

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIENCIAS NATURAIS E EXATAS
CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA E MODELAGEM
QUANTITATIVA

Vinicius Pinheiro Nunes

**APLICAÇÃO DA METODOLOGIA DE BOX-JENKINS PARA ANÁLISE
DA EVOLUÇÃO DOS ÓBITOS POR NEOPLASIAS MALIGNAS DE
PELE NAS REGIÕES SUL E SUDESTE DO BRASIL**

Santa Maria, RS
2017

Vinicius Pinheiro Nunes

**APLICAÇÃO DA METODOLOGIA DE BOX-JENKINS PARA ANÁLISE DA
EVOLUÇÃO DOS ÓBITOS POR NEOPLASIAS MALIGNAS DE PELE NAS
REGIÕES SUL E SUDESTE DO BRASIL**

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

Orientadora: Angela Pellegrin Ansuj

Santa Maria, RS
2017

Vinicius Pinheiro Nunes

APLICAÇÃO DA METODOLOGIA DE BOX-JENKINS PARA ANÁLISE DA EVOLUÇÃO DOS ÓBITOS POR NEOPLASIAS MALIGNAS DE PELE NAS REGIÕES SUL E SUDESTE DO BRASIL

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

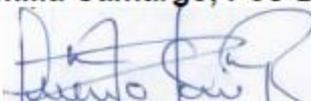
Aprovado em 21 de Agosto de 2017



**Angela Pellegrin Ansuji, Dra. (UFSM)
(Presidente/Orientadora)**



Maria Emilia Camargo, Pós-Dra. (UCS)



Fernando Arturo Peña Ramirez, Dr. (UFSM)

Santa Maria, RS
2017

DEDICATÓRIA

Dedico este trabalho primeiramente ao meu esforço e minha persistência. Dedico também a minha orientadora da especialização (Angela Pelegrin Ansuji) e a minha orientadora do Bacharelado em Estatística (Angela Isabel dos Santos Dullius) por todo o conhecimento que me foi ensinado, pela paciência e principalmente pela motivação em seguir em frente nesta dura trajetória na academia e na vida.

AGRADECIMENTOS

Este trabalho não poderia ser realizado sem a contribuição de muitas pessoas, seja pelo conhecimento que me foi ensinado, pela motivação e incentivo no desenvolvimento no deste estudo e também pelo apoio psicológico e emocional, por isso, de uma maneira especial, agradeço:

- À minha colega da especialização Virginia Margareth Passatti Rocha pelas horas dedicadas à estudar e aprofundar o conhecimento na área, e também pela forte amizade que construímos ao longo desta jornada.

- Aos professores da Especialização em Estatística e Modelagem Quantitativa, Roselaine Ruviano Zanini, Dra. Luciane Flores Jacobi e Dr. Ivanor Muller, pelos ensinamentos relacionados a Estatística;

- À minha mãe Ana Circe Pinheiro Nunes por todo amor, carinho e paciência durante todo este período;

- Às minhas amigas que tanto amo Gyanna Marramon de Brum, Ana Paula Fiuza, Karolyne Nish e Ingrid Hardok Fuchs;

- À banca, por ter aceitado o convite e pelas importantes considerações realizadas para a conclusão deste trabalho;

- À Universidade Federal de Santa Maria (UFSM), pela oportunidade de realizar esta pesquisa e mais uma vez poder contribuir aos preceitos da ciência estatística.

Enfim, a todos aqueles que, de uma certa forma, possam ter contribuído para realização desta monografia.

If I could offer you only one tip for the future,
"sunscreen" would be it.

The long-term benefits of sunscreen have been
proved by scientists, whereas the rest of my advice
has no basis more reliable than my own meandering experience.

(Mary Schmich)

RESUMO

APLICAÇÃO DA METODOLOGIA DE BOX-JENKINS PARA ANÁLISE DA EVOLUÇÃO DOS ÓBITOS POR NEOPLASIAS MALIGNAS DE PELE NAS REGIÕES SUL E SUDESTE DO BRASIL

AUTOR: Vinicius Pinheiro Nunes
ORIENTADORA: Angela Pellegrin Ansuji

Nos conjuntos de dados cujas observações encontram-se ordenadas em instantes de tempo, eventualmente em intervalos equidistantes, indica-se a análise de séries temporais. Esse tipo de análise permite identificar padrões como: sazonalidade, tendência, ciclos e mudanças de nível. Além disso, é possível gerar modelos explicativos capazes de descrever o comportamento da sequência de observações, e ainda obter previsões acerca da série analisada. Na análise de séries temporais univariadas tem-se a metodologia de Box-Jenkins. Essa, possibilitou analisar não apenas processos estacionários utilizando os modelos ARMA, mas também os não estacionários, através dos modelos ARIMA. Diante disso, esta pesquisa propõe-se a realizar um estudo utilizando a metodologia de Box-Jenkins nas séries de óbitos por neoplasias malignas de pele entre o período de 1996 e 2014, ocorridos nas regiões Sul e Sudeste do Brasil, afim de identificar padrões, gerar modelos explicativos e obter previsões. Este estudo possibilitou evidenciar algumas características como presença de tendência crescente, ausência de sazonalidade e comportamento não estacionário nas séries de óbitos por neoplasias malignas de pele tanto na região Sul, quanto na região Sudeste. A metodologia de Box-Jenkins demonstrou boa qualidade de predição para as séries da região Sul e Sudeste as quais apresentaram, os modelos ARIMA (4,1,0) com constante e ARIMA (0,1,1) com constante como modelos melhores ajustados aos dados. Além disso, a modelagem das séries resultou no ajuste de modelos com componentes diferentes, os óbitos por neoplasias malignas de pele nas regiões Sul foram explicados apenas por componentes AR e a região Sudeste apenas por componentes MA, ambas com componente de integração.

Palavras-chave: Neoplasias malignas de pele. Séries temporais. Metodologia de Box-Jenkins.

ABSTRACT

METHODOLOGY FOR ANALYSIS OF THE EVOLUTION OF DEATHS CAUSED BY MALIGNANT NEOPLASMS OF THE SKIN IN THE SOUTH AND SOUTHEAST REGIONS OF BRAZIL

AUTHOR: Vinicius Pinheiro Nunes

ADVISER: Angela Pellegrin Ansuji

In data sets whose observations are ordered in moments of time, possibly at equidistant intervals, the analysis of time series is indicated. This type of analysis allows identifying patterns such as: seasonality, trend, cycles and level changes. In addition, it is possible to generate explanatory models capable of describing the behavior of the sequence of observations, and also to obtain predictions about the analyzed series. In the analysis of univariate time series we have the Box-Jenkins methodology. This allowed us to analyze not only stationary processes using ARMA models, but also non stationary processes through ARIMA models. Therefore, this study proposes to carry out a study using the Box and Jenkins methodology in the series of deaths due to malignant skin neoplasms between 1996 and 2014, in the South and Southeast regions of Brazil, in order to identify patterns, generate explanatory models and obtain forecasts. This study made it possible to show some characteristics such as presence of increasing tendency, absence of seasonality and non-stationary behavior in the series of deaths due to malignant skin neoplasms in both the South and Southeast regions. The Box-Jenkins methodology showed good prediction quality for the series of the South and Southeast regions, which presented ARIMA models (4,1,0) with constant and ARIMA (0,1,1) with constant as best adjusted models to the data. In addition, modeling of the series resulted in adjustment of models with different components, deaths due to malignant skin neoplasms in the South were explained only by AR components and the Southeast region only by MA components, both with integration component.

Keys words: Neoplasms of the skin. Time series. Box-Jenkins methodology.

LISTA DE FIGURAS

Figura 1 – Ciclo iterativo utilizado na metodologia de Box-Jenkins.....	32
Figura 2 – Série trimestral de óbitos por neoplasias malignas de pele nas regiões Sul (a) e Sudeste (b) do Brasil entre 1996 e 2014.....	51
Figura 3 – Função de autocorrelação (FAC) e Função de autocorrelação parcial (FACP) da Série trimestral de óbitos por neoplasias malignas de pele nas regiões Sul (a) e Sudeste (b) do Brasil entre 1996 e 2014.....	52
Figura 4 – Primeira diferença da Série trimestral de óbitos por neoplasias malignas de pele nas regiões Sul (a) e Sudeste (b) do Brasil entre 1996 e 2014. ..	53
Figura 5 – Função de autocorrelação (FAC) e Função de autocorrelação parcial (FACP) da primeira diferença da série trimestral de óbitos por neoplasias malignas de pele em primeira diferença nas regiões Sul (a) e Sudeste (b) do Brasil, entre 1996 e 2014.	54
Figura 6 – Resíduos dos modelos ajustados para as séries de óbitos por neoplasias malignas de pele nas regiões Sul (a) e Sudeste (b) do Brasil.	56
Figura 7 – Função de autocorrelação (FAC) e Função de autocorrelação parcial (FACP) dos resíduos dos modelos ARIMA ajustados para os óbitos por neoplasias malignas de pele nas regiões Sul (a) e Sudeste (b).....	56
Figura 8 – Gráfico de previsões para as séries trimestral de óbitos por neoplasias malignas ocorridos nas regiões Sul (a) e Sudeste (b).	58

LISTA DE QUADROS

Quadro 1 – Resumo das condições de estacionariedade e invertibilidade dos modelos AR, MA e ARMA..	28
Quadro 2 – Descrição dos tipos de melanomas e outras (os) neoplasias (tumores) malignas (os) da pele (C43-C44) de acordo com a CID-10.....	43
Quadro 3 – Critério de informação Bayesiano (BIC) dos três modelos de melhor ajuste as séries de óbitos por neoplasias malignas de pele nas regiões sul e sudeste, entre 1996 e 2014.....	55
Quadro 4 – Parâmetros estimados dos modelos ARIMA(4,1,0) com constante (Região Sul) e ARIMA(0,1,1) com constante (Região Sudeste).....	55
Quadro 5 – Previsões dos óbitos por neoplasias malignas de pele para os quatro trimestres do ano de 2015 nas regiões Sul e Sudeste considerando, respectivamente, os modelos ARIMA(4,1,0) com constante e ARIMA (0,1,1) com constante	57

LISTA DE ABREVIATURAS E SIGLAS

ADF	Dickey Fuller aumentado
AIC	Critério de informação de Akaike
AR	Modelo Autorregressivo
ARIMA	Modelo Autorregressivo Integrado de Médias Móveis
ARMA	Modelo Autorregressivo e de Médias Móveis
BIC	Critério de informação Bayesiano ou de Schwarz
CEP	Comitê de Ética em Pesquisa
CID-10	Classificação estatística internacional de doenças e problemas relacionados à saúde – Décima revisão
CID-9	Classificação estatística internacional de doenças e problemas relacionados à saúde – Nona revisão
DF	Dickey Fuller
DO	Declaração de Óbito
DATASUS	Departamento de Informática do Sistema Único de Saúde
FAC	Função de Autocorrelação
FACP	Função de Autocorrelação Parcial
FE	Erro de previsão
HCV	Vírus da Hepatite C
IBGE	Instituto Brasileiro de Geografia e Estatística
<i>i.i.d</i>	Independentes e identicamente distribuídos
INCA	Instituto Nacional de Câncer José Alencar Gomes da Silva
JB	Jarque-Bera
LM	Multiplicador de Lagrange
KPSS	Kwiatkowski, Phillips, Schmidt e Shin
MA	Modelo de Médias Móveis
MAE	Erro absoluto médio
MAPE	Erro absoluto percentual médio
MLG	Modelos Lineares Generalizados
MPE	Erro percentual médio
MSE	Erro quadrático médio
OMS	Organização Mundial da Saúde
RMSE	Raiz do erro quadrático médio
SARIMA	Sazonal Autorregressivo Integrado de Médias Móveis
SARS	Síndrome respiratória aguda severa
SIM	Sistema de Informações Sobre Mortalidade
SOE	Sonda Oroenteral
SUS	Sistema Único de Saúde
SVS	Sistema de Vigilância em Saúde

SUMÁRIO

1 INTRODUÇÃO	14
1.1 DELIMITAÇÃO DO TEMA	15
1.2 PROBLEMÁTICA	15
1.3 OBJETIVOS	16
1.3.1 Objetivo geral	16
1.3.2 Objetivos específicos	16
1.4 JUSTIFICATIVA	16
1.5 ESTRUTURA DO TRABALHO	17
2 SÉRIES TEMPORAIS	19
2.1 CONCEITOS BÁSICOS	19
2.2 TIPOS DE ANÁLISES	20
2.3 TIPOS DE ENFOQUE	20
2.3.1 Análise no domínio temporal	20
2.3.2 Análise no domínio da frequência	20
2.4 OBJETIVOS DA ANÁLISE DE SÉRIES TEMPORAIS	21
2.5 COMPONENTES NÃO OBSERVÁVEIS DE UMA SÉRIE TEMPORAL	22
2.6 PROCESSOS ESTOCÁSTICOS	23
2.6.1 Estacionariedade	23
2.6.2 Ergodicidade	24
2.7 RUÍDO BRANCO	25
3 MODELOS LINEARES PARA SÉRIES TEMPORIAS ESTACIONÁRIAS	26
3.1 MODELOS AUTORREGRESSIVOS	26
3.2 MODELOS DE MÉDIAS MÓVEIS	27
3.3 MODELOS AUTORREGRESSIVOS DE MÉDIAS MÓVEIS	27
3.4 ESTACIONARIEDADE E INVERSIBILIDADE	28
3.5 FUNÇÃO DE AUTOCORRELAÇÃO E FUNÇÃO DE AUTOCORRELAÇÃO PARCIAL	29
4 MODELOS LINEARES PARA SÉRIES TEMPORIAS NÃO ESTACIONÁRIAS ...	31
4.1 INTRODUÇÃO	31
4.2 IDENTIFICAÇÃO DE MODELOS ARIMA	32
4.2.1 Determinação das “d” diferenças	33
<i>4.2.1.1 Testes de raiz Unitária</i>	33
4.2.2 Determinação das ordens p e q	34
4.3 ESTIMAÇÃO DE MODELOS ARIMA	34
4.3.1 Critérios de seleção de modelos	35
4.4 ANÁLISE DE DIAGNÓSTICO	36
4.5 PREVISÃO	38
4.5.1 Medidas de precisão de previsão	38
4.6 ESTUDOS APLICADOS UTILIZANDO A METODOLOGIA DE BOX-JENKINS..	40
5 NEOPLASIAS MALIGNAS DA PELE	43
5.1 DEFINIÇÃO DE NEOPLASIAS MALIGNAS DE PELE	43
5.2 A IMPLANTAÇÃO DA CID-10 NO BRASIL	44
5.3 FONTE DE DADOS OFICIAIS SOBRE MORTALIDADE NO BRASIL	45
5.4 INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA....	46
5.4.1 Estimativas feitas pelo INCA	46
6 MATERIAIS E MÉTODOS	48
6.1 CARACTERIZAÇÃO DA PESQUISA	48

6.2 DESCRIÇÃO DA VARIÁVEL ANALISADA.....	49
6.3 DESCRIÇÃO DO PERÍODO ANALISADO.....	49
6.4 DESCRIÇÃO DA REGIÃO ANALISADA	49
6.5 FONTE DOS DADOS.....	49
6.6 PROCEDIMENTOS METODOLÓGICOS.....	50
7 RESULTADOS E DISCUSSÃO	51
8 CONCLUSÃO	59
REFERÊNCIAS.....	61

1 INTRODUÇÃO

O uso de técnicas estatísticas em estudos e pesquisas tem se tornado cada vez mais importante em diversas áreas do conhecimento como: economia, biologia, agronomia, administração, epidemiologia, física, entre outras áreas. O processo de escolha da(s) técnica(s) que será(ão) utilizada(s), entre as possíveis existentes, dependerá dos objetivos e interesses de cada pesquisador e das especificidades da variável ou das variáveis que se pretende analisar.

Nos conjuntos de dados cujas observações encontram-se ordenadas em instantes de tempo, eventualmente em intervalos equidistantes, indica-se a análise de séries temporais. Esse tipo de análise permite identificar padrões como: sazonalidade, tendência, ciclos e mudanças de nível. Além disso, é possível gerar modelos explicativos capazes de descrever o comportamento da sequência de observações, e ainda obter previsões acerca da série analisada (MORETTIN; TOLOI, 2004; SOUZA; CAMARGO, 1996; BOX; JENKINS; REINSEL, 1994).

Do ponto de vista analítico, os modelos utilizados na análise de séries de tempo se diferenciam dos modelos de regressão linear e dos modelos lineares generalizados (MLG) devido à sua capacidade de analisar a dependência temporal presente no conjunto de dados. Em outras palavras, essa dependência pode ser compreendida como a influência que as observações passadas exercem sobre as observações presentes.

Na análise de séries temporais univariadas têm-se os modelos autorregressivo (AR), de médias móveis (MA), e autorregressivo e de médias móveis (ARMA), descrito na sua forma geral por Whittle (1951) e, posteriormente, popularizado por Box e Jenkins (1970). O sucesso da metodologia de Box e Jenkins se deve a generalidade que os autores apresentaram a análise de modelos ARMA. Essa, possibilitou analisar não apenas processos estacionários utilizando os modelos ARMA, mas também os não estacionários.

A análise dos processos não estacionários tornou-se possível a partir da inclusão de uma ordem integrada (I), que consiste em realizar sucessivas diferenças até tornar o processo estável, originando, assim, o modelo autorregressivo integrado de médias móveis (ARIMA), amplamente utilizado na análise de séries de tempo (MADDALA, 2001).

Diante disso, esta pesquisa propõe-se a realizar um estudo utilizando a metodologia de Box e Jenkins nas séries de óbitos por neoplasias malignas de pele, ocorridos nas regiões Sul e Sudeste do Brasil.

1.1 DELIMITAÇÃO DO TEMA

Este estudo delimitou-se a analisar a ocorrência de óbitos por neoplasias malignas de pele nas regiões Sul e Sudeste do Brasil, pois essas representam as regiões de maior incidência desse problema de saúde pública no país. Quanto a periodicidade analisada, considerou-se as séries de óbitos ocorridos entre 1996 a 2014. Esses períodos foram definidos por se tratarem, respectivamente, do marco inicial da utilização da classificação estatística internacional de doenças e problemas relacionados à saúde – Décima revisão (CID-10) no Brasil, e do penúltimo ano de dados divulgados e revisados.

Os óbitos por neoplasias malignas de pele analisados referem-se aos casos de melanomas malignos de pele e também a outros tipos de neoplasias malignas de pele, também conhecidas como não melanomas. A análise dos dados foi realizada de forma agregada, considerando as frequências absolutas trimestrais por local de residência dos indivíduos.

É importante ressaltar que não foram utilizadas as taxas, ou frequências relativas, referentes aos óbitos, devido à indisponibilidade de dados oficiais quanto ao tamanho populacional em cada período analisado. Essa limitação deve-se a inexistência de levantamentos e informações confiáveis quanto ao tamanho da população em curtos períodos de tempo. As divulgações de informações de maior precisão ocorrem após a realização dos censos demográficos, comumente, realizados a cada 10 anos.

1.2 PROBLEMÁTICA

Tomando-se como base as informações apresentadas na introdução, as delimitações apresentadas na seção 1.1 e a justificativas elucidadas na seção 1.4, surgem os seguintes questionamentos:

Quais padrões e/ou características as séries de óbitos por neoplasias malignas de pele nas regiões Sul e Sudeste do Brasil apresentam?

Quais modelos ajustados podem ser utilizados na explicação e realização de previsões acerca das séries de óbitos por neoplasias malignas de pele nas regiões Sul e Sudeste do Brasil, utilizando a metodologia de Box-Jenkins?

1.3 OBJETIVOS

1.3.1 Objetivo geral

Analisar as séries temporais trimestrais de óbitos por neoplasias malignas de pele, entre o período de 1996 e 2014, nas regiões Sudeste e Sul do Brasil, utilizando a metodologia de Box-Jenkins.

1.3.2 Objetivos específicos

- Descrever e avaliar o comportamento das séries óbitos por neoplasias malignas de pele das regiões Sul e Sudeste do Brasil entre 1996 e 2014;
- Modelar as séries analisadas afim de identificar o modelo ARIMA de maior parcimônia para cada região;
- Obter previsões para os trimestres do ano de 2015 a partir do modelo escolhido para cada região, e avaliar a qualidade dessas previsões;
- Comparar as características da análise dos óbitos por neoplasias malignas de pele das regiões Sul e Sudeste a fim de evidenciar semelhanças e diferenças entre as regiões.

1.4 JUSTIFICATIVA

Sabe-se que há uma ampla base de dados relacionada a saúde disponível no Brasil. Grande parte desses dados são disponibilizados pelo Ministério da Saúde

através do departamento de informática do Sistema Único de Saúde (DATASUS) e possuem fácil acesso. Apesar disso, verifica-se uma subutilização dessas informações em estudos e pesquisas no processo de melhorias da saúde pública do país.

Em concordância com o que foi dito anteriormente, está a baixa quantidade de estudos relacionados as neoplasias malignas de pele, identificada como a classe de doença de maior incidência entre todos os tipos de neoplasias no Brasil. Segundo estimativas do INCA, esse problema de saúde pública representa cerca de 30% do esperado entre os casos de câncer nos últimos anos. Além disso, os óbitos relacionados a esse tipo de neoplasias tem demonstrado alto crescimento nas últimas décadas.

Vale destacar também que não há propriamente um programa consolidado que busque o desenvolvimento de estratégias para o controle e combate ao câncer de pele, mesmo sendo esse o tipo de neoplasia de maior incidência no Brasil. Por isso, fica evidente a importância de promover ações em saúde voltadas a prevenção do câncer de pele, embora já sejam observadas ações focadas no diagnóstico precoce desse problema.

Assim, esta pesquisa justifica-se pela necessidade de realizar estudos quanto a evolução dos óbitos por neoplasias malignas de pele nas regiões Sul e Sudeste, as quais apresentam as maiores incidências comparado as demais regiões, com o propósito de evidenciar e estimular a implementação de novas políticas públicas no combate e prevenção dessa doença.

1.5 ESTRUTURA DO TRABALHO

Este trabalho está estruturado em oito capítulos: O primeiro capítulo apresenta uma introdução acerca da temática estudada, a delimitação do tema, a problemática, os objetivos, geral e específicos, assim como a justificativa para a realização deste estudo. O segundo capítulo apresenta uma revisão de literatura introdutória sobre séries temporais.

O terceiro capítulo apresenta uma revisão bibliográfica sobre modelos lineares para séries temporais estacionárias. O quarto capítulo apresenta uma

revisão de literatura sobre modelos lineares para séries temporais não estacionárias, a metodologia de Box-Jenkins e seus procedimentos para realização da análise de dados. O quinto capítulo apresenta uma revisão de literatura relacionada às neoplasias malignas de pele;

O sexto capítulo apresenta os Materiais e métodos, tais como: a caracterização da pesquisa, a descrição da variável, do período e das regiões analisadas, a fonte dos dados e os procedimentos metodológicos utilizados na análise de séries temporais. O sétimo capítulo apresenta os resultados e a discussão das análises realizadas. O oitavo capítulo apresenta as principais conclusões evidenciadas no estudo realizado nesta monografia. A monografia é finalizada com a apresentação das referências bibliográficas utilizadas.

2 SÉRIES TEMPORAIS

Neste capítulo é apresentada uma revisão de literatura dos conceitos básicos e definições importantes relacionadas a séries temporais.

2.1 CONCEITOS BÁSICOS

Uma série temporal é qualquer conjunto de observações ou sequência de dados numéricos, ordenados no tempo (MORETTIN; TOLOI, 2004). Para Box *et al.* (2015), uma característica intrínseca de uma série temporal diz respeito a dependência temporal entre as observações e a análise dessas séries está relacionada a análise dessa dependência.

As séries temporais podem ser classificadas como discretas ou contínuas quanto a sua característica de mensuração. Classificam-se como discretas quando o conjunto de observações ou a coleção de dados for finita ou infinita numerável, e como contínuas, quando for infinita não-numerável (SOUZA; CAMARGO, 1996; MORETTIN; TOLOI, 2004).

Uma série temporal é dita discreta quando Y_t representa uma sequência de observações da variável aleatória Y no instante t , em que t varia do instante 1 até n . As observações destas séries são geralmente ordenadas em instantes de tempo equiespaçados, ou seja, em intervalos equidistantes. Para este tipo de série, estas periodizações dos dados coletados podem usualmente ser: diárias, semanais, mensais, trimestrais, anuais, etc., conforme o interesse de estudo de cada pesquisador (SOUZA; CAMARGO, 1996).

De acordo com Souza e Camargo (1996), as séries contínuas podem ser denotadas por $Y(t)$, e se diferenciam das discretas na forma em que a coleta é realizada. O monitoramento da variável aleatória acontece de forma ininterrupta, ou seja, não há como enumerar cada observação pois a coleta é realizada continuamente. Sendo assim, t pode ser representado no intervalo de $\{0 < t < n\}$. Morettin e Toloí (2004) ressaltam ainda que, as séries contínuas podem ser discretizadas acumulando-se ou agregando-se observações em intervalos de tempo iguais, conforme explicito nas séries discretas (MORETTIN; TOLOI, 2004).

2.2 TIPOS DE ANÁLISES

Segundo Maddala (2001) e Sousa e Camargo (1996), a análise de uma série temporal pode ser classificada como univariada, quando uma única sequência de dados é analisada, e multivariada, quando as observações de várias variáveis são analisadas conjuntamente, em uma mesma sequência de períodos de tempo. Neste caso, tem-se um vetor de observações $\{Y(t), t \in n\}$ de ordem $r \times 1$.

2.3 TIPOS DE ENFOQUE

Morettin e Toloi (2004) citam dois tipos de enfoques utilizados na análise de séries temporais: a análise no domínio temporal e a análise no domínio da frequência.

2.3.1 Análise no domínio temporal

A análise no domínio temporal baseia-se na modelagem direta das relações entre uma série e seus períodos passados, ou seja, é considerada a evolução temporal do processo em determinados instantes de tempo. Este tipo de análise é baseado em modelos paramétricos, com números finitos de parâmetros (MADDALA, 2001; MORETTIN; TOLOI, 2004; SOUZA; CAMARGO, 1996)

2.3.2 Análise no domínio da frequência

A análise no domínio do tempo baseia-se em modelos não paramétricos e a técnica de análise utilizada é a espectral, a qual consiste em decompor a série em componentes de frequência, expresso por curvas de cosseno e de seno, variando independentemente com amplitudes aleatórias. Nesse caso, a existência do espectro é considerada a característica fundamental da análise (MADDALA, 2001; MORETTIN; TOLOI, 2004).

Esse tipo de análise é útil quando os componentes harmônicos da série apresentam alguma explicação física, ou quando os efeitos práticos são mais convenientes medidos por suas componentes separadamente (SOUZA E

CAMARGO, 1996). Para Maddala (2001), o objetivo da análise consiste em verificar como a variância da série Y_t se distribui entre oscilações das várias frequências.

Salienta-se que este enfoque não é o objetivo deste estudo. Sendo assim, para maiores informações recomenda-se a leitura de Jenkins e Watts (1968) e Morettin (1979).

2.4 OBJETIVOS DA ANÁLISE DE SÉRIES TEMPORAIS

Os principais objetivos descritos na literatura para a análise de séries temporais variam de acordo com a necessidade de cada estudo ou pesquisa e podem estar intimamente relacionados às seguintes possibilidades:

- descrever o comportamento da série: refere-se ao processo inicial da análise de séries temporais, e consiste em obter medidas descritivas acerca das principais características observadas. Há também a representação gráfica da série, evidenciando a forma na qual uma ou mais variáveis evoluem no tempo (SOUZA; CAMARGO, 1996);
- obter previsões: apresenta-se como um dos principais objetivos da análise de séries temporais e fundamenta-se em gerar valores futuros da variável analisada (MORETTIN; TOLOI, 2004; SOUZA; CAMARGO, 1996; BOX; JENKINS; REINSEL, 1994).
- investigar o mecanismo gerador da série: baseia-se na determinação de um modelo de entrada-saída dinâmico que pode mostrar o efeito na saída de um sistema de qualquer série de entradas (BOX; JENKINS; REINSEL, 1994). A operação que relaciona a série de entrada à série de saída é denominada filtro e a função que relaciona este processo é denominada função de transferência (SOUZA; CAMARGO, 1996);
- controlar o processo gerador da série: utiliza-se quando uma série temporal mede qualidade. Consiste em modelar a série, gerar previsões e a partir de disso, realizar medidas corretivas afim de evitar que a série oscile fora dos limites estabelecidos (SOUZA; CAMARGO, 1996). Box, Jenkins e Reinsel, (1994) descrevem estas medidas corretivas relacionando-as ao design de esquemas de controle simples pelo qual os possíveis desvios da saída

podem, na medida do possível, ser compensados pelo ajuste dos valores da série de entrada.

- Encontrar periodicidades relevantes na série: Morettin e Tolo (2004) reiteram que o encontro destas periodicidades pode ser realizado a partir da análise espectral.

2.5 COMPONENTES NÃO OBSERVÁVEIS DE UMA SÉRIE TEMPORAL

Segundo Makridakis e Wheelwright (1998) e Morettin e Tolo (2004), a decomposição clássica é realizada a partir de um modelo univariado que descreve o comportamento da série através de quatro componentes não observáveis: a tendência, a sazonalidade, o ciclo e o termo aleatório ou resíduo do modelo. Esses quatro componentes podem ser definidos como:

- tendência (T_t): trata-se de um movimento contínuo dos dados numa determinada direção que dura por um determinado período de tempo. Essa componente ocorre quando há um movimento crescente ou decrescente na série ao longo do tempo. (MAKRIDAKIS; WHEELWRIGHT, 1998; VASCONCELLOS; ALVES, 2000);
- sazonalidade (S_t): refere-se a um movimento oscilatório que pode ocorrer em intervalos de periodicidade fixa (semanal, mensal, trimestral, etc.). De modo geral, quando há um movimento periódico de curta duração, normalmente inferiores a um ano (MAKRIDAKIS; WHEELWRIGHT, 1998; VASCONCELLOS; ALVES, 2000);
- ciclo (C_t): caracteriza-se por movimentos oscilatórios parecidos aos da sazonalidade. A distinção em relação ao padrão sazonal ocorre pois, nesse caso, assume-se comprimentos de periodicidades variáveis. Além disso, o comprimento médio de um ciclo geralmente é maior do que o da sazonalidade (vários anos ou décadas, por exemplo) (MAKRIDAKIS; WHEELWRIGHT, 1998);
- variações aleatórias ou resíduos (ε_t): Trata-se de variações esporádicas ou aleatórias que ocorrem em decorrência de fatores não explicáveis. Essas variações também podem ser chamadas de resíduos, pois não são causadas

por quaisquer outros componentes e são consideradas não previsíveis (VASCONCELLOS; ALVES, 2000).

Cabe lembrar que, não necessariamente a série analisada apresentará todos estes componentes citados. A ocorrência desses fatores dependerá das especificidades de cada conjunto de dados (VASCONCELLOS; ALVES, 2000).

De acordo com Brockwell e Davis (2009), a decomposição clássica possibilita que a série seja escrita na forma aditiva ou multiplicativa dos componentes não observáveis. Admite-se um modelo aditivo quando os componentes que formam a série são somados, e um modelo multiplicativo quando os componentes são multiplicados, respectivamente, conforme as equações a seguir:

$$Z_t = T_t + C_t + S_t + \varepsilon_t \quad (1)$$

$$Z_t = T_t \cdot C_t \cdot S_t \cdot \varepsilon_t \quad (2)$$

2.6 PROCESSOS ESTOCÁSTICOS

Conforme Souza e Camargo (1996), processos estocásticos são denominados sistemas que evoluem no tempo e/ou espaço a partir de leis probabilísticas, ou também, modelos que descrevam a estrutura de probabilidade de um conjunto sequencial de observações. Bueno (2008) destaca que a estimação de uma série temporal depende das restrições de estacionariedade e ergodicidade deste processo gerador da série.

2.6.1 Estacionariedade

Define-se um processo estacionário como aquele que apresenta média (μ) e variância (σ^2) constantes ao longo do tempo e que o valor da covariância (γ_k) entre dois períodos dependa apenas da distância, do intervalo ou da defasagem entre períodos e não do tempo real. Matos (2000) salienta que a violação do pressuposto

de estacionariedade pode acarretar em resultados espúrios, sendo assim, podem não retratar verdadeiramente a realidade.

Uma série temporal é dita estritamente estacionária quando todos os momentos de uma distribuição de probabilidade são invariantes ao longo do tempo, e não apenas o primeiro e o segundo momento, representados pela média e a variância. Em linhas gerais, as séries costumam ser fracamente estacionárias, ou estacionárias de segunda ordem (GUJARATI; PORTER, 2011). Esse processo satisfaz as seguintes condições:

$$E(Y_t) = \mu(t) = \mu \quad (3)$$

$$E(Y_t^2) < \infty \quad (4)$$

$$\gamma(t_1, t_2) = Cov(Y_{t_1}, Y_{t_2}) \quad (5)$$

2.6.2 Ergodicidade

Quando o mecanismo gerador de uma série temporal é realizado uma única vez e essa é suficiente para obter todas as estatísticas necessárias, é estabelecida uma condição de ergodicidade, ou propriamente, chamado de processo ergódico. De modo geral, todo processo dito ergódico também é estacionário, pois, em contrapartida, processos não estacionários não são capazes de gerar todas as especificações necessárias para a realização do processo (SOUZA; CAMARGO, 1996).

Segundo Bueno (2008) esta propriedade permite usar uma série para calcular as médias em cada instante de tempo, e devido ao fato das médias serem constantes, basta uma única e particular realização, r , do processo estocástico. Sendo assim, a média temporal pode ser definida como:

$$\bar{Y}^{(r)} = \frac{1}{n} \sum_{t=1}^n Y_t^{(r)} \quad (6)$$

Um processo é considerado ergódico para a média se as autocovariâncias, quando somadas, resultarem em um valor finito, conforme provado em Brockwell e Davis (2009), e definido como:

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty \quad (7)$$

Bueno (2008) complementa que se essa autocovariância vai a zero de forma rápida, quando aumenta o número de defasagens (k), então, é possível provar que a série é ergódica.

2.7 RUÍDO BRANCO

Uma sequência ε_t de variáveis aleatórias é dita ruído branco quando se distribuem de forma independente e identicamente distribuída (*i.i.d.*), com média zero e variância constante, isto é, $\varepsilon_t \sim i.i.d. (0, \sigma_\varepsilon^2)$. Pode-se dizer também que a condição de independência residual implica na ausência de autocorrelação, e de modo geral, um ruído branco também é caracterizado como sendo estacionário. (BOX *et al.*, 2015; BUENO, 2008),

Este processo é denominado ruído branco pois é comparado de forma análoga a "luz branca", na qual todas as cores estão presentes em quantidades iguais, ou seja, de forma homogênea (SOUZA; CAMARGO, 1996).

3 MODELOS LINEARES PARA SÉRIES TEMPORIAS ESTACIONÁRIAS

Neste capítulo é apresentada uma revisão bibliográfica quanto aos modelos lineares úteis na modelagem de séries temporais estacionárias.

3.1 MODELOS AUTORREGRESSIVOS

Segundo Maddala (2001), os processos autorregressivos (AR) são descritos como aqueles que os valores correntes de uma série Y_t dependem apenas de seus valores passados e de seus resíduos ε_t , distribuídos como um ruído branco. A equação geral do modelo autorregressivo de ordem p , denotado por $AR(p)$, é definida como:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (8)$$

essa equação pode ser reescrita como (BOX *et al.*, 2015):

$$\tilde{Y}_t = \phi_1 \tilde{Y}_{t-1} + \phi_2 \tilde{Y}_{t-2} + \dots + \phi_p \tilde{Y}_{t-p} + \varepsilon_t \quad (9)$$

Em que p representa o número de defasagens de Y_{t-p} e ϕ_p , μ representa o parâmetro da média e $\tilde{Y}_t = Y_t - \mu$.

O modelo $AR(p)$ também pode ser expresso na forma do operador de retardo (B) como (BOX *et al.*, 2015):

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \tilde{Y}_t = \varepsilon_t \quad (10)$$

ou simplesmente,

$$\phi(B) \tilde{Y}_t = \varepsilon_t$$

$$\tilde{Y}_t = \frac{1}{\phi(B)} \varepsilon_t = \phi(B)^{-1} \varepsilon_t \quad (11)$$

3.2 MODELOS DE MÉDIAS MÓVEIS

De acordo com Hamilton (1994), nos processos de médias móveis (MA) as Y_t observações de uma série temporal são representadas como uma soma ponderada, semelhante a uma média, dos resíduos $\varepsilon_t \sim (0, \sigma^2)$. Essas somas caracterizam-se como combinações lineares desses dos termos de ruído branco, assim o modelo MA de ordem q , $MA(q)$ pode ser expresso como:

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (12)$$

ou ainda,

$$\tilde{Y}_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (13)$$

Na forma do operador de retardo o modelo $MA(q)$ também pode ser descrito como (BOX *et al.*, 2015):

$$\tilde{Y}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (14)$$

ou de forma simplificada,

$$\tilde{Y}_t = \theta(B) \varepsilon_t \quad (15)$$

3.3 MODELOS AUTORREGRESSIVOS DE MÉDIAS MÓVEIS

Os processos autorregressivos de médias móveis (ARMA) nada mais são que a combinação dos processos AR e MA (BUENO, 2008). De modo geral, o modelo $ARMA(p,q)$ é denotado como:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (16)$$

ou como,

$$\tilde{Y}_t = \phi_1 \tilde{Y}_{t-1} + \phi_2 \tilde{Y}_{t-2} + \dots + \phi_p \tilde{Y}_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (17)$$

Onde q representa o número de defasagens de ε_{t-p} e θ_q .

Na forma do operador de retardo o modelo ARMA(p, q) é também representado como (BOX *et al.*, 2015):

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \tilde{Y}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (18)$$

ou simplesmente,

$$\phi(B) \tilde{Y}_t = \theta(B) \varepsilon_t$$

$$\tilde{Y}_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t = \theta(B) \cdot \phi(B)^{-1} \varepsilon_t \quad (19)$$

3.4 ESTACIONARIEDADE E INVERSIBILIDADE

De acordo com Souza e Camargo (1996) a estacionariedade e a invertibilidade dos modelos AR(p), MA(q) e ARMA(p, q) podem ser resumidas conforme o quadro a seguir:

Quadro 1 – Resumo das condições de estacionariedade e inversibilidade dos modelos AR, MA e ARMA.

Modelo	Equação	Características
AR(p)	$\tilde{Y}_t = \phi(B)^{-1} \varepsilon_t$	Raízes de $\phi(B) = 0$ fora do círculo unitário <ul style="list-style-type: none"> • Estacionário e Trivialmente inversível
MA(q)	$\tilde{Y}_t = \theta(B) \varepsilon_t$	Raízes de $\theta(B) = 0$ fora do círculo unitário <ul style="list-style-type: none"> • Inversível e Trivialmente estacionário
ARMA(p, q)	$\tilde{Y}_t = \theta(B) \cdot \phi(B)^{-1} \varepsilon_t$	Raízes de $\phi(B) = 0$ fora do círculo unitário <ul style="list-style-type: none"> • Estacionário Raízes de $\theta(B) = 0$ fora do círculo unitário <ul style="list-style-type: none"> • Inversível

Fonte: Adaptado pelo autor com base em Souza e Camargo (1996).

3.5 FUNÇÃO DE AUTOCORRELAÇÃO E FUNÇÃO DE AUTOCORRELAÇÃO PARCIAL

A Função de autocorrelação (FAC) de defasagem k , denotada por ρ_k , representa a autocovariância (γ_k) padronizada. Essa medida apresenta resultados entre -1 e +1 e serve para medir o comprimento e a memória de um processo, ou seja, a dependência temporal existente na série analisada (GUJARATI; PORTER, 2011; SOUZA; CAMARGO, 1996).

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t) \cdot Var(Y_{t+k})}} \quad (20)$$

Onde:

$Var(Y_k) = Var(Y_{t+k}) = \gamma_0 =$ Variância do processo.

Como na prática têm-se apenas a realização amostral de um processo estocástico, logo há apenas a possibilidade de se obter a função de autocorrelação (FAC) amostral, conforme a Equação 21 (GUJARATI; PORTER, 2011; SOUZA; CAMARGO, 1996).

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (21)$$

Em que: $\hat{\gamma}_k$ e $\hat{\gamma}_0$ representam, respectivamente, a covariância amostral e variância amostral com k defasagens, e \bar{Y} representa $\frac{1}{n} \sum_{t=1}^n Y_t$.

A autocorrelação parcial corresponde a correlação entre duas observações seriais Y_t e Y_{t+k} , desconsiderando a dependência dos termos intermediários

$(Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1})$, e pode ser denominada pela equação 22. De modo geral, a função de autocorrelação parcial (FACP), ϕ_{kk} , pode ser obtida a partir do sistema de equações de Yule-Walker (Equação 23), e a FACP amostral, $\hat{\phi}_{kk}$, pode ser obtida substituindo os ρ_i por $\hat{\rho}_i$ nas equações de Yule-Walker (SOUZA; CAMARGO, 1996; MORETTIN; TOLOI, 2004).

$$\text{Cor}(Y_t, Y_{t+k} | Y_{t+1}, \dots, Y_{t+k-1}) \quad (22)$$

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (23)$$

De acordo com Souza e Camargo (1996) e Morettin e Toloí (2004), a FAC de um modelo $\text{AR}(p)$ apresenta comportamento decrescente quando k tende ao infinito, e a FACP apresenta comportamento finito para $k \leq p$, anulando-se de forma brusca após as defasagens $k > p$.

Em contrapartida, a FAC de um $\text{MA}(q)$ apresenta comportamento finito e anula-se após $k > p$, e a FACP apresenta comprimento infinito com decrescimento geométrico sem anular-se bruscamente. Já o modelo $\text{ARMA}(p, q)$ tem FAC e FACP infinita que decaem de acordo com exponencias e/ou senoides amortecidas, respectivamente, após $k > p - q$ e $k > q - p$, sem anularem-se bruscamente. (SOUZA; CAMARGO, 1996; MORETTIN; TOLOI, 2004).

4 MODELOS LINEARES PARA SÉRIES TEMPORIAS NÃO ESTACIONÁRIAS

Neste capítulo é apresentada uma revisão bibliográfica quanto aos modelos lineares úteis na modelagem de séries temporais não estacionárias.

4.1 INTRODUÇÃO

Sabe-se que, na prática, as séries temporais raramente se comportam de forma não estacionária. Mesmo assim, muitas destas apresentam certa homogeneidade de modo que, além do nível local, ou talvez nível local e tendência, uma parte da série se comporta como qualquer outra. Esses modelos capazes de descrever comportamentos não estacionários homogêneos podem ser obtidos a partir de transformações como o operador de diferenças, ∇^d , ou algum outro tipo de transformação para tornar o processo estacionário (BOX *et al.*, 2015).

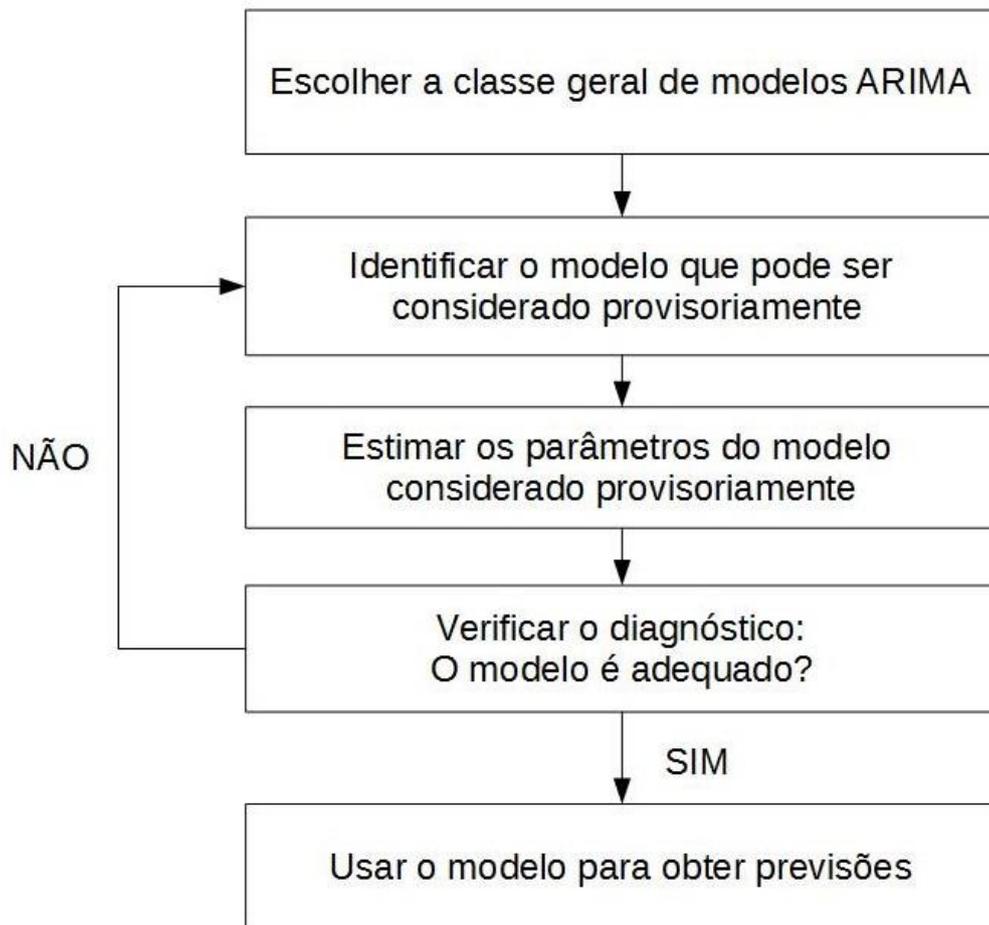
Nesse contexto, Box e Jenkins (1970) propuseram uma metodologia na qual possibilitou analisar séries com comportamentos estacionários e não estacionários, e também com ou sem a presença de termos sazonais. Maddala (2001) considera que este tipo de abordagem se tornou bastante popular, justamente, por essa capacidade de generalização na análise de dados, destacando-se como uma das metodologias mais utilizadas na análise de séries temporais. Devido a esse destaque, os modelos ARMA passaram a ser, frequentemente, confundidos como modelos de Box-Jenkins.

A metodologia de Box-Jenkins pode ser descrita em quatro etapas baseadas na identificação de modelos, estimação dos parâmetros dos modelos, verificação da adequação dos modelos e obtenção de previsões a partir desses modelos (MATOS, 2000; GUJARATI; PORTER, 2011; BOX *et al.*, 2015).

Maddala (2001) descreve ainda uma etapa anterior a identificação de modelos, na qual consiste em analisar a estacionariedade da série, e realizar sucessivas diferenças para tornar a série estacionária, caso necessário. Em Matos (2001), Gujarati e Porter (2011) e Box *et al.* (2015) a análise da estacionariedade e a identificação de modelos são descritas conjuntamente.

Sendo assim, de modo geral, a metodologia de Box-Jenkins pode ser resumida como um ciclo iterativo conforme descrito na Figura 1.

Figura 1 – Ciclo iterativo utilizado na metodologia de Box-Jenkins.



Fonte: Adaptação de Box *et al.* (2015).

Para Souza e Camargo (1996) um princípio importante da metodologia de Box-Jenkins leva em consideração o princípio da parcimônia. Segundo Tukey (1961) modelos parcimoniosos são aqueles que se utilizam do menor número de parâmetros possíveis para explicar o comportamento de uma série temporal.

4.2 IDENTIFICAÇÃO DE MODELOS ARIMA

A etapa de identificação de modelos ARIMA (p, d, q) pode ser descrita em dois estágios: a determinação das d diferenciações necessárias para tornar a série

estacionária e a determinação dos número de parâmetros p e q do modelo ARMA. Em ambos os estágios, a identificação das ordens p , d e q do modelo ARIMA pode ser realizada a partir da FAC e da FACP e os correlogramas resultantes dessas funções GUJARATI; PORTER, 2011; SOUZA; CAMARGO, 1996).

Pode-se observar que os modelos ARIMA são representados pela combinação de três componentes, também denominados “filtros”: o filtro AR estacionário, associado aos p parâmetros ϕ , o filtro de Integração não estacionário (I) associado as d diferenciações, e filtro de MA, associado aos q parâmetros θ (BOX *et al.*, 2015; MORETTIN; TOLOI, 2004)

Além disso, salienta-se que nem todas as séries apresentam estes três componentes citados. Há também a possibilidade da existência de apenas parte desses filtros, assim sendo, a metodologia de Box-Jenkins resulta em diferentes modelos passíveis de análise (VASCONCELLOS; ALVES, 2000).

4.2.1 Determinação das “ d ” diferenças

Em séries consideradas estacionárias não há a necessidade da realização de diferenciações, pois a série apresenta de forma estável. Nos casos em que as séries são não estacionárias homogêneas, conforme já comentado anteriormente, é possível torná-las estacionárias utilizando o operador de diferenças tomando-se um número finito de diferenças d (MORETTIN; TOLOI, 2004; SOUZA; CAMARGO, 1998). Em geral, conforme observado na prática, e também de acordo com Makridakis, Whellwright e Hyndman (1998) e Morettin e Tolo (2004), uma ou duas diferenças são suficientes para tornar uma série temporal estacionária.

4.2.1.1 Testes de raiz Unitária

Para avaliar a estacionariedade de uma série temporal podem-se realizar testes de raiz unitária. Na literatura existem diversos testes, entre estes estão: os testes de Dickey e Fuller (1979,1981), conhecidos respectivamente, como teste Dickey-Fuller (DF) e teste Dickey-Fuller aumentado (ADF), o teste de Phillips e Perron (1984), conhecido como teste Phillips-Perron (PP), o qual faz uma correção não paramétrica no teste ADF e o teste proposto por Kwiatkowski *et. al* (1992), o

qual recebe em seu nome as iniciais do nome dos autores Kwiatkowski, Phillips, Schmidt e Shin (KPSS).

Segundo Bueno (2008) os testes DF e ADF possuem baixo poder um baixo poder, por isso não estão aqui apresentados. Para maiores informações, a formulação matemática desses testes pode ser consultada em Gujarati e Porter (2011), Bueno (2008) e Mackinnon (1996).

O teste KPSS possui maior poder de teste em relação ao DF, ADF e PP. As hipóteses desse teste são:

H_0 : A série é estacionária (Ausência de raiz unitária)

H_1 : A série não é estacionária (Presença de raiz unitária)

O teste KPSS pode ser avaliado constante e tendência, apenas com constante, ou com constante e tendência. A estatística do teste KPSS é um multiplicador de Lagrange (LM), conforme equação 24, e os valores críticos são aqueles tabulados por KPSS (1992) de acordo com os resultados assintóticos descobertos pelos autores.

$$LM_{KPSS} = \frac{1}{n^2} \sum_{i=1}^n \frac{J_i^2}{\hat{v}^2} \quad (24)$$

Em que: $J_i = \sum_{i=1}^n \hat{\varepsilon}_i$ e \hat{v}^2 é a variância de longo prazo.

4.2.2 Determinação das ordens p e q

Após a determinação das d diferenças da ordem Integrada $I(d)$ do modelo, são determinadas as ordens p e q dos termos AR e MA. Este processo pode ser realizado com o auxílio da FAC e FACP e segue as mesmas características descritas na seção 3.5.

4.3 ESTIMAÇÃO DE MODELOS ARIMA

O segundo passo do ciclo iterativo da metodologia de Box-Jenkins consiste em estimar os parâmetros do modelo ARIMA (p, d, q) identificado.

Conforme Morettin e Tolo (2004) e Box et al. (2015) o modelo ARIMA (p, d, q) geral pode ser definido por:

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (25)$$

onde $w_t = \nabla^d Y_t = (1 - B)^d Y_t$.

E na forma do operador de retardo pode ser escrito como:

$$\phi(B)w_t = \theta(B)\varepsilon_t \quad (26)$$

Em que:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \text{ e } \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q.$$

A estimação de parâmetros pode ser realizada utilizando o método dos mínimos quadrados (ordinários ou não lineares) ou pelo método da máxima verossimilhança (condicional ou incondicional). A ênfase nesse trabalho refere-se à estimação pelo método da máxima verossimilhança Condicional e incondicional. Para maiores informações consultar Bueno (2008).

4.3.1 Critérios de seleção de modelos

O processo de escolha do modelo mais parcimonioso pode ser realizado utilizando critérios de seleção de modelos. Esses critérios permitem identificar um modelo candidato a melhor ajuste aos dados, pois possibilitam detectar o número ideal de parâmetros a serem ajustados.

Os critérios mais utilizados são o critério de informação de Akaike (AIC) e critério de informação de Schwarz ou Bayesiano (BIC), desenvolvidos respectivamente por Akaike (1973) e Schwarz (1978), cujas equações são representadas pelas equações 27 e 28.

$$AIC = \log(\hat{\sigma}_\varepsilon^2) + \frac{2(p+q)}{n} \quad (27)$$

$$BIC = \log(\hat{\sigma}_\varepsilon^2) + \frac{(p+q)\log(n)}{n} \quad (28)$$

onde \log equivale ao logaritmo na base e , $\hat{\sigma}_\varepsilon^2$ é a variância estimada dos resíduos ε_t , e n ao tamanho da amostra analisada.

De acordo com Bueno (2008), o resultado desses critérios penaliza o número de parâmetros e pela variância dos resíduos de cada modelo. O modelo escolhido deve ser aquele que apresenta os menores valores de AIC e BIC.

4.4 ANÁLISE DE DIAGNÓSTICO

A terceira etapa consiste em realizar uma análise de diagnóstico dos resíduos a fim avaliar a adequação do modelo ajustado. O desejado é que os resíduos sejam independentes e identicamente distribuídos (*i.i.d*) e não correlacionados entre si.

Conforme Matos (2001), a análise da ausência de autocorrelação pode ser avaliada a partir da FAC dos resíduos observando se estes encontram-se dentro dos limites estabelecidos. A verificação da ausência de autocorrelação dos resíduos também pode ser realizada utilizando o teste Q' de Ljung-Box proposto por Ljung e Box (1978). Esse teste é uma variação do teste de Box e Pierce (1970) e pode ser calculado através da equação 29.

$$Q'(k) = n(n+2) \sum_{k=1}^K \frac{\hat{\rho}_k^2(\hat{\varepsilon})}{n-k} \quad (29)$$

A estatística $Q'(k)$ tem distribuição χ^2 com $k - p - q$ graus de liberdade.

A hipótese H_0 do teste é de os resíduos são *i.i.d*, e comportam-se como um ruído branco, e a hipótese alternativa, H_1 , é de os resíduos não são *i.i.d* e, portanto,

não comportam-se como um ruído branco. Se a hipótese H_0 for rejeitada deve-se escolher um novo modelo e realizar novamente os procedimentos anteriores da metodologia de Box-Jenkins.

Deve-se também testar a normalidade dos resíduos do modelo ajustado. Segundo Razali e Wah (2011), há na literatura diversos testes de normalidade, entre estes os autores citam os testes Shapiro-Wilk, Anderson-Darling, Lilliefors, Jarque-Bera, Cramer-von Mises, D'Agostino Pearson e Kolmogorov-Smirnov.

Quanto ao poder dos testes, Carradori e Ramos (2014) consideram os testes de Shapiro-Wilk e Anderson-Darling os mais poderosos para distribuições simétricas com caudas curtas; Em distribuições simétricas com caudas longas, os testes de Jarque-Bera (JB) e D'Agostino-Pearson equiparam com o teste de Shapiro-Wilk; Para distribuições assimétricas o teste Shapiro-Wilk é o mais o mais poderoso, seguido do teste Anderson-Darling.

Entre os testes de normalidade citados, o proposto por Jarque e Bera (1987) é o mais recente. Este é um teste assintótico cujas hipóteses a serem testadas são:

H_0 : Não há evidências para rejeitar a normalidade dos resíduos do modelo.

H_1 : Há evidências para rejeitar a normalidade dos resíduos do modelo.

A estatística do teste JB baseia-se nos coeficientes estimados assimetria (\hat{b}_1) e curtose (\hat{b}_2) dos resíduos, conforme equação a seguir:

$$JB = n \left[\frac{\hat{b}_1}{6} + \frac{(\hat{b}_2 - 3)^2}{24} \right] \quad (30)$$

Em que:

$$\hat{b}_1 = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^3}{\left(\frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 \right)^{3/2}} \quad \text{e} \quad \hat{b}_2 = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^4}{\left(\frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 \right)^2} .$$

Sob a hipótese H_0 a estatística JB (Equação 30) segue uma distribuição qui-quadrado com 2 graus de liberdade. Essa hipótese é rejeitada se $JB > \chi^2_{(\alpha,2)}$, ou se o p-valor for menor que o nível de significância (α).

4.5 PREVISÃO

A última etapa consiste em gerar previsões acerca do modelo ajustado e avaliar a qualidade dessas previsões.

4.5.1 Medidas de precisão de previsão

Após estimação dos parâmetros do modelo escolhido, análise de diagnóstico dos resíduos e obtenção de previsões é conveniente avaliar a precisão ou qualidade das previsões obtidas. Para isso, Makridakis, Wheelwright e Hyndman (1998) apresentam o erro de previsão (FE) como a medida mais simples para avaliar a qualidade das previsões geradas pelo modelo. Essa medida é calculada, conforme a equação 31, a partir da diferença entre a observação Y em um instante de tempo t e respectivo valor previsto (ou valor estimado \hat{Y}_t) para o período considerado.

$$FE = Y_t - \hat{Y}_t \quad (31)$$

Observa-se que o FE é bastante limitado e restringe-se a medir os erros a um período específico da série analisada. Quando o interesse for avaliar previsões para n períodos de tempo, tem-se outras medidas. Nesse caso, Makridakis, Wheelwright e Hyndman (1998) citam o erro médio (ME), o erro absoluto médio (MAE) e o erro quadrático médio (MSE), respectivamente, representados por:

$$ME = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t) \quad (32)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (33)$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \quad (34)$$

Os autores reiteram que, é possível que o ME gere resultados pequenos pois, como os FE podem assumir resultados tanto positivos quanto negativos, então estes tendem a se anular ou subestimar interpretação desta medida. Em contrapartida, o MAE e o MSE solucionam esse problema de subestimação, tendo o MAE a vantagem da fácil interpretação em relação ao ME, e o MSE a vantagem de ser mais fácil de manipular matematicamente.

Hyndman e Koehler (2006) complementam apresentando a raiz do erro quadrático médio (RMSE), definindo-a na sequência. Essa medida tem a vantagem de apresentar o resultado na mesma unidade de medida da variável. Porém, os autores criticam a impossibilidade de comparação entre outras séries de escalas diferentes.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (35)$$

As medidas que apresentam melhores resultados são o erro percentual médio (MPE) (Equação 36) e o erro percentual médio absoluto (MAPE) (Equação 37), em relação as anteriores. Essas, encontram-se na forma de percentuais sendo passíveis de comparação com outras séries pois estão livres escala (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; HYNDMAN; KOEHLER, 2006).

$$MPE = \frac{1}{n} \sum_{t=1}^n \left[\left(\frac{Y_t - \hat{Y}_t}{Y_t} \right) \times 100 \right] \quad (36)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \left(\frac{Y_t - \hat{Y}_t}{Y_t} \right) \times 100 \right| \quad (37)$$

A última medida apresentada é a estatística U (Equação 38), desenvolvida por Theil (1966). Essa medida estatística permite uma comparação relativa dos métodos formais de previsão com as previsões ingênuas e também coloca os erros envolvidos, de modo que grandes erros recebem muito mais peso do que os pequenos erros. É desejado que o resultado do U de Theil seja menor que 1 para que as previsões do utilizado sejam melhores as previsões ingênuas (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998).

$$U = \sqrt{\frac{\sum_{t=1}^{n-1} \left(\frac{\hat{Y}_{t+1} - Y_{t+1}}{Y_t} \right)^2}{\sum_{t=1}^{n-1} \left(\frac{Y_{t+1} - Y_t}{Y_t} \right)^2}} \quad (38)$$

4.6 ESTUDOS APLICADOS UTILIZANDO A METODOLOGIA DE BOX-JENKINS

A metodologia de Box-Jenkins foi utilizada por Earnest *et al.* (2005) no monitoramento e previsão do número de leitos ocupados, durante um surto de síndrome respiratória aguda severa (SARS) ocorrido no ano de 2003, em um hospital terciário em Cingapura. Os autores identificaram um comportamento estacionário para a série analisada, sendo o modelo ARIMA (1,0,3) o mais adequado, e as previsões obtidas a partir desse modelo foram consideradas razoavelmente boas com um MAPE de 8,7%.

Earnest *et al.* (2005) salientam que este foi o primeiro estudo a utilizar os modelos ARIMA no controle e previsão de camas de isolamento durante o surto de SARS em Cingapura. Os autores reiteram a importância dos modelos ARIMA para administradores e clínicos no planejamento do uso de camas de isolamento durante surtos, não somente de SARS, mas também de outras doenças infecciosas, bem como na previsão de requisitos para outros recursos críticos. Quanto a estudos relacionados a cuidados em saúde utilizando a metodologia de Box-Jenkins, os

autores identificaram uma literatura limitada, com maiores aplicações no setor financeiro.

O estudo de Akhtar e Rozi (2009) tinha como objetivo utilizar a metodologia de Box-Jenkins para realizar previsões, no curto prazo sobre a soroprevalência do vírus da hepatite C (HCV) em doadores de sangue do sexo masculino em Karachi, no Paquistão entre 1998 e 2005. Os autores utilizaram o AIC para identificar o modelo de melhor ajuste, sendo o ARIMA (2,1,7) considerado o mais parcimonioso. Na análise de adequação do modelo a estatística $Q'(k)$ de Ljung-Box apresentou p-valor= 0,433, demonstrando um comportamento *i.i.d* aos resíduos, ou seja, uma boa adequação do ajuste.

Diferente do estudo realizado por Earnest *et al.* (2005), a série temporal analisada por Akhtar e Rozi (2009) apresentou um comportamento não estacionário, por isso foi necessária a utilização da série em primeira diferença. Quanto ao modelo de melhor ajuste, esse estudo apresentou maior complexidade em relação ao da ocupação de leitos, pois foi necessário a estimação de um número superior de parâmetros AR e MA para explicar a soroprevalência de HCV em doadores de sangue do sexo masculino.

Quanto as previsões obtidas pelo modelo ajustado por Akhtar e Rozi (2009) para os seis meses posteriores a série apresentaram 6,5% de MPE, sendo considerados satisfatórios. Contudo, essas demonstraram uma tendência crescente de soroprevalência de HCV entre os doadores de sangue em Karachi. Em consequência disso, os autores acreditam que esse aumento previsto da soroprevalência de HCV pode estar relacionado a esforços inconsistentes e ingênuos de controle de HCV por parte de funcionários de saúde pública no Paquistão.

Akhtar e Rozi (2009) destacam que essa é a primeira aplicação dos modelos ARIMA para monitorar e prever a soroprevalência do HCV em doadores de sangue em múltiplos bancos de sangue. Destacam ainda que modelo ARIMA ajustado poderia ser usado pelas autoridades de saúde, no controle e prevenção do HCV e outros agentes transmitidos pelo sangue em países com recursos limitados, incluindo o Paquistão.

Liu *et al.* (2011) utilizaram o para obter previsões da incidência de febre hemorrágica com síndrome renal na China. Este estudo considerou a série temporal entre 1978 e 2008. Os resultados obtidos pelos autores demonstraram que o modelo

ARIMA (0,3,1) apresentou menor AIC, e por isso considerado modelo mais parcimonioso entre os demais modelos comparados. A análise dos resíduos desse modelo evidenciou o comportamento *i.i.d* e a avaliação da precisão das previsões realizadas a partir do modelo estimado apresentou um MAPE de 12,2%.

Liu *et al.* (2011) consideram que o modelo ARIMA é uma ferramenta importante para a vigilância da incidência de febre hemorrágica com síndrome renal na China. Consideraram ainda que, o estudo realizado mostrou que é possível obter previsões precisas para a incidência analisada.

A metodologia de Box-Jenkins também foi utilizada para estimação e predição de prevalência de cárie da primeira infância na China continental. Neste estudo foram considerados os dados de 1988 a 2010 para estimação do modelo ARIMA e 2011 a 2013 para realização de previsões. O modelo de melhor ajuste nesse tudo foi o ARIMA(2,1,3) e o MAPE obtido para as previsões foi de 5,74%. Os resultados obtidos mostraram que o estado de saúde bucal na China melhorou ao longo do tempo e também que o modelo ARIMA apresentou bons resultados e pode ser usado no ajuste e previsão de cárie da primeira infância na China Continental (ZHANG *et. al*, 2017)

5 NEOPLASIAS MALIGNAS DA PELE

Apresenta-se nesta sessão uma revisão de literatura relacionada a neoplasias malignas de pele.

5.1 DEFINIÇÃO DE NEOPLASIAS MALIGNAS DE PELE

De acordo com a Classificação estatística internacional de doenças e problemas relacionados à saúde – Décima revisão (CID-10), estabelecida pela Organização Mundial da Saúde (OMS), as neoplasias malignas de pele correspondem aos melanomas e outras (os) neoplasias (tumores) malignas (os) da pele (C43-C44). Essas neoplasias malignas de pele estão descritas no Capítulo II da CID-10, relacionado as neoplasias (tumores), e dividem-se em dois grandes grupos: os melanomas malignos da pele (C43) e outros tipos de neoplasias da pele (C44), as quais compreendem a todas as doenças especificadas no quadro 1. (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 1994).

Quadro 2 – Descrição dos tipos de melanomas e outras (os) neoplasias (tumores) malignas (os) da pele (C43-C44) de acordo com a CID-10.

(continua)

Códigos	Melanomas malignos da pele (C43)
C43.0	Melanoma maligno do lábio
C43.1	Melanoma maligno da pálpebra, incluindo as comissuras palpebrais
C43.2	Melanoma maligno da orelha e do ducto auditivo externo
C43.3	Melanoma maligno de outras partes e partes não especificadas da face.
C43.4	Melanoma maligno do couro cabeludo e do pescoço
C43.5	Melanoma maligno do tronco
C43.6	Melanoma maligno do membro superior, incluindo ombro
C43.7	Melanoma maligno do membro inferior, incluindo quadril
C43.8	Melanoma maligno invasivo da pele
C43.9	Melanoma maligno de pele, não especificado
Códigos	Outros tipos de neoplasias malignas da pele (C44)
C44.0	Neoplasia maligna da pele do lábio
C44.1	Neoplasia maligna da pele da pálpebra, incluindo o canto
C44.2	Neoplasia maligna da pele da orelha e do ducto auditivo externo
C44.3	Neoplasia maligna da pele de outras partes e partes não especificadas da face.
C44.4	Neoplasia maligna da pele do couro cabeludo e do pescoço
C44.5	Neoplasia maligna da pele do tronco

Quadro 2 – Descrição dos tipos de melanomas e outras (os) neoplasias (tumores) malignas (os) da pele (C43-C44) de acordo com a CID-10.

(conclusão)

C44.6	Neoplasia maligna do membro superior, incluindo ombro
C44.7	Neoplasia maligna do membro inferior, incluindo quadril
C44.8	Neoplasia maligna de pele sobreposta
C44.9	Neoplasia maligna da pele, não especificado

Fonte: elaborado pelo autor com base nas informações da WHO (2007).

O grupo associado aos melanomas malignos de pele inclui os casos de morfologias classificáveis como nevos e melanomas (M872-M879) com comportamento biológico do tipo 3, isto é, os tumores malignos de localização primária. No entanto, desconsidera os melanomas malignos da pele dos órgãos genitais (C51-C52, C60 e C63). Na categoria C43.0 são excluídos os casos relacionados a área vermelha (vermelhão) do lábio (C00.0-C00.2). Exclui-se também os casos relacionados ao ânus Sonda oroenteral (SOE) (C21) em C43.5 (WORLD HEALTH ORGANIZATION, 2007).

O grupo associado a outros tipos de neoplasias malignas de pele inclui os casos relacionados as neoplasias malignas das glândulas sebáceas e sudoríparas e exclui o sarcoma de kaposi (C46) e, assim como em C43, os de pele de órgãos genitais (C51-C52, C60 e C63). São excluídos os casos de neoplasias do lábio (C00) na categoria 40.1, os casos ligados ao tecido conjuntivo da pálpebra (C49.0) na categoria 40.2, e também os casos ligados ao tecido conjuntivo da orelha (C49.0) na categoria 40.3. Exclui-se também os casos relacionados ao ânus SOE (C21) da categoria 44.5, do mesmo modo que em C43 (WORLD HEALTH ORGANIZATION, 2007).

5.2 A IMPLANTAÇÃO DA CID-10 NO BRASIL

A aprovação da CID-10 ocorreu em 1989 na Conferência Internacional para a Décima Revisão e definido pela Quadragésima Terceira Assembleia Mundial de Saúde para entrar em vigor em 1º de janeiro de 1993. Até então, utilizava-se oficialmente a Classificação estatística internacional de doenças e problemas relacionados à saúde – Nona revisão (CID-9) (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 1994).

A utilização desta classificação em território brasileiro ocorreu a partir de 1º de janeiro de 1996, para fins de notificações de óbitos, em cumprimento a uma portaria publicada em 3 de novembro de 1994 (GRASSI; LAURENTTI, 1998). Posteriormente, a partir de 1º de janeiro de 1998, o uso da CID-10 foi estendido a todo território nacional a morbidades hospitalares e ambulatoriais, após publicação da Portaria de número 1.311, em 12 de setembro de 1997 no Diário da União (MINISTÉRIO DA SAÚDE, 1997).

5.3 FONTE DE DADOS OFICIAIS SOBRE MORTALIDADE NO BRASIL

Há no Brasil o Sistema de Informações Sobre Mortalidade (SIM), desenvolvido pelo Ministério da Saúde, em 1975, para fins de coleta de dados sobre mortalidade. Esse sistema reúne dados quantitativos e qualitativos sobre os óbitos ocorridos no país, por isso se destaca como uma importante ferramenta de gestão na área da saúde (PORTAL DA SAÚDE, 2017).

A coleta dos dados do SIM baseia-se na Declaração de Óbito (DO), documento emitido sob a responsabilidade de um médico, conforme previsto no artigo 115 do Código de Ética Médica, no artigo 1º da Resolução nº 1779/2005 do Conselho Federal de Medicina e na portaria nº 116/2009 do Sistema de Vigilância em Saúde (SVS). As variáveis contidas e coletadas a partir das DO's permitem a obtenção de indicadores e análises epidemiológicas auxiliando no processo de melhorias em saúde (PORTAL DA SAÚDE, 2017).

Os dados coletados pelo SIM são gerenciados pelo Departamento de Informática do Sistema Único de Saúde (DATASUS), órgão da Secretaria de Gestão Estratégica e Participativa do Ministério da Saúde, criado em 1991, cuja missão é prestar suporte e manutenção das bases de dados nacionais. O acesso a estes dados está disponível no site do DATASUS através do TABNET, tabulador genérico de domínio público, de fácil e intuitiva utilização (DEPARTAMENTO DE INFORMÁTICA DO SUS, 2017).

Segundo estudo realizado por Mathers *et al.* (2005), publicado nos boletins da OMS, o SIM foi classificado com um sistema de qualidade intermediária. A partir desse estudo, o Ministério da saúde promoveu iniciativas como: o desenvolvimento de novos aplicativos informatizados e cursos de formação e capacitação relacionado

a codificação de causas básicas, a fim de fortalecer, aprimorar e melhorar a qualidade do sistema.

5.4 INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA

O Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) é um órgão brasileiro auxiliar do Ministério da Saúde, responsável por desenvolver e coordenar ações integradas no combate, prevenção e controle do câncer no Brasil. Entre as ações prestadas destaca-se a assistência médico-hospitalar, direta e gratuitamente aos pacientes com câncer, atuação áreas estratégicas voltadas a prevenção e diagnóstico precoce, formação de profissionais especializados, assim como o desenvolvimento de pesquisas e estudos epidemiológicos.

O INCA coordena diversos programas nacionais para controle do câncer como: Programa Nacional de Controle do Tabagismo, Controle do Câncer Colo do Útero, Controle do Câncer de Mama e Programa de Transplante de Medula Óssea, entre outros. Esses programas citados são aqueles que apresentam maior atenção do instituto. Apesar disso, não há um programa focado no controle do câncer de pele, que é o tipo de câncer de maior incidência do Brasil.

5.4.1 Estimativas feitas pelo INCA

Este item dedica-se exclusivamente a apresentar estimativas feitas pelo INCA (2014), para os anos de 2014 e 2015, e INCA (2015), para os anos de 2016 e 2017, quanto incidência de neoplasias malignas (câncer) de pele em nível nacional e também para as Regiões Sudeste e Sul.

De acordo com as estimativas para 2014 e 2015, esperavam-se 188.020 novos casos de câncer (neoplasias malignas) de pele no Brasil, sendo 101.380 casos para o sexo masculino e 86.640 casos para o sexo feminino. A taxa bruta brasileira geral (para cada 100 mil habitantes) esperada para os mesmos períodos citados anteriormente, era equivalente a 188,87 casos, sendo 103,78 para o sexo masculino e 85,09 para o sexo feminino.

Quanto as estimativas feitas para os anos de 2016 e 2017 comparando-as com 2014 e 2015, houve uma pequena redução nacional de 3,5%, de 188.020 para

181.430, no número esperado de novos casos. Considerando-se o esperado quanto ao sexo masculino, a redução representou cerca de 17,3%, de 101380 para 83850, e quanto ao sexo feminino houve um crescimento de cerca de 12,63%, de 86640 para 97580. A referida taxa bruta geral esperada para 2016 e 2017 recuou 5,09%, de 188,87 para 179,26, em nível nacional, 19,35% de redução, de 103,78 para 83,69, para o sexo masculino e 11,14% de aumento, de 85,09 para 94,57, para o sexo feminino.

No que diz respeito às estimativas para os anos de 2014 e 2015, nas Regiões Sudeste e Sul, eram esperados respectivamente 105830 e 36790 novos casos de câncer de pele, 55850 e 23350 referentes ao sexo masculino, 49980 e 13560 ao sexo feminino. As taxas brutas esperadas, respectivamente, para as regiões Sudeste e Sul foram superiores à média nacional, 252,44 (+33,66%) e 258,12 (+36,67%), assim quanto ao sexo masculino, 136,67 (+31,69%) e 166,06 (+60,01%) e quanto ao sexo feminino, 115,77 (+36,06%) e 92,06 (+8,19%).

Quanto as estimativas, para os anos de 2016 e 2017, nas Regiões Sudeste e Sul, são esperados respectivamente 99350 e 35500 novos casos de câncer de pele, 39590 e 20720 referentes ao sexo masculino, 59760 e 14780 ao sexo feminino. Observou-se uma redução de 7,93% (de 252,44 para 232,41) na taxa bruta esperada para Região Sudeste comparando as estimativas de 2014/2015 e 2016/2017, e uma redução de 4,77% (de 258,12 para 245,79) na taxa bruta esperada para a Região Sul.

Quando comparadas as taxas brutas esperadas, para os anos de 2016 e 2017, nas regiões Sudeste e Sul as esperadas em nível nacional, foram verificadas taxas superiores, respectivamente, em ambas as regiões: 232,41 (+29,65%) e 245,79 (+37,65%), assim quanto ao sexo masculino, 95,69 (+14,33%) e 145,71 (+74,10%) e quanto ao sexo feminino, 136,72 (+44,57%) e 100,08 (+5,83%).

Sendo assim, fica evidente que, apesar de ter ocorrido uma leve redução nas taxas esperadas para 2016 e 2017 em relação a 2014 e 2015, em ambas as regiões citadas, as mesmas apresentaram taxas esperadas bastante acima da média nacional deixando explícita a necessidade de maiores estudos nessas regiões do Brasil.

6 MATERIAIS E MÉTODOS

São apresentados neste capítulo elementos metodológicos relacionados a caracterização da pesquisa, descrição da variável, do período e da região analisada, fonte dos dados, assim com os procedimentos metodológicos utilizados nas análises estatísticas.

6.1 CARACTERIZAÇÃO DA PESQUISA

Realizou-se um estudo quantitativo, descritivo e inferencial, aplicado a um contexto epidemiológico cujo objetivo foi analisar as frequências absolutas de óbitos por neoplasias malignas de pele, nas regiões Sudeste e Sul do Brasil, entre os anos de 1996 e 2015. Tratou-se de um estudo de séries temporais univariado, pois as regiões foram analisadas em um espaço de tempo, cujo enfoque foi descrever o problema no período analisado, comparar as regiões analisadas, realizar a modelagem da série de dados e obter previsões de novos casos, explicadas unicamente pela dependência temporal existente entre seus períodos passados.

Além disso, pode ser definido como um estudo ecológico, pois as frequências absolutas referentes aos óbitos foram analisadas considerando o nível agregado, ou seja, a ocorrência de óbitos em grupos específicos, que nesse contexto, refere-se as regiões Sul e Sudeste do Brasil. As observações das séries foram agregadas em períodos trimestrais, equidistantes, ao longo do período analisado.

Realizou-se também uma pesquisa bibliográfica relacionada à análise de séries temporais, utilizando materiais impressos e eletrônicos, tais como: livros, artigos científicos e páginas da internet, os quais permitiram um maior conhecimento acerca do assunto estudado. Assim, a partir desse tipo de pesquisa foi possível obter subsídios substanciais para realização de uma análise correta, estatisticamente bem executada e que expresse de forma mais fidedigna possível o contexto estudado.

Quanto ao contexto relacionado as informações referentes as neoplasias malignas de pele, foi efetuada uma pesquisa documental baseada na consulta de documentos oficiais, registros e relatórios. Esse tipo de pesquisa possibilitou um maior aporte da mensuração de informações associadas a esse contexto

pesquisado, contribuindo no processo teórico do desenvolvimento de justificativas que reforçam a necessidade da realização deste estudo.

6.2 DESCRIÇÃO DA VARIÁVEL ANALISADA

Considerou-se todos óbitos que apresentaram como causa básica às neoplasias malignas de pele, ou seja, os dados populacionais relacionados a essa doença. Esses óbitos correspondem aos melanomas malignos de pele (C43) e a outros tipos de neoplasias malignas de pele (C44), e de forma mais específica, a todos os seus subtipos, conforme descrito no Quadro 1.

6.3 DESCRIÇÃO DO PERÍODO ANALISADO

Considerou-se 80 observações de periodicidade trimestral, entre os anos de 1996 a 2015. Utilizou-se apenas óbitos ocorridos a partir de 1996, ano que marca o início da utilização da CID-10 como critério de classificação de doenças e agravos relacionados a saúde no Brasil.

6.4 DESCRIÇÃO DA REGIÃO ANALISADA

Analisou-se as informações referentes às regiões brasileiras do Sudeste, formada pelos estados do Espírito Santo (ES), Minas Gerais (MG), Rio de Janeiro (RJ) e São Paulo (SP), e do Sul, formada pelos estados do Paraná (PR), Rio Grande do Sul (RS) e Santa Catarina (SC).

6.5 FONTE DOS DADOS

Os registros dos óbitos foram obtidos através da base de dados do SIM, disponibilizada pelo DATASUS no TABNET. Estes dados tratam-se de informações oficiais brasileiras e estão disponíveis de forma pública. Deste modo, não houve a necessidade de avaliação e aprovação do Comitê de Ética em Pesquisa (CEP) da Universidade Federal de Santa Maria (UFSM).

6.6 PROCEDIMENTOS METODOLÓGICOS

O estudo das séries de óbitos por neoplasias malignas de pele nas regiões Sul e Sudeste foi realizado utilizando os seguintes procedimentos metodológicos:

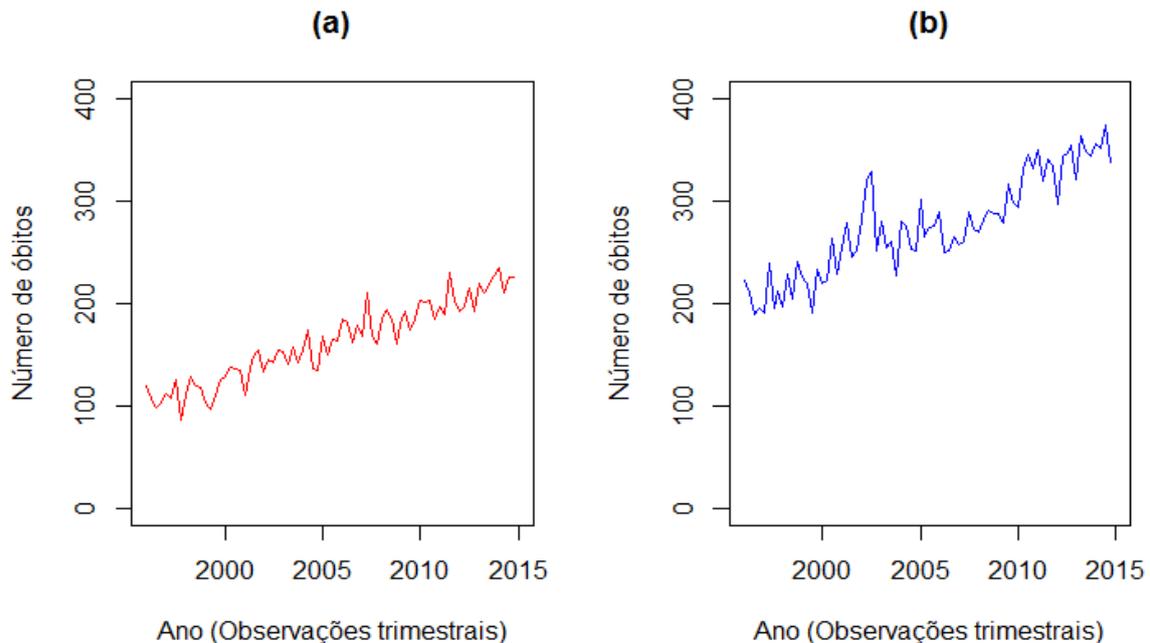
- Foram analisados os gráficos das séries originais afim de observar e evidenciar características a partir da evolução das observações ao longo do tempo;
- Foram analisadas a FAC e FACP das séries originais afim de evidenciar características quanto a estacionariedade e possíveis ordens do modelo de melhor ajuste;
- Foi realizado o teste de raiz unitária KPSS para a avaliar a estacionariedade das séries analisadas. Quando verificado a não estacionariedade, ou seja, presença de raiz unitária, foi necessário a realização de sucessivas diferenças, $I(d)$, até tornar o processo estacionário;
- Foram novamente analisadas a FAC e a FACP, considerando as séries de óbitos na ordem $I(d)$ em que foi evidenciada a ausência de raiz unitária (estacionariedade), com a intenção de identificar as ordens p e q dos termos AR e MA do modelo ARIMA (p,d,q) ;
- Foram ajustados modelos ARIMA (p,d,q) e comparados a partir dos critérios de informação, AIC e BIC, afim de identificar o modelo de maior parcimônia para cada região;
- Foi realizada a diagnóstico dos resíduos dos modelos mais parcimoniosos, analisando os pressupostos de ausência de autocorrelação, normalidade através do teste Jarque Bera, e utilização do teste $Q'(k)$ de Ljung-Box para verificar se os resíduos são *i.i.d*;
- Por fim, foram obtidas as previsões acerca do modelo de melhor ajuste, para as regiões Sul e Sudeste, e avaliada a qualidade das previsões por meio do MAPE e U de Theil;

Os gráficos, os testes realizados e os modelos ajustados foram obtidos utilizando o RStudio versão 1.0.136, software livre de ambiente de desenvolvimento integrado para R, utilizando os pacotes stats, graphics, astsa, urca, forecast e tseries. Quanto ao nível de significância, foi considerado 5% em todas as análises realizadas.

7 RESULTADOS E DISCUSSÃO

Analisou-se, inicialmente, os gráficos das séries trimestrais dos óbitos por neoplasias malignas de pele, entre 1996 e 2014, nas regiões Sul (a) e Sudeste (b), apresentados na Figura 2. Conforme observado, pode-se verificar uma possível tendência de crescimento em ambas as regiões, assim como a inexistência de sazonalidade nas séries. Essas mesmas características na evolução da série temporal foram observadas em Akhtar e Rozi (2009) e Liu *et al.* (2011).

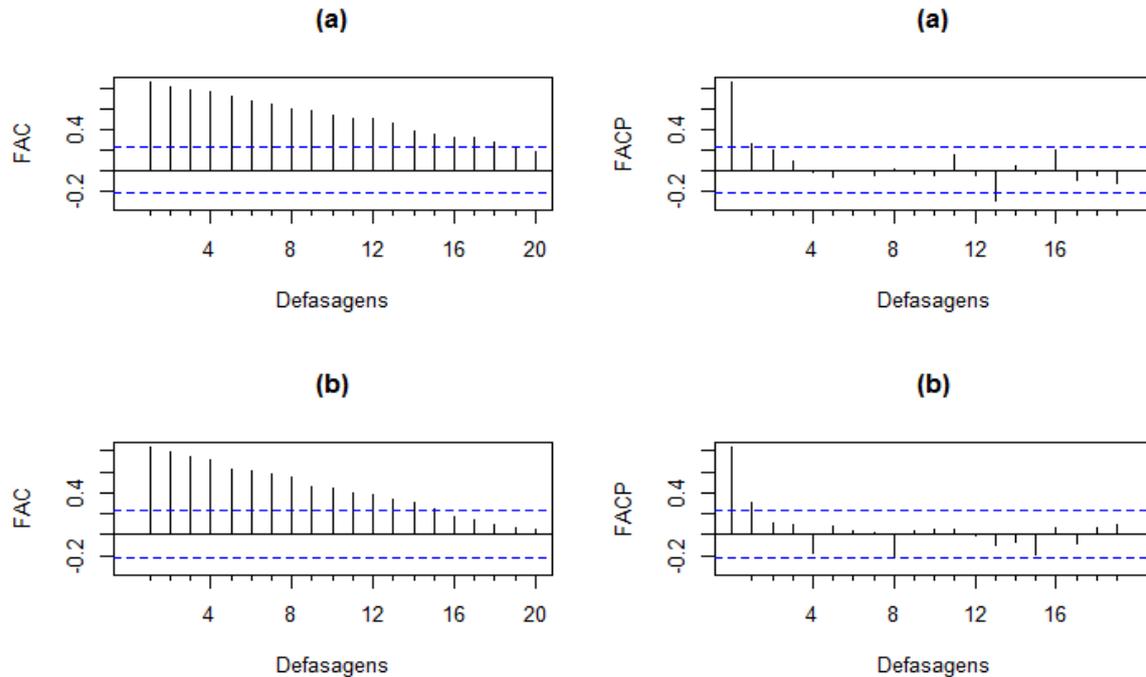
Figura 2 – Série trimestral de óbitos por neoplasias malignas de pele nas regiões Sul (a) e Sudeste (b) do Brasil entre 1996 e 2014.



Fonte: Elaborado pelo autor.

Quanto ao comportamento observado na FAC e na FACP (Figura 3) das séries das regiões analisadas, observou-se características de um processo não estacionário. Verificou-se um decaimento bastante lento na FAC, tanto na série da região Sul quanto na série da região Sudeste, cujas autocorrelações tornaram-se não significativas, respectivamente, apenas a partir das defasagens 20 e 16. Sendo assim, essas características reforçam a teoria de não estacionariedade na evolução das séries.

Figura 3 – Função de autocorrelação (FAC) e Função de autocorrelação parcial (FACP) da Série trimestral de óbitos por neoplasias malignas de pele nas regiões Sul (a) e Sudeste (b) do Brasil entre 1996 e 2014.



Fonte: Elaborado pelo autor.

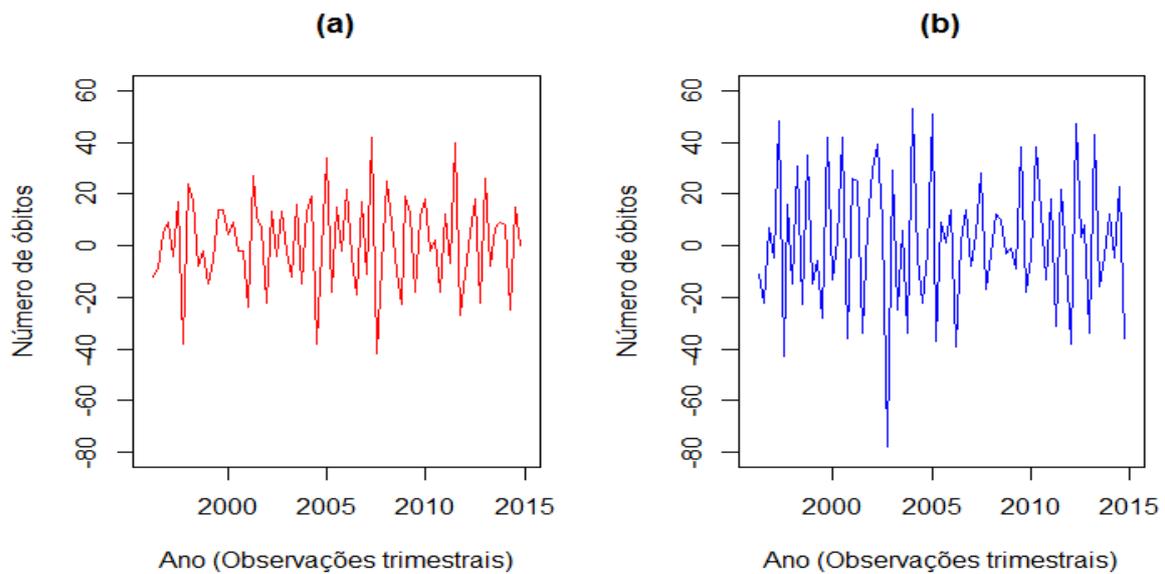
O Teste de raiz unitária KPSS truncado na segunda defasagem apresentou p-valor = 0,01 demonstrando que há evidências para rejeitar a estacionariedade da série da região Sul, ao nível de significância de 5%. Esse mesmo resultado foi obtido para a série da região Sudeste. Sendo assim, houve a necessidade de realização da primeira diferença em ambas as séries temporais afim de torna-las estacionárias e livres de tendência (Figura 4).

Analisando-se os gráficos da Figura 4 observou-se que a primeira diferença das séries de óbitos por neoplasias malignas de pele ocorridos nas regiões Sul (a) e Sudeste (b) apresentam um comportamento em torno de uma média e variância constante, ou seja, a análise gráfica indica que a séries se tornaram estacionárias após a realização da primeira diferença. Essa afirmação foi comprovada pelo teste KPSS cujo p-valor > 0,1 para ambas as séries em primeira diferença.

Assim como observado nesse estudo, nas séries analisadas por Akhtar e Rozi (2009) e Zhang et al (2017), a realização da primeira diferença também foi suficiente para tornar o processo estacionário. Esse resultado satisfaz a afirmação feita por Makridakis, Whellwright e Hyndman (1998) e Morettin e Tolo (2004), de que uma ou

duas diferenças, em geral, seriam suficientes para tornar uma série temporal estacionária. Porém, as vezes pode fugir à regra, como pode ser observado em Liu *et al.* (2011), na qual foram necessárias três diferenças para estacionarizar a série de incidência de febre hemorrágica com síndrome renal na China.

Figura 4 – Primeira diferença da Série trimestral de óbitos por neoplasias malignas de pele nas regiões Sul (a) e Sudeste (b) do Brasil entre 1996 e 2014.

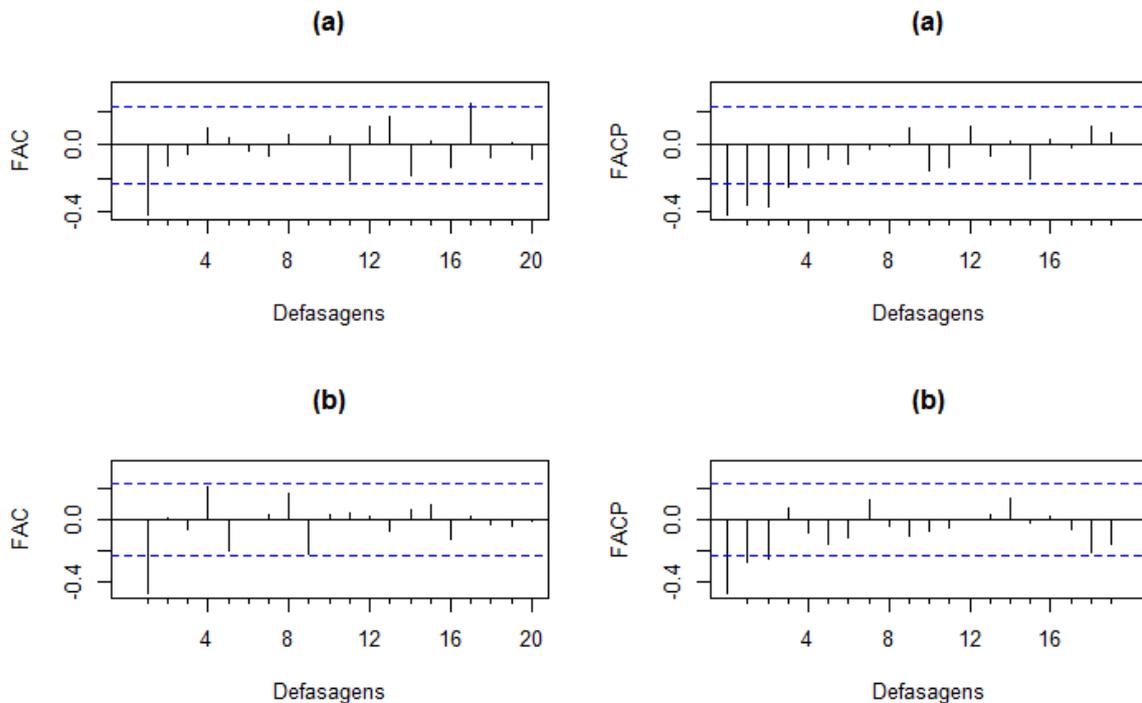


Fonte: Elaborado pelo autor.

Sabe-se, partir do que foi realizado até então, que os modelos ajustados para as regiões Sul e Sudeste deve representam um ARIMA com p componentes AR, uma componente I e q componentes MA, ou seja, ARIMA $(p,1,q)$. Utilizou-se a FAC e FACP (Figura 5) para indicar um possível modelo explicativo para as séries de óbitos por neoplasias malignas de pele ocorridos nas regiões Sul e Sudeste, em primeira diferença.

A FACP e a FAC (Figura 5) da região Sul apresentam, respectivamente, 1 e 3 defasagens fora dos limites de confiança. Desse modo, o modelo ARIMA $(3,1,1)$ é um possível candidato a melhor ajuste para a série dessa região. Para a região Sudeste, a FAC e a FACP (Figura 5), apresentam, respectivamente, 2 e 1 defasagens fora dos limites de confiança, indicando assim um ARIMA $(2,1,1)$ como possível modelo de melhor ajuste.

Figura 5 – Função de autocorrelação (FAC) e Função de autocorrelação parcial (FACP) da primeira diferença da série trimestral de óbitos por neoplasias malignas de pele em primeira diferença nas regiões Sul (a) e Sudeste (b) do Brasil, entre 1996 e 2014.



Fonte: Elaborado pelo autor.

Após o ajuste de diversos modelos com número de parâmetros em torno dos modelos indicados na análise da Figura 5, foi constatado que os modelos com melhores ajustes para as séries das regiões Sul e Sudeste foram os apresentados no Quadro 3. Entre os três modelos apresentados para cada região (Quadro 3) o modelo ARIMA (4,1,0) com constante mostrou-se o mais parcimonioso considerando o BIC para a série de dados da região Sul, e o modelo ARIMA (0,1,1) com constante apresentou maior parcimônia para a série de dados da região Sudeste.

Com isso, pode-se observar que os modelos ajustados (Quadro 4) considerados os mais parcimoniosos para as séries de óbitos das regiões Sul e Sudeste, são explicados por diferentes componentes. A região Sul não apresentou componentes de MA, enquanto a região Sudeste não apresentou componentes AR.

O modelo justado para a região Sudeste apresentou semelhanças com o modelo ajustado por Liu *et al.* (2011), apenas divergem no número de diferenças

necessárias para estacionarizar a série, mas em ambos os modelos uma componente MA foi suficiente para explicar a evolução das séries. O modelo ajustado para a região Sul não possui semelhanças com nenhum dos ajustados por Earnest *et al.* (2005), Akhtar e Rozi (2009) e ZHANG *et. al* (2017) descritos na seção 4.6, em todos esses estudos estão presentes tanto componentes AR, quanto componentes MA.

Quadro 3 – Critério de informação Bayesiano (BIC) dos três modelos de melhor ajuste as séries de órbitos por neoplasias malignas de pele nas regiões sul e sudeste, entre 1996 e 2014.

Modelos	Região Sul	Modelos	Região sudeste
ARIMA(4,1,0) com constante	619,46	ARIMA(0,1,1) com constante	677,60
ARIMA(3,1,0) com constante	620,66	ARIMA(1,1,1) com constante	681,92
ARIMA(5,1,0) com constante	621,66	ARIMA(0,1,2) com constante	681,92

Fonte: Elaborado pelo autor.

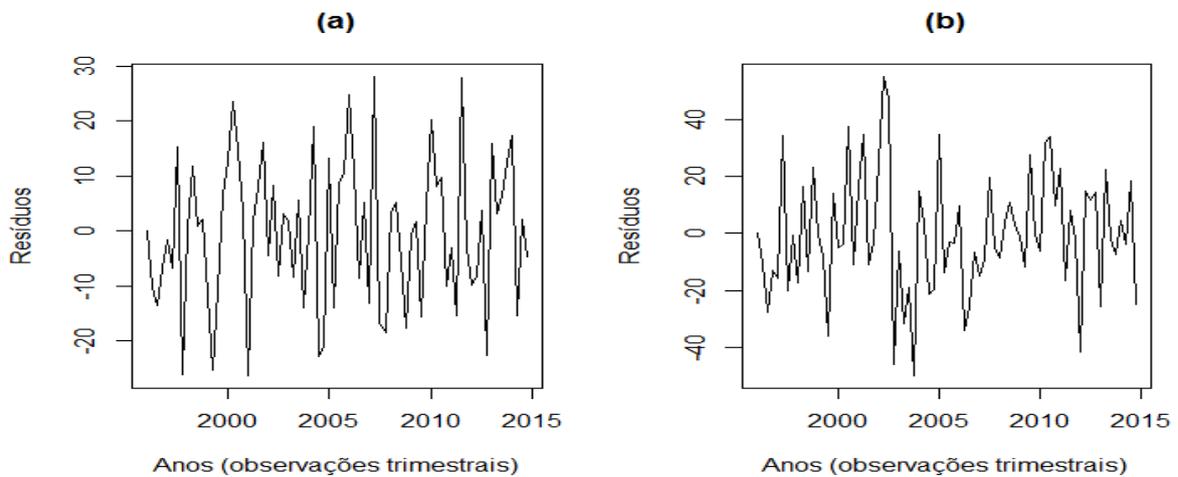
Quadro 4 – Parâmetros estimados dos modelos ARIMA(4,1,0) com constante (Região Sul) e ARIMA(0,1,1) com constante (Região Sudeste).

Modelo Ajustado (Região Sul)	Parâmetros estimados
ARIMA(4,1,0) com constante	Constante = 1,6014
	$\hat{\phi}_1 = -0,8150$
	$\hat{\phi}_2 = -0,7382$
	$\hat{\phi}_3 = -0,5957$
	$\hat{\phi}_4 = -0,2686$
Modelo Ajustado (Região Sudeste)	Parâmetros estimados
ARIMA(0,1,1) com constante	Constante = 2,0297
	$\hat{\theta}_1 = -0,7267$

Fonte: Elaborado pelo autor.

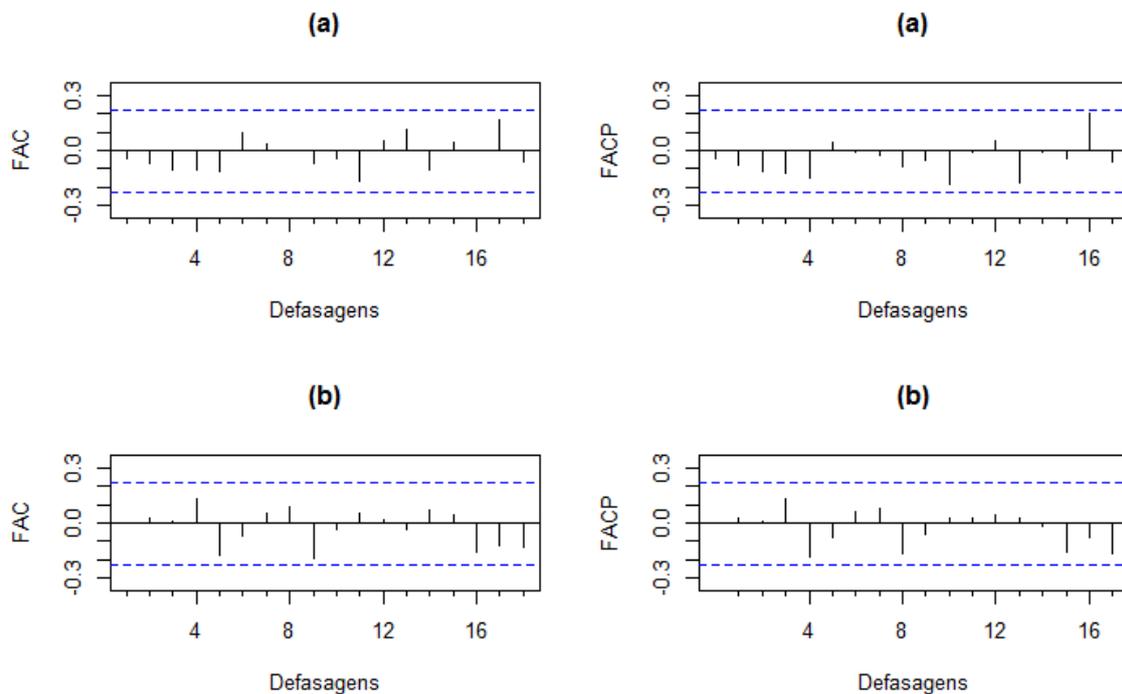
Após a escolha do modelo mais parcimonioso para cada região, foi realizada a análise de diagnóstico dos resíduos desses modelos. Na Figura 6 os resíduos dos modelos podem ser observados graficamente e na Figura 7 as FAC e FACP, onde verificou-se que os resíduos estão dentro dos limites de confiança.

Figura 6 – Resíduos dos modelos ajustados para as séries de óbitos por neoplasias malignas de pele nas regiões Sul (a) e Sudeste (b) do Brasil.



Fonte: Elaborado pelo autor.

Figura 7 – Função de autocorrelação (FAC) e Função de autocorrelação parcial (FACP) dos resíduos dos modelos ARIMA ajustados para os óbitos por neoplasias malignas de pele nas regiões Sul (a) e Sudeste (b)



Fonte: Elaborado pelo autor.

O resultado do teste Jarque-Bera, para os resíduos, apresentou p-valor = 0,5617 e p-valor = 0,8152, respectivamente, dos modelos ajustados para as regiões Sul e Sudeste. Logo, pode-se inferir que não há evidências para rejeitar a hipótese

de normalidade dos resíduos dos modelos ARIMA (4,1,0) com constante e ARIMA (0,1,1) com constante, considerando um nível de significância de 5%.

O resultado do teste Q' de Ljung e Box resultou em um p-valor igual a 0,7056 e 0,5511, para os respectivos resíduos dos modelos das regiões Sul e Sudeste. Com isso, também pode-se inferir que não há evidências para rejeitar a hipótese de que os resíduos são, *i.i.d*, ou seja, estão ausentes de autocorrelação. Sendo assim, conclui-se que os modelos ajustados são adequados para explicar as séries de óbitos por neoplasias malignas de pele nas regiões Sul e Sudeste.

O último procedimento da metodologia de Box-Jenkins diz respeito a obtenção de previsões acerca do modelo escolhido. As previsões de óbitos por neoplasias malignas de pele obtidas para os quatro trimestres de 2015 nas regiões Sul (a) e Sudeste (b) estão descritas no Quadro 5 e graficamente na Figura 8.

Quadro 5 – Previsões dos óbitos por neoplasias malignas de pele para os trimestres do ano de 2015 nas regiões Sul e Sudeste considerando, respectivamente, os modelos ARIMA(4,1,0) com constante e ARIMA(0,1,1).

Região Sul	Óbitos reais	Previsões	Limite inferior	Limite Superior
1º Trimestre	213	231,6174	204,7569	258,4779
2º Trimestre	229	229,7422	202,4261	257,0582
3º Trimestre	252	227,7973	200,3187	255,2758
4º Trimestre	235	232,3967	204,4132	260,3801
Região Sudeste	Óbitos reais	Previsões	Limite inferior	Limite Superior
1º Trimestre	377	357,9866	315,3216	400,6517
2º Trimestre	383	360,0163	315,7870	404,2456
3º Trimestre	413	372,0460	326,3059	417,7861
4º Trimestre	399	364,0757	316,8731	411,2783

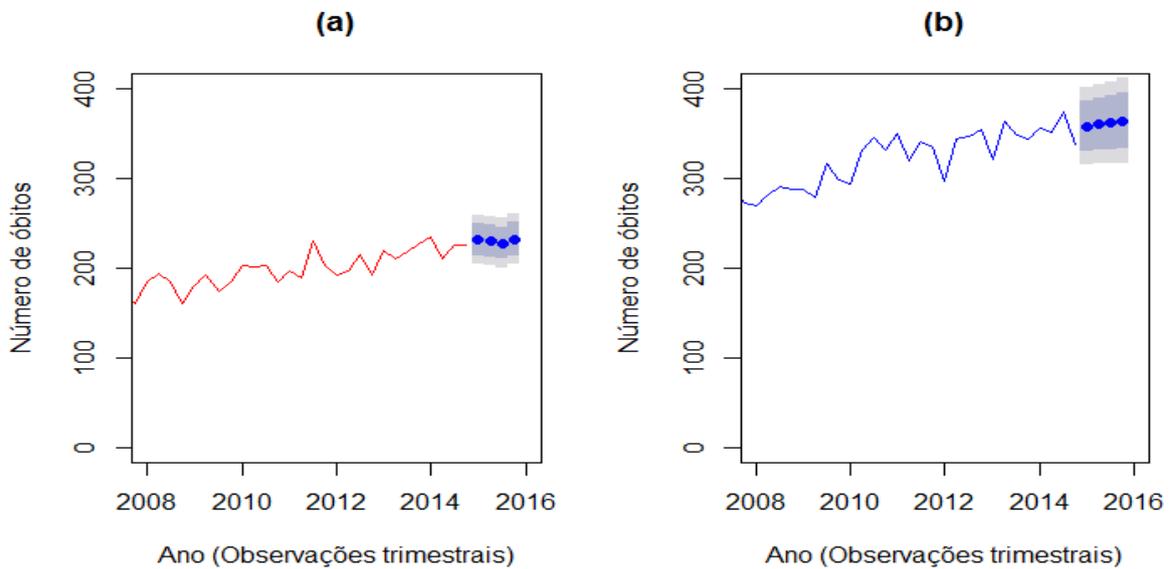
Fonte: Elaborado pelo autor.

Pode-se observar no Quadro 5 que os valores trimestrais reais dos óbitos de 2015 estão dentro dos limites de confiança. Quanto a qualidade das previsões para a região Sul, foi obtido um MAPE de 6,29% e U de Theil de 0,77, e para região Sudeste o MAPE foi de 7,13% e U de Theil de 0,80. Sendo assim, é observado um

MAPE um pouco menor para região Sul, mas ambos relativamente baixos para ambas as regiões analisadas, e os modelos estimados apresentaram previsões melhores que as previsões ingênuas.

Comparando-se com outros estudos, os MAPE's dos modelos preditivos da região Sul e Sudeste foram menores que o obtido pelo modelo de Earnest *et al.* (2005), com 8,7%, e também Liu *et al.* (2011), com 12,2%, porém foram levemente superiores aos 5,74% obtidos por ZHANG *et al.* (2017).

Figura 8 – Gráfico de previsões para as séries trimestral de óbitos por neoplasias malignas ocorridos nas regiões Sul (a) e Sudeste (b).



Fonte: Elaborado pelo autor.

8 CONCLUSÃO

Conforme observado, este estudo possibilitou evidenciar algumas características como presença de tendência crescente, ausência de sazonalidade e comportamento não estacionário nas séries de óbitos por neoplasias malignas de pele tanto na região Sul, quanto na região Sudeste.

O modelo de melhor ajuste e mais parcimonioso para os óbitos por neoplasias malignas de pele na região Sul foi o ARIMA (4,1,0) com constante e o modelo de melhor ajuste para a região Sudeste foi o ARIMA (0,1,1). Em ambos os modelos, as previsões obtidas apresentaram baixos MAPE's (respectivamente, 6,29% e 7,13%) e mostraram um contínuo crescimento nos óbitos.

Quanto a comparação dos modelos obtidos para as regiões Sul e Sudeste, observou-se a presença de diferentes componentes. Os óbitos por neoplasias malignas de pele na região Sul foram explicados apenas por componentes AR e os óbitos na região Sudeste apenas por componentes MA, sendo ambas com a necessidade de uma diferenciação para tornar a série estacionária.

Em geral, os modelos ARIMA apresentaram boa qualidade de ajuste e de previsões para as séries de óbitos por neoplasias malignas de pele nas regiões Sul e Sudeste do Brasil. A metodologia de Box-Jenkins mostrou-se bastante satisfatória, no monitoramento e detecção de padrões da evolução das séries analisadas. Sendo assim, os modelos ajustados podem ser utilizados como justificativa para a formulação de novas políticas públicas proporcionando combater, prevenir e/ou estabilizar a mortalidade por neoplasias malignas de pele, não somente nas regiões Sul e Sudeste, mas também nas demais regiões brasileiras.

É importante salientar que este foi um estudo pioneiro relacionado as neoplasias malignas de pele que se utiliza de modelos de previsão, mais especificamente os modelos ARIMA. Porém, é sabido que existem modelos de previsão mais recentes e que possivelmente consigam ajustar-se melhor as variações das séries analisadas. Por isso, sugere-se a modelagem dessas séries utilizando os modelos GARMA (p, q) e GLARMA (p, q) pois, como as séries temporais são de dados de contagem, esses podem reduzir os resíduos dos modelos e melhorar ainda mais a qualidade das previsões.

A continuidade de estudos desse tipo justifica-se pelo fato das neoplasias malignas de pele terem um impacto muito importante na área da saúde, e os

modelos de previsão podem direcionar à tomada de decisões em termos de saúde pública, bem como o conhecimento da evolução desse tipo de neoplasia. Pode também servir de alerta à população, e como consequência, estimular o aumento de investimentos em prevenção, evitando gastos em tratamentos de cura, pois estes são altamente dispendiosos.

REFERÊNCIAS

- AKAIKE, H. Maximum likelihood identification of Gaussian autoregressive moving average models. **Biometrika**, [S.l.], v.60, n.2, p.255–265, 1973.
- AKHTAR, S.; ROZI, S. An autoregressive integrated moving average model for short-term prediction of hepatitis C virus seropositivity among male volunteer blood donors in Karachi, Pakistan. **World Journal of Gastroenterology: WJG**, v. 15, n. 13, p. 1607, 2009.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis: forecasting and control**. San Francisco: Ed: Holden-Day, 575 p., 1970.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**, 3^o ed. Prentice Hall: New Jersey, 589 p., 1994.
- BOX, G. E. P *et al.* **Time series analysis: forecasting and control**. John Wiley & Sons, 709 p., 2015.
- BOX, G. E. P.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **Journal of the American Statistical Association**, v. 64, p. 1509-1526, 1970.
- BROCKWELL, P. J.; DAVIS, R. A. **Time series: theory and methods**. 2^o ed, 2^o printing. Springer Science & Business Media: New York, 589 p., 2009.
- BUENO, R.L.S. **Econometria de séries temporais** São Paulo: Cengage Learning, 299 p., 2008.
- CARRADORI, R. G.; RAMOS, P. de S. Avaliação de testes de normalidade implementados no programa r por simulação monte carlo. **Revista da Estatística da Universidade Federal de Ouro Preto**, v. 3, n. 2, p. 33-41, 2014.
- DEPARTAMENTO DE INFORMÁTICA DO SUS. **Informações de saúde (TABNET)**. Brasília, 2017. Disponível em: <<http://www2.datasus.gov.br/DATASUS/index.php?area=01>>. Acesso em: 1 jun. 2017.
- DEPARTAMENTO DE INFORMÁTICA DO SUS. **Informações de saúde (TABNET)**. Brasília, 2017. Disponível em: <<http://www2.datasus.gov.br/DATASUS/index.php?area=02>>. Acesso em: 1 jun. 2017.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit Root. **Journal of the American Statistical Association**, v.74, n. 366, p. 427-31, 1979.
- _____. DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. **Econometrica**, v.49, n.4, p. 1057-1072, 1981.

EARNEST, A *et al.* Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. **BMC Health Services Research**, v.5, n.1, p.36, 2005.

GRASSI, P. R.; LAURENTI, R.. Implicações da introdução da 10ª Revisão da Classificação Internacional de Doenças em análise de tendência da mortalidade por causas. **Informe Epidemiológico do SUS**, v. 7, n. 3, p. 43-47, 1998.

GUJARATI, D. M; PORTER, D. C. **Econometria básica**. 5ª Ed. São Paulo: Makron books, 924 p., 2011.

HAMILTON, J. D. **Time series analysis**. Princeton: Princeton university press, 1994.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International journal of forecasting**, v. 22, n. 4, p. 679-688, 2006.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Estimativas 2014: Incidência de câncer no Brasil. Rio de Janeiro, 124 p., 2014.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Estimativas 2016: Incidência de câncer no Brasil. Rio de Janeiro, 124 p., 2015.

JARQUE, C. M.; BERA, A. K. A test for normality of observations and regression residuals. **International Statistical Review**, Woodbury, v.55, p. 163-172, 1987.

JENKINS, G. M.; WATTS, D. G. Spectral analysis and its applications. Holden Day, 541 p., 1968.

KWIATKOWSKI, D. *et al.* Testing the null hypothesis of stationarity against the alternative of a unit root. **Journal of Econometrics**, v.54, p. 159-178, 1992.

LJUNG, G.; BOX, G. E. P. On a measure of lack off it in time series models. **Biometrika**, v. 65, p. 297-303, 1978.

LIU, Q. *et al.* Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. **BMC infectious diseases**, v. 11, n. 1, p. 218, 2011.

MADDALA, G. S. **Introduction to Econometrics**. 3.ed. New York: John Wiley & Sons, 664 p., 2001.

MACKINNON, J. G. Numerical distribution functions for unit root and cointegration tests. **Journal of Applied Econometrics**, v. 11, p. 601-618, 1996.

MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting: Methods and Applications**. 3rd ed. Ed. Wiley & Sons, 642 p., 1998.

MATHERS, C. D. *et al.* Counting the dead and what they died from: an assessment of the global status of cause of death data. **Bulletin of the world health organization**, v. 83, n. 3, p. 171-177c, 2005.

MATOS, K. S. L de; VIEIRA, S. L.; **Pesquisa educacional: o prazer de conhecer**. Fortaleza: Demócrito Rocha, 143p., 2001.

MATOS, O. C. **Econometria Básica: teoria e aplicações**. 3.ed. São Paulo: Atlas, 300 p., 2000

MINISTÉRIO DA SAÚDE. **Portaria MS/GM nº. 1.311 de 12 de setembro de 1997**. Determina a implantação do CID/10^o Revisão em todo território nacional a partir de janeiro de 1998. Diário Oficial da União. Brasília, 1997.

MORETTIN, Pedro A. **Análise harmônica de processos estocásticos**. Rio de Janeiro: IMPA, 186 p.,1979.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. São Paulo: Blücher, 535 p., 2004.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **CID-10: Classificação Estatística Internacional de Doenças com disquete Vol. 1**. Edusp, 1994.

RAZALI, N. M.; WAH, Y. B. Power Comparisons of Shapiro-Wilk , Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. **Journal of Statistical Modeling and Analytics**, v.2, n.1, p.21-33, 2011.

SCHWARZ,G. Estimating the dimension of a model. **The annals of statistics**, Hayward, v.6, n.2, p.461–464, Abr. 1978.

SOUZA, R. C.; CAMARGO, M. E. **Análise de Séries Temporais: Os Modelos ARIMA**, Ijuí: SEDIGRAF. 242 p., 1996.

THEIL, H. **Applied economic forecasting**. Amsterdam: North-Holland, 1966.

TUKEY, J. W. Discussion, emphasizing the connection between analysis of variance and spectrum analysis. **Technometrics**, v. 3, n. 2, p. 191-219, 1961.

VASCONCELLOS, M. A. S.; ALVES, D.. **Manual de Econometria**. 1 ed. São Paulo: Atlas, 2000.

WORLD HEALTH ORGANIZATION. ICD-10: international statistical classification of diseases and related health problems: tenth revision. 2007.

WHITTLE, P. **Hypotesis testing in the time series analysis**. Thesis, Uppsala University, Almqvist and Wiksell, Uppsala, 1951.

ZHANG, X. *et al.* Tendência preditiva das caries da primeira infância na China continental: uma abordagem combinada de modelagem meta-analítica e matemática baseada em pesquisas epidemiológicas. **Scientific Reports** , v. 7, n. 1, p. 6507, 2017.