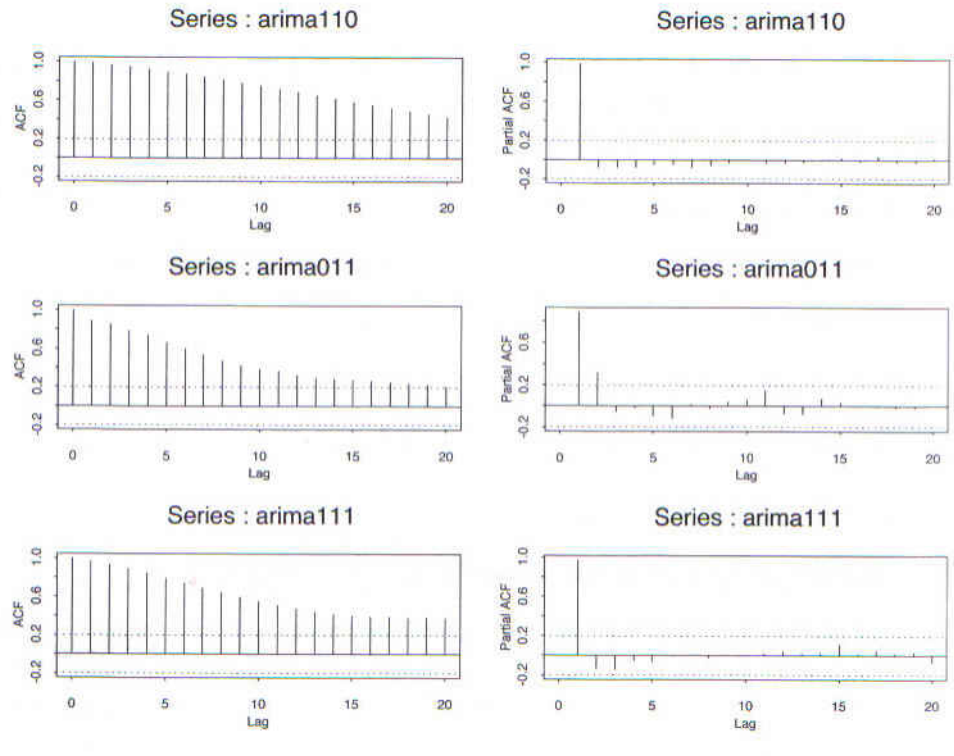
O item (i) é um caso importante, e é também chamado de modelo integrado de médias móveis, IMA (1, 1)

$$Z_t = (Z_{t-1} + a_t - \theta B) a_t. \quad 2.29$$

Pode-se demonstrar que esse modelo pode ser escrito na forma autorregressiva

$$Z_t = \lambda Z_{t-1} + \lambda(1-\lambda)Z_{t-2} + \lambda(1-\lambda)^2 Z_{t-3} + \dots + a_t \quad 2.30$$

onde, Z_t seja dado em termos de seu passado, através de uma ponderação exponencial. Na Figura 5, os gráficos da FAC e FACP amostrais das séries ARIMA (1,1,0), com $\phi = 0,8$; ARIMA(0, 1, 1), com $\theta = 0,3$; ARIMA(1, 1, 1), com $\phi = 0,8$ e $\theta = 0,3$.



Fonte: Morettin & Tolo, 2004.

Figura 5 – FAC e FACP amostrais da série ARIMA (1, 1, 0), com $\phi = 0,8$; ARIMA (0, 1, 1), com $\theta = 0,3$; ARIMA(1, 1, 1), com $\phi = 0,8$ e $\theta = 0,3$

Na Tabela 1, apresenta-se os comportamentos das funções de autocorrelações e autocorrelações parciais, juntamente com as estimativas de cada parâmetro de um modelo genérico ARIMA (p,d,q).

Tabela 1 – Comportamento das FAC e da FACP de um processo ARIMA (p, d, q)

Ordem	$(1, d, 0)$	$(0, d, 1)$
comportamento de ρ_k	decai exponencialmente	somente $\rho_1 \neq 0$
comportamento de ϕ_{kk}	somente $\phi_{11} \neq 0$	decaimento exponencial dominante
estimativas iniciais	$\phi = \rho_1$	$\rho_1 = -\frac{\theta}{1+\theta^2}$
região de admissibilidade	$-1 < \phi < 1$	$-1 < \theta < 1$
Ordem	$(2, d, 0)$	$(0, d, 2)$
comportamento de ρ_k	mistura de exponenciais ou ondas senóides amortecidas	somente $\rho_1 \neq 0$ e $\rho_2 \neq 0$
comportamento de ϕ_{kk}	somente $\phi_{11} \neq 0$ e $\phi_{22} \neq 0$	dominada por mistura de exponenciais ou senóides amortecidas
estimativas iniciais	$\begin{cases} \phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}, \\ \phi_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2} \end{cases}$	$\begin{cases} \rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}, \\ \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \end{cases}$
região de admissibilidade	$\begin{cases} -1 < \phi_2 < 1 \\ \phi_1 - \phi_2 < 1 \\ \phi_2 + \phi_1 < 1 \end{cases}$	$\begin{cases} -1 < \theta_2 < 1 \\ \theta_2 - \theta_1 < 1 \\ \theta_2 + \theta_1 < 1 \end{cases}$
Ordem	$(1, d, 1)$	
comportamento de ρ_k	decai exponencialmente após o lag 1	
comportamento de ϕ_{kk}	dominada por decaimento exponencial após o lag 1	
estimativas iniciais	$\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta}, \rho_2 = \rho_1\phi$	
região de admissibilidade	$-1 < \phi < 1, -1 < \theta < 1$	

Fonte: Morettin & Toloi, 2004

No item 2.4 apresenta-se as etapas para a efetivação da metodologia Box & Jenkins que será utilizada para a modelagem das séries em estudo.

2.4 Metodologia Box & Jenkins

A metodologia de Box-Jenkins refere-se ao método sistemático de identificação, ajuste, checagem e uso de modelos auto-regressivos integrados à média móvel ou, simplesmente, modelos ARIMA. O uso de modelos ARIMA é uma abordagem poderosa na solução de muitos problemas de previsão, pois pode proporcionar previsões, extremamente acuradas, de séries temporais.

O método é apropriado para séries de comprimento médio a longo, de no mínimo, 50 e, preferencialmente 100 observações. Um dos processos fundamentais na metodologia de Box-Jenkins é transformar uma série não-estacionária em uma

estacionária, cuja análise é mais simples, pois obtém-se, dessa maneira, a estabilidade dos parâmetros estimados.

Como uma série temporal tem dados coletados seqüencialmente ao longo do tempo, espera-se que ela apresente correlação seriada no tempo. Os modelos Box-Jenkins, genericamente conhecidos por ARIMA (*Auto Regressive Integrated Moving Average*) que em português é denominado de Autorregressivos Integrados de Médias Móveis. São modelos matemáticos que visam captar o comportamento da correlação seriada ou autocorrelação entre os valores da serie temporal, e, com base nesse comportamento, realizar previsões futuras.

A estratégia utilizada para a construção de modelos, pela metodologia Box & Jenkins, é baseada no ciclo iterativo, Figura 6, no qual a estrutura do modelo é formada pelos próprios elementos da série.



Figura 6 – Fluxograma do ciclo iterativo de Box & Jenkins

A descrição de cada uma das dessas etapas é apresentada a seguir:

2.4.1 Identificação

O processo de identificação consiste em determinar quais dos filtros AR, MA, ARMA ou ARIMA, compõem a série, bem como quais são suas respectivas ordens que, conforme explicitado o gráfico da série em 2.3., indicará, no máximo, se ela é, ou não, estacionária. A realização do processo de identificação necessita, portanto, de outros instrumentos, que são a função de autocorrelação e a função de autocorrelação parcial.

- Função de Autocorrelação (FAC)

O coeficiente de autocorrelação, ou correlação serial de ordem k , ou seja, a autocorrelação entre y_t e y_{t-k} , é dado por:

$$\rho_k = \frac{\text{cov}(y_t, y_{t-k})}{V(y_t)} = \frac{\gamma_k}{\gamma_0}. \quad 2.31$$

A seqüência de pares (k, ρ_k) , $k = 1, 2, \dots$ é denominada função de autocorrelação. Valores negativos de k não são considerados explicitamente, tendo em vista que $\rho_k = \rho_{-k}$. O coeficiente de autocorrelação ρ_k envolve parâmetros geralmente desconhecidos. Na prática, é necessário trabalhar com o coeficiente de autocorrelação amostral.

A análise estatística da FAC amostral requer o conhecimento da distribuição de r_k . Se $\rho_k = 0$ para $k \neq 0$, onde a variância de r_k é aproximadamente igual a $1/n$. Se n for grande, a distribuição de r_k é aproximadamente, Normal:

$$r_k \rightarrow N(0; 1/n). \quad 2.32$$

Bartellet (1946) obteve o seguinte resultado, para a variância de r_k , quando $\rho_k = 0$ para $k > q$:

$$V(r_k) = \frac{1}{n}(1 + 2\rho_1^2 + \dots + 2\rho_q^2), k = q + 1, q + 2, \dots \quad 2.33$$

Nesse caso, o cálculo efetivo da variância de $V(k)$ requer a substituição de ρ por r . Uma vez concluída a distribuição por r_k , pode-se construir os intervalos de confiança e realizar os testes de hipótese usuais, para verificar se cada coeficiente de autocorrelação (ρ_1, ρ_2, \dots) é nulo. Também é possível testar se os k , primeiros

coeficientes de autocorrelação são conjuntamente iguais a zero. Nesse caso, utiliza-se o teste de Ljung-Box, cuja estatística é dada por:

$$Q^*(k) = n(n+2) \sum_{k=1}^k \frac{r_k^2}{n-k} . \quad 2.34$$

$Q(k)$ tem distribuição χ^2 , com k graus de liberdade. Se $Q(k) > \chi^2_k$, rejeita-se a hipótese de k primeiros coeficientes de autocorrelação são nulos.

- Função de Autocorrelação Parcial (FACP)

O coeficiente de autocorrelação parcial de ordem k , usualmente representado por ϕ_{kk} , mede a correlação entre y_t e y_{t-k} , depois que a influência de $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$, sobre Y_t , for descontada. O coeficiente ϕ_{jj} , $j = 1, 2, \dots$ é dado pelo último coeficiente, β_{jj} , de cada uma das auto-regressões a seguir:

$$\begin{aligned} y_t &= \beta_{11}y_{t-1} + \varepsilon_t, \rightarrow \phi_{11} = \beta_{11} \\ y_t &= \beta_{11}y_{t-1} + \beta_{22}y_{t-2} + \varepsilon_t, \rightarrow \phi_{22} = \beta_{22} \\ &\vdots \\ y_t &= \beta_{k1}y_{t-1} + \beta_{k2}y_{t-2} + \dots + \beta_{kk}y_{t-k} + \varepsilon_t, \rightarrow \phi_{kk} = \beta_{kk} \end{aligned} \quad 2.35$$

os valores de ϕ_{jj} , podem também, ser obtidos a partir da solução do sistema de equações de Yule-Walker para sucessivos valores de j :

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2\rho_1 + \dots + \phi_k\rho_{k-1} \\ \rho_2 &= \phi_1\rho_1 + \phi_2 + \dots + \phi_k\rho_{k-2} \rightarrow \phi_{jj} = \phi_j, j = 1, 2, \dots \\ &\vdots \\ \rho_k &= \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \dots + \phi_k \end{aligned} \quad 2.36$$

A seqüência de pares (j, ϕ_{jj}) constitui a função de autocorrelação parcial.

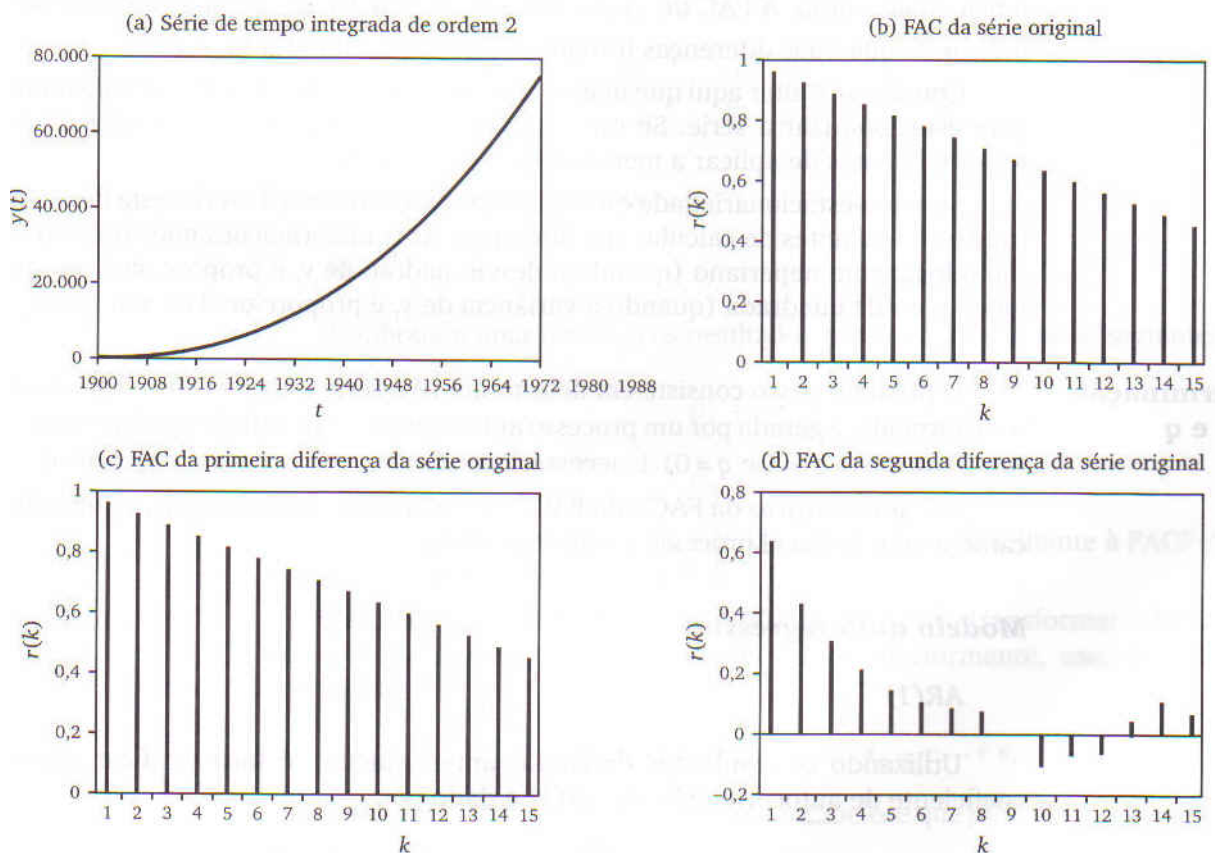
Estimativas de $\phi_{jj} - \hat{\phi}_{jj}$ podem ser obtidas aplicando-se mínimos quadrados ordinários às auto-regressões, para estimar β_{jj} , ou substituindo por ρ_j por r_j , no sistema de equações de Yule-Walker. Segundo Quenouille (1949), se o processo for auto-regressivo de ordem p , a variância de $\hat{\phi}_{jj}$ para $j > p$ é aproximadamente igual a $1/n$. Adicionalmente, para n moderado, $\hat{\phi}_{jj}$ tem-se distribuição normal. Portanto:

$$\hat{\phi}_{jj} = N(0; 1/n) \text{ para } j > p \quad 2.37$$

assim, os testes de hipótese, e os intervalos de confiança usuais, podem ser empregados para avaliar a significância da estatística de $\hat{\phi}_{jj}$. A seguir, será visto como a FAC e a FACP são importantes no processo de identificação dos modelos ARIMA. Como os modelos AR, MA e ARMA só se aplicam a séries estacionárias, ou estacionarizadas, a identificação começa pela determinação da ordem de integração da série. Se $d = 0$, a série é estacionária, e passa-se para a identificação dos filtros AR e MA. Caso contrário aplica-se quantas diferenças forem necessárias para torná-la estacionária, e trabalha-se com série resultante para a identificação dos componentes AR e MA.

- Identificação da Ordem de Integração (d)

O gráfico da série original pode dar uma primeira indicação da violação ou não da condição de estacionariedade. Considere a série y_t , representada na Figura 7a.



Fonte: Manual de Econometria Vasconcelos & Alves 2000.

Figura 7 – Série Y_t que representa um processo integrado de ordem 2 com suas respectivas funções de autocorrelações

Nota-se, facilmente, não se tratar de uma série estacionária. A média cresce com o tempo. Sabe-se, portanto, que d é diferente de zero. O valor de d será respondido com a ajuda da FAC.

Séries não-estacionárias apresentam fortes correlações seriais. Os valores dos coeficientes de autocorrelação declinam, muito lentamente, à medida em que k aumenta. Assim sendo, uma FAC amostral, com valores inicialmente altos, e que não declinam rapidamente para um valor estatisticamente igual a zero a medida que k aumenta, indica que a série é não estacionária e precisa, portanto, ser diferenciada.

Sugere-se a seguinte regra de bolso: “se para k maior do 5, o valor do coeficiente de autocorrelação ainda for maior 0,7, em módulo, a série dever ser considerada não estacionária. Caso isso ocorra, aplica-se a primeira diferença e observa-se a FAC da nova série, então obtida. Se ela também não declinar rapidamente, calcula-se a segunda diferença e observa-se a FAC correspondente. O procedimento deve ser repetido até que se obtenha uma série estacionária.” A Figura 7 traz a FAC da série original em (a). Todos os coeficientes de autocorrelação amostral apresentados (r_1 a r_{10}) são positivos e estatisticamente diferentes de zero (todos eles estão fora do intervalo de confiança). Esses resultados corroboram, portanto, a avaliação inicial: a série é não estacionária ($d > 0$). A Figura 7c apresenta a FAC da primeira diferença (Δy_t). Constata-se que também a nova série não é estacionária ($d > 1$), logo mais uma diferença é necessária. A FAC de $\Delta^2 y_t$, vista na Figura 13d, declina rapidamente indicando que duas diferenças tornam a série y_t estacionária ($d = 2$).

Convém ressaltar, aqui, que nem sempre a aplicação de diferenças é suficiente para estacionarizar a série. Se esta exibir tendência determinística, é necessário removê-la antes de aplicar a metodologia de Box & Jenkins.

2.4.2 Estimação dos Parâmetros

O próximo passo consiste em determinar se a série estacionária, original ou transformada, é gerada por um processo auto-regressivo ($p \neq 0$), de médias móveis ($q \neq 0$), e/ou mista. Neste caso é necessário determinar o valor de p e/ou q .

As características da FAC e da FACP em cada caso, é que indicarão qual o possível processo gerador da série.

- Modelo auto-regressivo - AR(1)

Utilizando os resultados derivados anteriormente, é fácil verificar que o coeficiente de autocorrelação do AR(1) é dado por:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k, \quad k = 1, 2, \dots \quad 2.38$$

sabe-se que $|\phi| < 1$, portanto, a FAC desse modelo declina exponencialmente à medida em que k aumenta.

No que diz respeito à FACP, tem-se o seguinte:

$$\begin{aligned} \phi_{11} &\neq 0 \\ \phi_{kk} &= 0, \end{aligned} \quad k = 2, 3, \dots \quad 2.39$$

portanto, a FACP é truncada em $k = 1 = p$. Essa característica é facilmente compreensível, quando se recorda que a FACP resulta da estimação de sucessivas auto-regressões. Como o AR (1) só tem y_{t-1} como “variável explicativa”, só o coeficiente associado a esta será diferente de zero.

- AR(p)

Para um modelo auto-regressivo genérico, ρ_k é expresso por:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, \quad k = 1, 2, \dots \quad 2.40$$

o comportamento da FAC não é óbvio como no AR (1). Continua declinando à medida que k aumenta, mas o declínio traduz-se agora em decrescimentos exponenciais e/ou ondas senoidais amortecidas. Em $k = p$, a FACP é truncada.

$$\begin{aligned} \phi_{11} &\neq 0, k = 1, 2, \dots, p \\ \phi_{kk} &= 0, k = p + 1, p + 2, \dots \end{aligned} \quad 2.41$$

- Modelo de médias móveis - MA (1)

Lembrando, uma vez mais, os resultados anteriores, chegamos ao seguinte:

$$\begin{aligned} \rho_1 &= \frac{-\theta}{1+\theta^2} ; \\ \rho_k &= 0, k > 1. \end{aligned} \quad 2.42$$

assim, o comportamento da FAC de um MA (1) é semelhante à FACP de um AR(1): ela é truncada em $k = 1 = q$.

Para definir a característica da FAC P é necessário transformar o MA (1) em um modelo auto-regressivo. Conforme visto anteriormente, esse modelo tem ordem infinita:

$$y_t = -\theta y_{t-1} - \theta^2 y_{t-2} - \theta^3 y_{t-3} - \dots + \varepsilon_t \quad 2.43$$

como, a condição de invertibilidade estabelece que $|\phi| < 1$, a FACP do MA (1) decresce exponencialmente.

- MA(q)

A generalização das características das funções de autocorrelação, e autocorrelação parcial, para um processo de médias móveis de ordem q, é:

a. a FAC é truncada em $k = q$, ver (2.42);

b. a FACP decresce à medida em que k aumenta, mas o decréscimo não segue nenhum padrão fixo.

- Modelo Autorregressivo de Médias Móveis - ARMA (1, 1)

Os coeficientes de autocorrelação, desse modelo, são definidos por:

$$\rho_1 = \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta} \quad 2.44$$

$$\rho_k = \phi\rho_{k-1}, k = 2, 3, \dots$$

como $|\phi| < 1$, a FAC do ARMA (1, 1) é declinante.

Tal como no caso do MA (1), a definição da FACP do modelo ARMA (1, 1) requer sua transformação em um AR(∞):

$$y_t = (\phi - \theta)y_{t-1} + \theta(\phi - \theta)y_{t-2} + \theta^2(\phi - \theta)y_{t-3} + \dots + \varepsilon_t \quad 2.45$$

A condição de invertibilidade $|\phi| < 1$, implica que a FACP do ARMA (1, 1) decresça quando k aumenta. Como nenhuma das funções é truncada, os valores de p e q não são indicados claramente.

- ARMA (p, q)

No caso geral, tem-se que:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, \quad k = p+1, p+2, \dots \quad 2.46$$

e conseqüentemente, a FAC é declinante a partir de $k = p$.

Esse comportamento é definido com base na FAC e na FACP teóricas. Na prática, será preciso trabalhar com as funções amostrais, que não serão tão bem comportadas quanto às funções teóricas, e que exigirão a análise da significância estatística dos coeficientes de autocorrelação r_k e $\hat{\phi}_{kk}$ obtidos. Em vista disso, não é recomendável trabalhar com séries curtas.

- Estimação

Uma vez determinados os valores de p , d e q , passa-se para a estimação dos p parâmetros ϕ , dos q parâmetros θ e da variância σ_ε^2 do modelo:

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad 2.47$$

onde:

$$w_t = \Delta^d y_t = (1 - B)^d y_t. \quad 2.48$$

A estimação pode ser por mínimos quadrados e por máxima verossimilhança.

A estimação por mínimos quadrados requer a minimização de:

$$S(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) = \sum_{t=1}^n \hat{\varepsilon}_t^2, \quad 2.49$$

onde:

$$\hat{\varepsilon}_t = \theta^{-1}(B)\phi(B)w_t. \quad 2.50$$

$S(\hat{\phi}, \hat{\theta})$ depende de valores de w_t e de ε_t , ou seja, valores anteriores ao período amostrado. Uma alternativa é atribuir aos ε 's o valor zero, para que corresponda a suas esperanças não condicionais. Aos w 's passados, atribui-se o valor médio da série \bar{w} . As estimativas obtidas por mínimos quadrados dependerão, portanto, desses valores atribuídos aos ε 's e aos w 's. Em vista disso, o método é chamado de mínimos quadrados condicionados. Outro aspecto, que cabe destacar,

é que, se o componente MA estiver presente, o modelo será não linear, o que exigirá a utilização do método de mínimos quadrados não lineares. Para a estimação da máxima verossimilhança, é necessário admitir, inicialmente, que os ruídos brancos ε_t têm a distribuição Normal. Como consequência, w_t também terá distribuição Normal e a distribuição conjunta de $w = (w_1, \dots, w_n)$ é:

$$w \sim N(0; \sigma_\varepsilon^2 \Omega). \quad 2.51$$

A função de verossimilhança a ser maximizada pode então ser definida:

$$L(\phi, \theta, \sigma_\varepsilon^2 / w) = [2\pi\sigma_\varepsilon^2]^{-\frac{n}{2}} |\Omega|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} w' \Omega^{-1} w / \sigma_\varepsilon^2\right]. \quad 2.52$$

A maximização de L é complexa, particularmente em virtude da presença da matriz Ω . Qualquer que seja o método adotado, o processo de estimação é extremamente trabalhoso e requer o uso do computador.

Experimentos de Monte Carlo, realizados por vários autores, indicam que o método de máxima verossimilhança é superior ao método de mínimos quadrados, quando o tamanho da série é pequeno e, sobretudo, quando os valores dos parâmetros aproximam-se dos limites da condição de invertibilidade.

Feita a identificação do modelo genérico ARIMA (p, d, q), que melhor representa a série em estudo, passa-se à obtenção de estimativas para os parâmetros do modelo identificado.

A estimação dos parâmetros significa a obtenção de “p” estimativas para os parâmetros $\phi_1, \phi_2, \dots, \phi_p$, e “q” estimativas para os parâmetros $\theta_1, \theta_2, \dots, \theta_q$, além de variância do ruído, σ_a^2 , ou seja, $p + q + 1$, parâmetros no vetor $\varepsilon = (\phi, \theta, \sigma_a^2)$.

Aqui, quando $d > 0$, supõe-se que $\mu = 0$. Caso contrário, μ é incluído como mais um parâmetro a ser estimados, e ter-se-á $p + q + 2$ parâmetros. Seja $\eta = (\phi, \theta)$.

Para determinar as estimativas de máxima verossimilhança (EMV) será necessário supor que o processo a_t seja normal, ou seja, para cada t , $a_t \sim (0, \sigma_a^2)$. Nessas condições, a EMV serão, aproximadamente os estimadores de mínimos quadrados (EMQ).

O método de máxima verossimilhança consiste em achar grandezas populacionais que gerem os valores que mais se assemelhem aos da amostra

observada, ou seja, o método consiste em selecionar aqueles estimadores que maximizam a probabilidade de se obter a amostra realmente observada.

O método de mínimos quadrados tem o objetivo de estimação de um modelo ARIMA (p,d,q). Para isso, passa-se a achar o vetor autorregressivo $\tilde{\phi}$ e o vetor dos parâmetros médias móveis $\tilde{\theta}$, tais que minimizem a soma das diferenças quadradas entre os pontos observados na amostra e o esperado pela estimativa obtida com os parâmetros estimados. Simbolicamente, deve-se achar $\tilde{\phi}$, $\tilde{\theta}$, de modo que $S(\tilde{\phi}, \tilde{\theta}) = \sum a_t^2$ seja um mínimo.

Tomando-se d diferenças, para alcançar estacionariedade, fica-se com $n = N - d$ observações Z_1, \dots, Z_n , onde $W_t = \Delta^d Z_t$. Como o modelo ARMA (p, q), resultante, é estacionário e inversível, pode-se escrever

$$a_t = \tilde{Z}_t - \phi_1 \tilde{Z}_{t-1} - \dots - \phi_p \tilde{Z}_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad 2.53$$

onde

$$\tilde{Z}_t = Z_t - \mu. \quad 2.54$$

2.4.3 Adequação do Modelo

A identificação é a etapa mais difícil da metodologia de Box & Jenkins. Não são raros os casos em que não se consegue identificar um único modelo, e sim vários modelos candidatos a gerador da série em estudo. Isso porque, trabalhando com a FAC e a FACP amostrais, fica difícil, muitas vezes, decidir se elas são decrescente ou se são truncadas.

Para escapar, desse impasse, muitos pesquisadores preferem utilizar um outro procedimento de identificação, que depende menos do julgamento de quem está analisando a série de tempo. Esse procedimento faz uso de critérios de seleção de modelos construídos com base na variância estimada ε_t , no tamanho da amostra e nos valores de p e q. Os mais utilizados são o critério AIC e o critério BIC.

$$AIC = \ln \hat{\sigma}_\varepsilon^2 + \frac{2(p+q)}{n};$$

$$BIC = \ln \hat{\sigma}_\varepsilon^2 + \frac{(p+q)\ln(n)}{n}.$$

Em vez de estabelecer p e q precisamente, estima-se os modelos correspondentes a vários pares (p, q) e escolhe-se aquela especificação que apresentar o menor valor para AIC ou para BIC.

A presença de p e q , nas fórmulas dos critérios AIC e BIC, tem por objetivo “penalizar” os modelos com muitos parâmetros, tendo em vista que modelos mais parcimoniosos devem ser privilegiados, por apresentarem menor número de parâmetros a serem estimados.

A crítica que se faz, a esses critérios, é que podem conduzir a modelos super especificados, ou seja, modelos com valores de p e/ou q maiores do que o correto.

Assim sendo, os critérios AIC e BIC deve ser usados como procedimento complementar, e não alternativo àquele baseado na FAC e na FACP.

- Verificação

Essa etapa da metodologia, de Box & Jenkins, consiste em verificar se o modelo identificado, e estimado, é adequado. Em caso positivo, pode-se adotá-lo para fazer previsão. Em caso negativo, outra especificação deve ser escolhida para modelar a série, o que implica refazer as etapas de identificação e estimação. As formas de verificação, mais comumente consideradas, serão agrupadas em dois itens: análise dos resíduos e avaliação da ordem do modelo.

- Análise dos Resíduos

s resíduos do modelo estimado, $\hat{\varepsilon}_t$, são estimativas do ruído branco, ε_t . Assim sendo, devem comportar-se como um ruído branco, se o modelo estiver adequadamente especificado. Em particular, seus coeficientes de autocorrelação devem ser, estatisticamente, iguais a zero. Portanto para verificar se isso ocorre, podem ser feitos testes individuais e testes conjuntos para os coeficientes de autocorrelação $r_k(\hat{\varepsilon})$, tal como na etapa de identificação. Esses coeficientes são dados por

$$r_k(\hat{\varepsilon}) = \frac{\sum_{t=k+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^n \hat{\varepsilon}_t^2}. \quad 2.56$$

Os resíduos se distribuem como uma Normal de média zero e variância $1/n$. Todavia, conforme Durbin (1970), para valores pequenos de k , a variância de $r_k(\hat{\varepsilon})$ pode ser bem menor que $1/n$, o que implica considerar $r_k(\hat{\varepsilon})$, estatisticamente igual a zero, quando isso pode não ser correto. Para valores moderados ou grandes de k , a distribuição é válida e podem-se realizar testes de hipóteses e construir intervalos de confiança para avaliar a significância de cada $r_k(\hat{\varepsilon})$. Para o teste conjunto, utiliza-se a estatística Q^* , de Ljung_Box, que expressa pela fórmula (2.18). $Q^*(k)$ tem distribuição qui-quadrado, com $k-p-q$ graus de liberdade. Convém, adicionalmente, analisar o gráfico dos resíduos, para verificar se há indícios de que a hipótese de variância constante no tempo não esteja sendo respeitada.

- Avaliação da ordem do modelo

O objetivo é verificar se o modelo não está superespecificado (p e/ou q maiores do que o devido), nem subespecificado (p e/ou q menores do que o devido). O modelo não deve conter parâmetros em excesso. É isso que estabelece o critério da parcimônia. A verificação da existência de parâmetros r redundantes é feita com base no erro-padrão dos coeficientes $\hat{\phi}$ e $\hat{\theta}$ e na correlação entre eles.

Se o valor de um coeficiente estimado for pequeno, em relação a seu erro-padrão, indicando sua não-significância estatística, é provável que haja superespecificação. Se for o coeficiente de maior ordem (por exemplo, $\hat{\phi}_4$ em um ARMA (4,2)), deve-se suprimi-lo, estimando-se, assim, um modelo de menor ordem. Se não for o de menor ordem, convém analisar a correlação entre ele e o outro coeficiente do modelo: alta correlação indica que um dos dois pode ser redundante.

Para verificar se está havendo subespecificação, deve-se introduzir parâmetros adicionais e analisar sua significância estatística. A cada nova estimação do modelo, apenas um parâmetro deve ser introduzido. Se após a verificação, pelas formas indicadas, houver “empate” entre os dois ou mais modelos, o desempate deve ser feito comparando as estimativas da variância de $\hat{\varepsilon}_t(\hat{\sigma}_\varepsilon^2)$ e os valores obtidos para os critérios AIC e BIC. Quanto menor o valor de $\hat{\sigma}_\varepsilon^2$, melhor, em princípio, o

modelo estimado e mais precisas as previsões feitas com base nele. Os critérios AIC e BIC devem, também, ser minimizados.

A estatística R^2 , muito utilizada na avaliação da qualidade do ajustamento de um modelo de regressão, não tem o mesmo papel para os modelos ARMA. Nesse caso, R^2 está relacionado aos valores dos ϕ 's e dos θ 's, e não a $\hat{\sigma}_\varepsilon^2$, conforme demonstrou Nelson 1976. Para o modelo MA(1), por exemplo, $R^2 = \theta^2/(1 + \theta^2)$. Respeitada a condição de invertibilidade, o valor máximo que R^2 pode assumir, nesse caso, está próximo de 0,5.

2.4.4 Previsão

Após o pesquisador haver selecionado entre os modelos estimados, aquele que se mostrar mais adequado e que tiver o menor AIC e BIC, chega-se à última etapa da metodologia de Box & Jenkins, que consiste na realização de previsões da série y_t em instantes de tempo posteriores a n .

A previsão pode ser pontual, ou por intervalo. No primeiro caso, é necessário definir o previsor ótimo da série; no segundo, é preciso conhecer, adicionalmente, a distribuição do erro de previsão.

- Previsão Pontual

O previsor ótimo de y é obtido respondendo à seguinte questão: tendo por base as informações disponíveis até o instante n , cristalizadas em $\{y_1, y_2, \dots, y_n\}$, qual o melhor previsor para y no instante $n + \ell$ ($y_{n+\ell}$)?

O previsor ótimo, “ ℓ períodos à frente”, representado por $\hat{y}_n(\ell)$, é aquele que minimiza o erro quadrático médio de previsão

$$E[y_{n+\ell} - \hat{y}_n(\ell)]^2 = E[e_n^2(\ell)], \quad 2.57$$

onde $e_n(\ell)$ é o erro de previsão ℓ instante à frente de n . Conforme demonstram Box & Jenkins (1976, p. 127-128), $\hat{y}_n(1)$ é dada pela esperança condicional y_{n+1} :

$$\hat{y}_n(\ell) = E[y_{n+\ell} / y_n, y_{n-1}, \dots, y_1]. \quad 2.58$$

suponha-se que o modelo selecionado seja um ARMA (1, 1):

$$y_t = \phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}. \quad 2.59$$

A previsão de y , nos instantes $n+1$, $n+2$ e $n+\ell$, será dada pela esperança condicional de y , conforme indicado a seguir.

1) instante $n+1$

equação do modelo:

$$y_{n+1} = \phi y_n + \varepsilon_{n+1} - \theta \varepsilon_n. \quad 2.60$$

equação de previsão:

$$\begin{aligned} \hat{y}_n(1) &= E[y_{n+1} / y_n, y_{n-1}, \dots, y_1]; \\ \hat{y}_n(1) &= E[(\phi y_n + \varepsilon_{n+1} - \theta \varepsilon_n) / y_n, y_{n-1}, \dots, y_1]; \\ \hat{y}_n(1) &= \phi y_n - \theta \hat{\varepsilon}_n. \end{aligned} \quad 2.61$$

2) instante $n+2$

equação do modelo:

$$y_{n+2} = \phi y_{n+1} + \varepsilon_{n+2} - \theta \varepsilon_{n+1}. \quad 2.62$$

equação de previsão:

$$\begin{aligned} \hat{y}_n(2) &= E[y_{n+2} / y_n, y_{n-1}, \dots, y_1]; \\ \hat{y}_n(2) &= E[(\phi y_{n+1} + \varepsilon_{n+2} - \theta \varepsilon_{n+1}) / y_n, y_{n-1}, \dots, y_1]; \\ \hat{y}_n(2) &= \phi \hat{y}_n(1). \end{aligned} \quad 2.63$$

3) instante $n+\ell$ ($\ell > 1$)

equação do modelo:

$$y_{n+\ell} = \phi y_{n+\ell-1} + \varepsilon_{n+\ell} - \theta \varepsilon_{n+\ell-1}. \quad 2.64$$

equação de previsão:

$$\begin{aligned} \hat{y}_n(\ell) &= E[y_{n+\ell} / y_n, y_{n-1}, \dots, y_1]; \\ \hat{y}_n(\ell) &= E[(\phi y_{n+\ell-1} + \varepsilon_{n+\ell} - \theta \varepsilon_{n+\ell-1}) / y_n, y_{n-1}, \dots, y_1]; \\ \hat{y}_n(\ell) &= \phi \hat{y}_n(\ell-1). \end{aligned} \quad 2.65$$

As previsões para y_t foram obtidas com base nas seguintes esperanças condicionais:

$$E[y_{n+j} / y_n, y_{n-1}, \dots, y_1] = \begin{cases} y_{n+1} \text{ para } j \leq 0 \\ \hat{y}_n(j) \text{ para } j > 0 \end{cases} \quad 2.66$$

$$E[\varepsilon_{n+j} / y_n, y_{n-1}, \dots, y_1] = \begin{cases} \hat{\varepsilon}_{n+j} = y_{n+j} - \hat{y}_{n+j}, \text{ para } j \leq 0 \\ 0, \text{ para } j > 0 \end{cases}, \quad 2.67$$

ou seja, a esperança condicional dos valores presente e passados de y_t é igual aos próprios valores, e a dos valores futuros é igual a suas previsões. A esperança condicional de ε_t é igual a zero no instante futuro (já que ε_t é um ruído branco) e igual a suas estimativas nos períodos corrente passados.

Generalizando o procedimento ARIMA (p, d, q), cuja a equação é:

$$\phi(B)\Delta^d y_t = \theta(B)\varepsilon_t \quad 2.68$$

ou

$$\begin{aligned} w_t &= \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} + \dots - \theta_q \varepsilon_{t-q} \\ w_t &= \Delta^d y_t. \end{aligned} \quad 2.69$$

tem-se a equação do modelo para $n+\ell$

$$w_{n+\ell} = \phi_1 w_{n+\ell-1} + \dots + \phi_p w_{n+\ell-p} + \varepsilon_{n+\ell} - \theta_1 \varepsilon_{n+\ell-1} - \dots - \theta_q \varepsilon_{n+\ell-q}. \quad 2.70$$

e a equação de previsão é dada por:

$$\hat{w}_n(\ell) = E[w_{n+\ell} / w_n, w_{n-1}, \dots, w_1]; \quad 2.71$$

$$\begin{aligned} \hat{w}_n(\ell) &= \phi_1 \hat{w}_n(\ell-1) + \phi_2 \hat{w}_n(\ell-2) + \dots + \phi_{\ell-1} \hat{w}_n(1) + \phi_\ell w_n + \phi_{\ell+1} w_{n-1} + \dots + \\ &+ \phi_p w_{n+\ell-p} - \theta_\ell \hat{\varepsilon}_n - \theta_{\ell+1} \hat{\varepsilon}_{n-1} - \dots - \theta_p \hat{\varepsilon}_{n+\ell-q}. \end{aligned} \quad 2.72$$

Se o horizonte da previsão for maior que a ordem do modelo ($\ell > p$ e $\ell > q$), a equação de $\hat{w}_n(\ell)$ fica reduzida a:

$$\hat{w}_n(\ell) = \phi_1 \hat{w}_n(\ell-1) + \phi_2 \hat{w}_n(\ell-2) + \dots + \phi_p \hat{w}_n(\ell-p). \quad 2.73$$

Portanto, a previsão para a série ℓ períodos a frente é baseada, apenas, em valores previstos de w . Isso revela que a metodologia de Box & Jenkins não é recomendável para previsões a longo prazo.

Nota-se que a equação é definida para a série estacionária w_t . Para retornar a série original y_t , basta utilizar a operação inversa à aplicação de diferenças. Se, por

exemplo, $d=1$, ou seja, $w_t = \Delta y_t = y_t - y_{t-1}$, as previsões para a série y_t serão obtidas da seguinte forma:

$$\begin{aligned}\hat{y}_n(1) &= \hat{w}_n(1) + y_n \\ \hat{y}_n(2) &= \hat{w}_n(2) + \hat{y}_n(1) = \hat{w}_n(2) + \hat{w}_n(1) + y_n \\ &\vdots \\ \hat{y}_n(\ell) &= \hat{w}_n(\ell) + \hat{y}_n(\ell-1) + \hat{w}_n(1) + y_n.\end{aligned}\tag{2.74}$$

- Previsão por intervalo

Para construir o intervalo de previsão, é necessário conhecer a distribuição do erro de previsão $e_n(\ell)$, que no modelo ARIMA (p, d, q) é dado por:

$$e_n(\ell) = w_{n+\ell} - \hat{w}_n(\ell).\tag{2.75}$$

A dedução da distribuição de $e_n(\ell)$ ficará facilitada se ele for expresso de forma conveniente. Essa forma resulta da transformação do ARMA (p,q), que descreve w_t em um MA(q):

$$\phi(B)w_t = \theta(B)\varepsilon_t;\tag{2.76}$$

$$w_t = \phi^{-1}(B)\theta(B)\varepsilon_t;\tag{2.77}$$

como, $\phi^{-1}(B)$ é um polinômio de ordem infinita, a equação de w_t passa a ser escrita da seguinte forma:

$$\begin{aligned}w_t &= \psi(B)\varepsilon_t; \\ w_t &= (\psi_0 + \psi_1 B + \psi_2 B^2 + \dots)\varepsilon_t; \\ w_t &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}.\end{aligned}\tag{2.78}$$

onde $\psi_0 = 1$.

Assim sendo, a equação $w_{n+\ell}$ é:

$$w_{n+\ell} = \varepsilon_{n+\ell} + \psi_1 \varepsilon_{n+\ell-1} + \dots + \psi_{\ell-1} \varepsilon_{n+1} + \psi_{\ell} \varepsilon_n + \psi_{\ell+1} \varepsilon_{n-1} + \dots.\tag{2.79}$$

A correspondente equação de previsão é:

$$\hat{w}_n(\ell) = \psi_{\ell} \varepsilon_n + \psi_{\ell+1} \varepsilon_{n-1} + \psi_{\ell+2} \varepsilon_{n-2} + \dots\tag{2.80}$$

o erro de previsão ℓ passos à frente passa, então, a ser representado por:

$$e_n(\ell) = \varepsilon_{n+\ell} + \psi_1 \varepsilon_{n+\ell-1} + \dots + \psi_{\ell-1} \varepsilon_{n+1}. \quad 2.81$$

Com essa expressão para $e_n(\ell)$, fica fácil deduzir sua média e sua variância, não esquecendo que ε_t é um ruído branco:

$$\begin{aligned} E[e_n(\ell)] &= 0 \\ V[e_n(\ell)] &= E[e_n^2(\ell)] = \sigma_\varepsilon^2 (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{\ell-1}^2) = V(\ell). \end{aligned} \quad 2.82$$

Supondo que ε_t tem distribuição normal:

$$e_n(\ell) \sim N[0; V(\ell)] \quad 2.83$$

Uma vez definida a distribuição do erro de previsão, pode-se construir o intervalo de previsão para $w_{n+\ell}$:

$$\hat{w}(\ell) \pm z \cdot \sigma_\varepsilon \sqrt{1 + \psi_1^2 + \dots + \psi_{\ell-1}^2} \quad 2.84$$

como a variância de $e_n(\ell)$ cresce à medida em que ℓ aumenta, o intervalo de previsão vai sendo ampliado. Isso indica que previsões de curto prazo são mais precisas do que as de longo prazo.

- Atualização da Previsão

Tendo em vista que a metodologia de Box & Jenkins tem sua capacidade de previsão comprometida a longo prazo, convém, sempre que possível, atualizar as previsões já realizadas.

À medida em que os valores y_{n+1} , y_{n+2} etc. vão se tornando conhecidos, pode-se proceder a atualização das previsões por dois caminhos alternativos:

1. substituindo-se $\hat{y}_n(1)$, $\hat{y}_n(2)$..., que aparecem nas equações de previsão por seus valores efetivos; 2. reestimando o modelo.

2.5 Análise de Intervenção

Os modelos univariados consideram apenas uma série no tempo. A metodologia de Box & Jenkins permite, contudo, que outras séries sejam incluídas no modelo, à semelhança dos modelos de regressão linear.

Quando as séries incluídas são variáveis quantitativas tradicionais, tem-se o modelo de função transferência. Quando são variáveis binárias (dummy), destinadas a captar o efeito de eventos mensuráveis, o modelo resultante é chamado de análise de intervenção (Vasconcelos e Alves, 2000).

A análise de intervenção é um modelo de função de Transferência Estocástica, onde é possível interpretar a maneira de incorporar seus efeitos ao modelo da série temporal. É possível fazer intervenções naturais e induzidas pelo homem, com o objetivo de avaliar o impacto de um evento no comportamento da série temporal, verificando se causa, ou não, uma mudança significativa no nível médio de uma série temporal.

Os maiores efeitos da intervenção são notados na mudança do nível na direção ou na inclinação da série em estudo, e também para alterar as variáveis dos erros e introduzir no modelo componentes que antes não haviam. Por exemplo, introduzir componentes autorregressivos em um processo de médias móveis. Outro exemplo é em relação à variabilidade da série, onde pode-se torná-la mais estável, ou mais variável, a partir das intervenções.

Quando a intervenção a ser feita é complexa, o seu efeito ocorre gradativamente, e se houver mais de uma intervenção é possível existir uma interação entre elas, confundindo parte de seus efeitos. Os efeitos de intervenção podem ser constatados ao longo do tempo, assim como pode ser variáveis. O seu efeito é determinado pela estrutura da Função Transferência, do seu modelo e dos seus parâmetros estimados. Três fontes de ruído podem lesar os efeitos de intervenção. Esses ruídos são: a tendência, o efeito sazonal e o erro aleatório. O erro a_t deve ser modelado por um ARIMA(p, d, q), pois este leva em conta as componentes de tendência, sazonalidade e erro.

A utilidade de Análise de intervenção pode manifestar-se nas mais diversas áreas: ciências sociais e políticas, economia, sociologia, história, psicologia, meio ambiente, entre outras. Pack mostra o desenvolvimento da construção de modelos para séries temporais e análise de regressão. Assim:

- a. Modelos de função de transferência de entrada simples podem ser comparados a modelos de regressão simples;
- b. Modelos de função de transferência de entrada múltipla são comparáveis a modelos de regressão múltipla.

- Modelo de Função de Transferência

É comum, em muitas áreas do conhecimento, modelar relações entre variáveis, no sentido de estabelecer padrões de causalidade ou “feedback” entre elas (MORETTIN & TOLOI, 2004). O caso mais simples é aquele em que se tem uma “variável de entrada” X_t e uma “variável de saída” Y_t , de modo a formar um sistema dinâmico.



Fonte: Manual de Econometria, 2000.

Figura 8 – Representação de um sistema dinâmico

Um dos objetivos do sistema dinâmico pode ser a previsão de série Y_t , com base em valores passados e presentes de X_t e Y_t , usando-se o modelo:

$$Y_t = \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_q X_{t-q} + \eta_t, \quad 2.85$$

onde $\alpha_1, \dots, \alpha_p$, β_0, \dots, β_q são parâmetros a serem estimados, e η_t é o ruído, em geral um ARIMA (p, d, q).

Considerando-se uma série temporal bivariada (X_t, Y_t) com $t \in Z = \{0, \pm 1, \pm 2, \dots\}$, de tal modo que X_t seja a série de entrada, e Y_t a série de saída do sistema dinâmico, a ligação entre X_t e Y_t pode ser de qualquer tipo, mas nos interessaremos no caso em que a relação for linear.

$$Y_t = \sum_{j=0}^{\infty} V_j X_{t-j}, \quad t \in Z. \quad 2.86$$

utilizando-se o operador retroativo B, $B X_t = X_{t-1}$, pode-se escrever (85) na forma

$$Y_t = (V_0 + V_1 B + V_2 B^2 + \dots) X_t, \quad e \quad 2.87$$

$$\varphi(B) (1-B)^d = \theta(B) a_t. \quad 2.88$$

onde $\varphi(B)$ é o operador autorregressivo de ordem p e $\theta(B)$ é o operador médias móveis de ordem q, e a_t é o ruído branco seguindo $N \approx (0, \sigma_a^2)$.

Freqüentemente, o modelo é não parcimonioso, no sentido de haver necessidade de estimar muitos V_j . Entretanto, se $V(B)$ for uma função racional, pode-se escrevê-la:

$$V(B) = \frac{W(B) B^b}{\delta(B)}, \quad 2.89$$

onde

$$W(B) = W_0 - W_1 B - W_2 B^2 - \dots - W_s B^s \quad \text{e} \quad 2.90$$

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r \quad 2.91$$

são polinômios finitos de graus s e r, respectivamente. b significa um atraso na resposta à entrada para admitir um efeito de retardo na entrada, ou seja, é um parâmetro de defasagem.

Se b diferir de zero, indicará que $V_0 = V_1 = \dots = V_{b-1} = 0$.

$$Y_t = V(B) X_t + \eta_t, \quad 2.92$$

onde η_t não é, necessariamente, ruído branco, usualmente segue um modelo ARIMA (p, d, q), ou seja:

$$\varphi(B) (1-B)^d = \theta(B) a_t; \quad 2.93$$

$$Y_t = \frac{W(B) B^b}{\delta(B)} X_t + \eta_t; \quad 2.94$$

$$\delta(B) Y_t = W(B) B^b X_t + \eta_t. \quad 2.95$$

Comparando-se (2.88) com (2.94) tem-se

$$V(B) \delta(B) = W(B) B^b. \quad 2.96$$

De (2.96), tem-se que o modelo de função de transferência discreto será estável se as raízes da equação característica $\delta(B)=0$ caírem fora do círculo unitário.

O número total de parâmetros de uma função de transferência é:

$d_1 \rightarrow$ diferenciações na saída ($\nabla_1^d Y_t$);

$d_2 \rightarrow$ diferenciações na saída ($\nabla_2^d Y_t$);

$r \rightarrow$ grau do polinômio $\delta(B)$;

$s \rightarrow$ grau do polinômio $W(B)$;

$p \rightarrow$ grau do polinômio $\varphi(B)$;

$q \rightarrow$ grau do polinômio $\theta(B)$;

$b \rightarrow$ lag do retardo.

Uma equação de diferença de ordem “ r ”, cuja solução é do tipo:

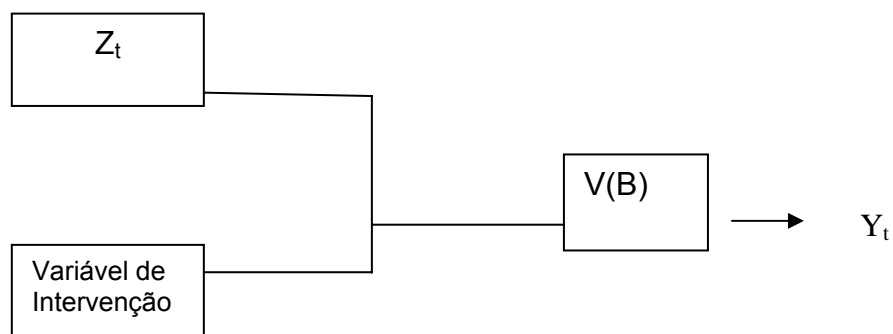
$r=0 \Rightarrow v_k=0$;

$r=1 \Rightarrow v_k$ tem um decrescimento exponencial;

$r=2 \Rightarrow v_k$ tem um comportamento de senóide amortecida (Raízes complexas).

- Modelos dinâmicos para intervenção

Basicamente, a construção de modelos de intervenção consiste em acrescentar, aos modelos ARIMA, os efeitos de variáveis exógenas, através de uma função de transferência.



Fonte: Manual de Econometria, 2000

Figura 9 - Representação de um modelo dinâmico com intervenção

Seja uma série temporal, para a qual verificou-se e estimou-se um modelo ARIMA com o qual vem se fazendo previsões há algum tempo. Num dado instante

ocorre um evento independente do fenômeno que originou a série temporal, mas cujos efeitos possam manifestar-se sobre ela.

Toma-se como exemplo uma série de produção anual de um determinado produto agrícola, para a qual se dispõe de um modelo adequado para fazer previsões. Em um determinado momento, ou intervalo de tempo, há uma ocorrência de alterações climáticas (como geada, seca ou enchentes) que possam afetar, temporariamente, ou permanentemente essa, produção agrícola e, conseqüentemente, o modelo utilizado para representá-la.

A esse evento externo, cujos efeitos influenciam a série em estudo, devem ser incorporados ao modelo, como uma informação adicional a série. A essa incorporação de informação chamamos de intervenção.

Na maioria dos casos tratamos de modelos dinâmicos onde uma variável pode ser endógena ou dependente; de uma ou mais variáveis chamadas exógenas ou independentes, e esta dependência é especificada através de modelos da forma:

$$Z_t = f(k, x, t) + b_t \quad 2.97$$

onde:

$Z_t = F(Z_t)$ é a série Z_t transformada ou não;

$f(k, x, t)$ é algum efeito determinística no tempo t , ou o efeito da variável exógena “ x ”, no caso, de intervenções

$$f(k, x, t) = \sum_{j=1}^k \frac{w_j(B)}{\sigma_j(B)} X_{tj}, \quad 2.98$$

$$f(k, x, t) = \sum_{j=1}^k V_j(B) X_{tj} + b_t, \quad 2.99$$

X_{tj} , $j=1, 2, \dots, k$ são k variáveis exógenas (intervenções), k é o conjunto de parâmetros desconhecidos que aparecem em $V_j(B)$ ou em $W_j(B)$, e $\delta_j(B)$.

A função (2.95) é uma função de transferência da j -ésima variável exógena, sendo $V_j(B)$, $W_j(B)$ e $\delta_j(B)$ polinômios em B , e b_t é ruído que poderá ser representado por um modelo ARIMA.

Em Análise de intervenção, supor-se-á que algumas das variáveis X_{tj} são variáveis binárias, fazendo o mesmo papel que variáveis “dummy” em regressão. As séries X_{tj} são chamadas indicadoras de intervenção. Como se sabe que intervenção

é a ocorrência de algum tipo de evento em dado instante de tempo, podendo manifestar-se por um intervalo de tempo subsequente, e que afeta temporariamente, ou permanentemente, a série temporal em estudo, a Análise de Intervenção tem por objetivo avaliar o impacto de tal evento no comportamento da série.

Usualmente, as séries indicadoras de intervenções podem ser representadas por três tipos de variáveis binárias.

1) Função impulso

$$X_{i,t} = I_{i,t}^{(T)} \begin{cases} 0, t \neq T \\ 1, t = T \end{cases} \quad 2.101$$

2) Função degrau ("step function")

$$X_{i,t} = S_{i,t}^{(T)} \begin{cases} 0, t < T \\ 1, t \geq T \end{cases} \quad 2.102$$

3) Função impulso sazonal

$$X_{i,t} = I_{i,t}^{(S)} \begin{cases} 0, t \neq S, 2S, 3S \\ 1, t = S, 2S, 3S \end{cases} \quad 2.103$$

No caso da função (2.101), o efeito da intervenção é temporário, no caso da função (2.102), o efeito da intervenção é permanente após o instante T, ao passo que a Função (2.103) o efeito é S e seus múltiplos.

Em geral, o efeito da Intervenção é mudar o nível da série, ou, então, a inclinação. Sabe-se que três fatores podem levar a falsas conclusões, ou seja, podem obscurecer o efeito da intervenção:

- (1) Tendência;
- (2) Sazonalidade;
- (3) Erro aleatório.

De fato, se a tendência existe, e uma Intervenção ocorre no instante T, o nível pós-intervenção é maior que o nível pré-Intervenção. Para isso utiliza-se os modelos ARIMA, pois os três elementos citados, acima, são levados em conta quando a componente residual b_t de (2.100) também for modelado por um ARIMA.

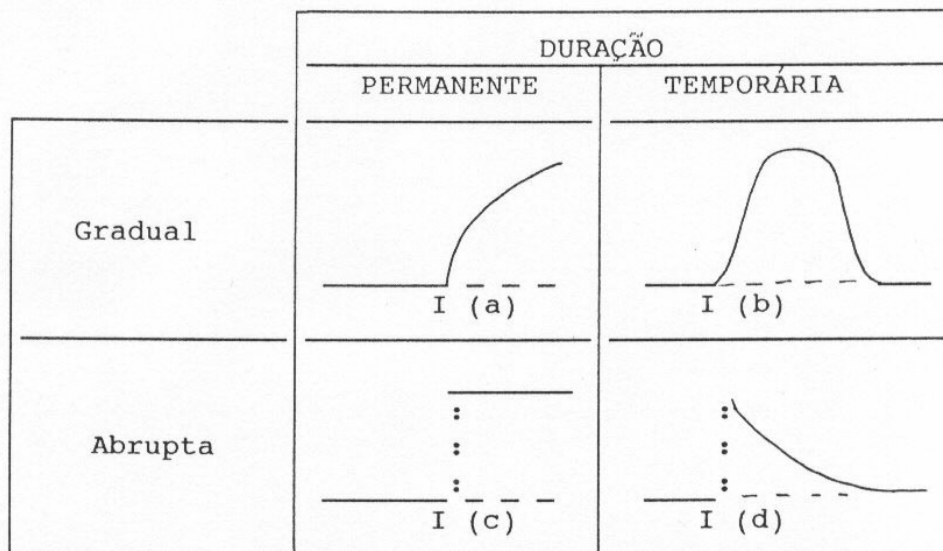
- Efeitos da Intervenção

Há muitas formas pela qual uma Intervenção pode afetar uma série temporal. As alterações mais comuns são as mudanças no nível da série e as mudanças na direção ou inclinação, da série. Estas últimas ocorrem quando a intervenção é complexa, e seu efeito manifesta-se gradativamente.

Além disso, a Intervenção pode alterar a variância dos erros e também introduzir no modelo componentes que antes não haviam. Por exemplo, introduzir uma componente auto-regressiva num processo de médias móveis.

A mudança pode ser abrupta (ou imediata), ou então, só ocorrer depois de algum tempo de iniciada a Intervenção (demorada ou defasada). Pode, ainda, ser temporária, ou permanente, podendo tornar a série mais estável, ou aumentar a variabilidade.

A série também pode ser afetada de várias maneiras, simultaneamente. O Figura 10 (adaptado de Glass et al., 1975, Mc Dowall et a.l, 1980) esquematiza os tipos mais comuns de efeitos de uma Intervenção sobre uma série de tempo.



Fonte: Manual de Econometria, 2000

Figura 10 – Efeitos de Intervenção

O efeito de evolução, mostrado na Figura 10, acontece quando a série decai inicialmente, como se extinguisse, para a seguir retomar o seu desenvolvimento até um novo nível. Este tipo de efeito aparece, por exemplo, quando se estuda a sobrevivência de uma espécie, após a mutação adaptada.

A mudança ocasionada na série temporal, pela ocorrência de uma Intervenção, pode ser:

(i) Quanto a manifestação

- abrupta;
- gradual.

(ii) Quanto a duração

- permanente;
- temporário.

Pode haver, também, mudança na variabilidade de série, após a intervenção, bem como um efeito de evolução pode aparecer. A série, decai inicialmente, e depois retoma o crescimento, até atingir um novo nível.

- Estrutura da função de transferência

O efeito de intervenção é determinado pela estrutura da Função de Transferência. Conhecendo-se a forma da Função de Transferência do modelo, e estimando-se seus parâmetros, conhece-se o tipo de efeito de Intervenção.

O conhecimento do problema pode sugerir o possível efeito da Intervenção, o que facilitará a identificação do modelo a ser usado. O conhecimento apriorístico, do problema, auxilia na identificação da função de transferência, enquanto os dados fornecem novas informações sobre efeitos desconhecidos a priori.

Caso de uma só Intervenção cujo modelo é dado por:

$$Z_t = V(B) X_t = \frac{W(B)}{\delta(B)} X_t. \quad 2.104$$

A seguir, descreve-se alguns dos casos mais comuns de Intervenção e suas respectivas funções de transferências, que são resumidas no Quadro 1, adequada por Pack, (1977) e Box & Jenkins, (1976), encontra-se considerações semelhantes para modelos de função de transferência.

i) Caso em que a função de transferência é:

$$V(B) = W_0 \quad 2.105$$

se ,

$$X_{i,t} = I_{i,t}^{(T)} \begin{cases} 0, t < T \\ 1, t \geq T, \end{cases} \quad 2.106$$

então:

$$r_t = \begin{cases} 0, t < T \\ W_0, t \geq T \end{cases} \quad 2.107$$

e tem uma mudança imediata e permanente no nível da série de uma quantidade W_0 , somente no instante T . É o caso (b) da Figura 10.

ii) Caso em que a função de transferência é

$$V(B) = W_0 / (1 - \delta_1(B)) \quad 2.108$$

Se,

$$X_t = \begin{cases} 0, t < T, \\ 1, t \geq T \end{cases} \quad 2.109$$

então,

$$r_t = \begin{cases} 0, t < T \\ W_0 \cdot \sum_{j=0}^k \delta_1^j, t = T + k, k = 0, 1, 2, \dots \end{cases} \quad 2.110$$

e obtém-se uma progressão geométrica em que

$$r_t \xrightarrow{t \rightarrow \infty} W_0 / (1 - \delta_1). \quad 2.111$$

portanto, uma mudança desacelerada de nível da série, até que atinja a assíntota $W_0 / (1 - \delta_1)$. É o caso (c) da Figura 10.

Se,

$$X_t = \begin{cases} 0, t < T \\ 1, t = T \end{cases}, \quad 2.112$$

então,

$$r_t = \begin{cases} 0, & t < T \\ \delta^k, W_0, & t = T + k, k = 0, 1, \dots \end{cases} \quad 2.113$$

Isso é a mudança em Z_t inicialmente é W_0 , depois exponencialmente, até desaparecer o efeito da intervenção, é o caso (d) da Figura 10.

iii) Caso em que a função de transferência é

$$V(B) = W_0 / (1 - (B)). \quad 2.114$$

Nesse caso, $\delta_1 = 1$ e, após a Intervenção, o modelo é não estacionário.

Se,

$$X_t = \begin{cases} 0, & t < T \\ 1, & t = T \end{cases}, \quad 2.115$$

então,

$$r_t = \begin{cases} 0, & t < T \\ (k+1) W_0, & t = T + k, k = 0, 1, \dots \end{cases}, \quad 2.116$$

e tem-se uma tendência determinística no modelo, introduzida pela Intervenção a partir de T . São os casos (e) do Quadro 1 e o casos (a) da Figura 10.

Se,

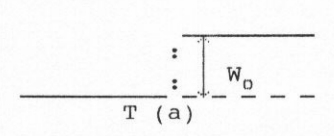
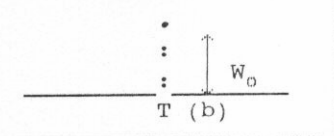
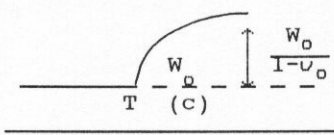
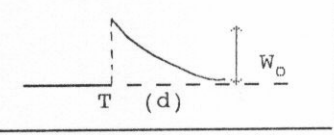
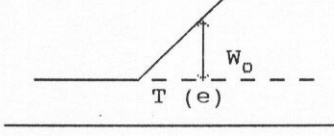
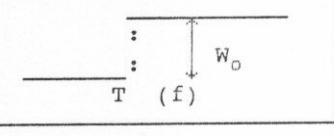
$$X_t = \begin{cases} 0, & t \neq T \\ 1, & t = T \end{cases}, \quad 2.117$$

então,

$$r_t = \begin{cases} 0, & t < T \\ W_0, & t \geq T \end{cases}, \quad 2.118$$

e tem-se, novamente, uma mudança fixa e permanente no nível da série de uma quantidade W_0 . São os casos (c) do Quadro 1 e (f) da Figura 10.

A seguir, são apresentados casos mais comuns dos efeitos da Intervenção e a estrutura da função de transferência, segundo Pack, 1977 e Box & Jenkins, 1976.

Função de Transf.	$X_t = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases}$	$X_t = \begin{cases} 0, & t \neq T \\ 1, & t = T \end{cases}$
W_0		
$\frac{W_0}{1 - \delta(B)}$		
$\frac{W_0}{1 - B}$		

Fonte: Manual de Econometria, 2000.

Figura 11 - Função de Transferência

Para garantir a estacionariedade é preciso que $|\delta| < 1$. Nesse caso,

$$r_t = \delta_1 r_{t-1} + W_0 X_t. \tag{2.119}$$

- Efeito de intervenção no modelo ARIMA

Seja um processo ARIMA (p, d, q), cujo nível inicial é μ :

$$Z_t = \mu + \sum_{j=0}^{t-1} \Psi_j a_{t-j}. \tag{2.120}$$

Suponha-se que ocorra uma Intervenção no instante T, e cujo efeito seja alterar o nível da série de uma quantidade δ a partir deste instante, então

$$Z_t = \begin{cases} \mu + \Psi(B) a_t, & t < T \\ \mu + \delta + \Psi(B) a_t, & t \geq T \end{cases}. \tag{2.121}$$

Para estimar e testar o efeito δ da intervenção, é necessário transformar o modelo ARIMA (p,d,q) num modelo linear, e então, aplicar os procedimentos usuais.

Seja o processo MA (1)

$$Z_t = \mu + a_t - \theta a_{t-1} \quad 2.122$$

que após a intervenção torna-se

$$Z_t = \mu + \delta + a_t - \theta a_{t-1} \quad 2.123$$

Transforme-se Z_t em uma série Y_t , que esteja na forma do modelo linear geral, isto é, na qual todos os erros, exceto a_t , tenham sido removidos, a primeira observação já estará na forma desejada:

$$Y_1 = Z_1 = \mu + a_1. \quad 2.124$$

A segunda observação também estaria, se não fosse o termo θa_1 , que deve ser removido.

$$\begin{aligned} Y_2 &= Z_2 + \theta Y_1 \\ &= \mu + a_2 - \theta a_1 + \theta \mu + \theta a_1 \\ &= (1+\theta) \mu + a_2. \end{aligned} \quad 2.125$$

De modo análogo, verifica-se que as outras observações podem ser transformadas através da relação recursiva.

$$\begin{aligned} Y_t &= Z_t + \theta Y_{t-1} \\ &= (1+\theta+\dots+\theta^{t-1}) \mu + a_t. \end{aligned} \quad 2.126$$

Após a intervenção, tem-se

$$\begin{aligned} Y_t &= Z_t + \theta Z_{t-1} \\ &= \mu + \delta + a_t - \theta a_{t-1} + \theta(1+\theta+\dots+\theta^{t-2}) \mu + \theta a_{t-1} \\ &= (1+\theta+\dots+\theta^{t-1}) \mu + a_t + \delta. \end{aligned} \quad 2.127$$

É fácil mostrar que, para $t \geq T$,

$$Y_t = (1 + \theta + \dots + \theta^{t-1}) \mu + (1 + \theta + \dots + \theta^{t-T}) a_t. \quad 2.128$$

Em forma matricial, pode-se escrever

$$\tilde{Y} = \tilde{x} \cdot \tilde{\beta} + \tilde{a}, \quad 2.129$$

onde

$$\tilde{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_{T-1} \\ \vdots \\ y_T \\ \vdots \\ y_N \end{bmatrix}, \tilde{x} = \begin{bmatrix} 1 & \dots & 0 \\ 1+\theta & \dots & 0 \\ \vdots & \dots & \vdots \\ 1+\dots+\theta^{T-2} & \dots & 0 \\ 1+\dots+\theta^{T-1} & \dots & 1 \\ \vdots & \dots & \vdots \\ 1+\dots+\theta^{N-1} & \dots & 1+\dots+\theta^{N-T} \end{bmatrix} \quad 2.130$$

$$\tilde{\beta} = \begin{bmatrix} \mu \\ \delta \end{bmatrix} \mathbf{e} \quad \tilde{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} \quad 2.131$$

os estimadores de mínimos quadrados μ e δ são dados por

$$\hat{\tilde{\beta}} = (\tilde{X}' \tilde{X})^{-1} \tilde{X}' \tilde{Y} \quad 2.132$$

as estimativas dos erros são dadas por

$$\hat{\tilde{a}} = \tilde{Y} - \tilde{x} \hat{\tilde{\beta}}, \quad 2.133$$

e as somas dos quadrados das estimativas dos erros, por

$$SQ(\theta, Z) = \hat{\tilde{a}}' \hat{\tilde{a}}. \quad 2.134$$

Na prática, calcula-se SQ para diversos valores de θ , tomando-se, como estimativa, $\hat{\theta}$, cujo valor minimiza SQ, o que equivale a minimizar a estimativa de variância residual do erro S_a^2

$$\frac{\mu - \hat{\mu}}{S_a \sqrt{C_{11}}} \sim t_{N-2}, \quad 2.135$$

e

$$\frac{\delta - \hat{\delta}}{S_a \sqrt{C_{22}}} \sim t_{N-2}. \quad 2.136$$

onde

$$S_a = a' \hat{\alpha} / (N-2). \quad 2.137$$

C_{jj} é o j-ésimo elemento da diagonal principal de $(X' X)^{-1}$, e t_{N-2} indica a distribuição t de Student com $N - 2$ graus de liberdade.

O intervalo de confiança, ao nível de α de significância, para efeito da intervenção, pode ser construído, obtendo-se

$$\hat{\delta} \pm t_{N-2} \left(1 - \frac{\alpha}{2}\right) S_a \sqrt{C_{22}}. \quad 2.138$$

Seja o processo AR (1)

$$Z_t - \mu = (Z_{t-1} - \mu) + a_t \quad 2.139$$

a transformação num modelo linear é feita por

$$\begin{aligned} Y_1 &= Z_1 = \mu + a_t \\ Y_2 &= Z_2 - \varphi Z_1 \\ &= (1-\varphi) \mu + \varphi Z_1 = a_t - \varphi Z_1 \\ &= (1-\varphi) \mu = a_t \end{aligned} \quad 2.140$$

e, de modo geral,

$$Y_t = Z_t - \varphi Z_{t-1} = (1-\varphi) \mu + a_t. \quad 2.141$$

Após a intervenção, o modelo torna-se

$$Z_t - (\mu + \delta) = \varphi [Z_{t-1} - (\mu + \delta)] + a_t \quad 2.142$$

ou

$$Z_t = [(1-\varphi) (\mu + \delta)] + \varphi Z_{t-1} + a_t. \quad 2.143$$

A transformação fica sendo

$$Y_T = (1-\varphi) \mu + \delta + a_T \quad 2.144$$

ou

$$Y_t = (1-\varphi) (\mu + \delta) + a_t, \quad t > T; \quad 2.145$$

$$\tilde{X} = \begin{bmatrix} \frac{1}{1-\varphi} & \dots & 0 \\ \dots & \dots & \dots \\ 1-\varphi & \dots & 0 \\ 1-\varphi & \dots & 1 \\ 1-\varphi & \dots & \frac{1}{1-\varphi} \\ \dots & \dots & \dots \\ 1-\varphi & \dots & 1-\varphi \end{bmatrix} \cdot \quad 2.146$$

Nesse capítulo, foi visto toda a teoria necessária para desenvolvimento da pesquisa, e, no próximo serão mostrados as etapas necessárias para a modelagem e previsão das séries em estudo.

3 METODOLOGIA

Neste item serão apresentadas as etapas necessárias para a modelagem das variáveis em estudo, assim como a previsão de seus valores.

3.1 Os Dados

Para a execução desta pesquisa foram consultados os relatórios elaborados no setor de estatística do HUSM, com a variável taxa de ocupação de leitos no pronto atendimento e hospital geral. Após a coleta dessas variáveis, foram analisadas as pressuposições básicas para a aplicação da análise de séries temporais.

Para a aplicação da metodologia de Box & Jenkins, conhecidos como Modelos ARIMA (modelos autoregressivos de médias móveis) e análise de Intervenção, utilizou-se uma amostra formada por 61 observações do ano de 2000 a 2004, do HUSM. O evento de interesse da pesquisa foi a taxa de ocupação dos leitos do PA e HG que são duas formas de ingresso no hospital a apresentarem maior demanda.

3.2 Análise Descritiva

As informações foram transpostas para uma tabela simples, com a utilização do *Excel*, totalizando 61 observações.

- Realizou-se a análise descritiva dos dados, por ano pois a mesma mostra o comportamento de uma situação, ou problema. Para tal, utilizaram-se as medidas descritivas: média, desvio-padrão e coeficiente de variação. A cada ano, investigando desta maneira, se houve alguma modificação do período analisado;

- Traçou-se o gráfico da série em estudo, de forma a perceber algumas características especiais como tendência, sazonalidade, desvios abruptos na série, assim como investigar a sua estacionariedade;

3.3 Modelagem das séries por meio da Metodologia Box & Jenkins

Nesta etapa, o fluxograma descrito na Figura 12 foi seguido, pois representa a metodologia a ser empregada neste estudo. Procedeu-se a estimação dos possíveis modelos concorrentes para representar a série. As FAC e FACP foram calculadas e grafadas, a fim de se investigar a possível categoria do modelo, e também a estacionariedade da série, sendo a série, não-estacionaria, transformações matemáticas, como descritas no item 2.3.2, devem ser realizadas de modo à estacionarizar a série para se estimar o modelo apropriado.

Encontraram-se vários modelos que poderiam descrever o fenômeno em estudo, os quais, investigando os seus resíduos, apresentaram ruído branco. Para optar pelo modelo mais adequado, foram utilizados os critérios do AIC e BIC, onde o menor valor do AIC e/ou BIC sinalizou para o modelo mais propício para as previsões desejadas. Também levou-se em consideração o modelo mais parcimonioso, facilitando, assim, a escolha do modelo mais adequado para as séries em estudo e tornar possível as previsões necessárias.

- Se a série em estudo apresenta alguma característica conforme o item 2 que se possa realizar uma análise de intervenção para melhorar as estimativas do modelo, esta análise será empregada, pois ela apresenta melhores resultados de previsões, explicando melhor os dados;
- Ao se realizar a análise de intervenção, procurar-se-á investigar as causas da ocorrência de um valor discrepante na série, ou as causas que levaram a série a mudar de nível;
- Encontrado, o melhor modelo, entre os modelos concorrentes, este será utilizado para realizar previsões, de forma a auxiliar no melhor desempenho das medidas gerenciais do hospital.

Dessa forma, pretende-se, ao desenvolver este estudo, mostrar a aplicabilidade dos modelos ARIMA na previsão de demanda dos leitos hospitalares. Pretende-se também, mostrar um roteiro de como essas séries foram modeladas.

4 RESULTADOS E DISCUSSÕES

Neste capítulo será aplicada a metodologia apresentada no capítulo 3, onde foram evidenciados os conceitos e revisões realizados no capítulo 2. Os resultados serão discutidos, oportunamente, quando as estatísticas e gráficos tenham sido mostrados.

4.1 Hospital Universitário de Santa Maria – HUSM

O Hospital Universitário de Santa Maria serve como base de atendimento primário dos bairros que o cercam; para o atendimento secundário à população no município sede e para o atendimento terciário da região centro e fronteira gaúcha. Tem sido referenciado até fora do Estado pela alta complexidade no tratamento de oncologia, incluindo transplantes de medula óssea. O hospital se constitui em centro de ensino e pesquisa no âmbito das ciências da saúde, centro de programação e manutenção de ações voltadas à saúde das comunidades local e regional, desenvolve programas específicos à comunidade devidamente integrada à rede regional de saúde. Também presta serviços assistenciais em todas as especialidades médicas, e serve de treinamento para alunos de graduação e pós-graduação em Medicina, Residência Médica, e de graduação em Farmácia, Fonoaudiologia, Fisioterapia e Enfermagem. (www.ufsm.br)

4.2 Variáveis e Período Analisados

Para a aplicação da metodologia de Box & Jenkins, conhecidos como Modelos ARIMA (modelos autoregressivos de médias móveis), e análise de intervenção, utilizou-se uma amostra mensal de janeiro de 2000 a dezembro de 2004 com coleta mensal, para o PA e para o HG do HUSM. A variável de interesse da pesquisa foi a taxa de ocupação dos leitos.

A taxa de ocupação hospitalar é a relação do percentual entre o número de pacientes-dia e o número de leitos-dia, num determinado período, multiplicado por cem. O que significa que se esse percentual estiver abaixo de cem por cento, o hospital ainda terá condições de receber pacientes, e se esse percentual estiver acima de cem por cento o hospital já estará com a sua capacidade esgotada e, a partir daí, revela uma superlotação do sistema. Por isso, o estudo dessa variável é muito importante.

4.3 Análise da Série de Taxa de Ocupação do Pronto Atendimento – PA-HUSM

A análise descritiva é sempre reveladora, pois é capaz de mostrar o comportamento de uma situação ou problema. Neste caso utiliza-se medidas descritivas, como média desvio-padrão e coeficiente de variação das variáveis em estudo. Para tal, essas medidas foram realizadas ano a ano, investigando, dessa maneira, se houve alguma modificação do período analisado. Na Tabelas 2, as estatísticas para PA, que são auto-explicativas.

Tabela 2 – Medidas descritivas das taxas de ocupação percentual do PA, por ano.

Estatísticas	2000	2001	2002	2003	2004
Média	212,07	194,87	222,12	129,15	119,33
Desvio padrão	22,17	21,33	54,76	20,21	27,35
Coeficiente de Variação	10,45	10,94	24,65	15,64	23,00

Taxa de ocupação= ((numero de pacientes dia)/(numero de leitos dia))*100

O coeficiente de variação de Pearson, para os nos de 2000 a 2004, é muito significativo, pois é bem abaixo de 50 %, o que revela ser a média da taxa de ocupação dos leitos do pronto atendimento muito significativa. Onde a média da taxa de ocupação oscila entre os anos analisados. Analisando as medidas descritivas percebe-se que o ano de 2000 a 2002 é média é bem superior a de 2003 a 2004, o que sugere uma investigação para tal atipicidade.

A Figura 12 mostra o gráfico da taxa de ocupação do PA do HUSM, e a Figura 11 os gráficos da FAC e FACP, respectivamente, onde os dados foram coletados, mensalmente, no período de 2000 a 2004.

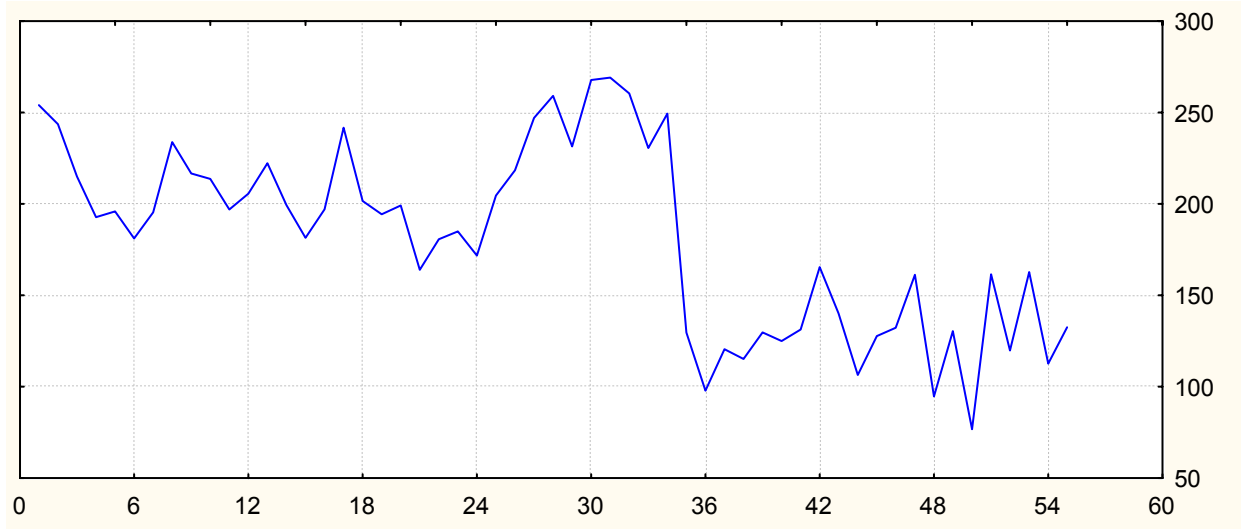


Figura 12 – Taxa de ocupação do PA

Analisando-se a Figura 12, verifica-se que a série não apresenta um comportamento estacionário, logo é necessário que se aplique diferenças à série em estudo para torná-la estacionária, desta forma, iniciando a estimação do modelo de acordo com a metodologia de Box & Jenkins.

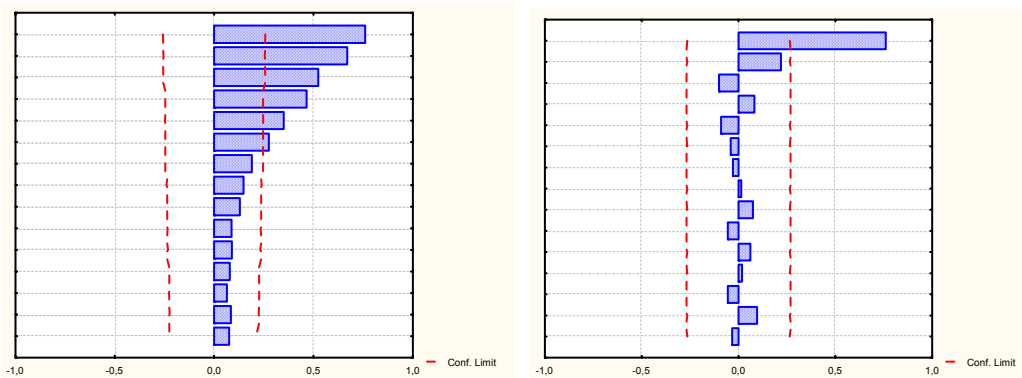


Figura 13 – Função de Autocorrelação e Autocorrelação Parcial, respectivamente da série do PA

Observando-se a Figura 13, verifica-se que a FAC apresenta um decaimento, rápido, para zero, o que é uma indicação de um modelo autoregressivo, e a FACP, neste caso, mostra a ordem do modelo, neste caso estaríamos trabalhando com um modelo AR (1), mas, pelas evidências mostradas na Figura 13, será feita uma diferença na série original, de forma que se estará, então, à procura de uma série integrada.

Na Figura 14, apresenta-se a série original e a série diferenciada.

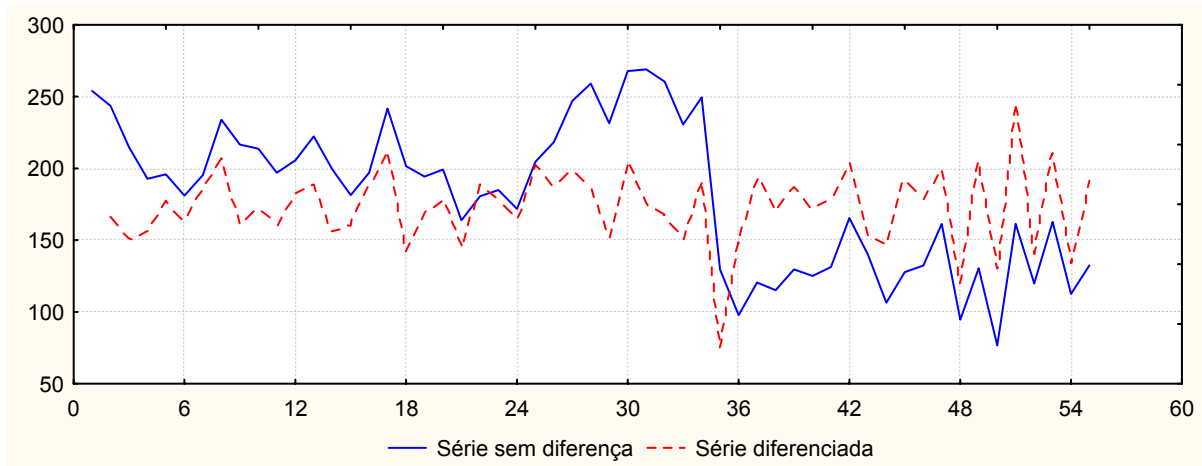


Figura 14 – Taxa de Ocupação PA – HUSM diferenciada

Observa-se, pela Figura 14, que a série diferenciada apresenta-se mais estável que a série original, logo ela deve ser utilizada para estimar-se o modelo que represente a série em estudo. Também observa-se que, no instante 36, há uma queda brusca na série original, sugerindo, então, uma intervenção nesse instante.

A Figura 15 mostra, respectivamente, FAC e FACP, da série diferenciada

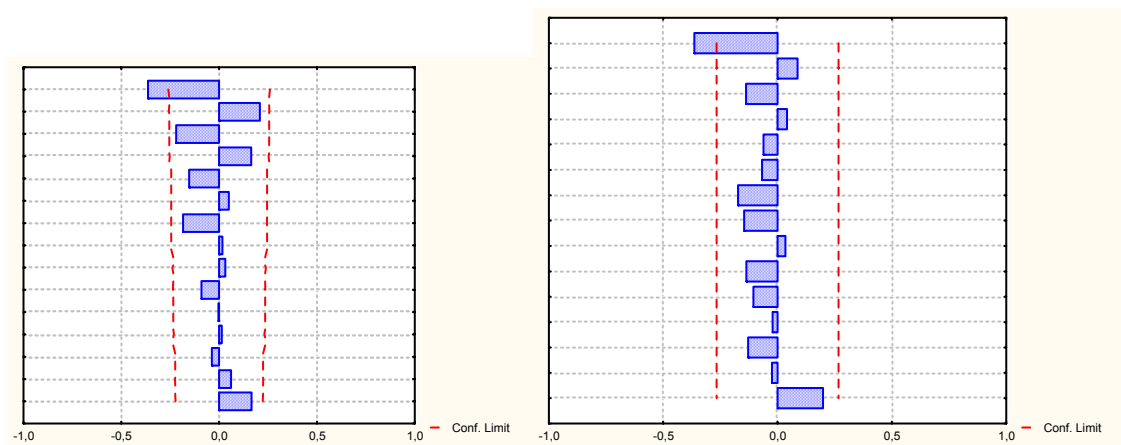


Figura 15 – Função de autocorrelação e autocorrelação parcial, respectivamente da série do PA diferenciada

Na Figura 15, se observa a FAC e FACP da série diferenciada a qual está dentro dos limites de confiabilidade.

Observa-se, na Tabela 3, os modelos encontrados para a taxa de ocupação do PA, onde o modelo mais adequado é um ARIMA (1,1,1), com uma intervenção abrupta permanente no lag 36.

Tabela 3 – Modelos propostos para o Pronto Atendimento

Modelo	Parâmetros	t(calculado)	Nível de Signif.	AIC	BIC	Intervenção
ARIMA(1,1,1)	$\theta=0,718103$ $\phi=0,880616$	3,719448 6.883.397	0,0005 0,0000	640,2	644,3	s/interv.
ARIMA(1,1,1)	$\theta=0,5132$ $\phi=0,8943$ $w=-74,1099$	267,868 949.533 -295.704	0,0009 0,0000 0,0005	578,3	587,3	Interv/Abrup/perm lag 36

Os modelos encontrados para representar a série em estudo serão descritos a seguir, mas pode-se antever pelas estatísticas de penalidade AIC e BIC que o modelo com intervenção apresentou melhores resultados. Estes resultados são corroborados pela análise dos resíduos descritos na seqüência.

- Modelo sem Intervenção

$$\Delta Z_t = 0,5132\Delta Z_{t-1} + a_t - 0,8943\varepsilon_{t-1}$$

$$\Delta Z_t (1 - 0,5132 B^1) = a_t + 0,8943\varepsilon_{t-1}$$

$$\Delta Z_t = a_t + \frac{0,8943\varepsilon_{t-1}}{(1 - 0,5132 B^1)}$$

- Modelo com Intervenção

Neste modelo com intervenção considera-se que até o instante 36 não tenha ocorrido nenhum evento externo e que a partir deste instante houve este evento que foi a implementação de quarenta leitos no hospital que foram disponibilizados para a população, logo antes do instante 36 atribui-se o valor zero e a partir dele considera-se valores iguais a 1.

$$\Delta Z_t = \delta(B)a_t + \frac{0,8943\varepsilon_{t-1}}{(1 - 0,5132 B^1)}$$

$$\Delta Z_t = -\frac{74,1099}{1 - B^{36}} + a_t + \frac{0,8943\varepsilon_{t-1}}{(1 - 0,5132 B^1)}$$

- Análise dos Resíduos do Modelo com Intervenção

Observando-se na Figura 16, os resíduos do modelo ARIMA (1,1,1), com uma intervenção encontrada, seguem uma distribuição normal, e na Figura 17 constata-se que autocorrelação e autocorrelação parcial estão entre os limites de confiabilidade, mostando assim que são ruído branco.

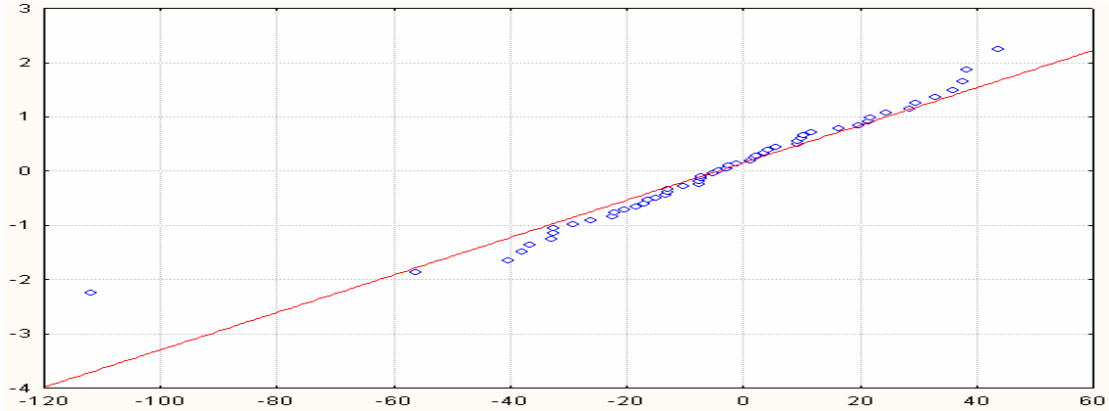


Figura 16 – Gráfico da Distribuição Normal dos resíduos do Modelo ARIMA (1, 1, 1) com intervenção

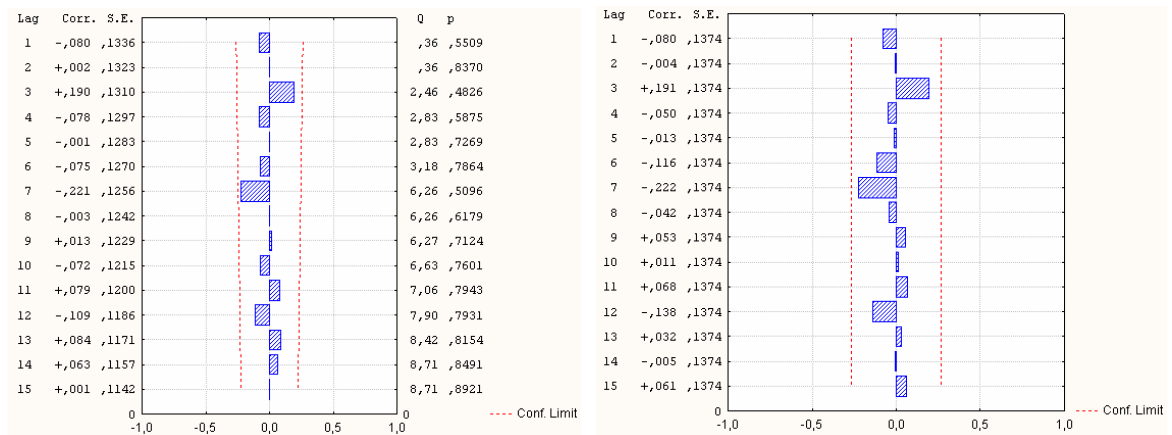


Figura 17 – FAC e FACP dos resíduos do modelo ARIMA (1, 1, 1) com intervenção

Para o ajuste com intervenção, observe-se novamente, as Figuras 20 e 21, notando-se que há um comportamento atípico na observação 36, podendo ser considerada uma possível intervenção. Assim, o modelo ARIMA (1,1,1), com uma intervenção w_1 , conforme Figura 10, no item 2, corresponde à observação 36.

Na Figura 18 apresenta-se, o gráfico da série diferenciada com a intervenção.

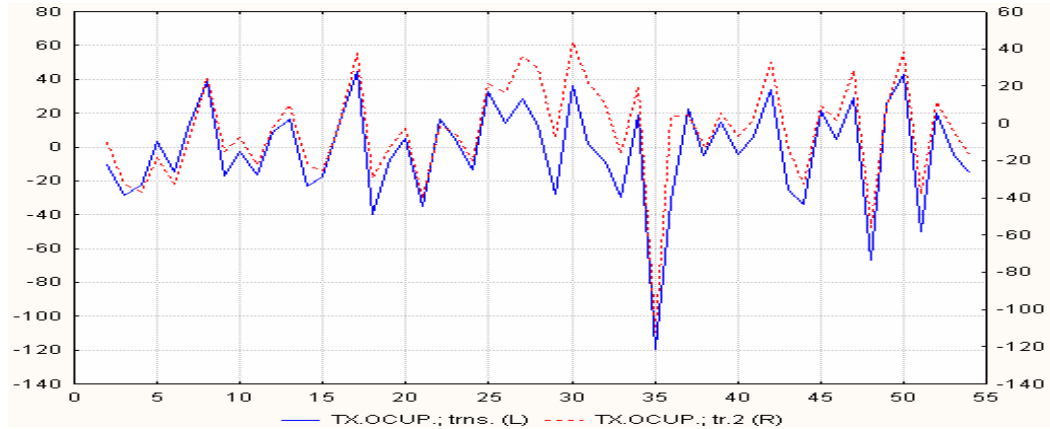


Figura 18 – Taxa de Ocupação PA – Intervenção

Na Tabela 4, observa-se os valores estimados para os coeficientes e na Tabela 4 observa-se as previsões para Pronto Atendimento do Hospital Universitário de Santa Maria, para os meses de janeiro a dezembro de 2005.

Tabela 4 – Previsões para PA – HUSM

Meses	Previsões	Inferior 95%	Superior 95%	Erro.
Janeiro	125,1	-11,6	260,8	81,2
Fevereiro	125,2	-16,4	265,6	84,2
Março	125,6	-21,5	270,4	87,0
Abril	124,6	-25,7	274,9	89,7
Maio	124,6	-30,2	279,4	92,4
Junho	125,2	-34,5	283,8	94,9
Julho	125,2	-38,7	287,9	97,5
Agosto	125,2	-42,8	292,0	99,9
Setembro	125,2	-46,8	296,0	102,3
Outubro	125,2	-50,7	299,9	104,6
Novembro	125,2	-54,5	303,8	106,9
Dezembro	125,2	-58,3	307,5	109,2

Na Figura 19 observa-se o gráfico das previsões do PA HUSM, para os meses de janeiro à dezembro de 2006, como representado pela Tabela 4.

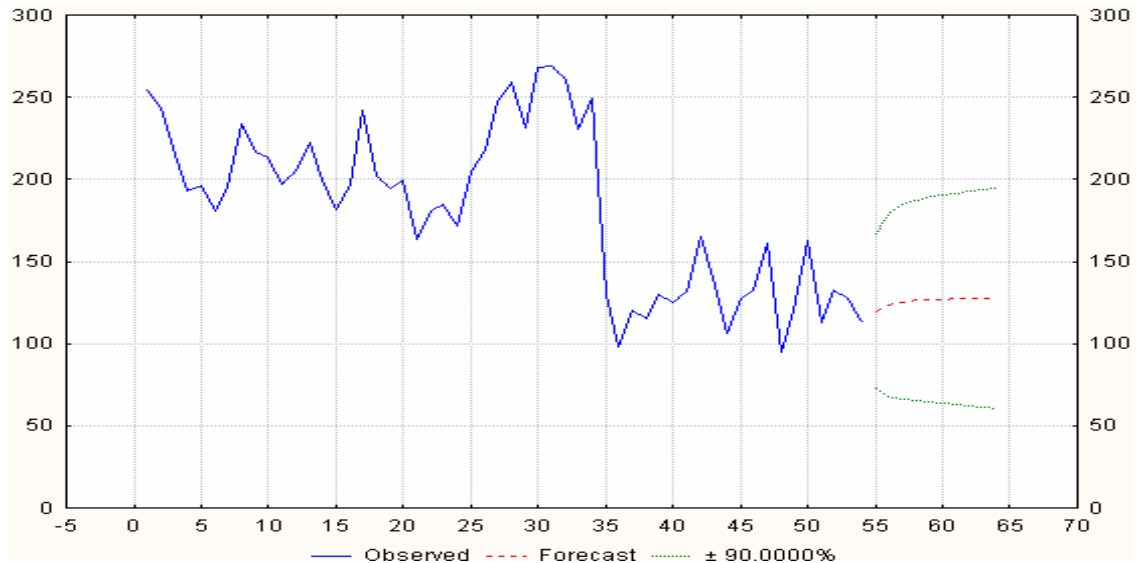


Figura 19 – Previsão da Demanda PA – HUSM

- Síntese do Item

Os modelos de séries temporais são úteis para descreverem a demanda dos leitos do PA do HUSM. Dessa forma, conseguiu-se estimar um modelo de Séries Temporais, entre os modelos ajustados. Observou-se que o modelo autoregressivo de ordem um, com uma diferença e uma intervenção no *lag* 36, ARIMA (1, 1, 1), foi o que melhor explicou a demanda dos leitos do PA HUSM, podendo assim fazer as previsões de demanda dos leitos ilustrada na Tabela 4, onde a intervenção realizada, no *lag* 36, reflete o ano de 2002, quando foi inaugurado o Pronto-Socorro Regional, aumentando sua capacidade para quarenta leitos, preenchendo, dessa forma, importante lacuna na assistência terciária, no ensino e educação permanente dos profissionais da rede do SUS, além de oportunizar linhas de pesquisa, o que explica a atipicidade encontrada na série taxa de ocupação PA – HUSM.

As previsões encontradas, para os meses vindouros, poderão ser úteis para a administração do hospital, pois as estimativas foram executadas por meio de um ferramental criterioso, os quais possibilitarão a organização de estoques de matérias de consumo, de medicamentos, mão-de-obra especializada entre outras rotinas que são envolvidas dentro de um hospital.

4.4 Análise da Série de Taxa de Ocupação do Hospital Geral – HG-HUSM

A análise descritiva é sempre reveladora, pois é capaz de mostrar o comportamento de uma situação, ou problema. Nesse caso utilizaram-se medidas descritivas, como média desvio-padrão e coeficiente de variação das variáveis em estudo. Para tal, essas medidas foram realizadas dentro de cada ano, investigando, dessa maneira, se houve alguma modificação do período analisado. A Tabelas 4 apresenta as estatísticas para o HG.

Tabela 5 – Medidas descritivas das taxas de ocupação percentual do HG, por ano

Estatísticas	2000	2001	2002	2003	2004
Média	82,00	74,11	80,00	85,29	91,06
Desvio padrão	17,15	20,40	5,76	14,00	13,00
Coeficiente de Variação	20,91	27,52	7,2	16,41	14,27

O coeficiente de variação de Pearson, para os anos entre 2000 e 2004 é muito significativo, pois é bem abaixo de 50 %, revelando que a média da taxa de ocupação dos leitos do pronto atendimento é muito significativa. Observa-se também que a média do HG oscila em torno de oitenta, mas em 2001 teve uma queda para 74 e a partir daí a taxa de ocupação do HG permaneceu crescente, mostrando que este setor estará próximo de sua capacidade total no próximo ano.

A Figura 20 mostra o gráfico da taxa de ocupação do HG do HUSM, e a Figura 21 os gráficos da FAC e FACP, respectivamente, onde os dados foram coletados, mensalmente, no período de 2000 a 2004.

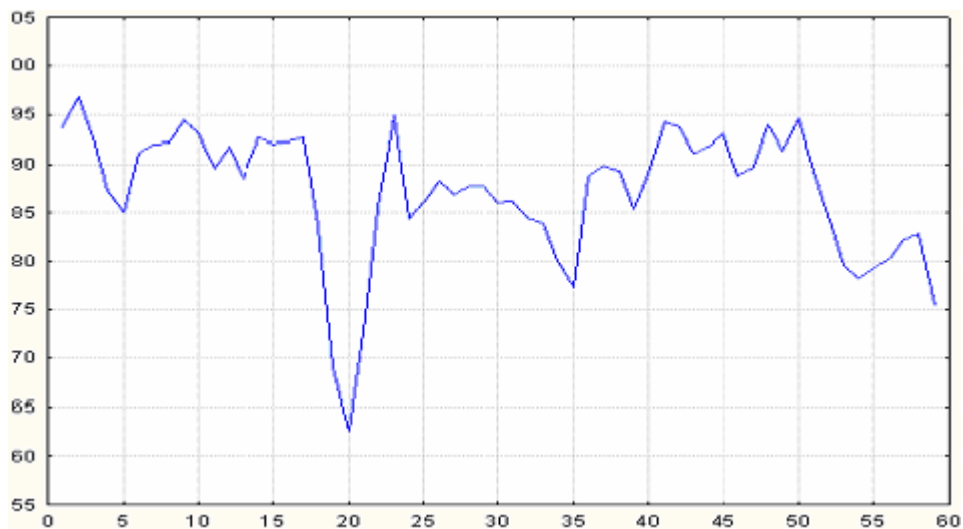


Figura 20 – Taxa de Ocupação do Hospital Geral

Analisando-se a Figura 20, verifica-se que a série não apresenta um comportamento estacionário, logo é necessário que se aplique diferença a série em estudo, para torná-la estacionária. Desta forma, iniciando a estimação do modelo, de acordo com a metodologia de Box & Jenkins.

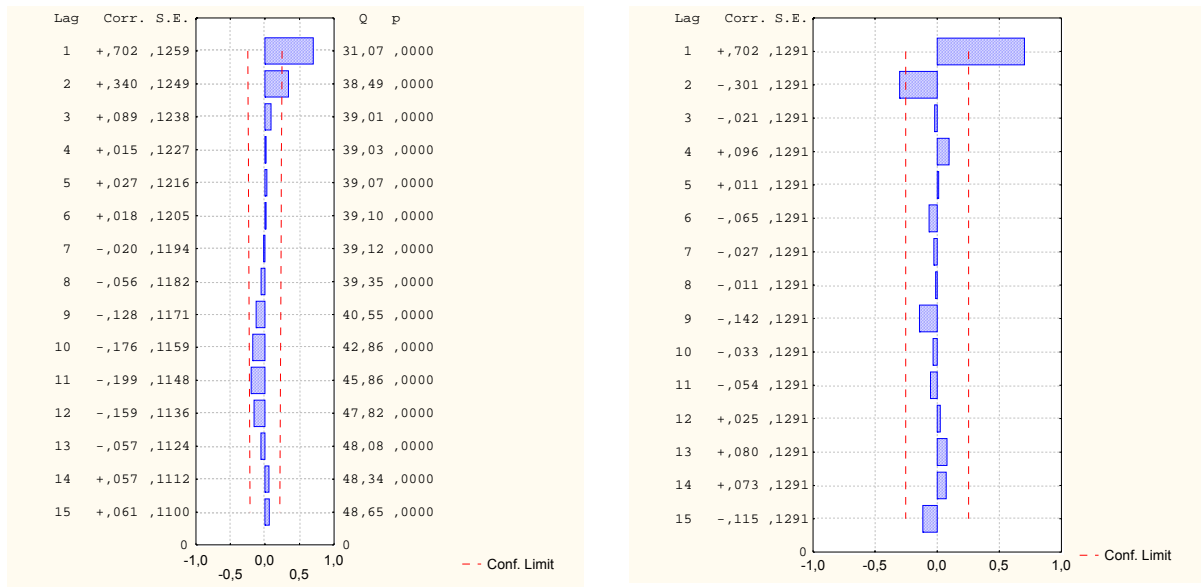


Figura 21 – Função de Autocorrelação e Autocorrelação Parcial, respectivamente

Observando-se a Figura 21, verifica-se que a FAC e a FACP sugerem um modelo de médias móveis, pois a FAC ao longo apresenta um comportamento próximo de uma senoide amortecida e a FACP sinaliza que pode haver dois parâmetros significativos.

Na Figura 22 apresenta-se a série original e a série diferenciada.

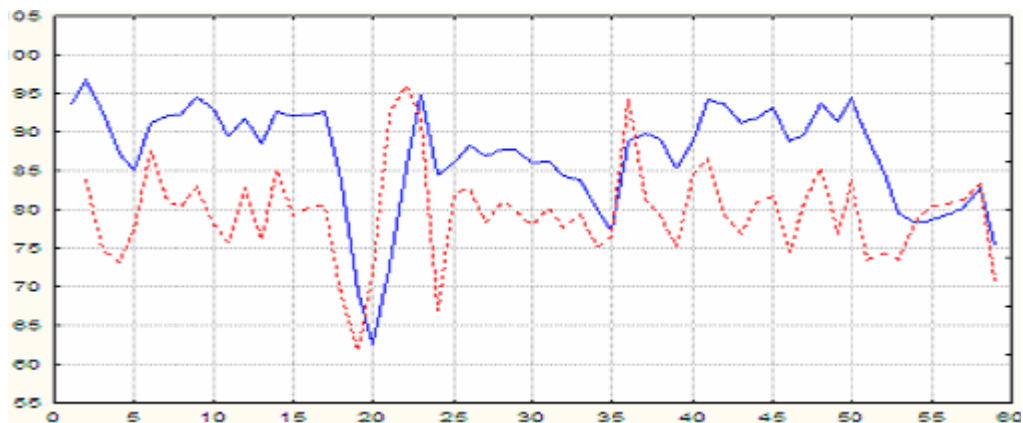


Figura 22 – Taxa de Ocupação HG – HUSM diferenciada

Observa-se, pela Figura 22, que a série diferenciada apresenta-se mais estável que a série original, logo deve ser utilizada para se estimar o modelo que represente a série em estudo. Observa-se, na Tabela 6 os modelos encontrados para a taxa de ocupação do HG, onde o modelo mais adequado é um ARIMA (1, 1, 2).

Tabela 6 – Modelos concorrentes para o pronto atendimento do HG

Modelo	Parâmetros	t(calculado)	Nível de Signif.	AIC	BIC
ARIMA(2,0,0)	$\Phi(1) = 1,05434$ $\Phi(2) = -0,0671$	0,133719 0,134764	0,000 0,654	539,64	543,78
ARIMA(1,1,2)	$\Phi(1) = 0,586974$ $\theta(1) = 0,489678$ $\theta(2) = -0,394723$	3,343278 2,951212 3,240990	0,001481 0,004616 0,002008	425,21	431,39

Escrevendo o modelo encontrado temos:

$$\Delta Z_t = \phi_1 \Delta Z_t + a_t + \theta_1 \Delta \varepsilon_{t-1} + \theta_2 \Delta \varepsilon_{t-2}$$

$$\Delta Z_t = \underset{(3,343278)}{0,586974} \Delta Z_t + a_t + \underset{(2,951212)}{0,489678} \Delta \varepsilon_{t-1} - \underset{(3,240990)}{0,394723} \Delta \varepsilon_{t-2}$$

onde os valores entre parênteses corresponde aos valores da estatística t de Student, o qual permite verificar que todos os parâmetros são significativos, também vale ressaltar que este modelo foi o que apresentou as melhores estatísticas AIC e BIC.

- Análise dos Resíduos

Observando-se, na Figura 23, a probabilidade dos resíduos encontrados através do modelo ARIMA (1, 1, 2), segue uma distribuição normal, e na Figura 24, constata-se que autocorrelação e autocorrelação parcial estão entre os limites de confiabilidade.

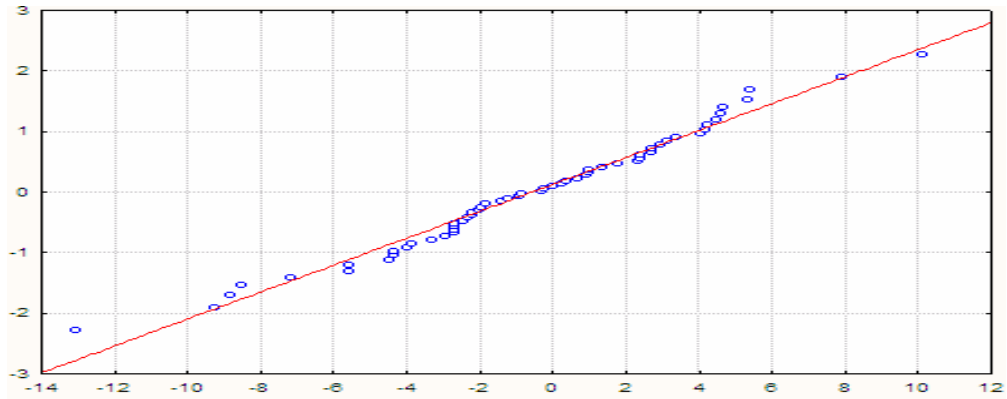


Figura 23 – Gráfico da Distribuição Normal do modelo ARIMA (1, 1, 2) do HG – HUSM

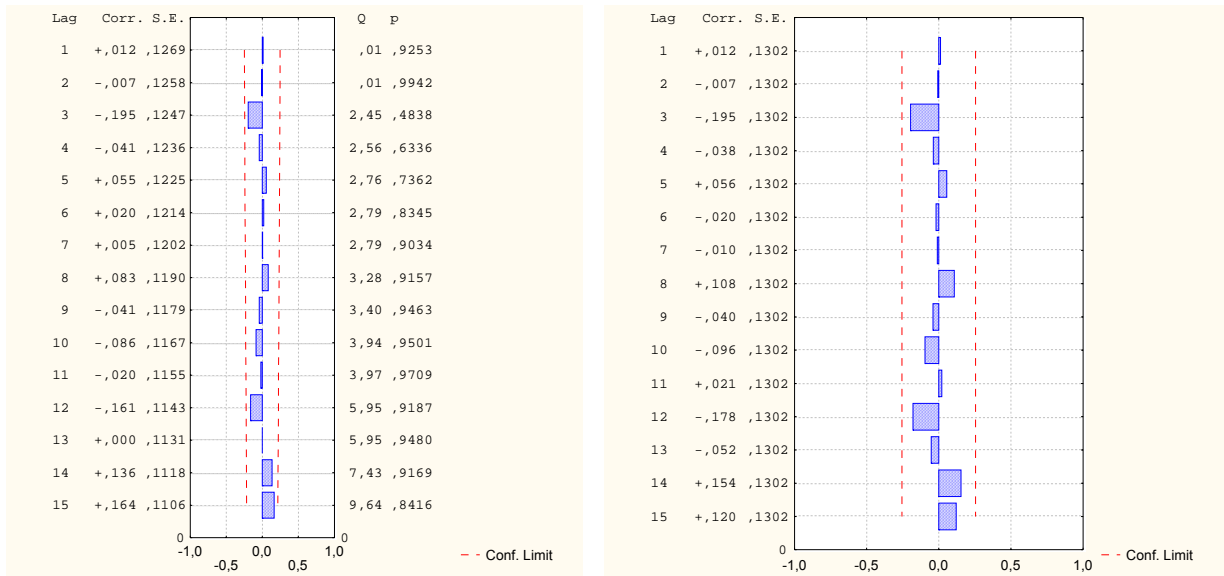


Figura 24 – FAC e FACP dos resíduos, respectivamente, do modelo ARIMA (1, 1, 2) HG – HUSM

Na Tabela 6 observa-se as previsões para HG do HUSM, para os meses vindouros.

Tabela 6 – Previsões para HG – HUSM, de janeiro a dezembro de 2005

Meses	Previsões	Inferior 95%	Superior 95%	Erro Padrão
Janeiro	75,37	67,84	82,90	4,50
Fevereiro	78,56	67,38	89,73	6,68
Março	80,43	67,87	92,98	7,50
Abril	81,53	68,28	94,78	7,92
Mai	82,17	68,51	95,84	8,17
Junho	82,55	68,60	96,51	8,34
Julho	82,77	68,59	96,96	8,48
Agosto	82,90	68,52	97,28	8,59
Setembro	82,98	68,43	97,54	8,70
Outubro	83,03	68,31	97,75	8,80
Novembro	83,05	68,17	97,93	8,89
Dezembro	83,07	68,04	98,10	8,98

Na Figura 25, observa-se o gráfico das previsões do Hospital Geral do Hospital Universitário de Santa Maria – HUSM, para os meses de janeiro à dezembro de 2006, como acima representado pela Tabela 6.

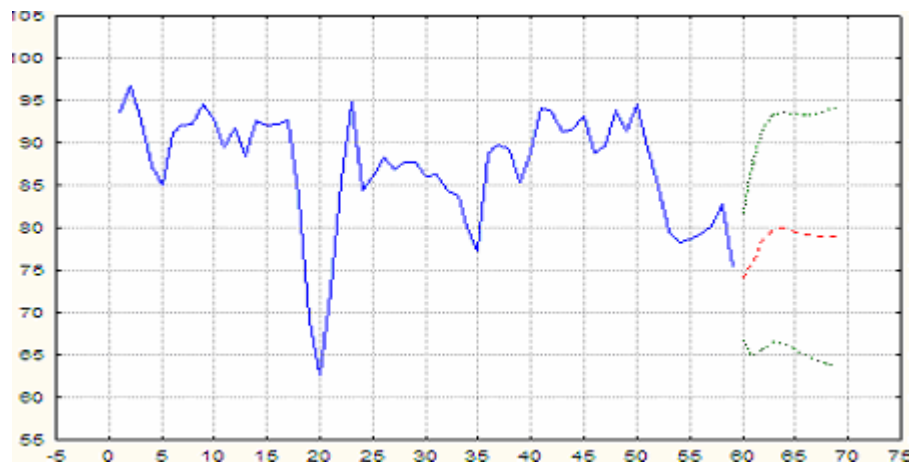


Figura 25 – Previsão da Demanda HG – HUSM

- Síntese do item

Os modelos de séries temporais são úteis para descrever a demanda dos leitos do HG do HUSM. Dessa forma, conseguiu-se estimar um modelo de Séries Temporais, entre os modelos ajustados. Observou-se que o modelo autoregressivo de ordem um com uma diferença, ARIMA (1,1,2), foi o que melhor explicou a demanda dos leitos do HG – HUSM, podendo-se, assim, fazer as previsões de demanda dos leitos ilustrada na Tabela 6.

As previsões encontradas, para os meses vindouros podem ser úteis para a administração do hospital, pois as estimativas foram executadas por meio de um ferramental criterioso, as quais possibilitarão a organização de estoques de matérias

de consumo, de medicamentos, mão-de-obra especializada entre outras rotinas que são envolvidas dentro de um hospital.

5 CONCLUSÃO

O HUSM localiza-se na região central do Rio Grande do Sul, e falar do HUSM é falar sobre a saúde pública do Estado e da região Centro Oeste do estado. “Dividimos com os gestores regionais, os problemas, as dificuldades e também as soluções para a implantação hierárquica do SUS” (HUSM, 2002). Logo, as necessidades do Hospital passam por diversos problemas, desde o alto fluxo de pessoas, que recorrem a este hospital devido ao que foi exposto anteriormente até, a falta de pesquisas que auxiliem na programação, na tomada de decisões para um melhor atendimento dos pacientes e desempenho de atividades de pesquisa.

Ao desenvolver um estudo sobre a previsão da taxa de ocupação dos leitos, desse hospital, procurou-se fornecer um componente a mais, para tentar dirimir as necessidades que o hospital enfrenta, pois no momento em que o conhecimento do comportamento destas variáveis que foram estudadas, puder ser antecipado, isto é, previsto, medidas gerenciais poderão ser tomadas.

Observando-se os períodos de 2000 a 2004 do PA e HG do HUSM, que são duas formas de ingresso muito importantes no hospital, devido a sua procura, onde a taxa de ocupação dos leitos disponíveis foi o principal variável em análise.

A metodologia empregada, para a análise, foi a de Box & Jenkins, onde a classe geral de modelo ARIMA (p,d,q) foi utilizada. Também foi possível empregar um recurso adicional para se obter melhores resultados nas previsões, que foi a análise de intervenção.

Para o setor do PA foi encontrado um modelo ARIMA (1,1,1), com uma intervenção no período 36, onde foi possível identificar que a série era não estacionária, e que o período de intervenção foi caracterizado pelo aumento da disponibilidade de quarenta leitos, onde a taxa de ocupação teve uma queda. Mas, ao longo do tempo, esses quarenta leitos passaram a não mais ser suficientes, pois a taxa de ocupação voltou a seu patamar inicial do período de análise. Logo a

análise de intervenção foi uma técnica capaz de captar o efeito ocorrido naquele período.

No setor do HG, onde se encontra a globalidade dos pacientes internados no HUSM, foi possível descrever a taxa de ocupação dos leitos por meio de um modelo ARIMA (1,1,2) que, embora seja um modelo com muitos parâmetros, foi o que melhor descreveu a série. Embora a série representativa do PA apresente-se mais instável que a do HG, o que é de se esperar, pois o PA apresenta maior fluxo de pacientes, e que muitas vezes não pode ser negado o ingresso desses pacientes, devido à gravidade da enfermidade, o que leva a uma alta taxa de ocupação, no HG é diferente, pois há a possibilidade de não aceitar o paciente devido a lotação do setor, onde muitas vezes o paciente pode esperar para se internado ou muitas vezes transferido para outro hospital, o que raramente acontece.

Espera-se que as previsões encontradas auxiliem, a direção do hospital, no dimensionamento da capacidade desses dois setores, assim como sirva para gestionar, junto aos órgãos competentes mais recursos para a ampliação, tanto dos recursos humanos quanto dos recursos físicos como materiais, salas e leitos. Sabe-se que apenas a modelagem, ou o estudo de dois setores, não retrata a realidade do hospital, pois existem outras variáveis que devem ser levadas em consideração, mas essas estudadas são as mais importantes. Logo, deixa-se como sugestão, para estudos futuros a modelagem do Hospital Psiquiátrico, pois este também é um setor de grande importância, por ser um dos únicos da região. Também o estudo da previsão dos estoques do hospital seria de grande interesse, pois como foi demonstrado, pelo estudo, há uma grande rotatividade de pacientes, e para que as condições de higiene e atendimento sejam mantidas, o controle de estoque é necessário.

Pode-se ver que a metodologia empregada foi capaz de retratar a realidade do HUSM e fornecer subsídios para melhor funcionamento e melhor atendimento aos pacientes.

6 BIBLIOGRAFIA

ABDEL-AAL, R.E.; AL-GARNI, Z. **Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate times-séries analysis.** Energy, v. 22, n.11, p.1059-1069. Novembro 1997.

BOX, G.E.P.; JENKINS, G.M.; REINSEL, G. **Time Series Analysis: Forecasting and Control.** Third Edition. Englewood: Prentice Hall, 1994.

BOX, G.E.P. & JENKINS, G.M. **Time Séries Analysis, forecasting and control,** San Francisco. Holden Day, 1970.

BOX, G.E.P. & TIAO, G.C. **Comparioson of forecast and actuality.** Applied Statistica, 25: 195 – 200, 1976.

BOX, G.E.P. & TIAO, G.C. Intervention analysis with aplications to economic and everinmental problens. **Journal of the Americam Statistical association,** 70, 70 – 79.

BOX, G.E.P. & PIERCE, D.A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **Journal of the American Statistical Association.** V. 65, p. 1509-1526, 1970.

GLASS, G.V.; WILLSON, V.L.; GOTTMAN, J.M. **Design and Analysis of Time Series Experiments.** Boulder: Colorado Associated Universities Press, 1975.

HAMILTON, D.J. **Times Series Analysis.** Library of Congress Cataloging-in-Publication Data, 1994

HARVEY, A.C. DURBIM, J. **The Effects of Seat Belt Legislation on British Casualties: A case Study in Structural Time Séries Modelling.** Journal Royal Statistical Society, 1986.

HOSPITAL UNIVERSITÁRIO DE SANTA MARIA, História de Dedicção, Competência e Profissionalismo, Edição Elaine Resner, Publicação do HUSM/Hospital Universitário de Santa Maria/RS – Jan./2002, Gráfica e edição Palloti.

JENKINS, J. et al. **Statistical Analysis of Multiple Time Series Associated With Air Quality Data.** New Jersey Co Data. Madison, University of Wisconsin, 1978.

MAKRIDAKIS, S; WHEELWRIGHT,S.; HYNDMAN,R J. **Forecasting methods and applications**. 3 Ed. New York; John Wiley e Sons, 1998.

McDOWALL, D.; McCLEARY, R.; MEIDINGER, E. E.; HAY, R.A. **Interrupted Time Series Analysis**. Beverly Hills: Sage Publications, 1980.

MONTGOMERY, D.C. & JOHNSON, L.A. **Forecasting and Time Series Analysis**. New York: McGraw-Hill, 1976.

MORETTIN, P.A . & TOLOI, C.M.C. **Análise de Séries Temporais**. São Paulo, Edgard Blücher, 2004

MORETTIN, P.A. & TOLOI, C.M.C. , **Modelos de Função de Transferência**, Rio de Janeiro, 3ª Escola e séries Temporais Econometria, 1989.

MORETTIN, P.A. & TOLOI, C.M.C. **Previsão de Séries Temporais**, São Paulo, 2ª ed.1987.

PACK, D.J. **Forecasting Time Series Affected by Identifiable Isolated Events**. Working Paper Series, College of Adm. Sciences, Ohio State University, 1977.

NELSON, H.L. **The Use of Box-Cox Transformations in Economic Time Series Analysis: An Empirical Study**. Doctoral Dissertation. Univ. of Califórnia, San Diego, 1976.

PELLEGRINI, F.R.; FOGLIATTO, F. **Estudo comparativo entre modelos de Winters e de Box & Jenkins para a previsão da demanda sazonal**. Revista Produto e Produção, v. 4, número especial, p. 72-85, Abril 2000.

PINO, F.A. **Análise de Intervenção e Séries Temporais** – Aplicações em Economia Agrícola. Dissertação de Mestrado, IME-USP, 1980.

QUENOUILLE, M.H. **Approximate tests of correlation in time series**. Journal of the Royal Statistical Society, Series B, 11, 68-84, 1949.

SOUZA, A.M. **Aplicação e Performance da Análise de Intervenção em Séries Macroeconômicas Brasileiras**. Monografia de Especialização, UFSM, 1991.

VASCONCELOS, M.A.S.; ALVES, D. Anual de Econometria, São Paulo, Atlas – SP, 2000.

www.ufsm.br, COPERVES/DERCA/PROGRAD; PRRH; Biblioteca Central, Prefeitura da Cidade Universitária. Atualizado: setembro/2005.