

ET-64



UNIVERSIDADE EDUARDO MONDLANE

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE MATEMÁTICA E INFORMÁTICA

CURSO DE LICENCIATURA EM ESTATÍSTICA

TRABALHO DE LICENCIATURA EM ESTATÍSTICA

Área de Interesse: Séries Temporais

Tema:

**ESTUDO DA TENDÊNCIA DOS CASOS DE INFECÇÃO DE HIV
EM MOÇAMBIQUE**

MAURÍCIO CHEMANE TIMECANE

**Maputo
2009**



UNIVERSIDADE EDUARDO MONDLANE

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE MATEMÁTICA E INFORMÁTICA

CURSO DE LICENCIATURA EM ESTATÍSTICA

TRABALHO DE LICENCIATURA EM ESTATÍSTICA

Área de Interesse: Séries Temporais

Tema:

**ESTUDO DA TENDÊNCIA DOS CASOS DE INFEÇÃO DE HIV
EM MOÇAMBIQUE**

Autor: Maurício Chemane Timecane

Supervisor: dr. Adelino Juga

**Maputo
2009**

À memória do meu pai, Maurício Bernardo Chemane Timecane,
com muitas saudades

Agradecimentos

Aos meus supervisores, Professor Doctor Humberto Mora e doctor Adelino Juga, pelo apoio e orientação neste trabalho e pela oportunidade de tê-los como supervisores, uma grande experiência para a vida;

Aos meus professores, e de facto, também supervisores, João Moreno e Tiago Devesse, que foram muito fundamentais para o desenvolvimento deste trabalho;

A todos os professores do curso de Estatística no departamento de Matemática e Informática pela contribuição fornecida durante cada semestre;

A todos integrantes do programa nacional de combate ao HIV e ITS's, em particular ao Dr. Celso Mazivila pela ajuda na obtenção dos dados;

Aos meus pais, Maurício Bernardo Chemane Timecane e Ivone Tembe, aos meus irmãos, sobrinhos, enfim, a todos os meus familiares pelo carinho, amor, paciência e confiança que me têm sido depositados durante todos esses anos. Saibam que esta é uma conquista conjunta e que devo tudo ao exemplo e integridade que venho aprendendo convosco durante todos esses anos de vida;

Aos meus amigos, dr. Elísio Mabasso, Paulo Ziaveia, Sérgio Fumo, ao "Mano Vena" (Venâncio Bata), Gabriel Cossa, Lúcio Matsimbe, Aizel Mutsenga, Manuel Cuambe pelo apoio, paciência e amizade prestados durante esses 4 anos de convívio;

A todos os meus colegas do curso que ao longo destes anos me acompanharam, ajudando me a chegar ao final deste trajecto;

Aos funcionários da Direcção Nacional dos Registos e Notariado na secção de estatística pelo companheirismo, apoio e incentivo concedidos a minha pessoa, fermento que me fez continuar firme nesta jornada;

E, finalmente, a todos que directa ou indirectamente contribuíram para a realização deste trabalho

Declaração de Honra

Declaro por minha honra que este trabalho é resultado da minha própria investigação e das orientações do meu supervisor. E que não foi apresentado para obtenção de nenhum outro grau acadêmico que não seja o indicado – Licenciatura em Estatística pela Faculdade de Ciências da Universidade Eduardo Mondlane.

Autor

Maurício Chemane Timecane

(Maurício Chemane Timecane)

“Se não der frutos, valeu a beleza das flores;

Se não der flores, valeu a sombra das folhas;

Se não der folhas, valeu a intenção da semente.”

Hengil

Resumo

Nos últimos anos tem se verificado uma tendência cada vez mais crescente dos casos de infecção de HIV, exigindo a introdução de novas formas de planificação socio-económica por parte do governo de Moçambique. O estudo da evolução dos casos de infecção de HIV em Moçambique é uma actividade importante para auxiliar na determinação dos recursos necessários para a mitigação da epidemia do HIV. Para este trabalho, a metodologia de Box-Jenkins, descrita por estes autores na década 70 foi utilizada para analisar dados históricos dos casos de infecção de HIV nas regiões Sul, Centro e Norte do país no período compreendido entre Agosto/2003 e Junho/2008, tendo se verificado que somente a série temporal da região Centro se apresentou estacionária em nível. A previsão foi direccionada a cada região, buscando representar o comportamento das mesmas por meio de modelos de Box-Jenkins. A obtenção dos modelos mais adequados foi baseada em análise de gráficos e testes estatísticos próprios da metodologia, os quais subsidiaram a decisão de adoptar o modelo ARIMA (1,1,0) para prever o número de casos de infecção de HIV na região Sul, o modelo ARIMA (1,0,1) para a região Centro e o modelo ARIMA (0,1,1) para prever os casos de infecção de HIV na região Norte. As previsões foram todas elas feitas para o período compreendido entre Julho e Dezembro de 2008, mostrando para as regiões Sul e Norte uma tendência crescente dos casos de infecção de HIV e decrescente para a região Centro, durante o período de previsão.

Palavras chaves: Casos de infecção, Séries temporais, Modelos ARIMA, Previsão dos casos de infecção

Lista de tabelas

Tabela 1. Identificação do processo ARIMA a partir do comportamento da FAC e FACP.....	17
Tabela 2. Identificação do processo SARIMA a partir do comportamento da FAC e FACP.....	17
Tabela 3. Medidas descritivas dos casos de infecção de HIV nas regiões Sul, Centro e Norte de Agosto/2003 a Junho/2008.....	27
Tabela 4. Região sul. Testes de raiz unitária para a série em níveis e em primeiras diferenças.....	29
Tabela 5. Modelo ARIMA (1,1,0). Região Sul.....	30
Tabela 6. Modelo ARIMA (0,1,1). Região Sul.....	31
Tabela 7. Teste de Ljung-Box para os resíduos da série dos casos de infecção de HIV na região Sul.....	33
Tabela 8. Critérios de comparação para verificação do melhor modelo.....	33
Tabela 9. Previsões de casos de infecção de HIV na região Sul.....	34
Tabela 10. Região Centro. Resultados dos testes ADF, PP e KPSS.....	37
Tabela 11. Modelo ARIMA (0,0,1). Região Centro.....	37
Tabela 12. Modelo ARIMA (1,0,1). Região Centro.....	37
Tabela 13. Teste de Ljung-Box para os resíduos da série dos casos de infecção de HIV na região Centro.....	39
Tabela 14. Critérios de comparação para escolha do melhor modelo.....	39
Tabela 15. Previsões de casos de infecção de HIV na região Centro.....	40
Tabela 16. Resultados dos testes ADF, PP e KPSS para a série em níveis e em 1ª diferença.....	43
Tabela 17. Modelo ARIMA (1,1,13). Região Norte.....	44
Tabela 18. Modelo ARIMA (0,1,1). Região Norte.....	44
Tabela 19. Teste de Ljung-Box para os resíduos da série dos casos de infecção de HIV na região Norte.....	46

Tabela 20. Critérios de comparação para escolha do melhor modelo.....	47
Tabela 21. Previsões de casos de infecção de HIV na região Norte.....	47

Lista de figuras

Figura 1. Processo iterativo do método de Box-Jenkins.....	20
Figura 2. Série cronológica dos casos de infecção de HIV na região Sul.....	28
Figura 3. FAC e FACP dos casos de infecção de HIV na região Sul.....	29
Figura 4. FAC e FACP dos casos de infecção de HIV na região Sul em primeira diferença.....	30
Figura 5. Cronograma, <i>Q-Q plot</i> normal, histograma, FAC e FACP da série dos resíduos do modelo ARIMA (1,1,0) para a série dos casos de infecção de HIV na região Sul.....	32
Figura 6. Previsões dos casos de infecção de HIV na região Sul.....	35
Figura 7. Série cronológica dos casos de infecção de HIV na região Centro.....	35
Figura 8. FAC e FACP dos casos de infecção de HIV na região Centro.....	36
Figura 9. Cronograma, <i>Q-Q plot</i> normal, histograma, FAC e FACP dos resíduos do modelo ARIMA (1,0,1) para a série dos casos de infecção de HIV na região Centro.....	38
Figura 10. Previsões dos casos de infecção de HIV na região Centro.....	40
Figura 11. Série cronológica dos casos de infecção de HIV na região Norte.....	41
Figura 12. FAC e FACP dos casos de infecção de HIV na região Norte.....	42
Figura 13. FAC e FACP dos casos de infecção de HIV na região Sul em primeiras diferenças.....	42
Figura 14. Cronograma, <i>Q-Q plot</i> normal, histograma, FAC e FACP dos resíduos do modelo ARIMA (1,1,13) para a série dos casos de infecção de HIV na região Norte.....	45
Figura 15. Previsões dos casos de infecção de HIV na região Norte.....	48

Lista de abreviaturas

ADF	-----	Teste Aumentado de Dickey-Fuller
ATS	-----	Atendimento e Testagem em Saúde
ARIMA	-----	Auto Regressive Integrated Moving Averages
CIA	-----	Critério de informação de Akaike
CIS	-----	Critério de informação de Schwarz
EQM	-----	Erro Quadrático Médio
FAC	-----	Função Autocorrelação
FACP	-----	Função Autocorrelação Parcial
HIV	-----	Vírus de Imunodeficiência Humana
KPSS	-----	Teste de Kwiatkowski-Phillips-Schmidt-Shin
MQO	-----	Mínimos Quadrados Ordinários
PP	-----	Teste de Phillips-Perron
SIDA	-----	Síndrome de Imune Deficiência Adquirida

Índice	Pág.
1 INTRODUÇÃO.....	1
1.1 Definição do problema e relevância do estudo.....	3
1.2 Objectivos.....	4
1.2.1 Objectivo geral.....	4
1.2.2 Objectivos Específicos.....	4
2 REVISÃO DA LITERATURA.....	5
2.1.3.1 Modelo auto-regressivo (AR).....	12
2.1.3.2 Modelo de médias móveis (MA).....	13
2.1.3.3 Modelo autoregressivo e de médias móveis (ARMA).....	14
2.1.3.4 Modelos autoregressivos integrados e de médias móveis (ARIMA).....	15
2.1.3.5 Modelos sazonais autoregressivos integrados e de médias móveis (SARIMA).....	15
(i) Identificação.....	16
(ii) Estimação.....	18
(iii) Verificação de Diagnóstico.....	19
3 METODOLOGIA.....	22
3.1 Material usado.....	22
3.1.1 Fonte de dados.....	22
3.2.1 Testes de raiz unitária.....	22
3.2.2 Metodologia de Box-Jenkins.....	24
4 DISCUSSÃO DOS RESULTADOS.....	26
4.1 Análise descritiva.....	26
4.2 Metodologia de Box-Jenkins para a região Sul.....	27
4.2.1 Etapa de identificação.....	28
4.2.2 Etapa de estimação.....	30
4.2.3 Etapa de verificação do diagnóstico.....	31
4.3 Metodologia de Box-Jenkins para a região Centro.....	35
4.3.1 Etapa de identificação.....	35
4.3.2 Etapa de estimação.....	37
4.3.3 Etapa de verificação de diagnóstico.....	38

4.4 Metodologia de Box-Jenkins para a região Norte	41
4.4.1 Etapa de identificação	41
4.4.2 Etapa de estimação	44
4.4.3 Etapa de verificação de diagnóstico.....	44
5 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES	49
5.1 Considerações finais.....	49
5.2 Recomendações	50
REFERENCIAS BIBLIOGRÁFICAS	51
ANEXOS	54

1 Introdução

Segundo Floriano (2005:4), o quadro epidemiológico de Moçambique é, em larga escala, medida de pré-transição, isto é, dominado por doenças transmissíveis infecciosas e parasitárias, designadamente a malária, o HIV/SIDA, tuberculose, etc. em rápida ascensão. Dentre elas, a epidemia do HIV/SIDA é a que actualmente tem preocupado toda a humanidade devido aos seus efeitos negativos, tanto na população como no desenvolvimento socio-económico, pois acomete uma boa parte da população economicamente activa, havendo no entanto, necessidade de desenvolver mecanismos de protecção e minimização das consequências negativas desta epidemia.

A SIDA (Síndrome de Imune deficiência Adquirida) é a fase mais avançada da infecção pelo vírus de imunodeficiência humana (HIV). Surge como epidemia nos finais dos anos 70, tendo sido registado o 1º caso em 1981 nos Estados Unidos da América (Médici, 2002). Médici argumenta ainda que, em países em vias de desenvolvimento o maior número de mortes é atribuído ao HIV/SIDA sendo que desde a eclosão da epidemia até 2005, já haviam sido registados cerca de 60 milhões de pessoas infectadas com o HIV em todo mundo, dos quais, aproximadamente 20 milhões já perderam a vida.

Segundo dados da ronda de vigilância epidemiológica, no ano 2000 estimava-se em 1.1 milhão o número de pessoas infectadas pelo HIV em todo o país, entre homens, mulheres e crianças. E as projecções para 2010, indicam que cerca de 2 milhões de pessoas estarão infectadas pelo vírus da SIDA em todo o país, com a região Centro a apresentar o maior número de pessoas infectadas (Barreto et al, 2002).

Neste contexto, emerge a possibilidade do uso de modelos estatísticos para o auxílio no conhecimento da situação actual ou futuro dessa epidemia que é o HIV/SIDA, baseando-se em dados passados (Séries Temporais ou cronológicas). As séries temporais são observadas em domínios variados; por exemplo: **economia** (Preços diários de acções, taxa mensal de desemprego, produção industrial), **medicina** (electrocardiograma, electroencefalograma), **epidemiologia** (número mensal de novos casos de meningite), **meteorologia** (precipitação pluviométrica, temperatura diária, velocidade do vento), etc. A característica mais importante deste tipo de dados é a dependência existente entre observações vizinhas e o interesse é modelar essa dependência.

Segundo Neto (2008:9), modelos de séries temporais têm desempenhado papel relevante na modelagem e previsão de variáveis associadas a problemas epidemiológicos. Neto refere também que estes modelos (de previsão epidemiológica), se formulados de maneira adequada e alimentados com dados que tenham qualidade, podem auxiliar os administradores da saúde pública na procura de estratégias para o combate às epidemias.

Uma das técnicas quantitativas mais difundida – para previsões- é a metodologia de Box-Jenkins, descrita por esses autores na década 70. Os modelos de Box-Jenkins, também conhecidos como ARIMA (*Auto Regressive Integrated Moving Averages*)¹ partem da ideia de que cada valor da série temporal pode ser explicado por valores prévios, apartir do uso da estrutura de correlação temporal que geralmente há entre os valores da série(Werner & Ribeiro, 2003). Nunes (2004) acrescenta ainda que caso essa estrutura de correlação seja bem modelada fornecerá boas previsões. Esses modelos ajustam parâmetros autoregressivos com parâmetros de médias móveis, e explicitamente inclui diferenciação na formulação do modelo.

Aplicando a metodologia proposta por Box-Jenkins, no presente trabalho serão analisadas as séries temporais dos casos de infecção de HIV das regiões Sul, Centro e Norte do país com o objectivo de se estudar o comportamento dos casos de infecção de HIV nessas regiões durante o período que se estende de Agosto de 2003 a Junho de 2008 e elaborar previsões para os 6 meses do segundo semestre de 2008. A utilização de previsões dos casos de infecção de HIV no nosso país, tem como factor importante a introdução de melhorias no que se refere a organização e sincronização das formas adoptadas para o combate a esta epidemia.

O trabalho está estruturado em 5 capítulos. O capítulo 1 que engloba a introdução, a relevância do estudo bem como os objectivos pretendidos com a realização deste trabalho.

No capítulo 2 é feita uma revisão da literatura que compreende o estudo dos conceitos básicos sobre séries temporais, que envolvem a caracterização da estacionaridade de uma série, testes de raízes unitárias, modelos autoregressivos, de médias móveis e sua combinação, modelos integrados e modelos sazonais. De seguida faz se uma apresentação da metodologia de Box-Jenkins e as três

¹ Especificamente os três tipos de parâmetros do modelo são: parâmetros auto-regressivos (p), o número de diferenças (d) e os parâmetros de médias móveis(q). Na notação introduzida por Box-Jenkins (1970) são indicados como ARIMA(p,d,q)

etapas do ciclo iterativo que a compõe. O capítulo 3 aborda a metodologia do trabalho, isto é, apresentação do material usado para o estudo do caso e aplicação das técnicas empregues.

No capítulo 4 faz-se a discussão dos resultados, geradas previsões para os casos de infecção de HIV nas regiões Sul, Centro e Norte do país e analisada a taxa de crescimento no período em análise. E finalmente, no capítulo 5 são apresentadas as conclusões acerca dos resultados obtidos. Além disso, neste capítulo são descritas as limitações do trabalho e sugestões para pesquisas futuras.

1.1 Definição do problema e relevância do estudo

Um dos grandes objectivos do milénio para vários países subdesenvolvidos é a redução da SIDA e outras enfermidades transmissíveis na proporção de dois terços das taxas existentes em 1990, até 2015 (Medici, 2002). Contudo Floriano (2005) alerta que apesar dos avanços científicos, sociais, clínicos e laboratoriais alcançadas no combate a epidemia do HIV, muitas questões ainda se colocam como desafios para a sociedade e governo, exigindo de pesquisadores, activistas e profissionais de saúde e educação novas posturas sociais e políticas, para o controle da epidemia.

No entanto, um conhecimento, o mais exacto possível da dimensão e distribuição geográfica do HIV no nosso país, contribuirá em grande medida para a tomada de decisões e desenho de políticas cada vez mais abrangentes e integradas para o combate a epidemia.

Dentro desse contexto, surge a possibilidade do uso de uma das áreas mais importantes da Estatística, a que trata da análise de séries temporais onde dentre as várias metodologias (para análise de séries temporais), encontra-se o método de Box-Jenkins, que se resume na identificação de um modelo *Autoregressive integrated and moving average (ARIMA)* para descrever o comportamento da série. Plaumer citado por Ribeiro e Paula, depois de vários trabalhos de pesquisa concluiu que os modelos *ARIMA* propostos por Box e Jenkins, podem produzir previsões tão precisas quanto os métodos tradicionais, e ainda apresentam a vantagem de considerar a incerteza da projecção. Ele realça ainda, o facto de tanto a produção quanto a utilização de previsões probabilísticas deste tipo de variáveis estarem ainda na sua infância.

Algumas perguntas de pesquisa

1. Qual tem sido o comportamento dos casos de infecção do HIV nas regiões Norte, Centro e Sul do país?
2. Como modelar a série temporal dos casos de infecção de HIV?

1.2 Objectivos

1.2.1 Objectivo geral

Aplicar os modelos ARIMA para prever os casos de infecção de HIV

1.2.2 Objectivos Específicos

- ✚ Analisar o comportamento dos casos de infecção de HIV
- ✚ Modelar a série temporal dos casos de infecção de HIV para cada região do país
- ✚ Prever os casos de infecção de HIV para os 6 meses do 2º semestre de 2008.

2 Revisão da literatura

Neste capítulo serão descritos todos aspectos relacionados às séries temporais, que irão auxiliar no uso da metodologia proposta por Box & Jenkins.

2.1 Série temporal

De acordo com Latorre e Cardoso (2001), uma série temporal (Y_t), também denominada série histórica, é uma sequência de dados obtidos em intervalos regulares de tempo t , durante um período específico².

Segundo Barbosa e Milone (2004) o estudo das séries temporais permite avaliar a evolução, classificar e descrever informações, projectar, predizer e controlar valores e comportamentos futuros. Em termos gerais, o estudo das séries temporais sustenta-se nos seguintes princípios (Pires, 2001):

- ✚ **Descrição:** este é um objectivo modesto, mas quase sempre presente no estudo de uma série e envolve: cálculo de estatísticas sumárias da série, representação gráfica dos dados da série e de funções (empíricas) dos mesmos e o ajustamento de características simples da série;
- ✚ **Modelação** a modelação consiste na explicação de uma série como tendo sido gerada a partir de um certo modelo e envolve a estimação de parâmetros e a avaliação da qualidade do diagnóstico, em termos de qualidade estatística e do ajustamento do (s) modelo (s) proposto (s);
- ✚ **Previsão** Consiste na estimação de valores futuros de uma série temporal e a sua importância advém do facto de ser habitualmente vantajoso, nomeadamente no plano económico, conhecer a evolução provável de uma série no futuro;
- ✚ **Controlo** Consiste na monitorização dos valores da série tendo em vista a detecção de alterações nas características da mesma, as quais podem indicar que o modelo proposto não é válido; o modelo que estava a ser usado deixou de ser válido (ex: a série correspondente a

² Contrariamente ao que se passa em relação a amostras aleatórias – em que o instante em que é efectuada cada uma das observações não desempenha nenhum papel especial e é, em geral, ignorado – nas séries temporais o instante em que é efectuada cada uma das observações necessita de ser registado e é usado na análise/modelação da série.

um índice pode sofrer uma rotura por mudança do modo do cálculo do índice); e alguns dos parâmetros do modelo mudaram de valor.

2.1.1 Componentes das séries temporais

Os componentes das séries temporais podem ser inicialmente classificados em sistemáticos (que apontam as regularidades) e não sistemáticos (que apontam as irregularidades). A principal componente sistemática é a tendência (seu comportamento no longo prazo), mas as séries também se sujeitam a variações cíclicas e sazonais. As demais ocorrências dos eventos, em geral irregulares costumam ser resumidas em uma única componente, denominada variação residual. Mais detalhadamente temos (Barbosa e Milone, 2004):

(i) **Tendência:** Ela é a componente macro da série temporal. Ela define a inércia do sistema, a direcção da variável em determinado intervalo de tempo;

(ii) **Ciclos e Sazonalidade:** São oscilações (mais ou menos) regulares em torno da tendência. Essas oscilações podem decorrer de usos e costumes ou de fenómenos naturais, socioculturais ou económicos, como ciclos vegetativos, períodos lectivos, competições desportivas, recessão e euforia económica. O que faz determinada oscilação ser chamada de ciclo ou de sazonalidade é seu período, isto é, diz se que o evento é cíclico quando sua oscilação for plurianual e quando o período for de curto prazo (menor que um ano), fala-se de evento sazonal. Pires (2001) acrescenta ainda que eventos sazonais, diferentemente dos cíclicos são de período conhecido e antecipável.

(iii) **Resíduos:** É a parcela não sistemática dos dados, a porção deles decorrentes de causas imprevistas, de ordem natural, social, económica, etc;

De acordo com Pires (2001), em séries temporais como é habitual possuir-se apenas uma única realização da série, para inferir correctamente a estrutura da dependência serial da mesma apartir dessa realização, a série terá de possuir uma estrutura de dependência serial simples, no sentido de esta se manter invariante ao longo do tempo. É por esse facto que convém introduzir noções de estacionaridade para séries temporais.

2.1.2 Estacionaridade

Uma importante classe de processos estocásticos é a dos chamados processos estacionários. Segundo Ehlers (2005), antes de se conduzir qualquer análise baseada em dados de séries temporais, é importante definir se a série temporal é ou não estacionária, para a partir daí estabelecer a estrutura do modelo probabilístico que estimará a série. A razão simples de se necessitar de dados estacionários, é que qualquer modelo que é inferido a partir desses dados pode ser interpretado como estacionário ou estável, fornecendo assim uma base válida para a previsão (Gujarati, 2004)

Segundo este mesmo autor, uma série temporal é dita (fracamente) estacionária se suas média (μ) e variância (σ^2) forem constantes ao longo do tempo e o valor da covariância entre dois períodos de tempo depender apenas da distância ou defasagem entre os dois períodos, e não do período de tempo efectivo em que a covariância é calculada, i.e.

$$i) E(Y_t) = \mu \quad (1)$$

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

$$iii) \gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)] \quad (3)$$

Igualando na equação (3), $k = 0$ obtêm-se a variância de Y_t , dada por $Var(Y_t) = \gamma_0$. O termo γ_k é uma função de k e é conhecido na literatura como função de autocovariância. Mas para efeitos de interpretação, é mais útil padronizar a função de autocovariância dando origem a uma função de autocorrelação dada por:

$$\rho_k = \frac{Cov(Y_t, Y_{t+k})}{Var(Y_t)} = \frac{\gamma_k}{\gamma_0} \quad (4)$$

que mede a correlação entre Y_t e Y_{t+k} .

A função de autocorrelação desempenha, a par da média e da variância um papel crucial no estudo da estacionaridade de um processo estocástico, uma vez que permite capturar a extensão temporal e robustez da memória do processo, ao medir a correlação dos valores actuais do processo com os seus valores passados (Menezes, 2004). Menezes acrescenta ainda que a função de autocorrelação tem a propriedade de se atenuar à medida que k aumenta e apresenta comportamentos semelhantes em muitos casos, sendo por isso difícil distinguir entre processos de ordem diferente, advindo deste modo a necessidade de analisar a função de autocorrelação parcial, introduzida a seguir.

Segundo Pires (2001) a função de autocorrelação parcial, $\{\phi_{kk}, k \in N_0\}$ de uma série estacionária Y_t de média nula é tal que o seu coeficiente de ordem k, ϕ_{kk} , representa o coeficiente de correlação parcial entre Y_t e Y_{t+k} uma vez eliminado o efeito de $\{Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}\}$, o qual se obtém com base no modelo de regressão linear com erros normais:

$$Y_{t+k} = \phi_{k1}Y_{t+k-1} + \phi_{k2}Y_{t+k-2} + \dots + \phi_{kk}Y_t + \varepsilon_{t+k} \quad (5)$$

com $\varepsilon_t \sim \text{Normal}(0, \sigma^2_{\varepsilon})$ e ε_{t+k} independente de $\{Y_{t+k-j}, j \geq 1\}$.

2.1.2.1 Testes de estacionaridade

Dado que a eficiência das estimações dos modelos que utilizam séries temporais depende da estacionaridade da série, são descritos alguns testes capazes de detectar se o processo estocástico é ou não estacionário, conhecidos como testes de raiz unitária. Tais testes são capazes de detectar se a série foi suficientemente diferenciada para se tornar estacionária (Franco et al, n.d). Ainda segundo eles, se uma série deve ser diferenciada d vezes antes de tornar-se estacionária, então ela contém d raízes unitárias e é dita ser integrada de ordem d , denotada por $I(d)$.

Para a apresentação dos testes de raiz unitária é necessário conhecer, inicialmente, a definição de um processo de dados auto-regressivos. Supondo um processo auto-regressivo de primeira ordem, AR (1) temos:

$$Y_t = \rho Y_{t-1} + \varepsilon_t \quad (6)$$

Onde: ρ = parâmetro que relaciona o valor presente a série ao seu valor no momento passado;

ε_t = termo de erro estocástico que satisfaz as hipóteses clássicas de i) média zero; ii) Variância constante e; iii) é não autocorrelacionado.

A partir da definição do processo AR (1) fica claro que a estacionaridade da série temporal depende do valor do parâmetro ρ , que representa a mediação entre os valores passados e presentes da série. Caso $|\rho| < 1$ a série temporal é dita estacionária, uma vez que os choques externos são dissipados ao longo do tempo. Segundo Resende (2006), com o passar do tempo e em virtude de $|\rho| < 1$, o efeito de um choque externo na variável, terá um impacto cada vez menor como apresentado a seguir:

$$Y_{t-1} = \rho Y_{t-2} + \varepsilon_{t-1}$$

$$Y_{t-2} = \rho Y_{t-3} + \varepsilon_{t-2}$$

$$Y_t = \rho^n Y_{t-n} + \sum_{n=1}^{\infty} \rho^n \varepsilon_{t+1-n} \quad (7)$$

Ele acrescenta ainda que caso um choque produza efeitos que se perpetuem ao longo do tempo, $\rho = 1$, diz se que a série é dita não estacionária e caso $|\rho| < 1$ a mesma é dita estacionária.

A despeito disso, é possível estacionarizar a série através do cálculo de sua 1ª diferença, ou seja, fazendo:

$\Delta Y_t = Y_t - Y_{t-1}$ e a partir da equação (6) temos que:

$\Delta Y_t = \rho Y_{t-1} + \varepsilon_t + Y_{t-1} = (\rho - 1)Y_{t-1} + \varepsilon_t$, de onde temos que:

$$\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t \quad (8)$$

onde: $\gamma = \rho - 1$ e como $\rho = 1$ então $\Delta Y_t = \varepsilon_t$

A intuição presente na equação acima nos leva a concluir que todo o processo AR (1) não estacionário em nível, o será em primeira diferença, pois a diferenciação gera uma série de dados igual aos resíduos - como dito na definição do processo AR(1) - que atendem às hipóteses clássicas e são estacionários.

Com base nas análises feitas, um processo estocástico AR (1), é considerado um processo integrado de ordem 1 ou alternativamente, de passeio aleatório (random walk). A série temporal descrita por um processo AR (1) com $\rho = 1$ é chamada de série temporal com raiz unitária.

Os testes mais usados para detectar a presença de raiz unitária (ou seja, não estacionaridade) e sua ordem de integração são o Dickey-Fuller(DF), Augmented Dickey-Fuller(ADF), Phillips-Perron(PP) e Kwiatkowski-Phillips-Schmidt-Shin(KPSS).

O teste de DF consiste no teste a partir de (6), das seguintes hipóteses:

$H_0: \rho = 1$, (A série tem raiz unitária, i.e., não é estacionária)

$H_1: |\rho| < 1$, (A série não tem raiz unitária, i.e., é estacionária)

A não rejeição da hipótese nula (H_0) indica a presença da raiz unitária na série testada e, portanto, a série não é estacionária. A hipótese nula de uma raiz unitária é baseada no teste t com uma distribuição não normal nem simétrica

De acordo com Enders (1995) citado por Resende (2006), o teste de DF pode ser estimado com base nos seguintes modelos:

$$\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t \quad (9)$$

$$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \varepsilon_t \quad (10)$$

$$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \alpha_2 t + \varepsilon_t \quad (11)$$

Note-se que, α_0 é uma constante, $\alpha_2 t$ captura uma eventual tendência determinística existente nos dados e γY_{t-1} modela a respectiva tendência estocástica.

O teste DF consiste na estimação por mínimos quadrados ordinários (MQO), de uma das três equações acima e com base no parâmetro estimado e no seu desvio padrão, compara-se a estatística t (denominada estatística tau) calculada ao valor crítico tabulado por Dickey e Fuller mediante simulações de Monte Carlo.

Mas como na maioria dos casos, não é ruído branco, utiliza-se um teste com uma correção paramétrica, denominado teste aumentado de Dickey-Fuller (ADF), que sugere aumentar a regressão adicionando termos suficientes em por forma branquear os resíduos (Franco et al, n.d). A inclusão dos termos supracitados dá origem aos seguintes modelos:

$$\Delta Y_t = \gamma Y_{t-1} + \sum_{i=2}^p \beta_i \Delta Y_{t-i+1} + \varepsilon_t \quad (12)$$

$$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \sum_{i=2}^p \beta_i \Delta Y_{t-i+1} + \varepsilon_t \quad (13)$$

$$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \alpha_2 t + \sum_{i=2}^p \beta_i \Delta Y_{t-i+1} + \varepsilon_t \quad (14)$$

Onde p representa o número de defasagens necessárias para tornar o termo de erro serialmente independente.

A realização do teste ADF segue os mesmos procedimentos que o teste DF e os valores críticos utilizados são os mesmos, já que suas estatísticas possuem a mesma distribuição (Camuri, 2005).

Contudo, Menezes (2004) alerta que apesar da sua popularidade, devida à simplicidade, os testes ADF não estão isentos de problemas. De acordo com ele vários autores analisaram os problemas da dimensão da amostra e potência dos testes de raiz unitária ADF, tendo observado que alguns processos de raiz unitária apresentam em amostras finitas um comportamento mais próximo de um

processo ruído branco do que de um processo passeio aleatório. Ao contrário, alguns processos estacionários em tendência comportam-se de forma aproximada ao passeio aleatório em amostras finitas.

Alternativamente aos testes propostos por Dickey e Fuller, alguns autores propuseram outros testes baseados numa filosofia diferente da usada por aqueles. Um desses testes é o teste de Phillips-Perron(PP), que contrariamente ao teste ADF, apresenta uma abordagem não paramétrica(Rua &Almas,1998). Segundo Resende (2006) o teste PP apresenta (em relação aos testes DF e ADF) a vantagem de permitir que os resíduos sejam autocorrelacionados, heterocedásticos, fracamente dependentes e/ou heterogeneamente distribuídos, o que suplanta as propriedades constituintes dos resíduos estritamente ruído branco³. Camuri (2005) acrescenta ainda que o teste PP pode ser visto como uma extensão do teste DF dado que assume as mesmas equações e os mesmos valores críticos no cálculo da estatística t. A raiz unitária é testada com base na estatística z, que é uma correcção da estatística t incorporando as possibilidades de resíduos não autocorrelacionados e/ou heterocedásticos. Necessário realçar contudo, que o teste PP pode apresentar distorções de tamanho em amostras finitas quando o processo gerador de dados apresentar predominância de autocorrelações negativas na 1ª diferença.

Os testes ADF e PP são baseados em teoria assintótica, e portanto, a escolha de qual o teste mais adequado dependerá de como as estatísticas aproximam a distribuição para amostra finita da estatística relevante (Holden&Perman, 1996, citado por Resende, 2006). Segundo este mesmo autor se termo de resíduo tiver um componente de média móvel positivo o teste PP é mais preferível que o ADF, porém, se o termo de média móvel for negativo, o poder do teste ADF é maior que o do teste PP. No entanto Enders (1995, citado por Camuri, 2005), aconselha ao uso dos dois testes de forma a verificar se um corrobora o outro.

Um outro teste comumente usado para a detecção de raiz unitária, foi sugerido em 1992 por Kwiatkowski, Phillips, Schmidt e Shin, cuja principal característica é a inversão das hipóteses em teste, isto é, a hipótese nula assume que a série é estacionária e sob a hipótese alternativa a série é

³ Resíduo estritamente ruído branco admite que além das hipóteses tradicionais do ruído branco-média zero, variância constante e não autocorrelação-sejam adicionadas as hipóteses de independência dos resíduos e distribuição idêntica (iid).

integrada de ordem um. Este teste baseia-se na representação em componentes da série temporal, sendo o processo gerador de dados representado pelo seguinte modelo:

$$\begin{aligned} Y_t &= \mu t + Z_t + u_t \\ Z_t &= Z_{t-1} + \varepsilon_t \end{aligned} \quad (15)$$

Onde Y_t representa a soma de uma tendência determinística (μ), de um passeio aleatório (Z_t) e de uma variável residual estacionária (u_t) e $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. A hipótese nula de estacionaridade da série é dada por $\sigma_\varepsilon^2 = 0$ com $Z_0 = \text{constante}$.

A estatística do teste KPSS é dada por:

$$LM = \frac{\sum_{t=1}^T S_t^2}{\hat{\sigma}_u^2} \quad (16)$$

onde $S_t = \sum_{r=1}^t \hat{u}_r$ é a soma dos resíduos da regressão de Y_t sobre uma constante e uma tendência determinística, ou seja $\hat{u}_t = x_t - \hat{\mu}t - \hat{z}_0$. Por outro lado, $\hat{\sigma}_u^2$ é um estimador da variância dos resíduos u_t . A vantagem deste teste resulta da falta de potência dos testes convencionais, isto é, tendem a não rejeitar a hipótese nula com demasiada frequência (Rua & Almas, 1998).

2.1.3 Modelos de séries temporais

2.1.3.1 Modelo auto-regressivo (AR)

Um modelo auto-regressivo é definido de maneira que os valores da série no tempo t dependem dos valores passados. Mais especificamente, o modelo de ordem p , AR(p) é:

$$Y_t = c + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t = c + \sum_{j=1}^p \alpha_j Y_{t-j} + \varepsilon_t \quad (17)$$

Onde ε_t é um processo puramente aleatório com média zero e variância constante e α_j são os parâmetros autoregressivos de ordem j .

Recorrendo ao operador atraso, a equação (17) pode ser escrita como:

$$(1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p) Y_t = c + \varepsilon_t \Leftrightarrow \alpha(B) Y_t = c + \varepsilon_t \quad (18)$$

Com $B^j Y_t = Y_{t-j}$

A condição de estacionaridade para modelos AR (p) é satisfeita quando as raízes do polinómio $\varphi(B)$ estiverem fora do círculo unitário.

Por simplicidade, considere-se um modelo autoregressivo de primeira ordem, AR(1), i.e.

$$(1 - \alpha_1 B)Y_t = c + \varepsilon_t \quad (19)$$

Nesse caso, o valor previsto no período t é simplesmente alguma proporção ($= \alpha_1$) de seu valor ao tempo (t-1) mais uma perturbação ou choque aleatório no período t e uma constante c.

As autocorrelações são obtidas a partir de: $\rho_k = \alpha^k$

A FAC de um processo autoregressivo apresenta um comportamento que é uma mistura de um decaimento exponencial e/ou sinusoidais e a FACP para o mesmo processo apresenta um corte para zero para defasagens maiores que a ordem p.

2.1.3.2 Modelo de médias móveis (MA)

Segundo Jonston & Dinardo (2001), o modelo de médias móveis de ordem q, ou seja MA(q), (do inglês *moving average*), descreve um conjunto de dados como sendo resultado da combinação linear dos choques aleatórios que se comportam como ruído branco, ocorridos no período corrente e nos períodos passados, i.e.:

$$Y_t = v + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} = v + \sum_{i=1}^q \alpha_i \varepsilon_{t-i} \quad (20)$$

Alternativamente, (20) pode ser escrito como:

$$Y_t = v + \beta(B)\varepsilon_t, \text{ com } \beta(B) = 1 + \beta_1 B + \beta_2 B^2 + \dots + \beta_q B^q \text{ a representar o}$$

polinómio do operador atraso.

Segundo Barros (n.d), a condição de estacionaridade em processos MA (q) é sempre satisfeita e portanto em geral é desejável impor restrições para que ele satisfaça uma condição chamada invertibilidade semelhante à condição de estacionaridade.

Um processo MA (q) é invertível se a equação $\theta(B) = 0$ tem todas as raízes fora do círculo unitário, isto é, se $\theta(B) \neq 0$ para $|B| \leq 1$.

O processo MA (1),

$$Y_t = v + \varepsilon_t + \beta_1 \varepsilon_{t-1} \quad (21)$$

é caso particular de MA (q) quando $q=1$. Caso $|\beta_1| < 1$ o processo é dito estacionário e invertível.

As autocorrelações são obtidas como:

$$\rho_k = -\frac{\beta}{(1 + \beta^2)} \quad (22)$$

A FAC é zero para defasagens maiores que q e a FACP apresenta uma combinação de decaimentos exponenciais e sinusoidais.

2.1.3.3 Modelo autoregressivo e de médias móveis (ARMA)

Como descrito anteriormente, um processo estacionário e invertível pode representar-se na forma autoregressiva ou na forma médias móveis. Contudo pode bem acontecer que o processo gerador da série observada tenha uma estrutura que resulte de combinação de ambos processos. Segundo Ehlers (2005), processos autoregressivos e de médias móveis (ARMA) formam uma classe de modelos muito úteis e parcimoniosos para descrever dados de séries temporais.

Um processo Y_t diz-se ARMA (p,q) se possuir representação da forma:

$$Y_t = r + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (23)$$

ou ainda na forma polinomial:

$$\alpha(B)Y_t = r + \beta(B)\varepsilon_t \quad (24)$$

A condição de estacionaridade em processos ARMA (p,q) é satisfeita se as raízes de $\phi(B) = 0$ estiverem fora do círculo unitário e a da invertibilidade se as raízes $\beta(B) = 0$ estiverem fora do círculo unitário.

O modelo ARMA (1,1) é particularmente muito importante entre os modelos ARMA (p,q) e é dado por:

$$Y_t = r + \alpha_1 Y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1} \quad (25)$$

e sua função autocorrelação definida por:

$$\rho_k = \frac{(\alpha - \beta)(1 - \alpha\beta)}{1 - 2\alpha\beta + \beta^2} \alpha^{k-1}, \quad k \in \mathbb{N}$$

A FAC de um processo ARMA (p,q) comporta-se como a de um processo AR puro - decai gradualmente para zero a partir da defasagem q e a FACP prolonga-se indefinidamente comportando-se eventualmente de modo muito semelhante à dos processos médias móveis puros, isto é, apresenta

um decaimento gradual para zero majorado por uma soma de exponenciais e/ou sinusoidais amortecidas (Murteira, Müller & Turkman, 1993).

2.1.3.4 Modelos autoregressivos integrados e de médias móveis (ARIMA)

Os modelos anteriormente discutidos, são apropriados para séries temporais estacionárias, ou seja, que variam no tempo em torno de uma média constante (séries estacionárias). Assim, para ajustar modelos a séries temporais não estacionárias é necessário remover as fontes de variação não estacionárias, desde que as mesmas não tenham um comportamento explosivo⁴ (Luz et al, 2005).

A remoção da tendência consiste em tomar um número finito de diferenças d , da série temporal original. A série que se torna estacionária através deste procedimento é chamada de não estacionária homogênea.

Denotando a série diferenciada por,

$$X_t = \Delta^d Y_t = (1 - B)^d Y_t$$

o processo autoregressivo integrado e de médias móveis denotado ARIMA (p,d,q) é dado por:

$$X_t = a + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (26)$$

ou de forma equivalente,

$$\alpha(B)(1 - B)^d Y_t = \beta(B)\varepsilon_t$$

onde ε_t é ruído branco com média zero, $\phi(B)$ e $\beta(B)$ são respectivamente os polinômios autoregressivo e médias móveis.

2.1.3.5 Modelos sazonais autoregressivos integrados e de médias móveis (SARIMA)

A classe de modelos SARIMA é uma generalização da classe de modelos ARIMA que permite modelar séries com componente sazonal. Por exemplo, com dados mensais e $S=12$ tipicamente espera-se que Y_t dependa de Y_{t-12} e talvez de Y_{t-24} além de Y_{t-1} , Y_{t-2} ,...

Em geral, diz se que a série Y_t é uma série integrada mista autoregressiva e medias móveis estritamente sazonal de período S com ordens P , D e Q e escreve-se $Y_t \sim SARIMA(P, D, Q)_s$, se satisfaz uma equação às diferenças estocásticas da forma:

$$\phi(B^s)(1 - B^s)^D Y_t = \theta(B^s)\varepsilon_t \quad (27)$$

⁴ Séries que mesmo depois de várias vezes diferenciadas não ganham comportamento estacionário

De modo mais geral, podem ainda considerar-se modelos integrados mistos multiplicativos, isto é, modelos com uma ou mais componentes integradas mistas entre componentes sazonais e componentes simples. No entanto, a série Y_t integrada mista multiplicativa $SARIMA(p, d, q)x(P, D, Q)$, possui uma representação da forma:

$$\Phi(B^s)\phi(1-B^s)^D(1-B)^d Y_t = \Theta(B^s)\theta(B)\varepsilon_t \quad (28)$$

com os polinómios de cada um dos seguintes pares

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \text{ e } \theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

$$\Phi(B) = 1 - \Phi_1 B - \dots - \Phi_p B^p \text{ e } \Theta(B) = 1 - \Theta_1 B - \dots - \Theta_Q B^Q$$

sem raízes no círculo unitário e sem raízes comuns.

2.3 Método de previsão

2.3.1 Metodologia de Box-Jenkins

O método proposto por Box & Jenkins (1976) tem recebido muita atenção nos últimos anos. O objectivo da metodologia apresentada por estes, é identificar e estimar um modelo estatístico da classe ARIMA que possa ser interpretado como tendo gerado os dados amostrais (Gujarati, 2004).

Segundo Souza & Camargo (citado por Ilgner, Russo & Miranda, 1998) a filosofia da modelagem Box-Jenkins se utiliza de duas ideias: o princípio da parcimônia e a construção de modelos através de um ciclo iterativo. A abordagem clássica para a construção e ajuste dos modelos Box-Jenkins passa por 3 etapas, nomeadamente (Nemec, 1996):

(i) Identificação

Esta etapa envolve subjectividade e julgamento pessoal. É considerada a mais difícil e delicada e não há consenso sobre qual a melhor estratégia a ser seguida (Granger & Newbold, 1986 citado por De Melo, 2001). De acordo com Murteira et al (1993), nesta etapa está-se mais interessado em reduzir o número de possíveis modelos do que em escolher o “melhor” modelo. Barros (n.d) acrescenta ainda que no passo da identificação (também chamado de identificação estrutural) procura-se encontrar procedimentos que permitem identificar a estrutura do modelo, isto é, conhecer os valores apropriados de p , d e q (ou P , D e Q em caso de modelos sazonais) que caracterizam a estrutura do

modelo ARIMA (ou SARIMA) para a série temporal em estudo, sendo condição indispensável para uma correcta identificação o conhecimento profundo das FAC e FACP dos modelos mais usuais.

As tabelas 1 e 2 apresentam de forma resumida e simplificada, as propriedades das FAC e FACP dos vários modelos ARIMA e SARIMA:

Tabela 1. Identificação do processo ARIMA a partir do comportamento da FAC e FACP

	AR(p)	MA(q)	ARMA(p,q)
FAC	Exponenciais e/ou senóides amortecidas	Corte após o lag q	Exponenciais e/ou senóides amortecidas
FACP	Corte após o lag p	Exponenciais e/ou senóides amortecidas	Exponenciais e/ou senóides amortecidas

Tabela 2. Identificação do processo SARIMA a partir do comportamento da FAC e FACP

	(P,D,0)	(0,D,Q)	(P,D,0) ou (0,D,Q)	(P,D,0) e (0,D,Q)	Nenhum operador sazonal
FAC	Decaimento	Picos nas defasagens $s, 2s, \dots, Qs$ e corte após Qs	Picos nas defasagens $s, 2s, \dots, Qs$ e corte após Qs	Decaimento rápido na defasagem sazonal	Valores pequenos em todas as defasagens sazonais (não há picos)

FACP	Picos nas defasagens, $2s, \dots, P_s$ e corte após P_s	Decaimento	Picos nas defasagens, $2s, \dots, P_s$ e corte após P_s	Decaimento rápido na defasagem sazonal	Valores pequenos em todas as defasagens sazonais (não há picos)
------	---	------------	---	--	---

(ii) Estimação

O passo que se segue à identificação de um possível modelo para uma série temporal é o da estimação dos parâmetros desse modelo, isto é, estimar os p parâmetros autoregressivos (α) e os q parâmetros de médias móveis (β), utilizando o método dos mínimos quadrados ordinários.

Fazendo $\varphi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p$ e $\theta(B) = 1 + \beta_1 B + \beta_2 B^2 + \dots + \beta_q B^q$ pode se escrever a equação (24) como (sem a constante):

$$\alpha(B)\Delta^d Y_t = \alpha(B)W_t = \beta(B)\varepsilon_t \tag{29}$$

onde $W_t = \Delta^d Y_t$

Reescrevendo (29) em termos de erro ε_t , temos:

$$\varepsilon_t = \beta^{-1}(B)\alpha(B)W_t \tag{30}$$

A estimação tem por objectivo a obtenção de um conjunto de α e β que minimize a soma de quadrados dos erros, ou seja:

$$S(\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q) = \sum_t \varepsilon_t^2 \tag{31}$$

No caso do modelo puramente autoregressivo, o método de estimação por mínimos quadrados ordinários (MQO) é consistente. Se existem termos de médias móveis, a equação (31) é não linear em seus parâmetros e, assim, faz se necessário o uso de um método de estimação não linear na minimização (Abitante, 2007).

(iii) Verificação de Diagnóstico

Esta etapa consiste em verificar se o modelo identificado e estimado descreve adequadamente o comportamento dos dados da série. Caso o modelo não seja adequado, o ciclo é recomeçado e um outro modelo alternativo é escolhido, para depois passar-se a última etapa da metodologia, que constitui o objectivo principal da metodologia: realizar previsões, daí Gujarati (2004) considerar a metodologia proposta por Box & Jenkins (1976) como sendo mais uma arte do que uma ciência (vide Fig. 1).

Uma técnica que pode ser utilizada caso se suspeite que um modelo mais elaborado (contendo parâmetros) é necessário, é o superajustamento que consiste em incluir parâmetros extras no modelo e verificar sua significância estatística.

A verificação da adequabilidade do modelo é efectuada nas autocorrelações amostrais dos erros (ε_t), os quais seguem assintoticamente uma distribuição normal, com média zero e desvio padrão $n^{-\frac{1}{2}}$ se forem provenientes de um ruído branco. As autocorrelações dos resíduos são calculadas por (Moretin, 2004):

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^n \hat{\varepsilon}_t^2} \quad (31)$$

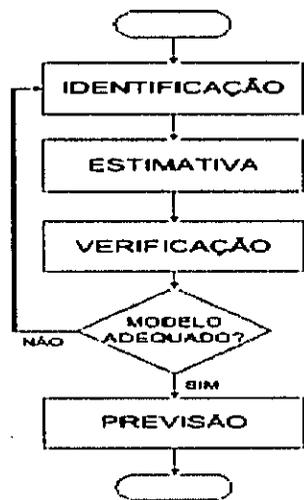
e seus valores comparados com limites $\pm \frac{2}{\sqrt{n}}$ fornecendo uma indicação geral de possível quebra de comportamento de ruído branco em ε_t , com a condição de que seja lembrado que para k pequeno, estes limites subestimarão a significância de qualquer discrepância.

Em adição ao exame das autocorrelações individuais dos resíduos, em 1970 Box e Pierce sugeriram um teste para a verificação da significância estatística de todo o conjunto das autocorrelações. Este teste apesar de não detectar quebras específicas no comportamento de ruído branco, pode indicar se esses valores são muito altos. No entanto, este teste mostrou não ter bom desempenho em amostras pequenas no sentido de que a distribuição se afasta da qui-quadrado. Mais tarde Ljung e Box (concretamente no ano de 1978) propuseram um melhoramento do teste que é apresentado a seguir:

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{r}_k^2}{(n-k)} \quad (32)$$

Sua distribuição amostral também é aproximadamente qui-quadrado com m-p-q graus de liberdade, rejeitando-se a hipótese nula (de que o modelo é adequado) para valores de Q maiores que o valor crítico assintótico. Na prática o número m de autocorrelações amostrais é tipicamente escolhido entre 15 e 30.

Figura 1. Processo iterativo do método de Box-Jenkins



Fonte: Luz et al (2005)

2.3.2 Critérios de selecção de modelos

Dada a possibilidade de existência de vários modelos que podem descrever de forma satisfatória uma série temporal, a escolha do melhor modelo torna-se mais facilitado recorrendo se a diversos critérios baseados em estatísticas construídas a partir da sucessão residual correspondente a um determinado modelo ajustado ou nos erros de previsão produzidos por esse mesmo modelo. (Murteira et al, 1993). Esses critérios são:

- **Critério de informação de Akaike (CIA)** $(CIA) = \frac{2k}{n} + \ln\left(\frac{\sum \hat{\epsilon}_t}{n}\right)$ (33)

onde k é o número de regressores (incluindo o intercepto) e n é o número de observações.

- **Critério de informação de Schwarz (CIS)** $(CIS) = \frac{k}{n} \ln n + \ln\left(\frac{\sum \hat{\epsilon}_t}{n}\right)$ (34)

$$\bullet \text{ Erro quadrático médio (EQM)} = \frac{\sum \hat{\epsilon}_i^2}{n} \quad (35)$$

Os critérios CIA e CIS ajudam a resolver o problema de discriminação entre os modelos que foram aceites na fase de verificação do diagnóstico, seleccionando-se como melhor modelo o qual os correspondentes CIA e CIS tenham o menor valor.

Ainda segundo esses autores, pela forma como são definidos os critérios verifica-se que são penalizados os modelos com maior número de parâmetros e com maior variância residual, isto é, um modelo menos parcimonioso só é seleccionado se os seus resíduos tiverem uma variância significativamente mais pequena (melhor ajustamento) do que os modelos concorrentes mais parcimoniosos.

No entanto, como a ocorrência de erros numa previsão é esperado e um dos objectivos dos métodos de previsão é minimizar a ocorrência desses erros, o critério do erro quadrático médio (EQM) é sugerido por vários autores como um bom critério de precisão (Luz et al, 2005).

3 Metodologia

Neste capítulo é feita uma apresentação dos aspectos relacionados com a metodologia apresentada por Box-Jenkins baseada nos modelos da classe ARIMA.

3.1 Material usado

Para a concretização dos objectivos almejados foram usados os casos de infecção mensais do HIV registados em cada região do país no período compreendido entre Agosto de 2003 a Junho de 2008 em mais de 200 unidades de atendimento e testagem em Saúde (ATS).

Para o processamento e análise dos dados foram usados dois pacotes estatísticos, nomeadamente: EViews V.4.1 e o SPSS V. 13.0.

Nos testes realizados o nível de significância estabelecido foi de 5%.

3.1.1 Fonte de dados

A base de dados usada foi fornecida pelo programa nacional de combate às ITS/HIV/SIDA com sua sede no edifício do Ministério da Saúde.

3.2 Análise de dados

A análise de dados começou primeiramente pela análise gráfica de forma a mostrar a evolução da série ao longo do tempo, isto é, a tendência da série, a existência ou não da sazonalidade, a ocorrência ou não de valores aberrantes, etc.

3.2.1 Testes de raiz unitária

Para que as conclusões advindas do processo estocástico tenham confiabilidade a série temporal utilizada deverá apresentar a característica da estacionaridade, isto é, variar ao longo do tempo em torno de uma média constante. A detecção da presença de não estacionaridade da série foi mediante o uso de testes de raiz unitária que revelaram o número de diferenças necessárias para tornar cada série estacionária. Para tanto foram usados os testes seguintes:

✓ **Teste ADF**

O teste ADF (Aumentado de Dickey-Fuller), também conhecido por teste τ (Dickey e Fuller, 1979, 1981), baseia-se na seguinte regressão para a variável Y_t

$$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \alpha_2 t + \sum_{i=2}^p \beta_i \Delta Y_{t-i+1} + \varepsilon_t \quad (36)$$

O procedimento usual nos testes ADF consiste em testar a hipótese nula $\gamma = 0$ (existência de uma raiz unitária) para a variável em níveis (Y_t) e em primeiras diferenças (ΔY_t) contra a hipótese alternativa $|\gamma| < 0$ (série estacionária). Se a hipótese nula não for rejeitada em níveis mas o for em primeiras diferenças, então a variável $Y_t \sim I(1)$. Se a hipótese nula for apenas rejeitada em segundas diferenças, a variável $Y_t \sim I(2)$ e assim sucessivamente. A hipótese nula de uma raiz unitária ($H_0 : \rho - 1 = \gamma = 0$) é baseada no teste t com uma distribuição não normal:

$$t_\gamma = \frac{\hat{\rho} - 1}{S_{\hat{\rho}-1}} \quad (37)$$

Fuller (1976), Dickey e Fuller (1981) e mais tarde MacKinnon (1991, 1996) fornecem os valores críticos adequados a esta situação. Mas a verdadeira distribuição dos testes de Dickey-Fuller sob a hipótese nula é dada por:

$$t_{\rho}^{(u_0)} \sim \frac{W^2(1) - 1}{2 \sqrt{\int_0^1 W^2(s) ds}} \quad (38)$$

com $W(\cdot)$ a representar um processo de Wiener no intervalo $[0, 1]$.

Caso o valor absoluto de t calculado exceda o valor crítico, rejeita-se a hipótese nula e aceita-se a hipótese de estacionaridade para a série em causa.

✓ **Teste PP**

O teste PP diferentemente do teste ADF apresenta uma abordagem não paramétrica e a vantagem de permitir que os resíduos sejam autocorrelacionados, heterocedásticos, fracamente dependentes e/ou heterogeneamente distribuídos, o que suplanta as propriedades constituintes dos resíduos estritamente ruído branco.

Este teste pode ser visto como uma extensão do teste DF dado que assume as mesmas equações e os mesmos valores críticos no cálculo da estatística t . A raiz unitária é testada com base na estatística z , que é uma correção da estatística t incorporando as possibilidades de resíduos não autocorrelacionados e/ou heterocedásticos. Neste teste é rejeitada a hipótese nula (a série não é estacionária) nos casos em que a estatística de teste (estatística de P-Peron) apresente um valor absoluto superior ao valor crítico.

✓ Teste KPSS

Este teste apresenta-se mais vantajoso dada a falta de potência dos testes convencionais, isto é, tendem a não rejeitar a hipótese nula com demasiada frequência (Rua & Almas, 1998). Este teste contrariamente aos outros assume como hipótese nula que a série é estacionária e sob a hipótese alternativa a série é integrada de ordem um. A estatística do teste é dada pela equação 16.

3.2.2 Metodologia de Box-Jenkins

Para a realização das previsões, foram utilizados modelos univariados do tipo ARIMA (veja com mais detalhes o capítulo anterior) apresentados por Box & Jenkins (1976). A metodologia apresentada por aqueles autores permite a previsão de valores futuros de uma série tomando por base apenas seus valores passados, e também choques ocorridos no presente e no passado, considerando a correlação temporal existente entre os valores exibidos pela série. A aplicação desta metodologia foi realizada em 3 etapas (Nemec, 1994):

✓ Identificação

O primeiro passo neste método, é descobrir os valores apropriados de p , d e q (ou P , D e Q em caso de modelos sazonais). Para tanto foram elaboradas as FAC (que mostra o coeficiente de correlação entre uma observação no período t e as observações em $t-1, t-2, \dots, t-k$ sendo $k \geq 0$) e FACP (que apresenta a correlação entre Y_t e Y_{t+k} uma vez eliminado o efeito de $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$) da série.

✓ Estimação

O segundo passo é a estimação dos parâmetros dos termos AR e MA incluídos no modelo (apartir dos valores de p e q identificados). Para tanto, o método mais comumente usado para a estimação de parâmetros é o Método dos Mínimos Quadrados (do inglês *ordinary least square*-OLS)

✓ **Verificação de diagnóstico**

Depois de identificado o modelo e estimados os respectivos parâmetros, passou-se a verificação da adequabilidade do modelo. Para tanto, foi verificado se:

(i) ε_t possuem distribuição normal

(ii) $Var[\varepsilon_t]$ é constante

(iii) ε_t é uma série de ruído branco

Para os dois primeiros casos foi analisada a série das estimativas dos resíduos padronizados através do seu cronograma, histograma e do *Q-Q plot* normal. E para o terceiro caso foram analisadas as FAC e FACP empíricas dos resíduos (mais precisamente das suas estimativas). E de forma a testar se as k autocorrelações eram estatisticamente significativas foi usada a estatística Q de Ljung-Box (equação 32 com m graus de liberdade), no qual é rejeitada a hipótese nula (o modelo é adequado), caso o valor de Q calculado apresente-se superior ao valor do Q crítico.

3.2.3 Critérios de selecção de modelos

De forma a escolher entre os modelos cuja avaliação de diagnóstico foi positiva foram analisados os critérios 33, 34 e 35, que se baseiam nas estimativas dos valores dos resíduos. Estes critérios permitem a comparação de diferentes modelos para a série, e a consequente escolha do melhor modelo, o qual minimiza a variância dos resíduos (Barros, n.d).

Os critérios de Akaike e de Schwarz estabelecem que o modelo óptimo para uma mesma série é aquele que minimiza as quantidades CIA e CIS respectivamente.

O terceiro critério (EQM) foi usado para comparar a qualidade da previsão (ou ajuste) para a mesma série em diferentes modelos; quanto menor o erro, melhor a habilidade preditiva do modelo.

4 Discussão dos resultados

Os resultados das análises deste trabalho são apresentados e discutidos neste capítulo. As estimativas dos modelos de séries temporais escolhidos, bem como as previsões geradas a partir dos mesmos também são apresentados neste capítulo. No presente trabalho optou-se pela estimação de modelos para cada região do país (Sul, Centro e Norte) referentes aos casos de infecção do HIV dado o facto de cada região ter suas peculiaridades.

4.1 Análise descritiva

De acordo com o objectivo do trabalho, pretende-se prever os casos de infecção de HIV nas 3 regiões do país, ou seja, prever a evolução dos casos de infecção de HIV nas regiões Sul, Centro e Norte. Para isso dispõe-se de 3 séries cronológicas que relacionam o histórico de casos de infecção mensais desde Agosto de 2003 até Junho de 2008, com o intuito de prever os casos de infecção para os meses de Julho a Dezembro do ano de 2008.

Da tabela 3. é possível observar que:

- **Região Sul**

As 59 observações apresentaram uma média mensal de casos de infecção de HIV de cerca de 4800 com uma variabilidade de aproximadamente 556 casos. O menor número de casos de infecção observado durante o período em análise foi de 3281 e um máximo de 5431 casos. De salientar ainda que 50% das observações mensais analisadas apresentaram casos equivalentes a 4998. Importa acrescentar ainda que durante o período de Agosto/2003 a Junho/2008, nesta região verificou-se uma variação mensal dos casos de infecção de HIV de cerca de 0.4% (Tabela A1 em anexo)

- **Região Centro**

A região Centro apresentou casos médios mensais de cerca de 4760 com um desvio de quase 449 casos. O máximo de casos observado em um mês durante o período em análise foi de 5362 casos e um mínimo de 3807 casos. Por outro lado, 50% das observações apresentaram casos de infecção de HIV equivalentes a 4893. Acrescentar ainda que foi verificada uma variação mensal de cerca de 0.7% (superior ao verificado na zona Sul) durante o período em análise (Vide Tabela A1 em anexo). Esta constatação pode ser devida ao facto de esta mesma região apresentar o

maior número de ATS's comparativamente as restantes regiões ou mesmo o facto de fazer fronteira com países que apresentam os maiores índices de infectados na zona Austral como é o caso do Zimbabwe.

- **Região Norte**

E por fim, a região Norte apresentou um número médio mensal de casos de infecção de cerca de 412 com um desvio padrão de quase 287. Ainda é possível constatar que o menor número de casos observados foi de 147 com um máximo de 1114 casos. Das 59 observações, 50% delas apresentaram um valor inferior a média mensal (237 casos). De referir que esta região diferentemente das restantes, apresentou uma variação mensal superior a 3% durante o período de Agosto/2003 a Junho/2008, mas apresentando os menores casos de infecção entre as três regiões do país, facto este que pode ser justificado pelos baixos índices de infecção nos países vizinhos (Veja tabela A1 em anexo).

Tabela 3. Medidas descritivas dos casos de infecção de HIV nas regiões Sul, Centro e Norte de Agosto/2003 a Junho/2008

	CASOSCENRO	CASOSNORTE	CASOSSUL
Média	4760.119	412.0678	4809.763
Mediana	4893.000	237.0000	4998.000
Máximo	5362.000	1114.000	5431.000
Mínimo	3807.000	147.0000	3281.000
Desvio padrão	448.9371	287.3617	556.2778
Observações	59	59	59

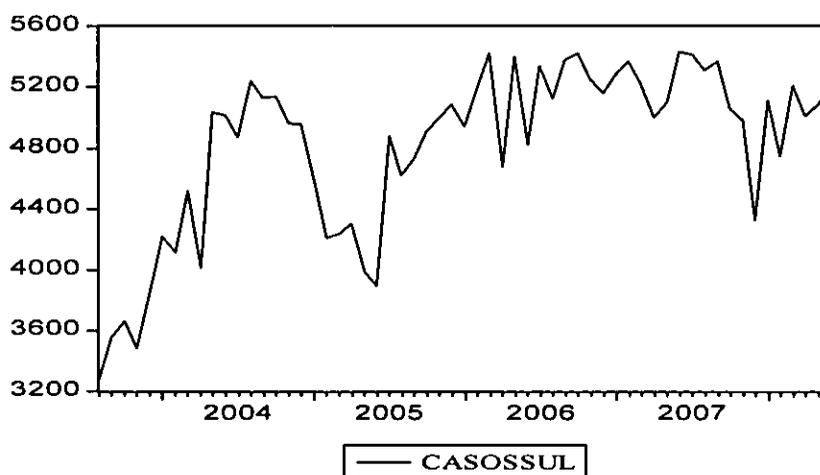
4.2 Metodologia de Box-Jenkins para a região Sul

Conforme foi apresentado no capítulo 2, os dados serão submetidos a um ciclo iterativo, cujo processo consiste em aplicar aos dados a três etapas distintas, que são: identificação, estimação e verificação de diagnóstico.

4.2.1 Etapa de identificação

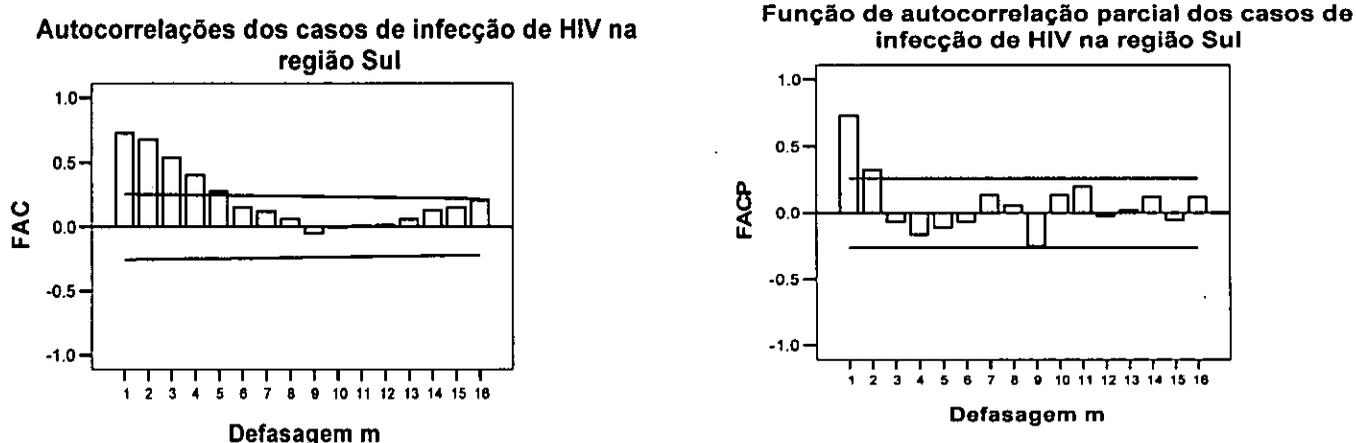
Ao se analisar uma ou mais séries temporais a representação gráfica dos dados sequencialmente ao longo do tempo é fundamental e pode revelar padrões de comportamento importantes. Tendências de crescimento (ou decrescimento), alterações estruturais, observações aberrantes, etc. são muitas vezes facilmente identificados. Sendo assim, o gráfico temporal deve ser sempre o primeiro passo e antecede qualquer análise. Observando a figura 2 é possível notar uma tendência (crescente) da série, apresentando uma redução dos casos de infecção no mês de Julho de 2007 e voltando a elevar-se nos meses subsequentes indiciando não estacionaridade da mesma. De referenciar que do gráfico não se notam regularidades sazonais.

Figura 2. Série cronológica dos casos de infecção de HIV na região Sul



Feito isto, o passo a seguir é a análise das FAC (função autocorrelação) e FACP (função autocorrelação parcial) cujos comportamentos devem imitar os comportamentos das respectivas quantidades teóricas (tabelas 1 e 2) que irão auxiliar na verificação da estacionaridade e na definição do modelo mais adequado. A figura 3 apresenta as FAC e FACP estimadas da sucessão dos casos de infecção de HIV na região Sul revelando um decaimento algo lento para zero (FAC) e alguns picos significativos na FACP corroborando com o constatado na análise gráfica indiciando presença de raiz unitária (não estacionaridade), havendo portanto necessidade de calcular diferenças para neutralizar a tendência. Neste estágio, a utilização de testes para verificar a existência de raízes unitárias pode ser de grande utilidade.

Figura 3. FAC e FACP dos casos de infecção de HIV na região Sul



4.2.1.1 Testes de raiz unitária

Analisando a tabela 4 que contém os testes de raiz unitária descritos no capítulo 2 (ADF, PP e KPSS), constata-se que nos testes ADF e PP a hipótese nula não é rejeitada em níveis mas é rejeitada em primeiras diferenças ao nível de significância de 5%, conclusão contrária no teste KPSS, ou seja, a hipótese nula não é rejeitada ao mesmo nível de significância. No entanto a série em análise é considerada integrada de ordem 1, isto é, ela só se torna estacionária após efectuada a primeira diferença (Nos três testes foram inclusas a constante e uma tendência determinística)

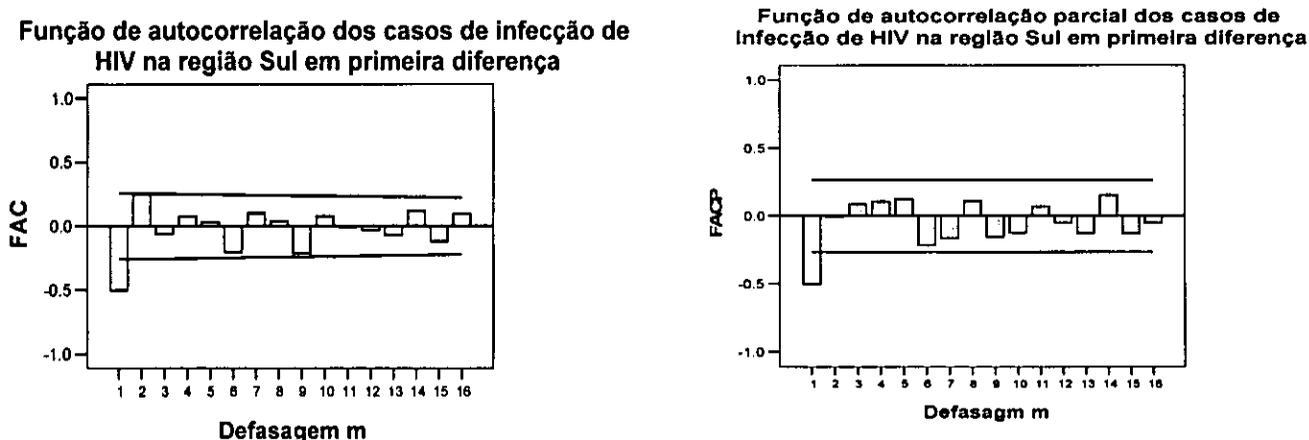
Tabela 4. Região sul. Testes de raiz unitária para a série em níveis e em primeiras diferenças

	ADF	PP	KPSS
CASOSSUL	-2.592893	-3.303575	0.650412
Δ CASOSSUL	-13.07208	-13.14409	0.056638

Nota: ADF: -3.496962 (5%), PP: -3.489228 (5%) e KPSS: 0.46300 (5%) em níveis e ADF: -3.490662(5%), PP: -3.490662 (5%) e KPSS: 0.146000 (5%) em 1^{as} diferenças

Alcançada a estacionaridade da série e da análise das FAC e FACP da série já diferenciada observa-se que (figura 4):

Figura 4. FAC e FACP dos casos de infecção de HIV na região Sul em primeira diferença



- A FAC apresenta um comportamento que decai exponencialmente rápido e apresenta autocorrelações significativas nos dois primeiros lag's (uma positiva e outra negativa);
- A FACP apresenta um decaimento mais brusco com autocorrelação negativa e significativa no 1ºlag

Neste caso, as evidências anteriores sugerem que o modelo a ser adoptado pode ser o ARIMA (1,1,0) ou ARIMA (0,1,1)

Os demais modelos não relacionados neste trabalho ou apresentavam coeficientes não significativos ou os resíduos não se comportavam em forma de ruído branco.

4.2.2 Etapa de estimação

Após verificada a ordem de integração d e os seus respectivos parâmetros p e q , passa-se para a estimativa dos parâmetros dos modelos propostos (sem a constante dado a mesma não ser significativa).

Como os modelos propostos apresentaram-se primeiramente como não estacionários e foi necessário tomar-se a 1ª diferença da série original para torná-la estacionária, isso permite que seja usado o modelo dado pela equação 26 como apresentado nas tabelas 4 e 5.

Tabela 5. Modelo ARIMA (1,1,0). Região Sul

Variável dependente: $\Delta Casos_{sul}$

Variável	Coefficiente	Std. Error	t-Statistic	Prob.
AR(1)	-0.492557	0.115584	-4.261457	0.0001

Fonte: Output do EViews

Tabela 6. Modelo ARIMA (0,1,1). Região Sul

Variável dependente: Δ Casossul

Variável	Coefficiente	Std. Error	t-Statistic	Prob.
MA(1)	-0.390279	0.121352	-3.216099	0.0021

Fonte: Output do EViews

As tabelas 5 e 6 mostram para além dos parâmetros estimados, o cálculo das estatísticas t, que medem a significância estatística dos mesmos, para serem comparadas com o valor de 1.96 que corresponde ao nível assintótico da distribuição normal a 95% de significância. Avaliando as duas tabelas pode se constatar que a um nível de significância de 5% não se rejeita a hipótese de que o verdadeiro parâmetro populacional seja igual ao apresentado nas tabelas, tanto para o modelo ARIMA (1,1,0) assim como para o modelo ARIMA (0,1,1)⁵.

4.2.3 Etapa de verificação do diagnóstico

Esta etapa consiste em verificar se o modelo identificado e estimado é adequado. Se o modelo estiver adequadamente especificado, os resíduos do modelo estimado serão estimativas de ruído branco e, assim, os coeficientes de autocorrelação dos resíduos devem ser estatisticamente iguais a zero. Caso o modelo não seja adequado, será necessário identificar outro modelo e identificar as etapas de estimação e verificação (Costa, 2005).

Para simplificação e para evitar repetição de procedimentos, nesta etapa, dos dois modelos identificados, apenas será apresentada a figura de um modelo e analisada profundamente.

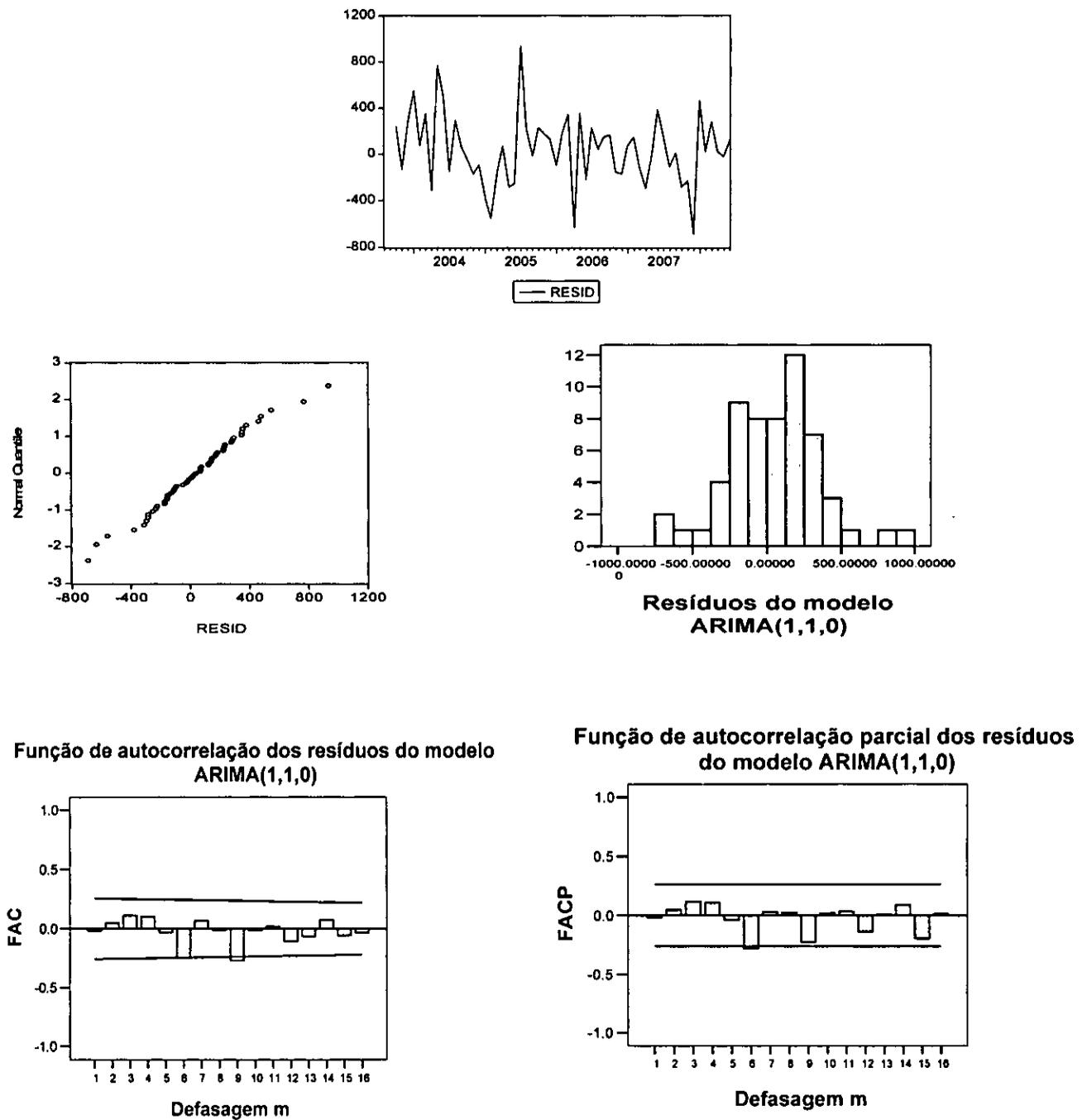
A série das estimativas dos resíduos do modelo ARIMA (1,1,0) ajustado é caracterizada na figura 5, sendo apresentados o cronograma, o histograma, o *Q-Q plot* normal e as FAC e FACP empíricas. Da análise da informação contida nessa figura conclui-se que:

- Não são revelados desvios da série das estimativas dos resíduos do modelo ajustado relativamente à hipótese de os resíduos possuírem distribuição normal e variância constante;
- A série das estimativas dos resíduos do modelo ajustado não aparenta qualquer desvio relativamente à nulidade de valores individuais da FAC e FACP;

⁵ Ho: $\beta=0$

H1: $\beta \neq 0$ Caso $t_{calculado} > t_{critico}$, rejeita-se a hipótese nula segundo a qual o valor estimado não é o verdadeiro parâmetro populacional

Figura 5. Cronograma, *Q-Q plot* normal, histograma, FAC e FACP da série dos resíduos do modelo ARIMA (1,1,0) para a série dos casos de infecção de HIV na região Sul.



De forma a complementar a análise, a série da estimativa dos resíduos do modelo ajustado foi usada para efectuar o teste Q de Ljung-Box, tendo gerado os seguintes resultados:

Tabela 7. Teste de Ljung-Box para os resíduos da série dos casos de infecção de HIV na região Sul

Lag	Autocorrelation	Std.Error	Box-Ljung Statistic		
			Value	df	Sig.
12	-.104	.115	12.212	12	.429
13	-.063	.114	12.521	13	.485
14	.072	.112	12.936	14	.532
15	-.055	.111	13.180	15	.588
16	-.033	.110	13.271	16	.653

A comparação do valor de teste com o valor tabelado da distribuição Qui-quadrado com 14 graus de liberdade ($m=15$, $p=1$, $q=0$), revela que os resíduos podem ser considerados ruído branco, pois o teste não rejeita a hipótese de erros não correlacionados. ($Q_{\text{calculado}} = 13.180 < Q_{\text{crítico}} = 23.6848$).

A tabela 8 apresenta uma comparação dos modelos propostos ARIMA (1,1,0) e ARIMA (0,1,1) para os casos de infecção de HIV na região Sul, sendo usados para tal os critérios de Akaike e de Schwarz e o critério do erro quadrático médio.

Em face dos resultados expostos é natural escolher o modelo ARIMA (1,1,0), dado o mesmo apresentar os menores valores de CIA, CIS e EQM quando comparado com o outro modelo alternativo. A escolha do modelo ARIMA (1,1,0) é reforçada pela avaliação do valor do desvio padrão residual que revela que o modelo escolhido apresentou o menor desvio padrão residual em relação ao modelo ARIMA (0,1,1).

Tabela 8. Critérios de comparação para verificação do melhor modelo

Modelo	CIA	CIS	Desvio padrão residual	EQM
ARIMA (0,1,1)	14.40579	14.44132	322.2771	100341.76
ARIMA (1,1,0)	14.34393	14.37978	312.4155	92640.56

Como o modelo ARIMA (1,1,0), foi o que melhor se apresentou para modelar a série que representa os casos de infecção de HIV na região Sul, o mesmo foi usado para fazer previsões com base na

equação apresentada abaixo, para o período compreendido entre Julho e Dezembro de 2008, como mostra a tabela 9:

$$\Delta Casossul_t = -0.492557 \Delta Casossul_{t-1}$$

$$Casossul_t - Casossul_{t-1} = -0.492557(Casossul_{t-1} - Casossul_{t-2})$$

$$Casossul_t = 0.507443 Casossul_{t-1} + 0.492557 Casossul_{t-2}$$

$$Casossul_{2008/7} = 0.507443 * 5175 + 0.492557 * 5085 = 5131$$

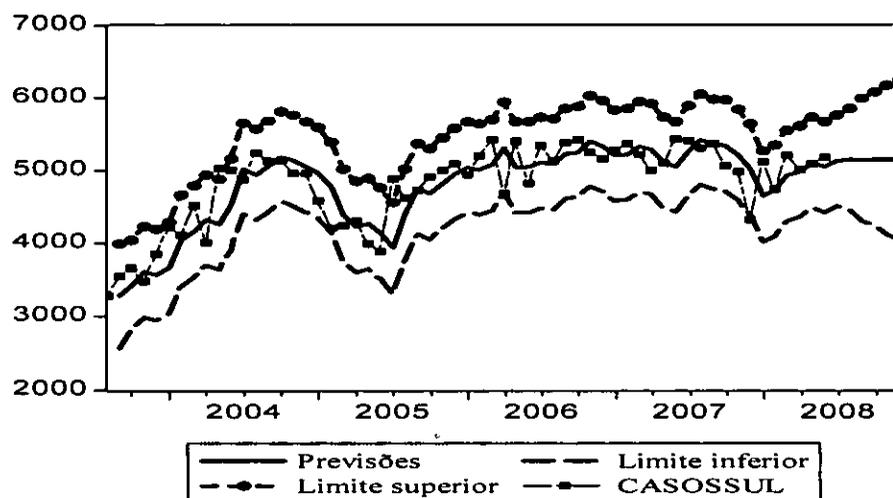
Nota: Os restantes valores são obtidos substituindo os valores correspondentes, na mesma equação

Tabela 9. Previsões de casos de infecção de HIV na região Sul

Período	Previsão	L. Inferior	L. Superior
Julho	5131	4508	5754
Agosto	5153	4453	5852
Setembro	5142	4300	5984
Outubro	5147	4217	6077
Novembro	5145	4120	6169
Dezembro	5146	4041	6250

O gráfico com as previsões e o intervalo de 95% de confiança para as previsões é apresentado a seguir (Figura 6). Da tabela 9 e da figura 6 é possível observar-se uma tendência crescente dos casos de infecção de HIV na região Sul durante o segundo semestre de 2008.

Figura 6. Previsões dos casos de infecção de HIV na região Sul

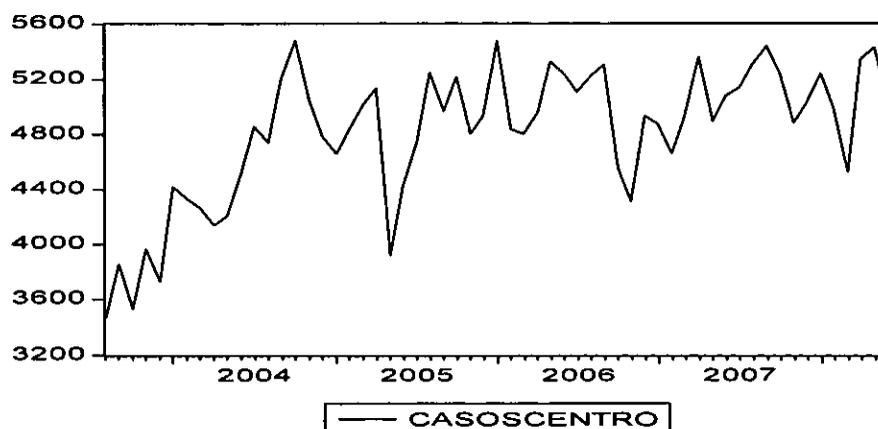


4.3 Metodologia de Box-Jenkins para a região Centro

4.3.1 Etapa de identificação

Para identificar o modelo apropriado, inicialmente, foi analisada a figura 7 que apresenta o gráfico de tempo dos casos de infecção de HIV na região Centro. A análise dessa figura indica uma fraca tendência crescente (a partir de 2004) com uma elevação e uma redução nos meses de Maio e Março de 2004 e 2005 respectivamente, indiciando uma estacionaridade da série. Da mesma é possível a não verificação de indícios de sazonalidade.

Figura 7. Série cronológica dos casos de infecção de HIV na região Centro



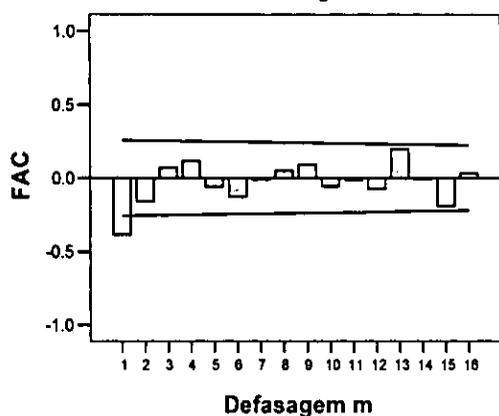
De forma a complementar a análise da série, foram analisadas as FAC e FACP amostrais. O comportamento dessas funções indica qual o modelo a ser utilizado, bem como auxilia no uso dos testes de raízes unitárias para confirmar a estacionaridade.

A figura 8, apresenta o correlograma e o correlograma parcial da série temporal dos casos de infecção de HIV na região Centro das quais constata-se que:

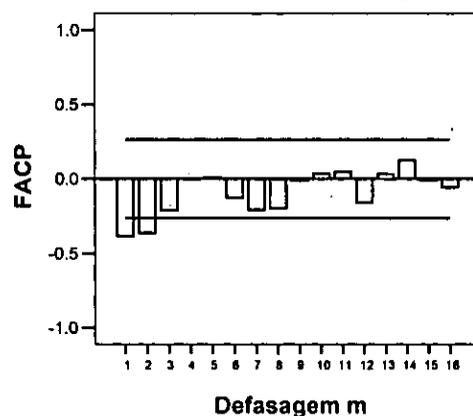
- A FAC da série temporal em estudo revela um decaimento exponencial rápido e apresenta correlações estatisticamente significativamente diferentes de zero nos primeiros 4 lag's;
- Na FACP da série, observa-se que a autocorrelação parcial na primeira defasagem é significativa, indicando um modelo de primeira ordem.

Figura 8. FAC e FACP dos casos de infecção de HIV na região Centro

Função de autocorrelações dos casos de infecção de HIV na região Centro



Função de autocorrelações parciais dos casos de infecção de HIV na região Centro



De acordo com os factos constatados nas FAC e FACP estimadas, a série parece sugerir estacionaridade em nível. De forma a complementar a análise, foram realizados os testes de raiz unitária.

4.3.1.1 Testes de raiz unitária

Para verificar a existência de uma raiz unitária, foram utilizados os testes ADF, PP e KPSS (com a constante incluída em cada teste), sendo os resultados dos mesmos apresentados na tabela 10.

Tabela 10. Região Centro. Resultados dos testes ADF, PP e KPSS

	ADF	PP	KPSS
CASOSENTRO	-4.567762	-4.448958	0.175002

Nota: valores Críticos: ADF: -2.912631(5%), PP: -2.912631(5%) e KPSS: 0.46300(5%)

Os resultados dos testes fornecem evidências de que a série temporal em estudo não possui raiz unitária a um nível de significância de 5%. Visto que a série pode ser considerada estacionária, a ordem de integração é zero, isto é, $d=0$.

Feitas as análises às FAC e FACP estimadas e constatada a estacionaridade, dois modelos possíveis foram identificados: o modelo ARIMA (1,0,1) \Leftrightarrow ARMA (1,1) e o modelo ARIMA (1,0,0).

4.3.2 Etapa de estimação

Uma vez indicados os valores de p , d e q , passa-se para a estimação dos parâmetros do modelo proposto. Os parâmetros estimados para os dois modelos identificados são apresentados nas tabelas 11 e 12.

Tabela 11. Modelo ARIMA (1,0,0).Região Centro

Variável dependente: Casoscentro

Variável	Coefficiente	Std. Error	t-Statistic	Prob.
r	4882.245	127.4231	38.31524	0.0000
AR(1)	0.646740	0.090079	7.179710	0.0000

Fonte: EViews

Tabela 12. Modelo ARIMA (1,0,1). Região Centro

Variável dependente: Casoscentro

Variável	Coefficiente	Std. Error	t-Statistic	Prob.
r	5044.633	26.91937	187.3979	0.0000
AR(1)	0.0883823	0.020444	43.23176	0.0000
MA(1)	-0.997459	0.058538	-17.03945	0.0000

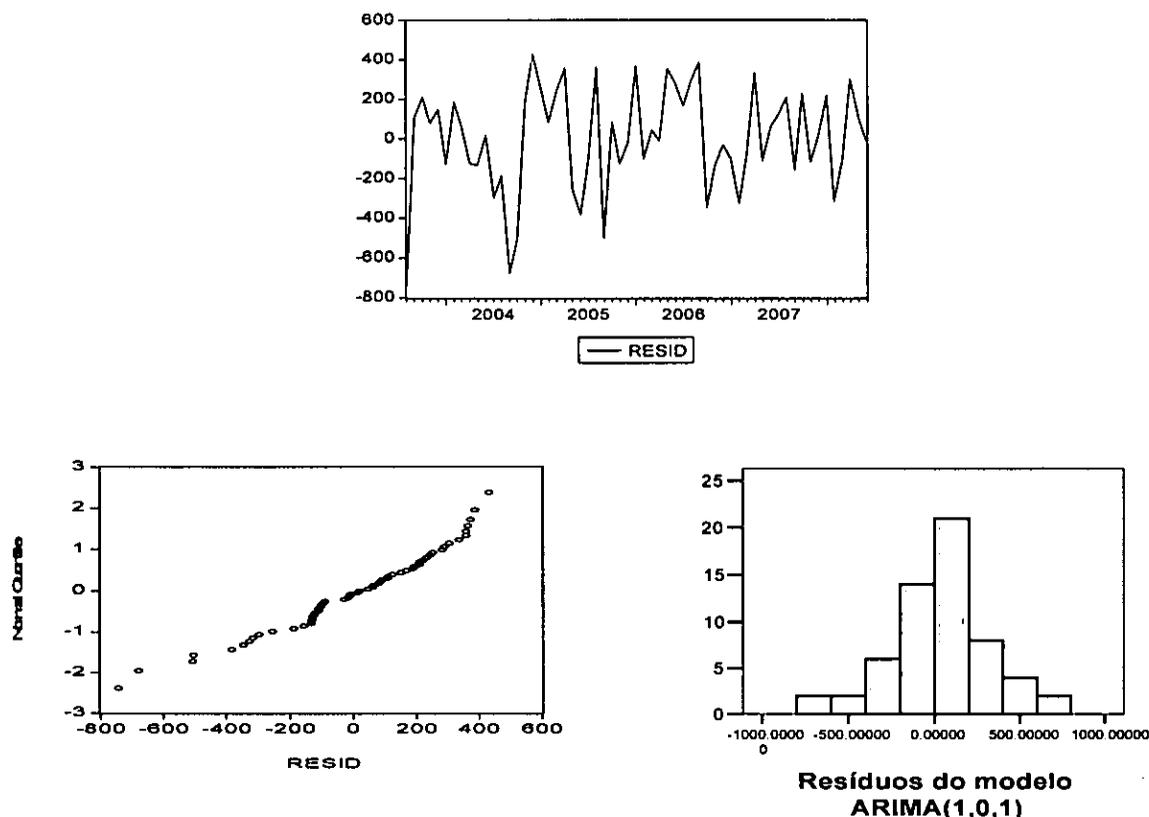
Fonte: EViews

Nos três modelos propostos, a um nível de significância de 5% não se rejeita a hipótese de que os verdadeiros parâmetros populacionais sejam os dados nas tabelas 11 e 12, observando-se uma significância muito forte dos parâmetros em todos os modelos em análise.

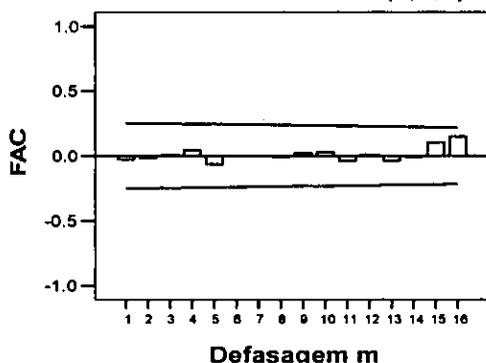
4.3.3 Etapa de verificação de diagnóstico

Para que o modelo seja considerado adequadamente especificado é necessário que os resíduos do modelo estimado $\hat{\varepsilon}_t$ sejam estimativas de ruído branco, isto é, suas autocorrelações devem se comportar de maneira aleatória e devem ser não significativas (Nunes, 2004). Para testar essa suposição, verificou-se o comportamento cronograma, do histograma, o *Q-Q plot* e as FAC e FACP empíricas. A análise da **figura 9** revela que os resíduos apresentam comportamento aleatório, ou seja, não revelam padrão específico. Assim, a análise dos resíduos confirma a adequação do modelo.

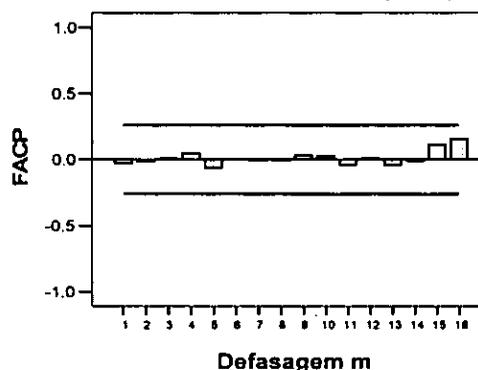
Figura 9. Cronograma, *Q-Q plot* normal, histograma, FAC e FACP dos resíduos do modelo ARIMA (1,0,1) para a série dos casos de infecção de HIV na região Centro.



Função de autocorrelação dos resíduos do modelo ARIMA(1,0,1)



Função de autocorrelação parcial dos resíduos do modelo ARIMA(1,0,1)



O uso do teste de Ljung-Box reforça essa afirmação. Esse teste compara o valor de Q calculado com o valor Q crítico com m-p-q graus de liberdade. Aplicando o teste para m =16 sobre os erros estimados pela aplicação do modelo ARMA (1,1), obtém-se uma estatística de teste Q=18.70, que não é significativa quando comparada ao valor Q crítico com m-p-q=14 graus de liberdade (23.6848).

Tabela 13. Teste de Ljung-Box para os resíduos da série dos casos de infecção da região Centro

Lag	Autocorrelações	Std.Error	Box-Ljung Statistic		
			Value	df	Sig.
12	.000	.114	8.059	12	.780
13	.229	.113	12.148	13	.516
14	-.025	.112	12.197	14	.590
15	-.276	.111	18.420	15	.241
16	-.058	.109	18.700	16	.285

Assim, pela análise do teste de Ljung-Box os resíduos podem ser considerados ruído branco. A análise se estendeu para o modelo ARIMA (1,0,0) tendo se verificado a adequabilidade do mesmo. A tabela 14 apresenta uma comparação entre os dois modelos propostos para os casos de infecção de HIV na região Centro de forma a escolher dentre eles, o que melhor ajusta-se aos dados.

Tabela 14. Critérios de comparação para escolha do melhor modelo

Modelos	CIA	CIS	Desvio padrão residual	EQM
ARIMA (1,0,0)	14.52321	14.59425	307.9960	90038.05
ARIMA (1,0,1)	14.31984	14.42642	254.4432	60352.12

Dos resultados constantes da tabela 14, conclui-se que o modelo ARIMA (1,0,1) é de facto o melhor modelo entre vários possíveis, dado o facto do mesmo ter apresentado os menores valores de Akaike, de Schwarz, do erro quadrático médio (EQM) e o menor desvio padrão residual.

Feitos todos os testes e encontrado o melhor modelo, o mesmo foi usado para prever os casos de infecção de HIV na região Centro para os últimos 6 meses do ano 2008 (vide tabela 15 e um simples exemplo de obtenção dos resultados que constam da tabela). O gráfico com as previsões e o intervalo de 95% de confiança para as previsões é apresentado na figura 10.

$$Casoscentro_t = 4624.516 + 0.970 * Casoscentro_{t-1} + 0.593 * \varepsilon_{t-1}$$

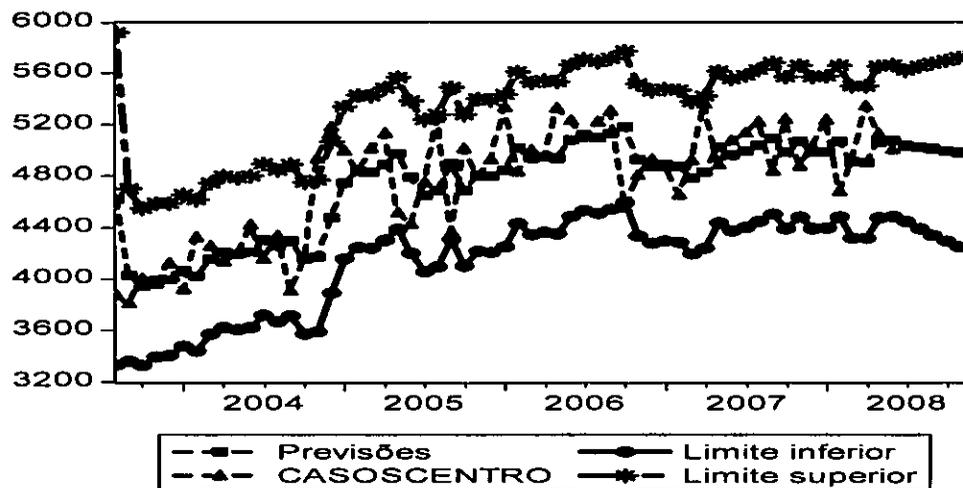
$$Casoscentro_{2008/7} = 5036$$

Analisando os valores previstos dos casos de infecção de HIV na região Centro, dos mesmos constata-se que durante o período de previsão se observa uma tendência decrescente dos casos.

Tabela 15. Previsões de casos de infecção de HIV na região Centro

Período	Previsão	L. Inferior	L. Superior
Julho	5036	4447	5625
Agosto	5023	4392	5655
Setembro	5011	4342	5680
Outubro	4999	4296	5703
Novembro	4988	4253	5723
Dezembro	4977	4213	5741

Figura 10. Previsões dos casos de infecção de HIV na região Centro



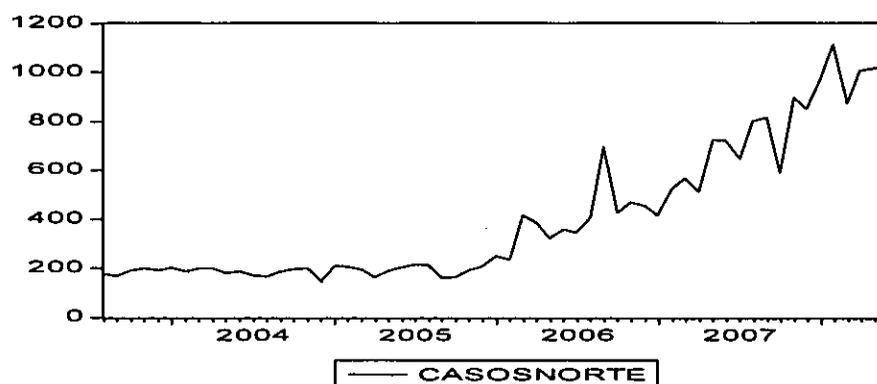
4.4 Metodologia de Box-Jenkins para a região Norte

4.4.1 Etapa de identificação

Para identificar os modelos apropriados, inicialmente deve ser analisada a série temporal em estudo. A análise da série pode indicar a presença ou ausência de tendência o que revelaria se a série é ou não estacionária.

O gráfico da série temporal mostrada na **figura 11** representa os casos de infecção de HIV na região Norte. Nele é possível observar uma elevação dos casos nos meses de Agosto de 2006 e Janeiro de 2008 e uma tendência (crescente) que revela uma série não estacionária. Do gráfico é ainda possível constatar a não verificação de regularidades sazonais.

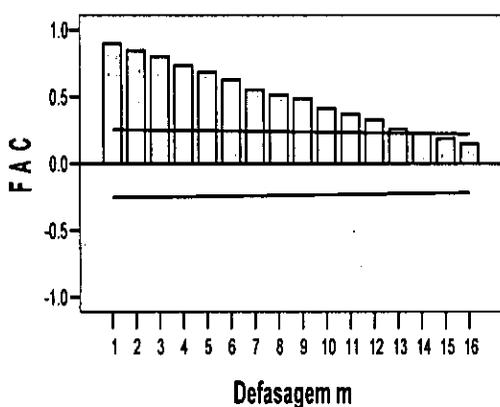
Figura 11. Série cronológica dos casos de infecção de HIV na região Norte



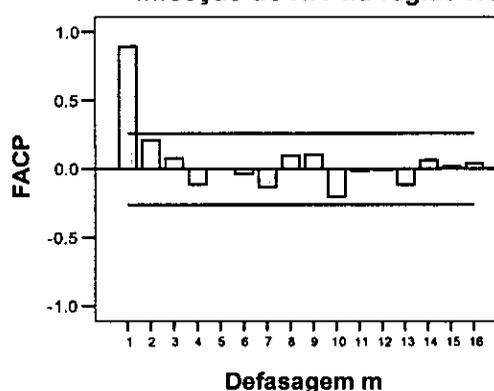
Para se prever uma série temporal através dos modelos ARIMA, torna-se necessário identificar a ordem dos parâmetros p , d e q . O primeiro parâmetro a ser identificado é o grau de diferenciação d necessário à estabilização dos dados. Isto é feito através de um exame do correlograma, ou seja, da FAC e FACP estimadas (Figura 12). Podemos notar das mesmas (FAC e FACP) que as autocorrelações decrescem de forma exponencial e muito lentamente, corroborando com análise gráfica feita anteriormente (a série sugere não estacionaridade), havendo necessidade de tomar diferenças sucessivas por forma a que ela se estabilize e descobriremos a ordem de diferenciação d do modelo.

Figura 12. FAC e FACP dos casos de infecção de HIV na região Norte

Função de autocorrelação dos casos de infecção de HIV na região Norte



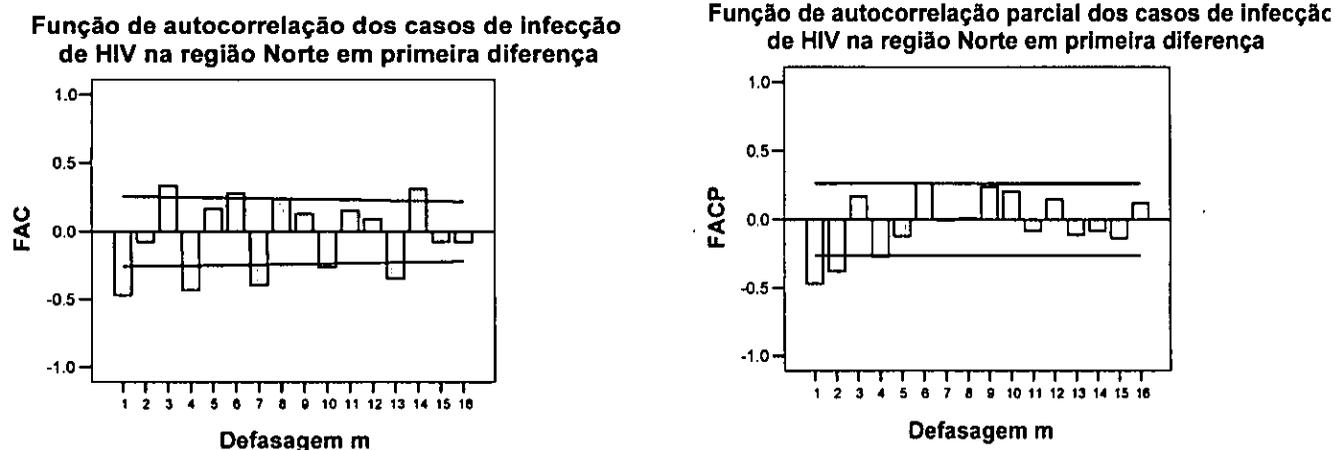
Função de autocorrelação parcial dos casos de infecção de HIV na região Norte



Após calculada uma diferença dos dados originais, da análise das FAC e FACP nota-se que (Figura 13):

- A FAC possui um decaimento rápido apresentando algumas correlações estatisticamente significativamente diferentes de zero;
- Na FACP observa-se que apenas as autocorrelações das defasagens 1, 2, 4 e 6 são significativas.

Figura 13. FAC e FACP dos casos de infecção de HIV na região Sul em primeiras diferenças



Das análises feitas às FAC e FACP estimadas, a série sugere uma estacionaridade em primeiras diferenças.

São de seguida analisados os testes de raiz unitária por forma a complementar o processo de identificação do modelo.

4.4.1.1 Testes de raiz unitária para a série em níveis e em primeiras diferenças

De acordo com os resultados dos testes ADF, PP e KPSS constantes da tabela 16 é possível verificar que a série em estudo só se torna estacionária após aplicada a primeira diferença aos dados originais. Isto é, nos testes ADF e PP só se rejeita a hipótese de existência de raiz unitária (hipótese nula) após ter sido feita uma diferença aos dados originais e no teste KPSS a hipótese nula (a série é estacionária) só não é rejeitada depois da série original ter sido diferenciada, havendo no entanto uma corroboração entre todos os testes.

Tabela 16. Resultados dos testes ADF, PP e KPSS para a série em níveis e em 1ª diferença

	ADF	PP	KPSS
CASOSNORTE	-0.187186	-2.688372	0.238236
Δ CASOSNORTE	-6.742725	-15.10616	0.069021

Nota: ADF: -3.496960 (5%), PP: -3.489228 (5%) e KPSS: 0.46300 (5%) em níveis e ADF: -3.490662(5%), PP: -3.490662 (5%) e KPSS: 0.146000 (5%) em 1ªs diferenças

Dos factos acima relatados chega-se a conclusão de que a ordem de integração d é um, isto é, $d=1$. Sendo assim, dois modelos são propostos: o modelo ARIMA (0,1,1) e o modelo ARIMA (4,1,1).

4.4.2 Etapa de estimação

Uma vez indicados os valores de p , d e q , passa-se para a estimação dos parâmetros do modelo proposto. Encontrou-se, portanto, para o modelo ARIMA (0,1,1) como estimativas significativas o coeficiente não sazonal $\beta_1 = -0.574401$ e a constante $r = 14.93876$ (que também representa o nível médio do processo) e para o modelo ARIMA (4,1,1) verificou-se como estimativas significativas os seguintes: $\alpha_4 = -0.363516$, $\beta_1 = -0.459725$ e $r = 15.78769$. As tabelas 17 e 18 apresentam de forma mais detalhada todos os aspectos relacionados aos parâmetros estimados.

Tabela 17. Modelo ARIMA (0,1,1). Região Norte

Variável dependente: Δ Casosnorte

Variável	Coefficiente	Std. Error	t-Statistic	Prob.
r	14.93876	4.850083	3.080104	0.0032
MA(1)	-0.574401	0.109216	-5.259316	0.0000

Tabela 18. Modelo ARIMA (4,1,1). Região Norte

Variável dependente: Δ Casosnorte

Variável	Coefficiente	Std. Error	t-Statistic	Prob.
r	15.78769	4.594923	3.435899	0.0012
AR(4)	-0.363516	0.140014	-2.596286	0.0123
MA(1)	-0.459725	0.129069	-3.561852	0.0008

Avaliando os dois modelos estimados, podemos notar que todos os parâmetros são estatisticamente significativos a um nível de significância de 5%, podendo considerar desta forma que os parâmetros estimados não difiram dos verdadeiros parâmetros populacionais.

4.4.3 Etapa de verificação de diagnóstico

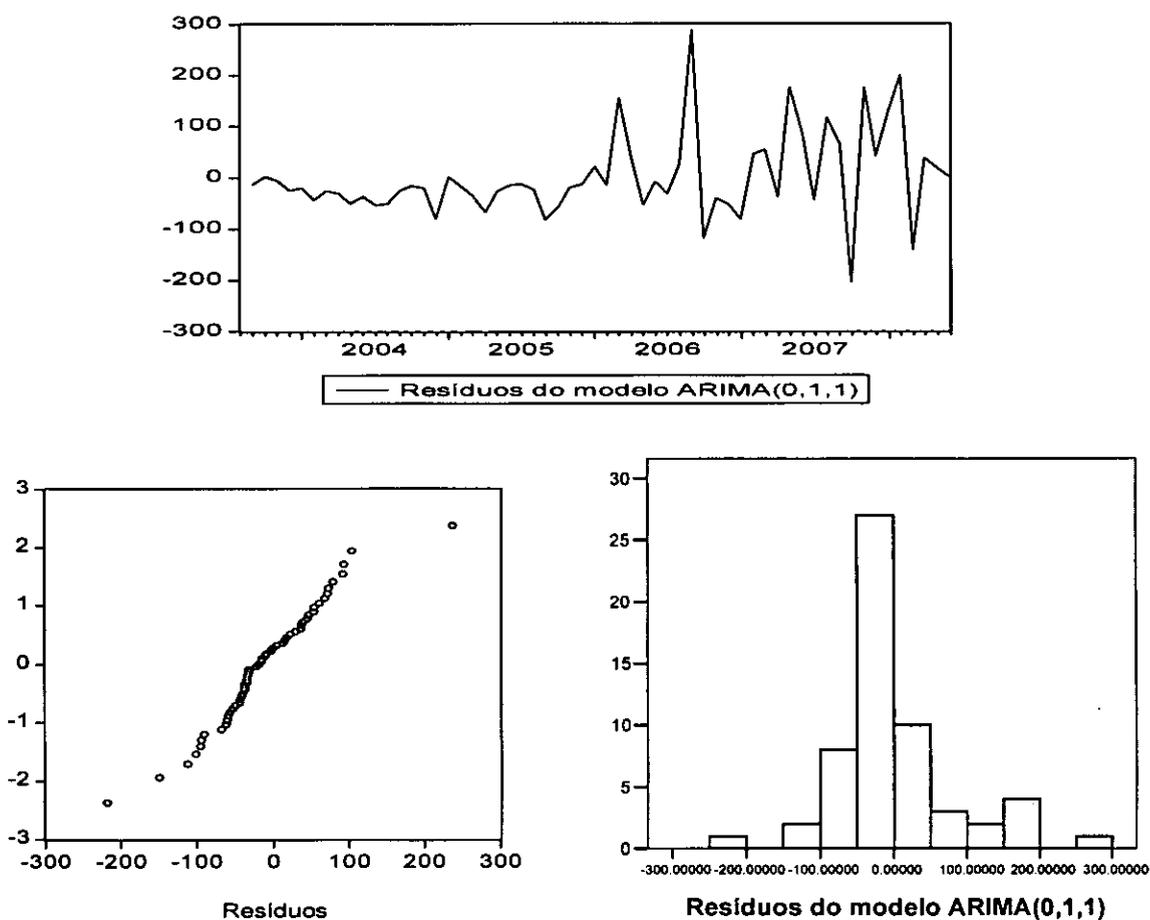
Após estimar os modelos é necessário verificar se ele representa, ou não, adequadamente, os dados. Para tanto a série das estimativas dos resíduos do modelo ARIMA (0,1,1) ajustado é caracterizada na

figura 14, sendo apresentados para a mesma: um cronograma, um histograma, o *Q-Q plot* e as FAC e FACP estimadas.

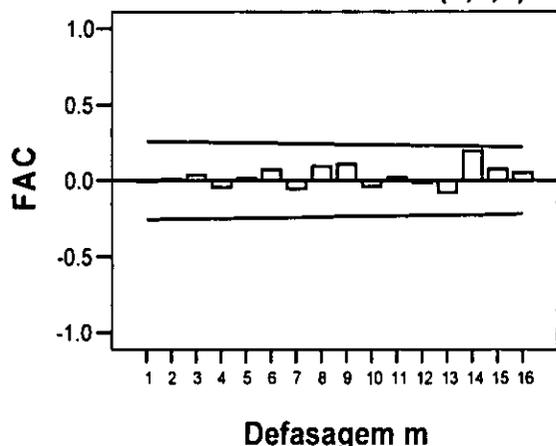
Dessa figura constata-se que:

- da análise do cronograma da série das estimativas dos resíduos padronizados, do histograma e do *Q-Q plot* não se revelam desvios acentuados relativamente ao modelo normal;
- as FAC e FACP estimadas dos resíduos (mais precisamente das suas estimativas) dão impressão de que os resíduos estimados são puramente aleatórios.

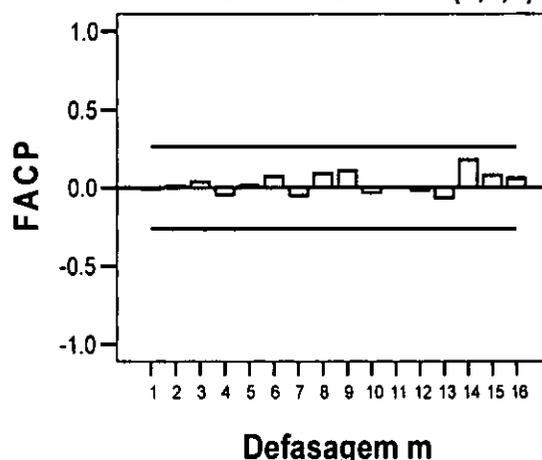
Figura 14. Cronograma, *Q-Q plot* normal, histograma, FAC e FACP dos resíduos do modelo ARIMA (0,1,1) para a série dos casos de infecção de HIV na região Norte



Função de autocorrelação dos resíduos do modelo ARIMA(0,1,1)



Função de autocorrelação parcial dos resíduos do modelo ARIMA(0,1,1)



O teste de Ljung-Box aplicado às séries das estimativas dos resíduos dos dois modelos confirma o verificado nas análises anteriormente feitas, ou seja, os resíduos do modelo estimado $\hat{\varepsilon}_t$ são estimativas de ruído branco. $Q_{calculado} = 6.388 < Q_{crítico} = 24.9958$ ($m=16, p=0$ e $q=1$).

Tabela 19. Teste de Ljung-Box para os resíduos da série dos casos de infecção de HIV da região Norte

Lag	Autocorrelation	Std.Error	Box-Ljung Statistic		
			Value	df	Sig.
12	-.015	.115	2.288	12	.999
13	-.079	.114	2.767	13	.999
14	.193	.112	5.726	14	.973
15	.074	.111	6.171	15	.977
16	.051	.110	6.388	16	.983

Pelo exposto acima, a escolha entre modelos alternativos (concorrentes) estimados será feita tomando por base, conjuntamente, os seguintes elementos enquadrados na metodologia Box-Jenkins: CIA, CIS, EQM e o desvio padrão residual.

Analisando esses elementos (patentes na tabela 20) constata-se que o modelo ARIMA (0,1,1) é que apresentou os melhores resultados (os menores valores de CIA, CIS, EQM e do desvio padrão

residual) em comparação com o modelo ARIMA (0,1,1), podendo desta forma ser usado para fazer previsões que é o principal objectivo da metodologia Box-Jenkins (Gujarati, 2004).

Tabela 20. Critérios de comparação para escolha do melhor modelo

Modelos	CIA	CIS	Desvio padrão residual	EQM
ARIMA (0,1,1)	11.74484	11.81589	83.64010	60471.02
ARIMA (4,1,1)	11.74488	11.85537	84.48240	67743.63

Como o modelo ARIMA (0,1,1) apresentou-se adequado para modelar os casos de infecção de HIV na região Norte, com base no mesmo far-se-ão as previsões para os 6 últimos meses de 2008 como mostra a tabela 21. O gráfico com as previsões e o intervalo de 95% de confiança para as previsões encontra-se na figura 15.

Exemplo da obtenção dos valores previstos com base no modelo escolhido.

$$\Delta Casosnorte_{t-1} = 14.93876 - 0.574401 * \varepsilon_{t-1}$$

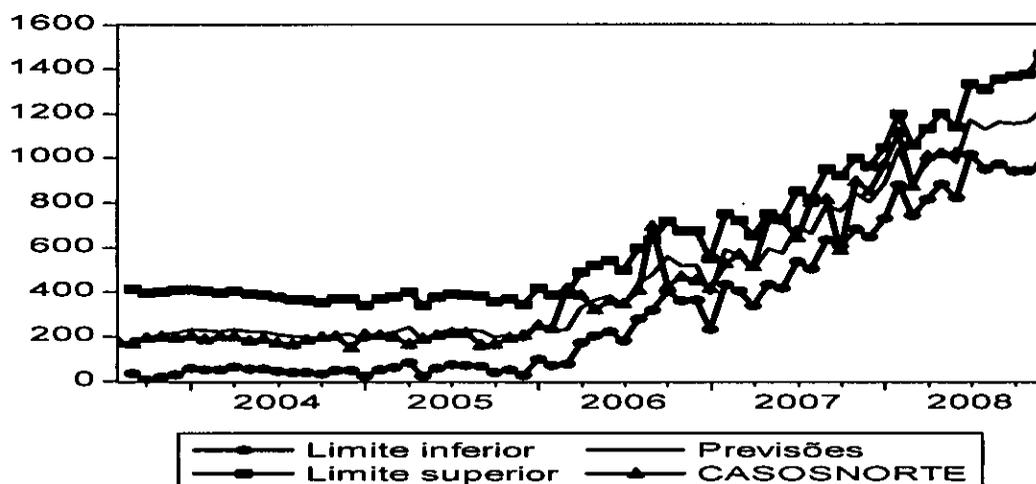
$$Casosnorte_t = Casosnorte_{t-1} + 14.93876 - 0.574401 * \varepsilon_{t-1}$$

$$Casosnorte_{2008/7} = 1038.93876 - 0.574401 * 1.98022 = 1038$$

Tabela 21. Previsões de casos de infecção de HIV na região Norte

Período	Previsão	L. Inferior	L. Superior
Julho	1038	867	1209
Agosto	1053	865	1240
Setembro	1067	864	1271
Outubro	1082	864	1301
Novembro	1097	865	1330
Dezembro	1112	865	1359

Figura 15. Previsões dos casos de infecção de HIV na região Norte



Da análise da tabela das previsões e do gráfico correspondente as mesmas, constata-se uma tendência crescente dos casos de infecção de HIV na região Norte.

5 Considerações finais e recomendações

5.1 Considerações finais

O objectivo proposto neste trabalho foi prever os casos de infecção de HIV nas regiões Sul, Centro e Norte, através da análise de séries temporais utilizando para tal a metodologia de Box-Jenkins, que se baseia nos modelos da classe ARIMA.

Com a aplicação desta metodologia, apoiada em testes estatísticos para subsidiar a modelagem, foi possível (dada a peculiaridade da mesma) definir modelos apropriados para as três regiões do país:

- Das três regiões do país, a Norte tem apresentado o menor número de casos de infecção de HIV;
- Apartir do ano de 2004 começa-se a verificar uma tendência cada vez mais crescente dos casos de infecção de HIV nas regiões Sul e Centro do país, sendo que na região Norte os casos começam a ter uma tendência cada vez mais crescente a partir dos finais de 2005;
- Durante o período em análise a região Norte apresentou a maior taxa de crescimento mensal e a menor taxa foi verificada na região Sul;
- Das 3 séries dos casos de infecção de HIV analisadas, apenas a série correspondente a região Centro apresentou estacionariedade em nível, sendo que as outras houve necessidade de diferencia-las uma vez para torná-las estacionárias e a partir daí produzir os modelos;
- Para a região Sul foi utilizado o modelo ARIMA (1,1,0) para a previsão dos casos de infecção, do qual a constante se apresentou não significativa;
- Para a região Centro a série dos casos de infecção de HIV mostrou-se estacionária em nível tendo sido definido o modelo ARIMA (1,0,1) como apropriado para prever os casos de infecção de HIV para esta região;
- Para modelar os casos de infecção de HIV na região Norte, foi utilizado o modelo ARIMA (0,1,1) e;
- Para as três séries estudadas as previsões foram feitas para um prazo de 6 meses, tendo se verificado para as séries da região Sul e Norte uma tendência crescente durante os 6 meses de previsão e decrescente para a região Centro.

5.2 Recomendações

Para outros estudos envolvendo séries temporais epidemiológicas, recomenda-se:

- Utilizar a metodologia de Box-Jenkins para estudar a evolução de outro tipo de doenças;
- Utilizar outro tipo de métodos para fazer a previsão dos casos de infecção de HIV e compará-los com os obtidos com base na metodologia de Box-Jenkins;
- Investigar o mecanismo da série em função de diversas variáveis (multivariada), relacionadas aos casos de infecção de HIV.



Referencias bibliográficas

Abitante, K.G. (2007). *Modelos de séries temporais aplicados á análise prospectiva de concessão de crédito bancário*. Dissertação apresentada para a obtenção do título de mestre em ciências.Piracicaba.

Barreto,A. et al. (2002). *Impacto demográfico do HIV/SIDA em Moçambique*. Edição financiada pela UNICEF. Moçambique.

Barros, P.(n.d). *Introdução ás séries temporais e aos modelos ARIMA*. Brasil

Barbosa, D.R.R. & Milone,G. (2004). *Estatística aplicada ao turismo e hotelaria*. Pioneirra Thomson Learning. São Paulo.

Camuri,P.A.(2005). *Dívida pública, política fiscal e restrição externa no Brasil: 1994-2004*. Belo Horizonte.

Costa, R.S.(2005). *Aplicação de modelos Box-Jenkins para previsão de volume de cartas e cartões postais importados à ECT/DR/PA*.Belém.

De Melo, B.S. (2001). *Modelo de previsão Para arrecadação tributária*. Brasil

Ehlers, R.S. (2005). *Análise de séries temporais*. Departamento de estatística, 3ªEd.Paraná.

Floriano,A. (2005).*Descrição custo-benefício do programa de cuidados domiciliários a pessoas com VIH/SIDA em Moçambique: evidências de um projecto de médicos do mundo em bairros da Matola*.

Franco, C.G.(n.d). *Testes de raízes unitárias através de estimadores semiparamétricos do parâmetro de memória longa*. Brasil.

Gujarati, D.N (2004); *Econometria básica*, 4ª edição (tradução). São Paulo. Editora Campus

Johnston, Jack & Dinardo, John (2001). *Métodos Econométricos* 4ª Edição. Editora McGrawHill.

Latorre, M.R. & Cardoso, M.R.A. (2001). *Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos*. Rev. Bras. epidemiol. vol.4

Luz, A.C.C. et al (2005). *Gerência pró-ativa de redes neurais usando método de previsão baseado em séries temporais (Modelo de Box-Jenkins)*. Campinas.

Médici, A.C. (2002). *Globalização, Doenças transmissíveis e Desigualdade mundial*. Nota técnica para discussão. Washington.

Menezes, Rui (2004). *Ensaio sobre a globalização dos mercados bolsistas internacionais: realidade ou mito? Uma abordagem econométrica sobre integração de cinco mercados bolsistas. Lição de síntese, Provas de agregação, Lisboa*.

Murteira, B.J.F, Müller, D.A. & Turkman, K.F. (1993). *Análise de sucessões cronológicas*. Editora McGrawHill. Portugal.

Nemec, Amand F. L. (1996). *Analysis of repeated measures and time series: an introduction with forestry examples*. Biometric information handbook N° 6. Res. Br. B. C. Min. For.

Neto, A. (2008). *Como trabalhar com séries temporais epidemiológicas?* Minas Gerais.

Nunes, N.L. (2005). *Previsão de demanda: Aplicação de modelos de Box-Jenkins para itens de venda de restaurante*. Belém.

Pires, A.P. (2001). *Notas de séries temporais*. Portugal

Resende, J.L. (2006). *Metas de inflação, câmbio flexível e autonomia de política monetária*. Belo Horizonte.

Ribeiro, L.C. & Paula, A.V. (n.d). *Previsão de população através dos modelos ARIMA de Box e Jenkins: Um exercício para Brasil.*

Rua, A. & Almas, D (1998). *A função de Consumo em Portugal. Uma Abordagem Econométrica.* Lisboa

Werner, L. & Ribeiro, J.L.D. (2003). *Previsão de demanda: Uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais.* Brasil.

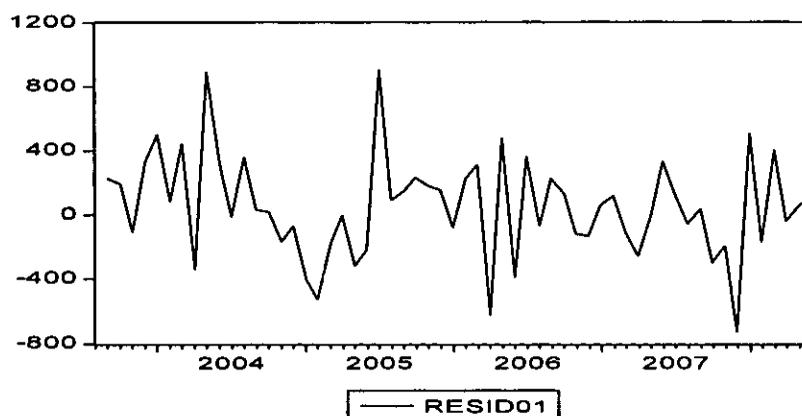
Anexos

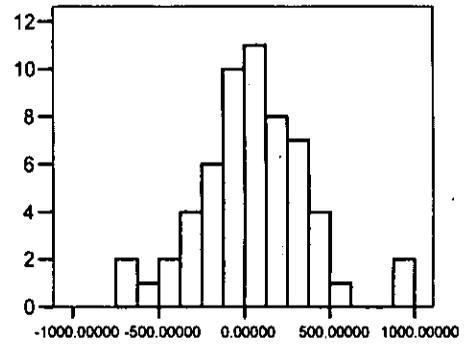
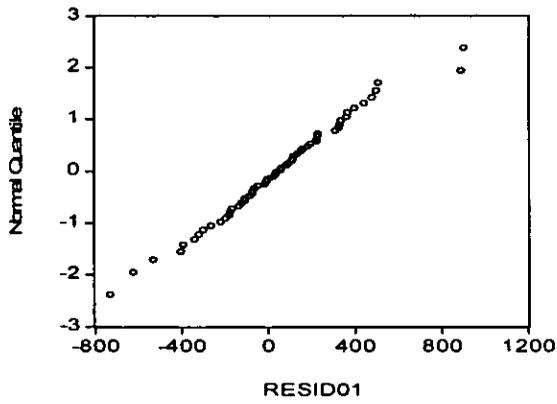
Tabela A1. Modelos para análise da taxa de crescimento dos casos de infecção de HIV nas regiões Sul, Centro e Norte

Variável dependente: ln(casossul)				
	Coeficiente	Std. error	t-statistic	Prob
T	0.004683	0.000738	6.343797	0.0000
C	8.335314	0.024825	335.7600	0.0000
Variável dependente: ln(casocentro)				
	Coeficiente	Std. error	t-statistic	Prob
T	0.004164	0.000514	8.095714	0.0000
C	8.342693	0.017297	482.3338	0.0000
Variável dependente: ln(casosnorte)				
	Coeficiente	Std. error	t-statistic	Prob
T	0.034569	0.001968	17.56641	0.0000
C	4.802316	0.066181	72.56332	0.0000

Região Sul

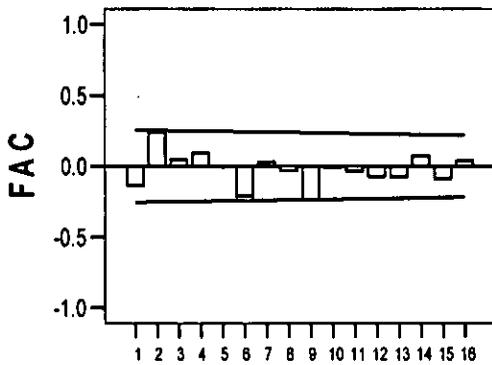
Figura A1. Cronograma, Q-Q plot, histograma, FAC e FACP dos resíduos do modelo ARIMA (0,1,1)





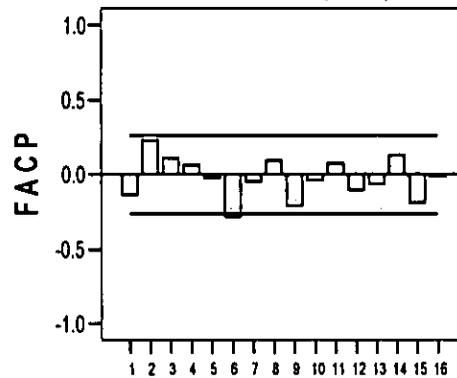
Resíduos do modelo
ARIMA(0,1,1)

Função de autocorrelação dos resíduos do modelo
ARIMA(0,1,1)



Defasagem m

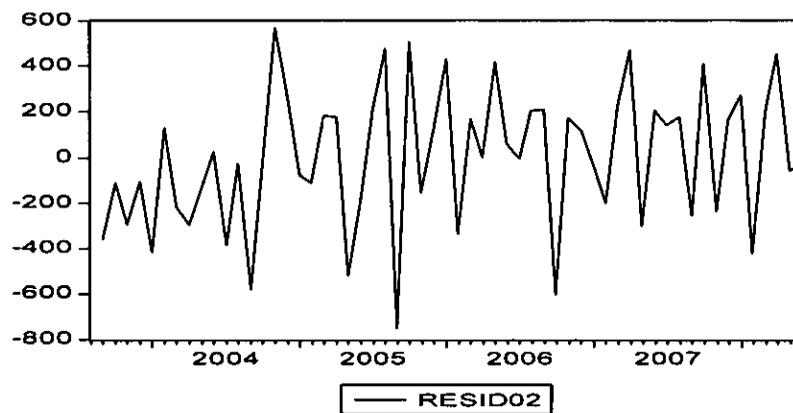
Função de autocorrelação dos resíduos do modelo
ARIMA(0,1,1)

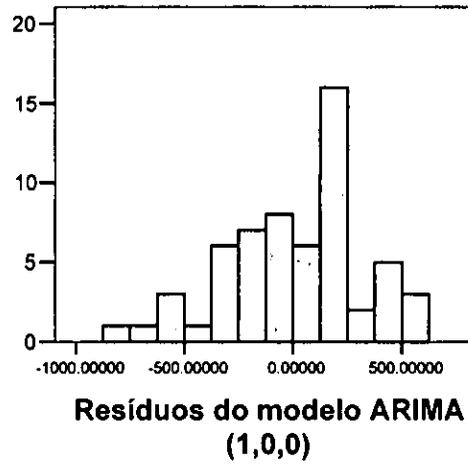
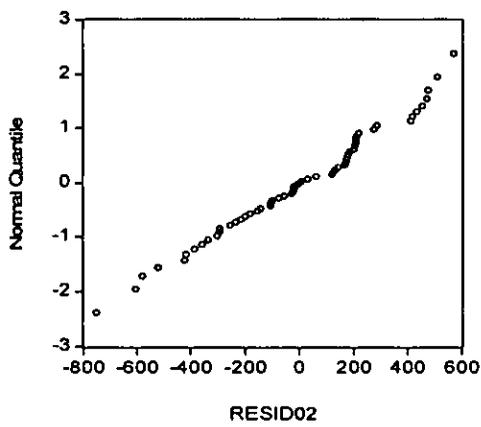


Defasagem m

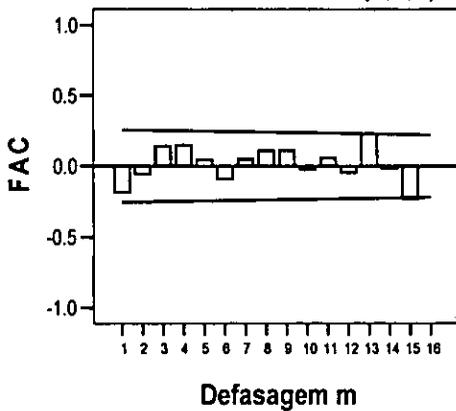
Região Centro

Figura A2. Cronograma, *Q-Q plot*, Histograma, FAC e FACP dos resíduos do modelo ARIMA (1,0,0)

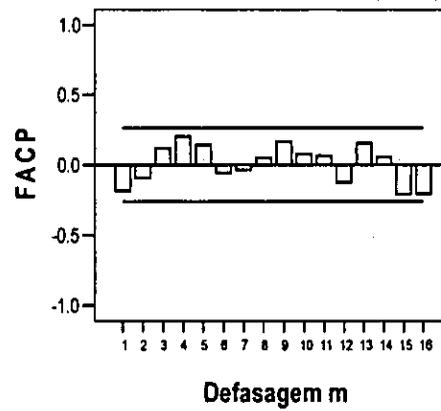




Função de autocorrelação dos resíduos do modelo ARIMA(1,0,0)

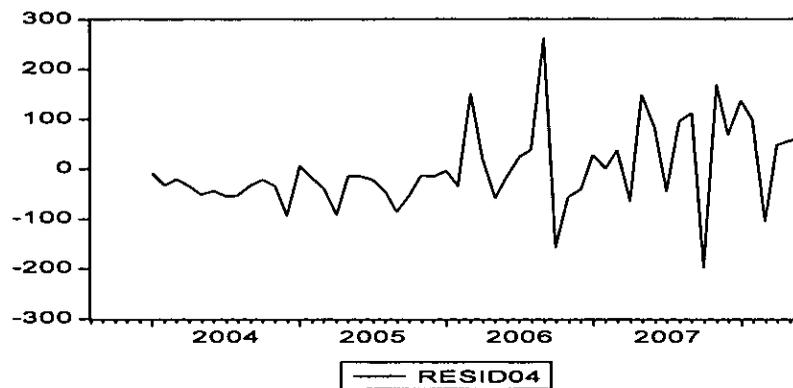


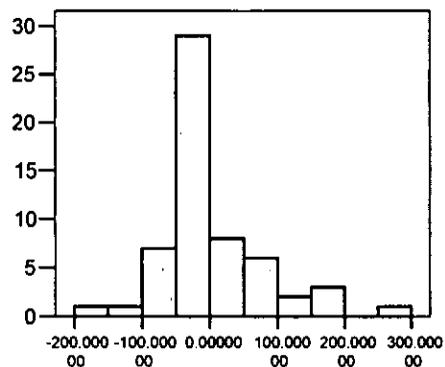
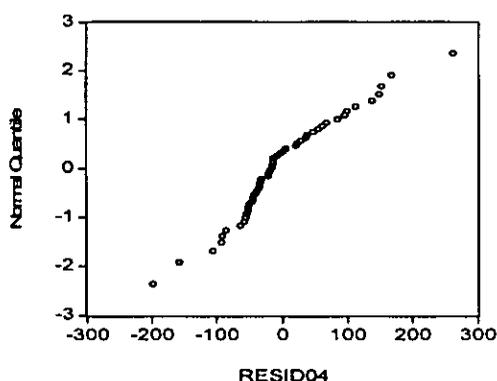
Função de autocorrelação parcial dos resíduos do modelo ARIMA(1,0,0)



Região Norte

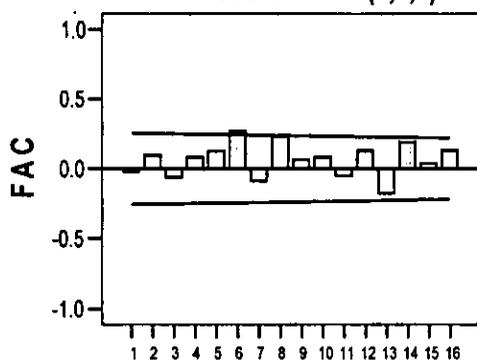
Figura A4. Cronograma, *Q-Q plot*, Histograma, FAC e FACP dos resíduos do modelo ARIMA (4,1,1)





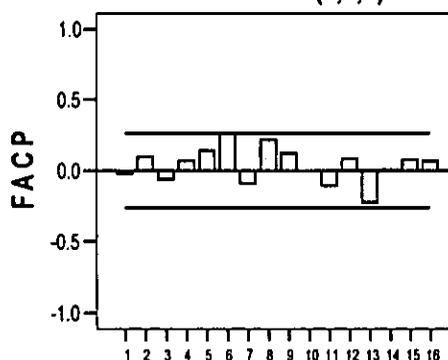
Resíduos do modelo ARIMA(4,1,1)

Função de autocorrelação dos resíduos do modelo ARIMA(4,1,1)



Defasagem m

Função de autocorrelação dos resíduos do modelo ARIMA(4,1,1)



Defasagem m

Tabela A1. Casos Mensais de infecção de HIV.Região Sul

Mês/Ano	2003	2004	2005	2006	2007	2008
Janeiro		4220	4581	4943	5277	5114
Fevereiro		4111	4210	5200	5370	4744
Março		4521	4238	5422	5217	5211
Abril		4010	4302	4679	4998	5010
Maio		5032	3987	5399	5098	5085
Junho		5011	3893	4821	5431	5175
Julho		4871	4879	5338	5412	
Agosto	3281	5237	4621	5124	5308	
Setembro	3556	5129	4733	5376	5370	
Outubro	3660	5138	4907	5421	5056	
Novembro	3479	4962	4998	5249	4979	

Dezembro	3852	4957	5085	5163	4325	
----------	------	------	------	------	------	--

Fonte: Programa Nacional de Combate ao HIV e ITS's

Tabela A2. Casos de infecção de HIV. Região Centro

Mês/Ano	2003	2004	2005	2006	2007	2008
Janeiro		3919	4997	5330	4875	5237
Fevereiro		4327	4835	4837	4658	4685
Março		4261	5014	5002	4925	4926
Abril		4138	5133	4950	5362	5346
Maiο		4205	4515	5327	4893	5129
Junho		4423	4427	5232	5077	5006
Julho		4157	4752	5103	5139	
Agosto	3876	4339	5249	5222	5218	
Setembro	3807	3906	4367	5307	4839	
Outubro	4010	4166	5014	4553	5245	
Novembro	3964	4936	4801	4808	4879	
Dezembro	4124	5184	4932	4931	5028	

Fonte: Programa Nacional de Combate ao HIV e ITS's

Tabela A3. Casos Mensais de Infecção de HIV. Região Norte

Mês/Ano	2003	2004	2005	2006	2007	2008
Janeiro		203	210	250	416	974
Fevereiro		187	207	237	524	1114
Março		200	196	416	566	872
Abril		199	163	384	512	1006
Maiο		182	190	321	725	1017
Junho		189	205	359	723	1024
Julho		171	216	346	645	
Agosto	175	166	214	406	802	
Setembro	168	185	161	695	815	
Outubro	193	198	164	425	587	
Novembro	201	201	192	468	896	
Dezembro	194	147	206	453	851	

Fonte: Programa Nacional de Combate ao HIV e ITS's