



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE**  
**CENTRO DE EDUCAÇÃO E SAÚDE**  
**UNIDADE ACADÊMICA DE FÍSICA E MATEMÁTICA**  
**CURSO DE LICENCIATURA EM MATEMÁTICA**

**MICAELY RODRIGUES XAVIER**

**SÉRIES TEMPORAIS: UMA ANÁLISE DOS CASOS DE TUBERCULOSE NA  
PARAÍBA**

**CUITÉ - PB**

**2021**

MICAELY RODRIGUES XAVIER

**SÉRIES TEMPORAIS: UMA ANÁLISE DOS CASOS DE TUBERCULOSE NA  
PARAÍBA**

Monografia apresentada à Banca Examinadora, como exigência parcial à conclusão do Curso de Licenciatura em Matemática, da Universidade Federal de Campina Grande campus Cuité.

Orientador: Jorge Alves de Sousa

CUITÉ - PB

2021

X3s Xavier, Micaely Rodrigues .

Séries temporais: uma análise dos casos de tuberculose na Paraíba. / Micaely Rodrigues Xavier. - Cuité, 2021.

41 f.: il. color.

Trabalho de Conclusão de Curso (Licenciatura em Matemática) - Universidade Federal de Campina Grande, Centro de Educação e Saúde, 2021.

"Orientação: Profa. Jorge Alves de Sousa".

Referências.

1. Tuberculose. 2. Séries temporais. 3. Tuberculose - Paraíba - cenário futuro. 4. Modelo Box-Jenkins. 5. Séries temporais - análise. I. Sousa, Jorge Alves de. II. Título.

CDU 616.24-002.5(043)

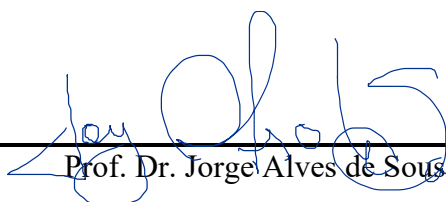
MICAELY RODRIGUES XAVIER

**SÉRIES TEMPORAIS: UMA ANÁLISE DOS CASOS DE TUBERCULOSE NA  
PARAÍBA**

Monografia apresentada à Banca Examinadora,  
como exigência parcial à conclusão do Curso de  
Licenciatura em Matemática, da Universidade  
Federal de Campina Grande Campus Cuité.

Aprovada em: 05/10/2021

**BANCA EXAMINADORA**



---

Prof. Dr. Jorge Alves de Sousa - UFCG

(Orientador)

 Assinado de forma  
digital por ALEXANDRO  
ALVES  
VIEIRA:02940732485

---

Prof. Dr. Alexandre Alves Vieira - UFCG

(Membro da Banca)

 Assinado de forma  
digital por ANSELMO  
RIBEIRO  
LOPES:86078038320

---

Prof. Msc. Anselmo Ribeiro Lopes - UFCG

(Membro da Banca)

CUITÉ - PB

2021

## **DEDICATÓRIA**

A minha família, em especial aos meus filhos  
Emanuel e Maria Elisa.

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus, por ter me guiado durante essa jornada.

Aos meus pais Cícero e Francisca por todo amor, dedicação, apoio e confiança depositada em mim, pois foram fundamentais para realizações dos meus sonhos.

Ao meu esposo Eugênio e meus filhos Emanuel e Maria Elisa por todo amor, cuidado, companheirismo e cumplicidades. Essa conquista é para vocês.

Aos meus irmãos Michelly e Michel por todo carinho, cuidados, ensinamentos e cumplicidade. Vocês são minha maior inspiração.

Ao meu orientador e Prof. Dr. Jorge Alves de Sousa por toda paciência, compreensão, dedicação e apoio na realização deste trabalho. Estendo esse agradecimento a todos os professores que passaram pela minha vida acadêmica.

Aos meus colegas pelos momentos de descontrações e incentivos. Sem dúvidas, são amigos que levarei por toda vida.

Por fim, agradeço a todos aqueles que direta ou indiretamente fizeram parte da minha formação acadêmica, meu muito obrigado.

*“Ainda que eu ande pelo vale da sombra da morte;  
não temerei mal algum, porque tu estás comigo;  
a tua vara e o teu cajado me consolam.”*

*(Salmo 23:4)*

## RESUMO

Este trabalho tem como objetivo descrever aspectos conceituais e sistematizar indicações metodológicas da análise de séries temporais, assim como sua aplicação para antever futuros cenários com respeito à evolução do número de casos da doença Tuberculose na Paraíba, evidenciando alguns fatores capazes de modificar a distribuição, ou comportamento, desta doença no estado. Antever o futuro é uma primeira e óbvia aplicação da análise de séries temporais. De fato, essa análise propicia a previsão dos valores futuros da série. Contudo, a incerteza sobre estimativas futuras é tanto maior quanto mais se distancia do presente, e intercorrências inesperadas são mais ou menos frequentes nas atividades humanas, o que implica risco adicional de erro não previsto. Destaca-se que o modelo Box-Jenkins foi adequado para estimar e prever os dados futuros de casos de tuberculose na Paraíba, e este modelo, obteve boa precisão, pois possui resíduos normalizados e não autocorrelacionados. Isso causou um aspecto de sequência no gráfico dos dados previstos e um valor considerável de intervalo de confiança.

**Palavras-chave:** Séries Temporais. Tuberculose. Modelo Box-Jenkins.



## **ABSTRACT**

This term paper aims to describe conceptual aspects and systematize methodological indications of time series analysis, as well as its application in order to predict future scenarios related to the evolution of the number of cases of Tuberculosis in Paraíba, by highlighting some factors that can change the distribution or behavior of this disease in the State. Predicting the future time is a first and clear application of time series analysis. In fact, this analysis provides a prediction of the future values of the series. However, uncertainty about future estimates is greater the further one gets from the present, and unexpected interurrences are more or less common in human activities, which imply additional risk of unanticipated error. It is noteworthy that the Box-Jenkins model was essential to estimate and predict the future data of tuberculosis cases in State of Paraíba. It resulted in a sequence aspect to the graph from the predicted data, in addition to a considerable value of confidence interval.

**Keywords:** Time Series. Tuberculosis. Box-Jenkins Model.

## LISTA DE FIGURAS

|  |    |
|--|----|
| 1.1 Um ciclo de negócios de quatro fases .....   | 8  |
| 1.1 Série de taxas de câmbio semanais BP / USD (1980-1993).....                                  | 9  |
| 1.2 Série mensal de passageiros de companhias aéreas internacionais.....                         | 9  |
| 2.1 A metodologia Box-Jenkins para a seleção do modelo ideal.....                                | 20 |
| 3.1 Série temporal de casos de tuberculose registrados na Paraíba no período de 2001 a 2020..... | 22 |
| 3.2 Série temporal de casos de tuberculose dessazonalizada e sua componente sazonal K ...        | 23 |
| 3.3 Série temporal de casos de tuberculose e sua série dessazonalizada.....                      | 23 |
| 3.4 Correlogramas da Função de Autocorrelação e Autocorrelação Parcial.....                      | 24 |
| 3.5 Previsão de casos de tuberculose na Paraíba para um período de 12 meses.....                 | 26 |

## LISTA DE SIGLAS

ACF - Funções de Autocorrelação

ACVF - Função de Autocovariância teórica

AIC – Critério de Informação Akaike

AR – Autoregressivo

ARCH - Condicional Autorregressiva Heteroscedasticidade

ARCH – Exponencial generalizada

ARFIMA - Média móvel fracionada integrada autorregressiva

ARIMA – Média Móvel Integrada Autorregressiva

ARMA - Média Móvel Autoregressiva

I.I.D - Independentes e identicamente distribuídas

MA - Média Móvel

NAR – Autoregressivo não linear

NMA - Média móvel não linear

PACF – Função de Autocorrelação Parcial

SARIMA – Média Móvel Integrada Autorregressiva Sazonal

TAR - Limiar Autorregressivo

## SUMÁRIO

|  |           |
|--|-----------|
| <b>INTRODUÇÃO .....</b>  | <b>1</b>  |
| <b>1 CAPÍTULO .....</b>  | <b>5</b>  |
| <b>1.1. Breve introdução à história da análise de Séries temporais.....</b>      | <b>5</b>  |
| <b>1.2. Definições de uma série temporal.....</b>                                | <b>6</b>  |
| <b>1.3. Componentes de uma série temporal.....</b>                               | <b>7</b>  |
| <b>1.4. Exemplos de dados de série temporal.....</b>                             | <b>8</b>  |
| <b>1.5. Introdução à Análise de Séries Temporais.....</b>                        | <b>10</b> |
| <b>1.6. Séries Temporais e Processo Estocástico.....</b>                         | <b>10</b> |
| <b>1.7. Conceitos de estacionariedade.....</b>                                   | <b>11</b> |
| <b>1.8. Modelos de Parcimônia.....</b>   | <b>12</b> |
| <b>2 CAPÍTULO.....</b>   | <b>13</b> |
| <b>2.1. Os Modelos Autorregressivos de Média Móvel (ARMA).....</b>               | <b>14</b> |
| <b>2.2. Análises de estacionariedade.....</b>                                    | <b>16</b> |
| <b>2.3. Funções de Autocorrelação e Autocorrelação Parcial (ACF e PACF).....</b> | <b>17</b> |
| <b>2.4. Alguns modelos de séries temporais não lineares.....</b>                 | <b>18</b> |
| <b>2.5. Metodologia Box-Jenkins.....</b>   | <b>19</b> |
| <b>3 CAPÍTULO.....</b>   | <b>21</b> |
| <b>3.1. Tratamento e Análise de Dados.....</b>                                   | <b>21</b> |
| <b>3.2. Validação e Aplicação de Teoria.....</b>                                 | <b>21</b> |
| <b>3.3. Diagnósticos dos resíduos.....</b>                                       | <b>25</b> |
| <b>3.4. Previsão usando Modelo Box-Jenkins.....</b>                              | <b>26</b> |
| <b>3.5. Considerações Finais.....</b>  | <b>26</b> |
| <b>REFERENCIAS .....</b>   | <b>28</b> |

## INTRODUÇÃO

A matemática surgiu há milhares de anos desde que o ato de contar tornou-se importante para o homem primitivo, antes mesmo de ser denominada e reconhecida como a ciência que relaciona as práticas do cotidiano, natureza, e raciocínio lógico. Todavia, na contemporaneidade, ela assume outras funções no campo do ensino, a saber: resolução de problemas; uso de jogos; trabalho com projetos; interdisciplinaridade; contextualização; modelagem matemática; questões culturais; uso da história da matemática, investigações matemáticas, dentre outras, (NACARATO, 2005).

Esta pesquisa tem como um de seus principais objetivos promover uma discussão acerca da importância da Estatística, antes considerada subárea da Matemática, ressaltando como ela tem um importante papel na construção do conhecimento nas várias áreas das ciências, que mesmo vivendo em um período marcadamente moderno e tecnológico e em um país em desenvolvimento, como o Brasil, muitas conquistas são realizadas e a influência desta ciência é inegável (D'AMBRÓSIO,2009).

Este estudo também pretende refletir sobre a importância das séries temporais, pois por intermédio delas, é possível e, a aplicação de vários métodos estatísticos de análise, tais como: a média móvel simples média ponderada exponencial, análises de regressões e projeções matemáticas. Tais métodos são muito presentes nos estudos atuais (NACARATO ,2009).

Em se tratando da ciência, na contemporaneidade, tem-se visto a necessidade do diálogo entre os campos de pesquisa, nesse sentido, os laços entre os métodos estatísticos e saúde devem ser celebrados (CARRAPATO,2017).

Assim, ao pesquisar sobre séries temporais e encontrar poucos trabalhos, percebi que parece ainda não existirem, no Brasil, sólidos grupos de pesquisa sobre a temática em questão, o que alimentou a minha decisão de direcionar esta investigação a partir de um levantamento de teóricos que se debruçaram acerca de séries temporais (RESENDE,2016).

Nessa perspectiva, esta pesquisa traz uma discussão acerca das séries temporais, buscando sua aplicabilidade no estudo da evolução do quantitativo de casos de tuberculose na Paraíba.

A tradição histórica nos mostra que os mais pobres sempre foram injustiçados em nosso país, alguns deles subjugados desde o momento de sua chegada, de forma arbitrária ao Brasil, nos casos dos negros escravizados, desta maneira tiveram pouco ou praticamente nenhum acesso à tratamentos de saúde em casos de enfrentamentos à endemias ou surtos epidêmicos (BELCULFINÉ,2009).

A tuberculose é uma das doenças mais antigas do mundo, contudo somente foi descoberta ao final do século XIX pelo alemão Robert Koch. Existem evidências da existência dela em ossos humanos pré-históricos na Alemanha e há registros datados de 8.000 antes de Cristo. Devido a ser pouco comum pessoas escolarizadas naquela época, atribuíam o surgimento de sintomas da referida doença à um castigo, provavelmente resultado de comportamentos considerados errôneos pelos sistemas de convenções. A desmistificação de tais pressupostos começou a partir dos estudos de Hipócrates, na Grécia antiga, quando ele afirmou que a tuberculose (denominada tísica) era algo natural, disseminada através do contato do indivíduo, ocorrendo o aumento no número de casos por meio do advento da guerra, somente em 1943 foi feita a descoberta principal no tratamento da tuberculose, os antibióticos, atualmente existem até mesmo vacinas preventivas (FIOCRUZ,2013).

Mesmo considerando esse histórico e passado séculos após as descobertas acerca do surgimento da tuberculose, diversos avanços para a cura de doenças e vacinas preventivas, a moléstia não foi superada. Em pleno século 21, 70 mil brasileiros continuam a ser diagnosticados todos os anos com a condição. Desses, 4 500 morrem. No mundo, são 10 milhões de casos a cada 12 meses e 1 milhão de óbitos. Números tão expressivos fazem da tuberculose o quadro infeccioso que mais mata no planeta (VEJASAÚDE,2019).

É evidente o quanto a tuberculose é perigosa, uma vez que mesmo em um mundo marcadamente moderno ela continua a fazer vítimas. Em anos remotos a situação era ainda mais crítica. A doença chegou ao Brasil durante a colonização portuguesa, através do contato frequente dos doentes com os índios, o resultado foi o adoecimento e a morte de muitos nativos (ANTUNES,2000).

Durante o movimento artístico Romântico brasileiro, a tuberculose era conhecida como “mal do século”, ou seja, prevalecia uma visão lírica da doença, pois considerava-se que a mesma permitia aos artistas expressarem seu sentimento de obscurantismo, pessimismo ora pelo sofrimento que a doença gerava, ora pela peculiaridade que a ela lhes garantia (QUEIROZ, 1949, Apud MONTENEGRO, 1971).

Este projeto deixa clara, a importância de conhecermos mais profundamente a perspectiva histórica da tuberculose na saúde pública, sobretudo a paraibana, pois através destes estudos, é possível começar a desmistificar a ideia de que ela é um problema superado, quando na realidade, tal doença é responsável por milhares de mortes no Brasil e no mundo diariamente (FUNDAÇÃO OSWALDO CRUZ, 2008).

Temos, pois, como pergunta norteadora de nossa pesquisa: *Qual o papel das séries temporais em relação à análise quantitativa do número de casos de tuberculose na Paraíba?*

Para o desenvolvimento da mesma, definimos os seguintes objetivos:

Objetivo Geral:

- Averiguar através de séries temporais a incidência e a evolução do número de casos de tuberculose na Paraíba.

Objetivos específicos:

- Compreender como se distribui os casos de forma temporal;
- Analisar a tendência e sazonalidade de casos de tuberculose no período de 2001 a 2020 na Paraíba;
- Averiguar e encontrar o modelo que melhor representa seu cateter preditivo;
- Estimar o quantitativo de casos para um período determinado utilizando modelos Box-Jenkin.

Em nossa sociedade é explícito o quanto os problemas sanitários continuam a fazer parte da realidade brasileira, provocadas por diversos fatores como ausência de saneamento básico, falta de tratamento de água, esgoto, podem fazer surgir ou agravar surtos de epidemias infecciosas (ROSEMBERG,2002).

Assim, este trabalho tem como um de seus principais objetivos promover uma averiguação por meio das séries temporais acerca da incidência de casos de tuberculose na Paraíba, pois mesmo vivendo em um país marcadamente moderno como o nosso, frequentemente temos que conviver com estereótipos e discursos que questionam a importância dos conhecimentos científicos e avanços nos campos da medicina, matemática, dentre outros (SHEPPARD,1985).

Nesse sentido, essa pesquisa faz uma discussão acerca de alguns aspectos da ciência matemática, a fim de que ela seja mais difundida e mostre representações positivas, sobretudo no espaço educacional. Para desenvolver essa pesquisa dividiremos as reflexões em dois capítulos. O primeiro deles, abordará conceitos teóricos acerca das séries temporais, à luz da teoria de, Box e Jenkins(1970) Hipell e Mcleod (1994), Cochrane(1997), dentre outros. O segundo, buscará promover uma discussão acerca dos dados sobre a tuberculose no Brasil.

Esta investigação trata-se de uma pesquisa-ação, de cunho quantitativo, bibliográfico e interpretativista, alocada no paradigma das pesquisas descritivistas (ENGEL, 2000), tendo como objeto de estudo as séries temporais. Por ter caráter de pesquisa bibliográfica, será executada a partir de dados bibliográficos, que foram coletados por intermédio de leituras de pesquisadores e estudiosos da matemática contemporânea. Além de estudar as séries temporais, este trabalho também tem por intento mostrar a importância do diálogo entre as ciências. A revisão bibliográfica em questão deu-se a partir de um levantamento histórico sobre a temática estudada. Desta forma, o presente trabalho é de caráter quantitativo, já que por meio desta pesquisa, analisamos, interpretamos e discutimos sobre o comportamento dos dados relativos à incidência da tuberculose através de séries temporais.



# 1 CAPÍTULO

## Conceitos básicos de modelagem de série temporal

Apresentaremos neste capítulo alguns conceitos e resultados preliminares para o desenvolvimento e compreensão deste trabalho.

### 1.1. Breve introdução à história da análise de Séries temporais

Assim como foi no princípio, os métodos de previsão analisam dados passados para prever os dados futuros. Com isso, alguns estudiosos começaram a observar alguns fenômenos. Logo essas aplicações contribuíram para o desenvolvimento de séries temporais. Há indícios, que há bastante tempo são realizadas as coletas e o desenvolvimento de séries temporais.

Uma das séries mais antigas, conhecida na literatura através de uma das publicações de Graunt (1662), intitulada por *Natural and Political Observations upon the Bills of Mortality*, retrata a distribuição das idades de morte no século XVII em Londres (KLEIN, 1997). Tufté (1983) apresenta curiosidades sobre o início do trabalho com séries temporais, e este autor mostra informações sobre os movimentos dos planetas e do sol do século X ou XI (BRILLINGER, 2001). Outra aplicação que motivou os estudos da análise de séries temporais, segundo os historiadores, refere-se à clássica série do número médio anual de manchas solares de Wolfer para o período de 1749 a 1924 (BRILLINGER, 2001; MORETTIN e TOLOI, 2006).

Ao que se refere ao desenvolvimento metodológico das séries temporais, à análise espectral é considerada como técnica exploratória. Tal técnica produziu a primeira aplicação datada em 1664 através do experimento de Newton, o qual envolvia a passagem da luz solar, decomposta em seus componentes, por um prisma (BRILLINGER, 2001).

Ao logo do tempo, surgiram outros desenvolvimentos técnicos na metodologia de séries temporais. Pode-se citar o *paper* de T.N. Thiele de 1880 que contém contribuições relacionadas à formulação e análise de um modelo para uma série temporal constituído de uma soma de componentes da regressão, movimento Browniano e um ruído branco, a partir do qual desenvolveu um procedimento recursivo para estimar a componente da regressão e prever o

movimento Browniano, procedimento conhecido a partir de 1961 por filtro de Kalman (KALMAN e BUCY, 1961; LAURITZEN, 1981; TSAY, 2000).

No entanto, o grande avanço no uso das técnicas ocorreu na década de 80 com diversos artigos publicados. E a partir daí, desenvolveu-se diversas formas de aplicabilidade da técnica, como por exemplo, o tratamento de valores faltantes ou *missings*, no desenvolvimento de novos métodos para a extração de sinal, para alisamento e ajuste sazonal, entre outros (TSAY, 2000).

O livro de Box and Jenkins (1976) é considerado uma referência na história das previsões com séries temporais. Tal livro agregou todo conhecimento existente relacionado às séries temporais, no que diz respeito; a estimação de parâmetros, identificação e previsão. Tais autores, citaram a construção de modelos estocásticos no domínio do tempo para séries temporais discretas, que eram o mais simples possível, com número mínimo de parâmetros e satisfazendo, portanto um critério de parcimônia. A justificativa que os autores utilizaram era que a escolha deste critério estava relacionada ao fato de estes modelos serem capazes de mostrar-se algo sobre a natureza dos processos que geram as séries temporais e à possibilidade de esses modelos serem utilizados para a obtenção de previsões ótimas de valores futuros das séries.

## **1.2. Definições de uma série temporal**

Uma série temporal é um conjunto sequencial de pontos de dados, medidos normalmente em tempos sucessivos. Isto é matematicamente definido como um conjunto de vetores  $x(t)$ ,  $t = 0, 1, 2, \dots$ . Onde  $t$  representa o tempo decorrido (HIPELL e MCLEOD., 1994; COCHRANE, 1997; RAICHAROEN et al., 2003). A variável  $x(t)$  é tratada como uma variável aleatória. As medições tomadas durante um evento em uma série de tempo são organizadas em uma ordem cronológica adequada.

Uma série temporal contendo registros de uma única variável é denominada uni variada. Mas, se os registros de mais de uma variável são considerados, é denominado como multivariada. Uma série temporal pode ser contínua ou discreta. Em uma série de tempo contínua, as observações são medidas em cada instância de tempo, enquanto uma série de tempo discreta contém observações medidas em pontos discretos de tempo. Por exemplo, leituras de temperatura, fluxo de um rio, concentração de um processo químico etc. podem ser gravados como uma série de tempo contínua. Por outro lado, a população de uma determinada cidade, produção de uma empresa, taxas de câmbio entre duas moedas diferentes, podem

representar séries temporais discretas. Normalmente, em uma série de tempo discreta, as observações consecutivas são registradas em intervalos de tempo igualmente espaçados, como separações horárias, diárias, semanais, mensais ou anuais. Conforme mencionado em Hipel (1994), a variável sendo observada em uma série de tempo discreta é considerada como uma variável contínua usando a escala de número real. Além disso, uma série de tempo contínua pode ser facilmente transformada em uma série discreta mesclando dados em um intervalo de tempo especificado.

### **1.3. Componentes de uma série temporal**

Uma série temporal em geral deve ser afetada por quatro componentes principais, que podem ser separados dos dados observados. Esses componentes são: Componentes de tendência, cíclicos, sazonais e irregulares. Uma breve descrição desses quatro componentes é fornecida aqui.

A tendência geral de uma série temporal para aumentar, diminuir ou estagnar por um longo período de tempo é denominada como tendência secular ou simplesmente tendência. Assim, pode-se dizer que tendência é um movimento de longo prazo em uma série temporal. Por exemplo, as séries relacionadas ao crescimento populacional e ao número de casas em uma cidade, mostram tendência de aumento, enquanto a tendência de queda pode ser observada em séries relacionadas a taxas de mortalidade, epidemias, etc.

As variações sazonais em uma série temporal são flutuações dentro de um ano durante a temporada. Os fatores importantes que causam as variações sazonais são: clima e condições meteorológicas, costumes, hábitos tradicionais, etc. Por exemplo, as vendas de sorvete aumentam no verão, as vendas de tecidos de lã aumentam no inverno. A variação sazonal é um fator importante para que empresários, lojistas e produtores façam planos futuros adequados.

A variação cíclica em uma série temporal descreve as mudanças de médio prazo na série, causadas por circunstâncias, que se repetem em ciclos. A duração de um ciclo se estende por um período de tempo mais longo, geralmente dois ou mais anos. A maior parte das séries temporais econômicas e financeiras mostra algum tipo de variação cíclica. Por exemplo, um ciclo de negócios consiste em quatro fases, são.

i) Prosperidade, ii) Declínio, iii) Depressão e iv) Recuperação

Esquemáticamente, um ciclo de negócios típico pode ser mostrado como abaixo:



**Fig. 1: Um ciclo de negócios de quatro fases**

**Fonte: Própria**

Variações irregulares ou aleatórias em uma série temporal são causadas por influências imprevisíveis, que não são regulares e também não se repetem em um determinado padrão. Essas variações são causadas por incidentes como guerra, greve, terremoto, inundação, revolução, etc. Não existe uma técnica estatística definida para medir as flutuações aleatórias em uma série de tempo.

Considerando os efeitos desses quatro componentes, dois tipos diferentes de modelos são geralmente usados para uma série de tempo são: Modelos multiplicativos e aditivos.

$$\text{Modelo Multiplicativo: } Y(t) = T(t) \times S(t) \times C(t) \times I(t).$$

$$\text{Modelo Aditivo: } Y(t) = T(t) + S(t) + C(t) + I(t).$$

Aqui  $Y(t)$  é a observação e  $T(t)$ ,  $S(t)$ ,  $C(t)$  e  $I(t)$  são respectivamente a tendência, a variação sazonal, cíclica e irregular no tempo  $t$ .

O modelo multiplicativo é baseado na suposição de que os quatro componentes de uma série temporal não são necessariamente independentes e podem afetar uns aos outros; enquanto no modelo aditivo é assumido que os quatro componentes são independentes uns dos outros.

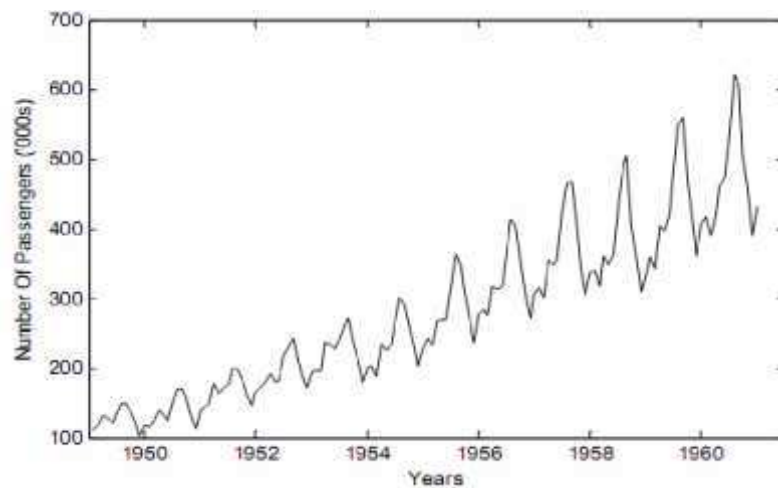
#### **1.4. Exemplos de dados de série temporal**

As observações de séries temporais são frequentemente encontradas em muitos domínios, como negócios, economia, indústria, engenharia e ciência, etc. (TONG,1983, ZHANG,2003, ZHANG,2007). Dependendo da natureza da análise e da necessidade prática, pode haver vários tipos diferentes de séries temporais. Para visualizar o padrão básico dos dados, geralmente uma série de tempo é representada por um gráfico, onde as observações são plotadas em relação ao tempo correspondente. Abaixo nas figuras 2 e 3, mostramos dois gráficos de série temporal:



**Fig. 2: Série de taxas de câmbio semanais BP / USD (1980-1993)**

**Fonte: Zhang (2003)**



**Fig. 3: Série mensal de passageiros de companhias aéreas internacionais (Jan. 1949-Dez. 1960).**

**Fonte: Box e Jenkins (1970)**

A primeira série temporal é tirada de Zhang (2003) e representa a taxa de câmbio semanal entre a libra esterlina e o dólar americano de 1980 a 1993. A segunda é uma série temporal sazonal, considerada em Box e Jenkins (1970), Faraway (1983) e Hamzacebi (2008) e

mostra o número de passageiros de companhias aéreas internacionais (em milhares) entre janeiro de 1949 a dezembro de 1960 em uma base mensal.

### **1.5. Introdução à Análise de Séries Temporais**

Na prática, um modelo adequado é ajustado a uma determinada série temporal e os parâmetros correspondentes são estimados usando os valores de dados conhecidos. O procedimento de ajustar um modelo adequado a uma série temporal é denominado como *Análise de séries temporais* (HIPEL e MCLEOD, 1994). Inclui métodos que tentam compreender a natureza da série e muitas vezes são úteis para previsões e simulações futuras.

Na previsão de séries temporais, as observações anteriores são coletadas e analisadas para desenvolver um modelo matemático adequado que captura o processo de geração de dados subjacente para a série ZHANG (2003,2008). Os eventos futuros são então previstos usando o modelo. Esta abordagem é particularmente útil quando não há muito conhecimento sobre o padrão estatístico seguido pelas observações sucessivas ou quando há falta de um modelo explicativo satisfatório. A previsão de séries temporais tem aplicações importantes em vários campos. Frequentemente, decisões estratégicas valiosas e medidas de precaução são tomadas com base nos resultados previstos. Assim, fazer uma boa previsão, ou seja, ajustar um modelo adequado a uma série de tempo é muito importante. Nas últimas décadas, muitos esforços foram feitos por pesquisadores para o desenvolvimento e melhoria de modelos de previsão de séries temporais adequados.

### **1.6. Séries Temporais e Processo Estocástico**

Uma série temporal é de natureza não determinística, ou seja, não podemos prever com certeza o que ocorrerá no futuro. Geralmente, uma série temporal  $\{x(t), t = 0, 1, 2, \dots\}$  é assumida como seguindo certo modelo de probabilidade (Cochrane (1997)), que descreve a distribuição conjunta da variável aleatória  $x_t$ . A expressão matemática que descreve a estrutura de probabilidade de uma série temporal é denominada como um processo estocástico. Assim, a sequência de observações da série é na verdade uma realização de amostra do processo estocástico que a produziu.

Uma suposição usual é que as variáveis da série temporal  $x_t$  são independentes e com distribuição idêntica (i.i.d) seguindo a distribuição normal. No entanto, como mencionado em

Cochrane (1997), um ponto interessante é que as séries temporais não são exatamente i.i.d; eles seguem mais ou menos algum padrão regular em longo prazo. Por exemplo, se a temperatura hoje de uma determinada cidade for extremamente alta, então se pode presumir que a temperatura de amanhã provavelmente também será alta. Esta é a razão pela qual a previsão de séries temporais, usando uma técnica adequada, produz resultados próximos do valor real.

## 1.7. Conceitos de estacionariedade

O conceito de estacionariedade de um processo estocástico pode ser visualizado como uma forma de equilíbrio estatístico (HIPEL e MCLEOD, 1994). As propriedades estatísticas, como média e variância de um processo estacionário, não dependem do tempo. É uma condição necessária para a construção de um modelo de série temporal útil para previsões futuras. Além disso, a complexidade matemática do modelo ajustado reduz com esta suposição. Existem dois tipos de processos estacionários que são definidos a seguir:

Definição 1: Um processo  $\{x(t), t = 0, 1, 2, \dots\}$  é *Fortemente Estacionário ou Estritamente Estacionário* se a função de distribuição de probabilidade conjunta de  $\{x_{t-s}, x_{t-s+1}, \dots, x_t, \dots, x_{t+s-1}, x_{t+s}\}$  é independente de  $t$  para todos os  $s$ .

Assim, para um processo estacionário forte, a distribuição conjunta de qualquer conjunto possível de variáveis aleatórias do processo é independente do tempo (COCHRANE, 1997, HIPEL e MCLEOD, 1994).

No entanto, para aplicações práticas, a suposição de forte estacionariedade nem sempre é necessária e, portanto, uma forma um pouco mais fraca é considerada.

Definição 2: Um processo estocástico é considerado *Fracamente Estacionário* de ordem  $k$  se os momentos estatísticos do processo até essa ordem dependerem apenas das diferenças de tempo e não do tempo de ocorrência dos dados usados para estimar os momentos (COCHRANE, 1997, HIPEL E MCLEOD, 1997, LEE, 2000). Por exemplo, um processo estocástico  $\{x(t), t = 0, 1, 2, \dots\}$  é estacionário de segunda ordem (HIPEL E MCLEOD, 1997, LEE, 2000) se tiver média e variância independente do tempo e os valores de covariância  $Cov(x_t, x_{t-s})$  dependendo apenas de  $s$ .

É importante notar que nem a estacionariedade forte nem fraca implica a outra. No entanto, um processo fracamente estacionário seguindo a distribuição normal também é fortemente estacionário (KOCHRANE, 1997a). Alguns testes matemáticos como o dado por Dickey e Fuller (KOCHRANE, 1997b) são geralmente usados para detectar estacionariedade em dados de série temporal.

Conforme mencionado em Box e Jenkins (1970) e Hipel e McLeod (1997), o conceito de estacionariedade é uma ideia matemática construída para simplificar o desenvolvimento teórico e prático de processos estocásticos. Para projetar um modelo, adequado para previsões futuras, espera-se que a série temporal subjacente seja estacionária. Infelizmente, nem sempre é o caso. Conforme afirmado por Hipel e McLeod (1997), quanto maior o intervalo de tempo das observações históricas, maior é a chance de que a série temporal apresente características não estacionárias. No entanto, para um intervalo de tempo relativamente curto, pode-se modelar razoavelmente a série usando um processo estocástico estacionário. Normalmente, as séries temporais, mostrando tendências ou padrões sazonais, são de natureza não estacionária. Nesses casos, a diferenciação e as transformações de potência costumam ser usadas para remover a tendência e tornar a série estacionária. No próximo capítulo, discutiremos sobre a técnica de diferenciação sazonal aplicada para tornar estacionária uma série temporal sazonal.

## **1.8. Modelos de Parcimônia**

Ao construir um modelo de série temporal adequado, temos que considerar o princípio da parcimônia (HIPEL E MCLEOD, 1994, CHATFIELD, 1996, ZANGH, 2003, ZANGH, 2008). De acordo com este princípio, sempre o modelo com o menor número possível de parâmetros deve ser selecionado de forma a fornecer uma representação adequada dos dados da série temporais subjacentes (CHATFIELD, 1996). Dentre uma série de modelos adequados, deve-se considerar o mais simples, ainda mantendo uma descrição precisa das propriedades inerentes às séries temporais. A ideia do modelo de parcimônia é semelhante ao famoso princípio da *Navalha de Occam* (HIPEL E MCLEOD, 1994a). Conforme discutido por Hipel e McLeod (HIPEL E MCLEOD, 1994b), um aspecto deste princípio é que quando confrontado com uma série de concorrentes e explicações adequadas, escolha a mais simples. A navalha de Occam fornece informações inerentes consideráveis, quando aplicada à análise lógica.



Além disso, quanto mais complicado o modelo, mais possibilidades surgirão para afastamento das suposições do modelo real. Com o aumento dos parâmetros do modelo, o risco de sobre ajuste também aumenta subsequentemente. Um modelo de série temporal super ajustado pode descrever os dados de treinamento muito bem, mas pode não ser adequado para previsões futuras. Como o super ajuste potencial afeta a capacidade de um modelo de prever bem, a parcimônia é frequentemente usada como um princípio orientador para superar esse problema. Assim, em resumo, pode-se dizer que, ao fazer previsões de séries temporais, deve-se dar atenção genuína para selecionar o modelo mais parcimonioso entre todas as outras possibilidades.

## 2 CAPÍTULO

### PREVISÃO DE SÉRIES TEMPORAIS USANDO MODELOS ESTOCÁSTICOS

No capítulo anterior, discutimos sobre os fundamentos da modelagem e previsão de séries temporais. A seleção de um modelo adequado é extremamente importante, pois reflete a estrutura subjacente da série e este modelo ajustado, por sua vez, é usado para previsões futuras. Um modelo de série temporal é considerado linear ou não linear, dependendo se o valor atual da série é uma função linear ou não linear de observações anteriores.

Em geral, os modelos de dados de séries temporais podem ter muitas formas e representar diferentes processos estocásticos. Existem dois modelos lineares de série temporal amplamente usado na literatura, são: *Modelos Autoregressivo* (AR) Box e Jenkins(1970), Lee (2000), Hipel e Mcleod(1997) e *Média Móvel* (MA) (BOX e JENKIS, 1970, HIPEL e MCLEOD,1994). Combinando estes dois, os modelos *Autoregressive Moving Average* (ARMA) Box e Jenkins(1970), Hipel e Mcleod (1994), Cochrane (1997) e J.Lee(2000) e *Autoregressive Integrated Moving Average* (ARIMA) Box e Jenkins(1970), Hipel e Mcleod (1994) e Cochrane (1997) têm sido propostos na literatura. O modelo *Autoregressive Fractionally Integrated Moving Average* (ARFIMA) Park (1999) e Galbraith e Zinde-Walsh(2001) generaliza os modelos ARMA e ARIMA. Para previsão de séries temporais sazonais, umas variações de ARIMA são: *O modelo de Média Móvel Integrada Autorregressiva Sazonal* (SARIMA) Box e Jenkins(1970), Hipel e Mcleod (1994) e Hamzacebi (2008) é usado. O modelo ARIMA e suas diferentes variações são baseados no famoso

princípio de Box-Jenkins(1970), Hipel e Mcleod (1994) J.Lee (2000) e Zangh (2008) e, portanto, são também amplamente conhecidos como modelos de Box-Jenkins.

Os modelos lineares têm chamado muita atenção devido à sua relativa simplicidade de compreensão e implementação. No entanto, muitas séries temporais práticas mostram padrões não lineares. Por exemplo, conforme mencionado por R. Parrelli (2001) modelos não lineares são apropriados para prever mudanças de volatilidade em séries temporais econômicas e financeiras. Considerando esses fatos, vários modelos não lineares foram sugeridos na literatura. Alguns deles são os famosos modelos *Autoregressive Conditional Heteroskedasticity* (ARCH) Park (1999) e Parrelli (2001) e suas variações como *Generalized ARCH* (GARCH), Park (1999) e Parrelli (2001), *Exponential Generalized ARCH* (EGARCH) Park(1999) etc., o modelo *Threshold Autoregressive* (TAR) Tong (1983) e Zangh (2008), o modelo não *linear autoregressivo* (NAR) Zhang (2003), o modelo *não linear de média móvel* (NMA) Parrelli (2001), etc.

Neste capítulo, discutiremos sobre os importantes modelos de séries temporais estocásticos lineares e não lineares com suas diferentes propriedades. Este capítulo fornecerá um pano de fundo para o próximo capítulo, no qual estudaremos outros modelos usados para previsão de séries temporais.

## 2.1 Os Modelos Autorregressivos de Média Móvel (ARMA)

Um modelo ARMA  $(p, q)$  é uma combinação de modelos AR  $(p)$  e MA  $(q)$  e é adequado para modelagem de série temporal univariada. Em um modelo AR  $(p)$ , o valor futuro de uma variável é assumido como uma combinação linear de  $p$  observações anteriores e um erro aleatório junto com um termo constante. Matematicamente, o modelo AR  $(p)$  pode ser expresso como (HIPEL e Mcleod,1994, J.LEE, 2000).

$$y_t = C + \sum_{i=1}^p \varphi_i y_{t-i} + \varepsilon_t = C + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t$$

(3.1)

Aqui  $y_t$  e  $\varepsilon_t$  são respectivamente o valor real e o erro aleatório (ou choque aleatório) no período  $t$ ,  $\varphi_i$  ( $i = 1, 2, \dots, p$ ) são parâmetros do modelo e  $C$  é uma constante.

A constante inteira  $p$  é conhecida como a ordem do modelo. Às vezes, o termo constante é omitido para simplificar. Normalmente, para estimar os parâmetros de um processo de AR usando a série de tempo dada, as equações de Yule-Walker (HIPEL e MCLEOD,1994) são usadas.

Assim como um modelo AR ( $p$ ) regreda contra valores passados da série, um modelo MA ( $q$ ) usa erros passados como variáveis explicativas. O modelo MA ( $q$ ) é dado por Hipel e Mcleod(1994) , Cochrane (1997) e Lee (2000):

$$y_t = \mu + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

(3.2)

Aqui  $\mu$  é a média da série,  $\theta_j$  ( $j = 1, 2, \dots, q$ ) são os parâmetros do modelo e  $q$  é a ordem do modelo. Os choques aleatórios são assumidos como sendo um processo de ruído branco, (Hipel e Mcleod(1994) e Cochrane (1997)) , ou seja, uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d) com média zero e uma variância constante  $\sigma^2$ . Geralmente, os choques aleatórios são considerados como seguindo a distribuição normal típica. Assim, conceitualmente, um modelo de média móvel é uma regressão linear da observação atual da série temporal contra os choques aleatórios de uma ou mais observações anteriores. Ajustar um modelo MA a uma série temporal é mais complicado do que ajustar um modelo AR porque no primeiro os termos de erro aleatório não são previsíveis.

Os modelos autoregressivos (AR) e de média móvel (MA) podem ser combinados com eficácia para formar uma classe geral e útil de modelos de séries temporais, conhecidos como modelos ARMA. Matematicamente, um modelo ARMA ( $p, q$ ) é representado como (Hipel e Mcleod(1994) , Cochrane (1997) e Lee (2000)):

$$y_t = C + \varepsilon_t + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

Aqui, as ordens do modelo  $p, q$  referem-se a  $p$  autoregressivo e  $q$  termos de média móvel.

Normalmente os modelos ARMA são manipulados usando a notação do operador de atraso Hipel e Mcleod(1994) e Cochrane (1997). O operador de atraso ou retrocesso é definido como  $Ly_t = y_{t-1}$ . Polinômios do operador de atraso ou polinômios de atraso são usados para representar os modelos ARMA da seguinte forma (COCHRANE ,1997):

$$\text{Modelo AR}(p): \varepsilon_t = \varphi(L) y_t.$$

$$\text{Modelo MA}(q): y_t = \theta(L) \varepsilon_t.$$

$$\text{Modelo ARMA}(p, q): \varphi(L)y_t = \theta(L)\varepsilon_t.$$

Onde

$$\varphi(L) = 1 - \sum_{i=1}^p \varphi_i L^i$$

e

$$\theta(L) = 1 - \sum_{j=1}^q \theta_j L_j.$$

É mostrado em Hipel e Mcleod(1994) que uma propriedade importante do processo AR( $p$ ) é a invertibilidade, ou seja, um processo AR( $p$ ) pode sempre ser escrito em termos de um processo MA ( $\infty$ ). Considerando que para um processo MA ( $q$ ) ser invertível, todas as raízes da equação  $\theta(L) = 0$  devem estar fora do círculo unitário. Esta condição é conhecida como condição de invertibilidade para um processo MA.

## 2.2 Análises de estacionariedade

Quando um processo AR( $p$ ) é representado como  $\varepsilon_t = \varphi(L) y_t$ , então  $\varphi(L) = 0$  é conhecida como a equação característica do processo. É provado por Box e Jenkins (1970) que uma condição necessária e suficiente para que o processo AR( $p$ ) seja estacionário é que todas as raízes da equação característica devem cair fora do círculo unitário. Hipel e McLeod (1994) mencionaram outro algoritmo simples (de Schur e Pagano) para determinar a estacionariedade de um processo AR. Por exemplo, como mostrado em Lee (2000) o modelo AR(1)  $y_t = C + \varphi_1 y_{t-1} + \varepsilon_t$  está parado quando  $|\varphi_1| < 1$ , com uma média constante  $\mu = \frac{c}{1-\varphi_1}$  e variância constante  $\gamma_0 = \frac{\sigma^2}{1-\varphi^2}$ .

Um processo MA( $q$ ) é sempre estacionário, independente dos valores dos parâmetros MA (HIPEL E MCLEOD,1994) . As condições relativas à estacionariedade e invertibilidade dos processos AR e MA também são válidas para um processo ARMA. Um processo ARMA( $p, q$ ) é estacionário se todas as raízes da equação característica  $\varphi(L) = 0$  estão fora do círculo unitário. Da mesma forma, se todas as raízes da equação de atraso  $\theta(L) = 0$  estiverem fora do círculo unitário, então o processo ARMA ( $p, q$ ) é invertível e pode ser expresso como um processo AR puro.

### 2.3 Funções de Autocorrelação e Autocorrelação Parcial (ACF e PACF)

Para determinar um modelo adequado para dados de uma determinada série temporal, é necessário realizar a análise ACF e PACF. Essas medidas estatísticas refletem como as observações em uma série temporal estão relacionadas entre si. Para fins de modelagem e previsão, muitas vezes é útil representar graficamente o ACF e o PACF em relação a atrasos de tempo consecutivos. Esses gráficos ajudam a determinar a ordem dos termos AR e MA. Abaixo, damos suas definições matemáticas:

Para uma série temporal  $\{x(t), t = 0,1, 2, \dots\}$  a Autocovariância Cochrane(1997) e Hipel e Mcleod (1994) no atraso  $k$  é definida como:

$$y_k = Cov(x_t, x_{t+k}) = E[(x_t - \mu)(x_{t+k} - \mu)] \quad (3.4)$$

O coeficiente de autocorrelação Cochrane(1997) e Hipel e Mcleod (1994) no atraso  $k$  é definido como:

$$\rho_k = \frac{y_k}{y_0} \quad (3.5)$$

Aqui  $\mu$  é a média da série temporal, ou seja,  $\mu = E[x_t]$ . A autocovariância no atraso zero isto é,  $y_0$  é a variação da série temporal. Da definição fica claro que o coeficiente de autocorrelação  $\rho_k$  é adimensional e, portanto, independente da escala de medição. Também, claramente  $-1 \leq \rho_k \leq 1$ . Os estatísticos Box e Jenkins denominaram  $y_k$  como a *Função de Autocovariância teórica* (ACVF) e  $\rho_k$  como a *Função de Autocorrelação teórica* (ACF).

Outra medida, conhecida como *Função de Autocorrelação Parcial* (PACF), é usada para medir a correlação entre uma observação  $k$  período atrás e a observação atual, após controlar as observações em defasagens intermediárias (ou seja, em defasagens  $< k$ )

(LEE,2000). No atraso 1, PACF (1) é o mesmo que ACF (1). As fórmulas detalhadas para o cálculo do PACF são fornecidas em (BOX e JENKIS,1970, HIPEL E MCLEOD,1994).

Normalmente, o processo estocástico que rege uma série temporal é desconhecido e, portanto, não é possível determinar os valores reais ou teóricos de ACF e PACF. Em vez disso, esses valores devem ser estimados a partir dos dados de treinamento, ou seja, a série temporal conhecida disponível. Os valores estimados de ACF e PACF dos dados de treinamento são respectivamente denominados como amostra de ACF e PACF (BOX e JENKIS,1970, HIPEL E MCLEOD,1994). Conforme dado em Hipel e Mcleod (1997), a estimativa de amostra mais apropriada para o ACVF no atraso  $k$  é

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu) \quad (3.6)$$

Então, a estimativa para a amostra ACF no lag  $k$  é dada por:

$$r_k = \frac{c_k}{c_0} \quad (3.7)$$

Aqui  $\{x(t), t = 0,1,2, \dots\}$  é a série de treinamento de tamanho  $n$  com média  $\mu$ .

Conforme explicado por Box e Jenkins, o gráfico de ACF de amostra é útil para determinar o tipo de modelo para se ajustar a uma série temporal de comprimento  $N$ . Como o ACF é simétrico em relação ao atraso zero, é necessário apenas traçar o ACF de amostra para atrasos positivos, do atraso um em diante até um atraso máximo de cerca de  $N / 4$ . O gráfico de amostra PACF ajuda a identificar a ordem máxima de um processo AR. Os métodos para calcular ACF e PACF para modelos ARMA são descritos em Hipel e Mcleod(1994).

## 2.4 Alguns modelos de séries temporais não lineares

Até agora, discutimos sobre modelos de séries temporais lineares. Conforme mencionado anteriormente, os modelos não lineares também devem ser considerados para uma melhor análise e previsão de séries temporais. Campbell, Lo e McKinley (1997) fizeram contribuições importantes nessa direção. Segundo eles, quase todas as séries temporais não lineares podem ser divididas em dois ramos: um inclui modelos não lineares na média e outro

inclui modelos não lineares na variância (heteroscedástico). Como um exemplo ilustrativo, aqui apresentamos dois modelos de séries temporais não lineares de Parrelli (2001):

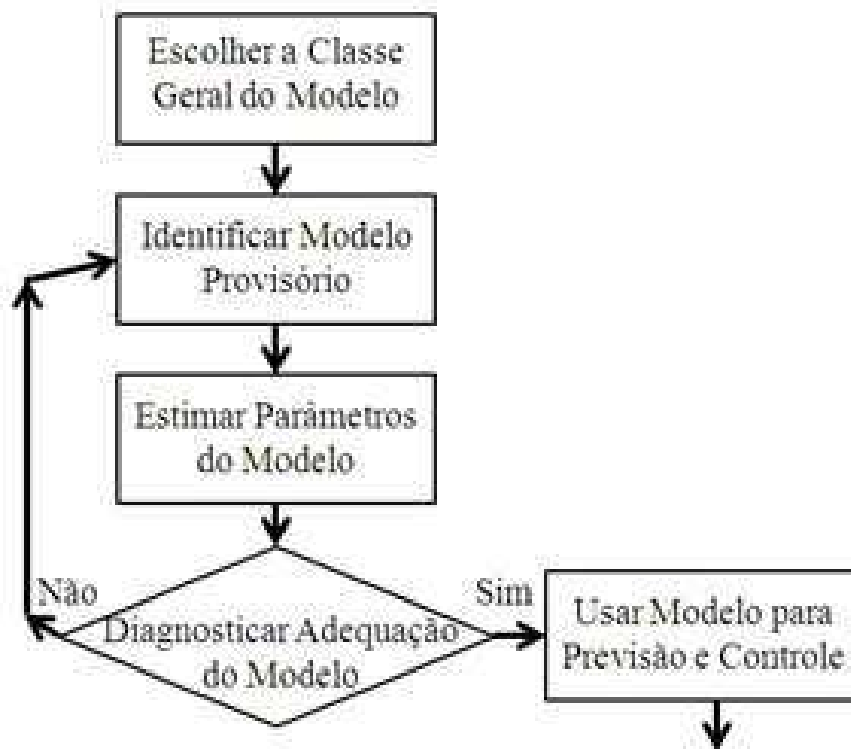
- Modelo de média móvel não linear (NMA):  $y_t = \varepsilon_t + \alpha \varepsilon_{t-1}^2$ . Este modelo não é linear em média, mas não em variância.
- Modelo ARCH de Eagle (1982):  $y_t = \varepsilon_t + \alpha \sqrt{\varepsilon_t^2}$ . Este modelo é heterocedástico, ou seja, não linear em variância, mas linear em média. Este modelo possui várias outras variações, como GARCH, EGARCH etc.

## 2.5 Metodologia Box-Jenkins

Depois de descrever vários modelos de série temporal, a próxima questão para nossa preocupação é como selecionar um modelo apropriado que pode produzir uma previsão precisa com base em uma descrição do padrão histórico nos dados e como determinar os pedidos de modelos ideais. Os estatísticos George Box e Gwilym Jenkins (1970) desenvolveram uma abordagem prática para construir o modelo ARIMA, que melhor se ajusta a uma dada série temporal e também satisfaz o princípio da parcimônia. Seu conceito tem importância fundamental na área de análise e previsão de séries temporais (ZHANG,2008, LOMBARDO e FLAHERTY,2000).

A metodologia Box-Jenkins não assume nenhum padrão particular nos dados históricos das séries a serem previstas. Em vez disso, ele usa uma abordagem iterativa de três etapas de identificação de modelo, estimativa de parâmetro e verificação de diagnóstico para determinar o melhor modelo parcimonioso de uma classe geral de modelos ARIMA (BOX e JENKINS,1970, LOMBARDO e FLAHERTY,2000, ZHANG,2003, LEE, 2000). Este processo de três etapas é repetido várias vezes até que um modelo satisfatório seja finalmente selecionado. Então, esse modelo pode ser usado para prever valores futuros da série temporal.

O método de previsão Box-Jenkins é mostrado esquematicamente na Fig. 2.1:



**Fig. 2.1: A metodologia Box-Jenkins para a seleção do modelo ideal.**

**Fonte: Adaptado de Box, Jenkins e Reinsel (1994, p. 17)**

Uma etapa crucial na seleção de um modelo apropriado é a determinação dos parâmetros ideais do modelo. Um critério é que a amostra de ACF e PACF, calculada a partir dos dados de treinamento, deve corresponder aos valores teóricos ou reais correspondentes (HIPEL E MCLEOD,1994, FARAWAY e CHATFIELD,1998, KIHORO et al. 2000). Outras medidas amplamente utilizadas para identificação de modelo são *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC) Faraway e Chatfield (1998) e KIHORO et al. (2000) , que são definidos abaixo Faraway e Chatfield (1998):

$$AIC(p) = n \ln(\sigma_e^2 / n) + 2p$$

$$BIC(p) = n \ln(\sigma_e^2 / n) + p + p \ln(n)$$

Aqui  $n$  é o número de observações efetivas, usadas para ajustar o modelo,  $p$  é o número de parâmetros no modelo e  $\sigma_e^2$  é a soma dos resíduos ao quadrado da amostra. A ordem ideal do modelo é escolhida pelo número de parâmetros do modelo, o que minimiza AIC ou BIC. Outros critérios semelhantes também foram propostos na literatura para a identificação do modelo ideal.



### 3 CAPÍTULO

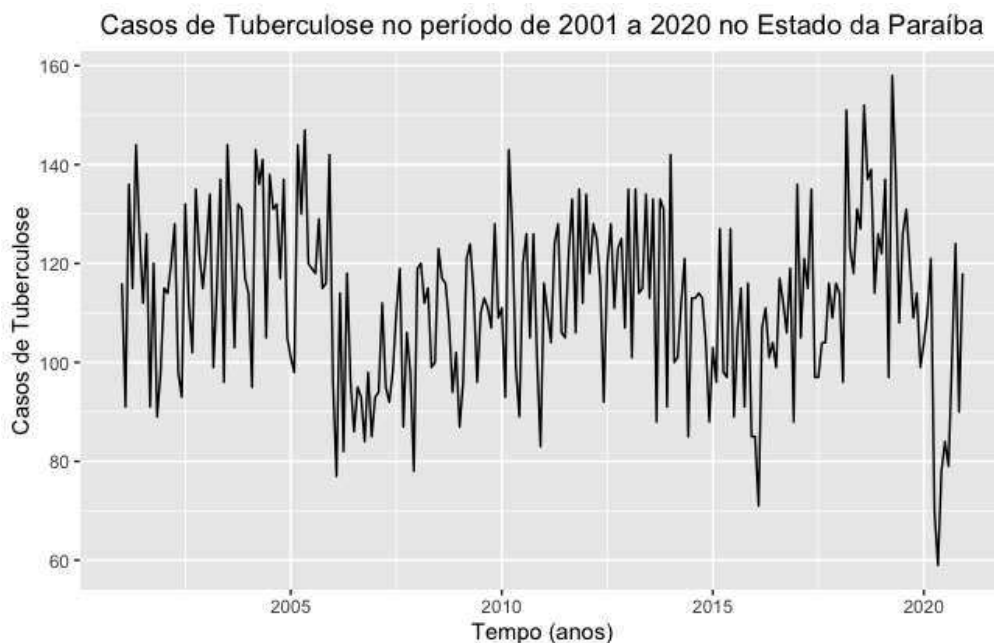
Neste capítulo mostraremos a aplicação de séries temporais e seus resultados, aos dados de tuberculose ocorridos no estado da Paraíba no período de 2001 a 2020.

#### 3.1. Tratamento e Análise de Dados

A partir dos dados de casos de tuberculose da Paraíba, extraídos da DATASUS/MS foram realizadas observações descritiva e estatística. Na avaliação descritiva foi observada a característica da variável caso de tuberculose durante o período estudado. Na avaliação estatística, primeiramente foi realizada a análise gráfica dos dados mensais para o período estudado, buscando encontrar comportamento estacionário, a partir do teste de teste de Dicky-Fuller e, em seguida, a decomposição da série para uma melhor interpretação das componentes da série temporal. Posteriormente ajustou-se a série ao modelo linear com a finalidade de investigar a presença de tendências estatisticamente significativas na série. E, assim através do ajuste e pelo teste F foi decidido se aceita ou rejeita  $H_0$ , ou seja, confirmar a hipótese de comportamento estável dos dados ou rejeitá-la a favor da presença de tendência. Em seguida, foi aplicado um ajuste sazonal com a finalidade de investigar e, caso encontre, remover a sazonalidade dos dados. Após a investigação da tendência e sazonalidade tornou-se necessário plotar os gráfico de Função de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) para através deles encontrar um modelo que melhor se ajuste aos dados. Daí, com os possíveis modelos Box e Jenkins escolhidos foi possível realizar a escolha do que melhor se ajusta por meio do Critério de Informação Akaike (AIC). Com a modelagem ARMA, foi feita a verificação da precisão do ajuste por meio da análise residual, com ajuda do teste de independência Ljung-Box e o teste de normalidade Shapiro-Wilk. Para realização dos ajustes e aplicação dos testes de hipóteses foi utilizado o software R Studios versão [Version 1.4.1106](#), usando os pacotes `forecast`, `urca`, `fpp2`, `ggplot2`, `readxl`, `dbplyr`.

#### 3.2. Validação e Aplicação da Teoria

Na Figura 3.1 podemos observar a distribuição dos casos de tuberculose anual fracionados em meses, no Estado da Paraíba, durante o período de 2001 a 2020. Observa-se que a menor ocorrência de casos ocorreu no ano de 2020, com apenas 59 casos, e a maior ocorreu no ano de 2019, com aproximadamente 158 casos de tuberculose registrados.



**Figura 3.1.** Série temporal de casos de tuberculose registrados na Paraíba no período de 2001 a 2020.

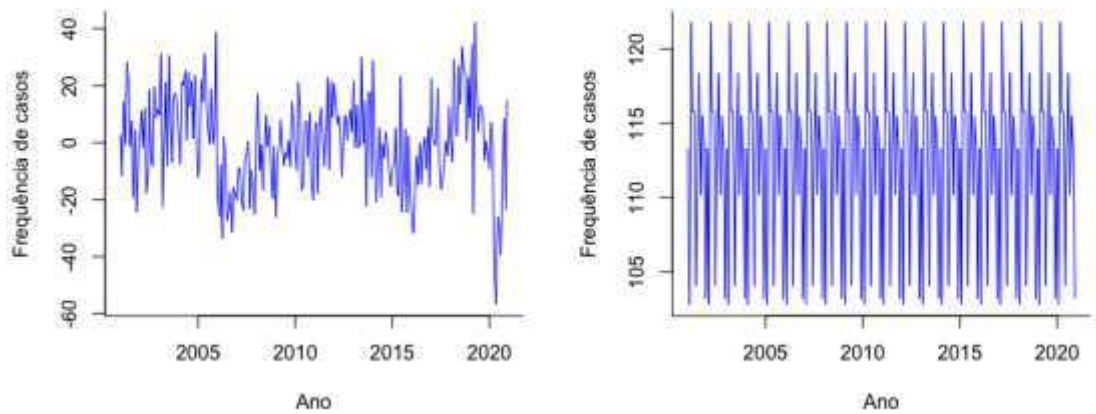
**Fonte:** DATASUS/MS - 2021

A maioria das séries temporais apresentam algum tipo de sazonalidade e, como nós estamos interessados no comportamento cíclico de nossas variáveis devemos tentar limpar nossas séries desse componente sazonal antes de tudo. Para o nosso caso, detectada a sazonalidade usaremos o método mais simples, através de *dummies sazonais* (Ferreira et al., 2015).

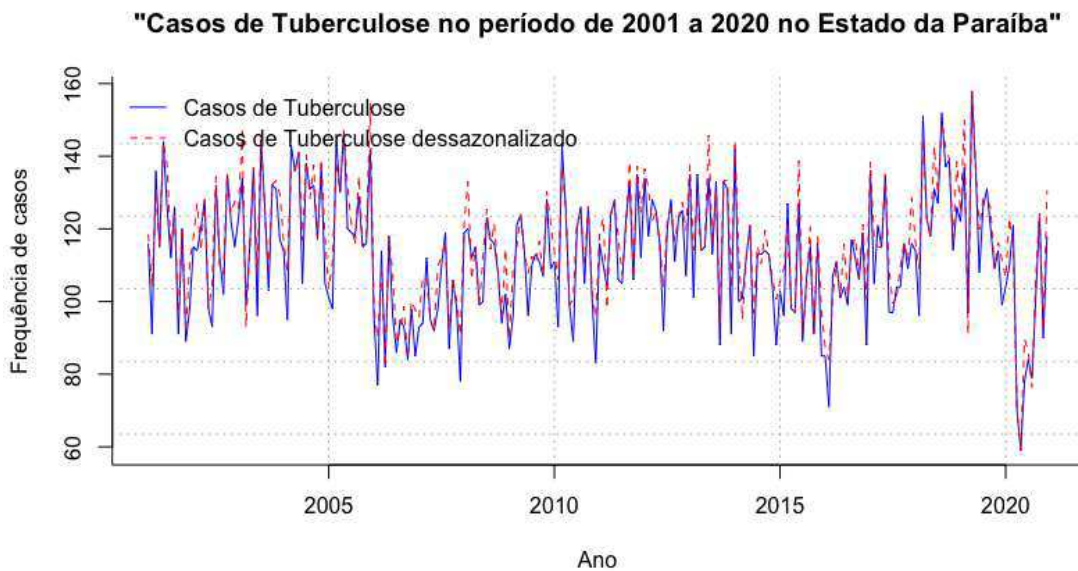
Na nossa análise, o resumo já nos deu o p-valor do teste que neste caso foi igual 0.002814. Logo, rejeitamos a hipótese nula  $H_0$  de que existe sazonalidade. Neste caso, continuaremos com o procedimento de dessazonalizar os dados. Agora, olhemos para os resíduos da regressão, ou seja, nossa série dessazonalizada, e o componente sazonal.

Agora, olhemos para a figura 3.2, onde temos os resíduos da regressão, ou seja, nossa série dessazonalizada, e o componente sazonal. Como podemos ver, a série dessazonalizada

começa em níveis negativos. Isto acontece, pois, ao subtrair  $Y_t$  da nossa série original estamos também retirando a sua média, captada por  $\hat{\alpha}$ . Logo, devemos normalizar nossa série adicionando sua média. Na Figura 3.3 podemos observar o gráfico com a série original e sua versão dessazonalizada.



**Figura 3.2.** Série temporal de casos de tuberculose dessazonalizada e sua componente sazonal



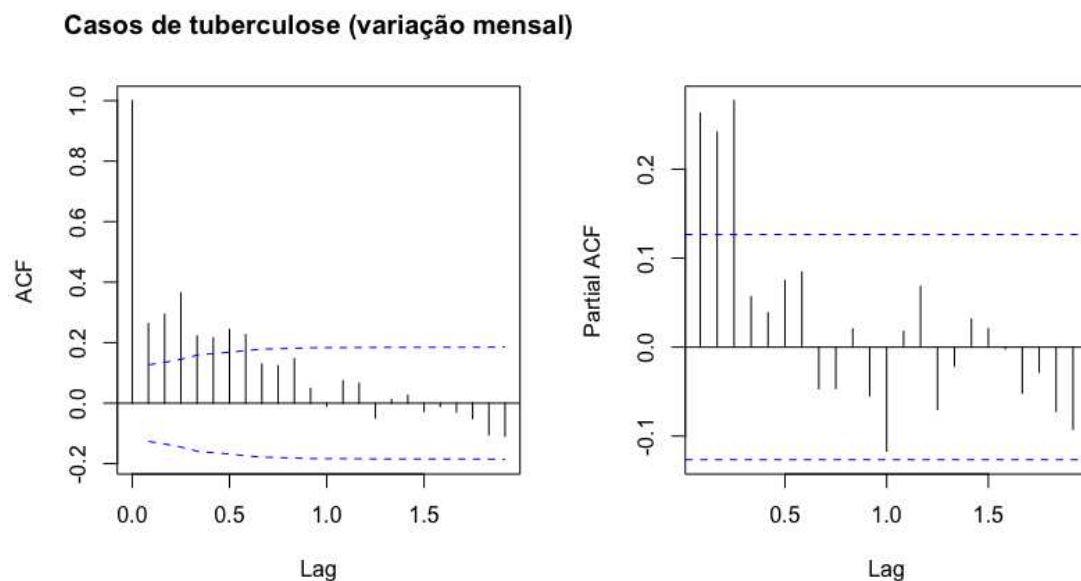
**Figura 3.3** Série temporal de casos de tuberculose e sua série dessazonalizada

A estacionaridade da série foi testada pelo Dickey-Fuller (1979), por meio da estatística t para a hipótese nula de  $\pi = 0$ , onde, o valor foi igual a 0,193 e logo não rejeitamos a hipótese nula ao seguir a distribuição padrão. Assim, concluímos que a série é não estacionária.

Olhando para os correlogramas funções de autocorrelação (FAC) e autocorrelação parcial (FACP) na figura 3.4 temos o que pode parecer, um decaimento exponencial em ambos os gráficos (bem de leve) sendo que a primeira defasagem do *ACF* é bastante significativa. Apesar de ser bem difícil identificar a partir dos correlogramas um modelo a ser estimado, é bem possível se tratar de algo como um *ARMA*. Segundo Morettin e Tolo (2006) os modelos devem ser combinações simples, logo foram escolhidos 4 modelos para serem aplicados, mostrados na Tabela 1.

Para escolher qual o melhor modelo ajustado aos dados e futuramente fazer previsões foi utilizado o Critério de Informação Akaike (AIC), AKAIKE, (1974), sendo o critério mais utilizado quando se trata de séries temporais. Em, seguida foi escolhido o modelo com menor valor AIC, com visto na Tabela 2, foi critério ARMA(2,0,2) ou apenas AR(2).

Como podemos ver na tabela 3 todos os coeficientes são estatisticamente significativos e diferentes de zero. Não sendo necessário reestimar o modelo r nem tampouco retirar a defasagens de maior ou menor valor.



**Figura 3.4.** Correlogramas da Função de Autocorrelação e Autocorrelação Parcial.

**Tabela 1.** Modelo ARMA propostos para Série de Casos de Tuberculose

| <b>Modelo</b> | <b>Configuração</b> |
|---------------|---------------------|
| <b>1</b>      | ARMA(1,0)           |
| <b>2</b>      | ARMA(0,1)           |
| <b>3</b>      | ARMA(1,1)           |
| <b>4</b>      | ARMA(2,2)           |

**Tabela 2.** Critério de AIC para determinação do modelo ARMA

| <b>Modelo</b> | <b>AIC</b> |
|---------------|------------|
| <b>1</b>      | 2023.158   |
| <b>2</b>      | 2013.094   |
| <b>3</b>      | 1981.188   |
| <b>4</b>      | 1980.349   |

**Tabela 3.** Coeficientes estimados para o modelo ARMA(2,2)

| <b>Variável</b> | <b>Estimativas</b> |
|-----------------|--------------------|
| ar1             | 1.59780***         |
| ar2             | -0.65679***        |
| ma1             | -1.49789***        |
| ma2             | 0.65295***         |
| Constante       | 112.46883***       |

\*\*\*significante a 1% de probabilidade pelo teste z

### 3.3. Diagnósticos dos resíduos

Para validar se nosso modelo realmente capta toda a estrutura do processo, verificamos inicialmente se os resíduos da regressão são não autocorrelacionados, para isto utilizou-se o teste de Box-Ljung (1978).

X-squared = 4.7852, df = 2, p-value = 0.09139

Como podemos ver a estatística de teste é igual a 4,78 com p-valor de 0,091, logo não rejeitamos a hipótese nula e podemos concluir que os resíduos são conjuntamente não correlacionados.

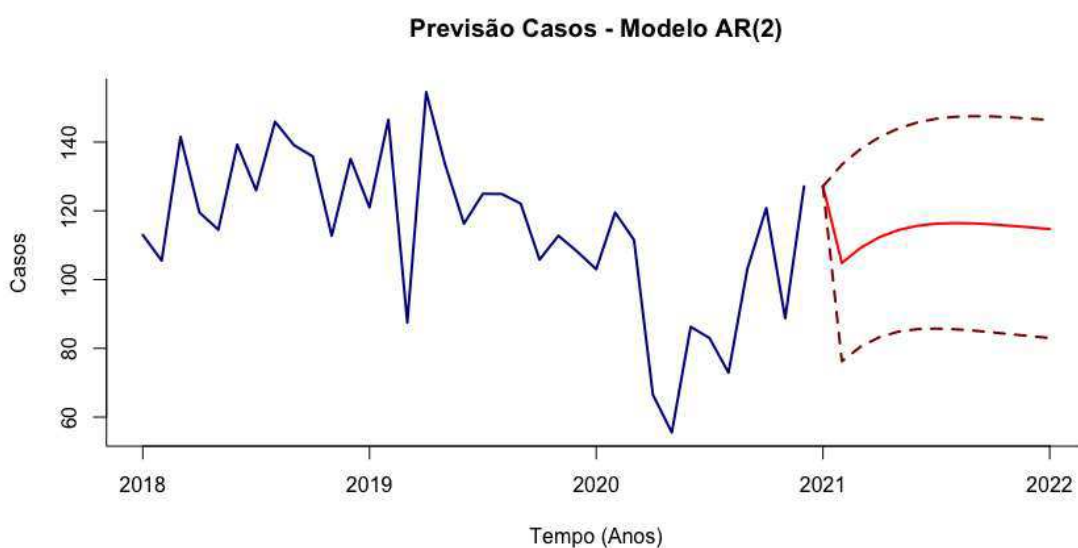
Por fim, verificou-se se os resíduos são normalmente distribuídos, para isto foi utilizado o teste de normalidade de Shapiro-Wilk (1965), onde a hipótese nula é de que a séries analisada é normalmente distribuída. Abaixo o resultado deste teste é apresentado.

$W = 0.99506$ ,  $p\text{-value} = 0.6324$

Como o p-valor é de 0.63, não podemos rejeitar que os resíduos desta regressão são normalmente distribuídos.

### 3.4. Previsão usando Modelo Box-Jenkins

Analisando-se a Figura 3.5 e possível pontuar que a média da casos de tuberculose na Paraíba para o período de 12 meses, apresenta-se em torno da média geral dos dados da série em questão. Ainda de acordo com a Figura 3.5, pode-se observar uma fraca tendência (decaimento) dos casos acumulados para os próximos 12 meses.



**Figura 3.5.** Previsão de casos de tuberculose na Paraíba para um período de 12 meses

### 3.5. Considerações Finais

Após especificar o modelo que melhor se ajustou aos dados, se constatou por meio da previsão que os valores obtidos se encontram num patamar esperado para os próximos 12 meses.

De acordo com os resultados obtidos neste trabalho, conclui-se que para as séries epidemiológicas, o modelo ARMA(2,2), se comporta de forma adequada para análise e previsão dos dados de contagem com variação no tempo, contornando o problema sazonalidade.

Conclui-se que os modelos Box-Jenkins foram adequados para estimar e prever os dados futuros de casos de tuberculose na Paraíba.

Usando os modelos Box-Jenkins, o AR(2) foi empregado e obteve boa precisão devido ao modelo possuir resíduos normalizados e não autocorrelacionados. Isso causou um aspecto de sequência no gráfico dos dados previstos e um valor considerável de intervalo de confiança.

## REFERENCIAS

- ADHIKARI, Ratnadip;R.k Agrawal. *An Introductory study on time modeling an forecasting*.10 ed Rev. Nova York, 2008, 245 p.
- AKAIKE, H. *A new look at the statistical model identification*. *IEEE Transactions on Automatic Control.*, Boston, v.19, n.6, p.716-723, Dec. 1974.
- Antunes JLF, Waldman EA, Moraes M. A tuberculose através do século: ícones canônicos e signos do combate à enfermidade. *Hist Ciênc Saúde-Manguinhos* 2000,5(2):367-79.
- Belculfiné D.C. Os sanatórios e seu tempo. *Pneumol Paulista* 2009;22(5):68-70
- BOX, G. E.; JENKINS, G. M. *Time series analysis: forecasting and control*, revised ed. [S.l.]: Holden-Day, 1976.
- BOX, G.E.P., G. J., “Time Series Analysis, Forecasting and Control”, Holden-Day, San Francisco, CA, 1970.
- BRILLINGER, D. R. *Time series: data analysis and theory*. Philadelphia. Siam, 2001.
- CARRAPATO, Pedro. P.C. *Determinante da saúde no Brasil: a procura da equidade na saúde*, saúde Soc. São Paulo, v.26, n.3, p.676-689, 2017.
- CHATFIELD, C. “Model uncertainty and forecast accuracy”, *J. Forecasting* 15 (1996), pages: 495–508.
- COCHRANE, John H. “Time Series for Macroeconomics and Finance”, Graduate School of Business, University of Chicago, spring 1997.
- D’AMBRÓSIO, U. **Etnomatemática e História da Matemática**. In: FANTINATO, Maria Cecília de Castello Branco (org.). *Etnomatemática – Novos desafios teóricos e pedagógicos*. Niterói: Editora da Universidade Federal Fluminense, 2009, p. 17-28.
- DICKEY, D. A. and FULLER, W. A. *series with a unit root. Distribution of the estimators for autoregressive time*, *Journal of the American statistical association*. 74(366a):427-431, 1979.
- ENGEL, G. I. **Pesquisa-ação**. *Educar*, Curitiba, n. 16, p. 181-191. 2000. Editora da UFPR Disponível em: <[http://www.educaremrevista.ufpr.br/arquivos\\_16/irineu\\_engel.pdf](http://www.educaremrevista.ufpr.br/arquivos_16/irineu_engel.pdf)> Acesso em: 03/04/2014.
- FARAWAY, J. CHATFIELD, C. “Time series forecasting with neural networks: a comparative study using the airline data”, *Applied Statistics* 47 (1998), pages: 231–250.



FERREIRA, P. Métodos de ajuste sazonal para séries de Business Tendency: um estudo de caso para a Sondagem da Indústria utilizando o método X13-ARIMASEATS. FGV: Núcleo de Métodos Estatísticos e Computacionais, 2015.

FIOCRUZ, Tuberculose. Disponível em: ><https://agencia.fiocruz.br/glossario-tuberculose> Acesso em 02 de Ago de 2021.

Fundação Oswaldo Cruz. Escola Nacional de Saúde Pública Sérgio Arouca. Educação a Distância. Controle da tuberculose: uma proposta de integração ensino-serviço. Fundação Oswaldo Cruz. Escola Nacional de Saúde Pública Sergio Arouca. Educação a Distância; coordenado por Maria José Procópio. Rio de Janeiro: EAD/ENSP, 2008.

GALBRAITH, J.W. ZINDE-WALSH, V. “Autoregression-based estimators for ARFIMA models”, CIRANO Working Papers, No: 2011s-11, Feb 2001.

HAMZACEB, C. “Improving artificial neural networks’ performance in seasonal time series forecasting”, Information Sciences 178 (2008), pages: 4550-4559.

HIPEL, K.W. MCLEOD, A.I. “Time Series Modelling of Water Resources and Environmental Systems”, Amsterdam, Elsevier 1994.

K.W. Hipel, A.I. McLeod, “Time Series Modelling of Water Resources and Environmental Systems”, Amsterdam, Elsevier 1994.

KIHHORO, J.M. OTIENO, R.O. WAFULA, C. “Seasonal Time Series Forecasting: A Comparative Study of ARIMA and ANN Models”, African Journal of Science and Technology (AJST) Science and Engineering Series Vol. 5, No. 2, pages: 41-49.

LEE, J. “Univariate time series modeling and forecasting (Box-Jenkins Method)”, Econ 413, lecture 4.

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

LOMBARDO, R. FLAHERTY, J. “Modelling Private New Housing Starts In Australia”, Pacific-Rim Real Estate Society Conference, University of Technology Sydney (UTS), January 24-27, 2000.

MONTENEGRO TH. Tuberculose e literatura: notas de pesquisa. 2ª ed. Rio de Janeiro: A Casa do Livro; 1971.

MORETTIN, P. A.; TOLOI, C.M.C. *Análise de séries temporais*. 2º ed. São Paulo: Blucher, 2006.

NACARATO, Adair Mendes. Eu trabalho primeiro no concreto. In: Revista de Educação Matemática, Ano 9., nos. 9-10, p. 1-6, 2004-2005. ISSN 1676-8868.

NACARATO, Adair Mendes; MENGALI, Brenda Leme da Silva; PASSOS, Cármen Lúcia Brancaglione. *A matemática nos anos iniciais: tecendo fios do ensinar e do aprender*. Belo Horizonte: Autêntica, 2009.

PARK, H. “Forecasting Three-Month Treasury Bills Using ARIMA and GARCH Models”, Econ 930, Department of Economics, Kansas State University, 1999.

PARRELLI, R. “Introduction to ARCH & GARCH models”, Optional TA Handouts, Econ 472 Department of Economics, University of Illinois, 2001.

RAICHARON, T. LURSINSAP, C. SANGUANBHOKI, P. “Application of critical support vector machine to time series prediction”, Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on Volume 5, 25-28 May, 2003, pages: V-741-V-744.

Resende, M. R. K.F. **Neurociência e o Ensino- Aprendizagem em ciências**. Disponível em: < [http://tede.unoeste.br/tede/tde\\_busca/arquivo.php?codArquivo=252](http://tede.unoeste.br/tede/tde_busca/arquivo.php?codArquivo=252)> acesso em 01/03/2016.  
Rios MZ. Sanatório Getúlio Vargas: medicina e relações sociais no combate da tuberculose no Espírito Santo (1942-1967). Dissertação (Mestrado em História). Centro de Ciências Humanas e Naturais. Universidade do Estado do Espírito Santo, UFES, 2009.  
Rosemberg J. Tuberculose – aspectos históricos, realidades, seu romantismo e transculturação. *Bol Pneumol Sanit* 1999;7(2):5-29. 4. Ruffino-Netto A. Tuberculose: a calamidade negligenciada. *Rev Soc Bras Med Trop* 2002;35(1):51-8.

SHAPIRO, S. S.; WILK, M. B. (1965). *An analysis of variance test for normality (complete samples)*. *Biometrika*. 52 (3–4): 591–611

SHEPPARD, D de S. ‘A literatura médica brasileira sobre a peste branca: 1870-1940’. *Hist CiêncSaúde-Manguinhos* 2001;8(1):172-92. 9. Alencar F, Ramalho LC, Ribeiro MV.

TONG, H. “Threshold Models in Non-Linear Time Series Analysis”, Springer-Verlag, New York, 1983.

VEJASAUDE, *Tuberculose: a infecção mais mortal do mundo* Disponível em: ><https://agencia.fiocruz.br/glossario-tuberculose> Acesso em 02 de Ago de 2021

ZHANG , G.P. “A neural network ensemble method with jittered training data for time series forecasting”, *Information Sciences* 177 (2007), pages: 5329–5346.

ZHANG, G.P. “Time series forecasting using a hybrid ARIMA and neural network model”, *Neurocomputing* 50 (2003), pages: 159–175.