



Universidade Estadual da Paraíba  
Centro de Ciências e Tecnologia  
Departamento de Estatística

**Erasnilson Vieira Camilo**

# **Modelos de previsão utilizando séries Temporais**

Campina Grande  
Dezembro de 2012

Erasnilson Vieira Camilo

# Modelos de previsão utilizando séries Temporais

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador:

Gustavo Henrique Esteves

Campina Grande  
Dezembro de 2012

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

C183m Camilo, Erasnilson Vieira.  
Modelos de previsão utilizando séries temporais  
[manuscrito] /Erasnilson Vieira Camilo. – 2012.  
37 f. : il. color.

Digitado.  
Trabalho de Conclusão de Curso (Graduação em  
Estatística) – Universidade Estadual da Paraíba, Centro de  
Ciências e Tecnologia, 2012.  
“Orientação: Prof. Dr. Gustavo Henrique Esteves,  
Departamento de Estatística”.

1. Séries temporais. 2. Metodologia Box e Jenkins.  
3.SARIMA. I. Título.

21. ed. CDD 551.55

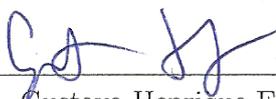
Erasnilson Vieira Camilo

# Modelos de previsão utilizando séries Temporais

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 05/12/2012.

## Banca Examinadora:



---

Prof. Dr. Gustavo Henrique Esteves  
Orientador



---

Profa. Drª. Divanilda Maia Esteves  
Universidade Estadual da Paraíba



---

Prof. Dr. Ricardo Alves de Olinda  
Universidade Estadual da Paraíba

# Agradecimentos

Agradeço a Deus, por sempre me fortalecer e sempre estar à frente de minhas vitórias.

Aos professores Ricardo e Gil, por seus ensinamentos, sendo exemplos de seriedade e profissionalismo.

Ao professor Juarez e a professora Diana, pelo grande apoio e incentivo ao longo na carreira acadêmica. Em especial, ao professor Gustavo por ter sido um ótimo orientador na realização deste trabalho, passando seriedade e confiança ao longo de todo o trabalho.

Aos amigos que me acompanharam durante os anos de minha graduação, em especial a Edlaine, Fernanda e Alessandra, companheiras de batalha em busca do conhecimento.

A minha mãe, Margarida, e ao meu irmão, Nildo, e meu pai, Erasmo (in memoriam), por todo apoio e compreensão, e acreditarem que o mundo sem educação é um conjunto vazio. A eles devo tudo que tenho.

# Resumo

No desenvolvimento e execução de determinadas estratégias, faz-se necessário um desenvolvimento prévio para tomada de decisão. Nesse contexto, entende-se que há necessidade de construir e utilizar ferramentas estatísticas para auxiliar os gestores na tomada de decisão. Contudo, o estudo de séries temporais a ser abordado nesse trabalho expõe uma metodologia bastante conhecida nessa área, ou seja, a utilização da metodologia Box e Jenkins com os respectivos passos para a identificação do modelo. Ademais, será utilizada uma única variável, ou seja, totais mensais de passageiros em linhas aéreas internacionais nos EUA, do banco de dados “*AirPassengers*”, bastante conhecida nos trabalhos de séries temporais e disponível no programa estatístico R. Através do *software* R versão 2.14.2 identifica-se os gráficos das funções de autocorrelação e autocorrelação parcial das diferenças simples e sazonais, entre outras coisas, destaca-se o teste de LJung-Box utilizado para verificação da autocorrelação residual na etapa de validação do modelo. Assim, com base no critério de seleção Bayesiano selecionou-se de forma iterativa o modelo com menor valor BIC, nesse caso, o modelo SARIMA(1, 1, 0) × (0, 1, 0), composto por um parâmetro autoregressivo com uma diferença simples e uma diferença sazonal, através do qual será utilizado para previsão dos meses posteriores.

**Palavras-chaves:** Séries temporais, metodologia Box e Jenkins, SARIMA.

# Abstract

In the development and implementation of specific strategies, it is necessary to advance a development decision making. In this context, it is understood that there is a need to build and use statistical tools to assist managers in making decisions. However, the study of time series to be addressed in this study exhibit a method well known in this field, namely the use of methodology Box and Jenkins with respective steps for identification of the model. Moreover, it will be used a single variable, namely, monthly totals of international airline passengers in the U.S., the series “*AirPassengers*”, well known in the work of series. Through the *software* R version 2.14.2 identifies the graphs of the autocorrelation functions and partial autocorrelation of simple and seasonal differences, among other aspects, there is the Ljung-Box test used to check residual autocorrelation in step validation model. Thus, based on Bayesian selection criterion was selected interactively model with the lowest BIC this case, the model SARIMA  $(1, 1, 0) \times (0, 1, 0)$ , comprising a autoregressive parameter difference with a simple and a seasonal difference, through which will be used for prediction of subsequent months.

**Keywords:** time series, Box and Jenkins methodology, SARIMA.

# Sumário

## Lista de Figuras

## Lista de Tabelas

|          |   |       |
|----------|---|-------|
| <b>1</b> | <b>Introdução</b>                                     | p. 11 |
| <b>2</b> | <b>Fundamentação Teórica</b>                          | p. 12 |
| 2.1      | Introdução à Análise de séries Temporais . . . . .    | p. 12 |
| 2.1.1    | Processo estocástico . . . . .                        | p. 13 |
| 2.1.2    | Processo estacionário . . . . .                       | p. 15 |
| 2.1.3    | Função de autocorrelação (FAC) . . . . .              | p. 16 |
| 2.2      | Modelos para séries temporais . . . . .               | p. 17 |
| 2.2.1    | Passeio aleatório . . . . .                           | p. 17 |
| 2.2.2    | Modelo autoregressivo (AR) . . . . .                  | p. 18 |
| 2.2.3    | Função de autocorrelação parcial (FACP) . . . . .     | p. 19 |
| 2.2.4    | Processo de média móveis (MA) . . . . .               | p. 19 |
| 2.2.5    | Modelos mistos (ARMA) . . . . .                       | p. 20 |
| 2.2.6    | Modelos mistos integrados (ARIMA) . . . . .           | p. 21 |
| 2.2.7    | Modelos Mistos Integrados Sazonais (SARIMA) . . . . . | p. 22 |
| 2.2.8    | Método de previsão . . . . .                          | p. 23 |
| 2.3      | Metodologia de Box e Jenkins . . . . .                | p. 24 |
| 2.3.1    | Identificação do modelo . . . . .                     | p. 25 |
| 2.3.2    | Estimação . . . . .                                   | p. 27 |

|          |                                     |       |
|----------|-------------------------------------|-------|
| 2.3.3    | Verificação e diagnóstico . . . . . | p. 27 |
| <b>3</b> | <b>Aplicação</b>                    | p. 29 |
| <b>4</b> | <b>Conclusão</b>                    | p. 36 |
|          | <b>Referências</b>                  | p. 37 |

# Lista de Figuras

|    |   |       |
|----|---|-------|
| 1  | Temperatura do ar, de dado local, durante 24 horas. . . . .                               | p. 12 |
| 2  | Um processo estocástico interpretado como uma família de variáveis aleatórias. . . . .    | p. 14 |
| 3  | Ciclo iterativo da metodologia de Box e Jenkins. . . . .                                  | p. 21 |
| 4  | Gráficos da série AirPassengers . . . . .   | p. 30 |
| 5  | Gráficos da série AirPassengers com uma diferença mensal. . . . .                         | p. 30 |
| 6  | Gráficos da série AirPassengers com diferença sazonal . . . . .                           | p. 31 |
| 7  | Gráfico da série AirPassengers com uma diferença simples e uma diferença sazonal. . . . . | p. 31 |
| 8  | Gráfico para validação do modelo. . . . .   | p. 33 |
| 9  | Gráfico do nível descritivo para o teste Ljung-Box. . . . .                               | p. 33 |
| 10 | Gráfico da normalidade dos resíduos com p-valor para o teste Shapiro Wilks. . . . .       | p. 34 |
| 11 | Histograma dos resíduos com teste Shapiro Wilks. . . . .                                  | p. 34 |
| 12 | Gráfico dos dados preditos no período de janeiro a dezembro de 1960. . . . .              | p. 35 |

# Lista de Tabelas

|   |   |       |
|---|---|-------|
| 1 | Identificação dos modelos ARIMA. . . . .  | p. 25 |
| 2 | Tabela do modelo ajustado para séries AirPassengers. . . . .  | p. 32 |
| 3 | Valores observados e previsto para a séries AirPassengers. . . . .  | p. 32 |
| 4 | Previsões e Limites inferior (Li) e Limite Superior (Ls) dos intervalos de previsões para os totais mensais de passageiros em linhas aéreas internacionais no EUA no período de janeiro de 1960 a dezembro de 1960. . . | p. 35 |

# 1 Introdução

O objetivo desse trabalho é estudar modelos de séries temporais, aplicado a série totais mensais de passageiros em linhas aéreas internacionais nos EUA disponível no pacote *datasets* do *Software R* versão 2.14.2. Dessa forma procura-se um modelo satisfatório com previsões cada vez melhores em relação às observações mensais. Nesse contexto, será exposta a metodologia dos processos autoregressivos e de médias móveis, além de abordar a metodologia de Box e Jenkins com seus respectivos passos para seleção e identificação do modelo.

É fato, que em diversos países os atrasos dos passageiros e excessos de filas vão aos poucos desestimulando os passageiros, que em muitos casos procuram outras formas de viajar, nesse sentido faz-se necessário investimentos e planejamento no sentido de melhorar o serviço e buscar a satisfação dos clientes. Dessa forma, a previsão de séries temporais pode servir como uma ferramenta a orientar os gestores das companhias aéreas na realização de um planejamento prévio contendo supostas previsões do quantitativo de passageiros em um determinado período de tempo, ou seja, com conhecimento de tais quantitativos, a empresa poderá planejar estratégias que atinjam os períodos de maior e menor fluxo de passageiros, visando-se a maximização do lucro e excelência na prestação dos serviços, diversificando-se em determinadas áreas da empresa, como por exemplo: a demanda necessária de funcionários para um determinado período ou pacotes aéreos promocionais.

Em síntese, será apresentada uma metodologia envolvendo assuntos como processos estocásticos, posteriormente funções de autocorrelação e autocorrelação parcial até os efeitos de sazonalidades para os modelos do tipo  $SARIMA(p, d, q) \times (P, D, Q)$ . Na aplicação o modelo será identificado por meio do processo iterativo com auxílio do pacote *forecast* e da função *auto.arima* que é capaz de auxiliar bastante na identificação do melhor modelo a ser usado, tal seleção é feita com base no Critério de Seleção Bayesiano (BIC).

## 2 Fundamentação Teórica

### 2.1 Introdução à Análise de séries Temporais

A ideia para aplicação das séries temporais consiste na possibilidade de retirar conclusões sobre determinado comportamento de uma série de dados do passado e inferir seu comportamento futuro, ou seja, estudar um conjunto de observações de forma ordenada em função do tempo ( $t$ ).

Segundo Fischer (1982), a análise de um modelo de série de tempo é um estudo que busca as características comportamentais e sistemáticas da série, capazes de construir modelos de previsão de uma série a partir de um conjunto de observações de uma variável.

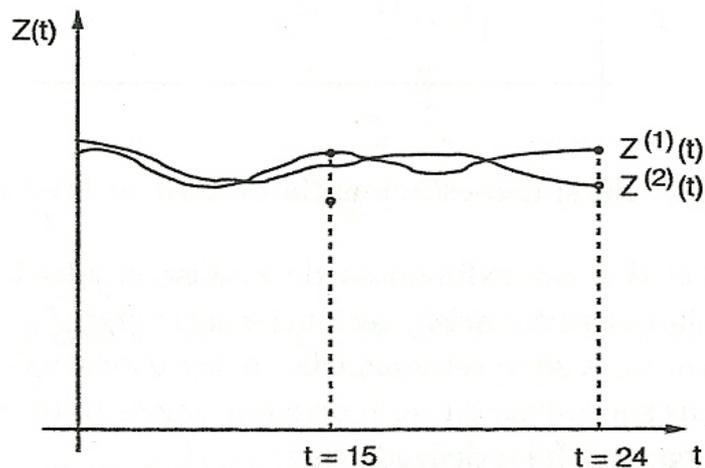


Figura 1: Temperatura do ar, de dado local, durante 24 horas.  
(MORETTIN; TOLOI, 2004)

Em Morettin e Toloi (2004) pode-se observar conforme Figura 1, duas curvas representando dois dias diferentes. Nesse caso, considera-se a função  $Z(t)$  a temperatura no instante  $t$  (tempo em horas). As curvas são trajetórias do processo físico que está sendo

observado, e o processo estocástico um conjunto de todas as possíveis trajetórias que pode-se observar, onde cada trajetória também é conhecida por uma série temporal. Ao considerar  $Z^{(1)}(15)$  o valor da temperatura no instante  $t = 15$ , para a primeira trajetória (primeiro dia de observação), tem-se um número real, para o segundo dia haverá outro número real  $Z^{(2)}(15)$ . Em geral, pode-se denotar uma trajetória qualquer por  $Z^{(j)}(t)$ . Para cada  $t$  fixo, existe uma variável aleatória  $Z(t)$ , com uma determinada distribuição de probabilidade.

De forma simplificada, Fischer (1982) afirma que a análise de uma série temporal pressupõe que exista um processo estocástico, onde cada realização aleatória da variável em estudo esteja associada a uma certa probabilidade de ocorrência da observação. Assim, procura-se descrever as características de sua aleatoriedade com finalidade de fornecer conclusões sobre as probabilidades associadas aos valores futuros alternativos da série.

Em síntese, uma série temporal significa um conjunto de variáveis aleatórias conjuntamente distribuídas no tempo. Sua análise baseia-se na suposição da existência de alguma função que assume probabilidades para todas as possíveis combinações dos valores da variável. Consequentemente, se é possível descrever numericamente como é a estrutura probabilística da variável no tempo, então poder-se-á inferir sobre a probabilidade de ocorrência de um outro futuro valor (FISCHER, 1982).

### 2.1.1 Processo estocástico

Os Processos estacionários são uma importante classe dos processos estocásticos podendo ser estritamente estacionária ou fracamente estacionária. O objeto da teoria dos processos estocásticos é o estudo daqueles mecanismos dinâmicos que proporcionam meios de análise de uma sequência de observações, vista conjunta e interdependentemente em uma sucessão de momentos de tempo, as quais são influenciadas por fatores aleatórios (FISCHER, 1982).

Segundo Fischer (1982) pode-se definir um processo estocástico como uma família de variáveis aleatórias  $\{X(t), t \in T\}$ , classificada mediante um parâmetro  $t$  que varia em um intervalo de tempo  $T$ . Portanto um processo estocástico caracteriza-se por ser uma função aleatória em  $t$ , para uma sequência de observações  $X(t)$ ,  $t = 1, 2, 3, 4, \dots$ , que corresponde a uma amostra de pontos no tempo, é chamada uma realização parcial do processo estocástico.

Assim como Fischer (1982), Morettin e Toloi (2004) definem processos estocásticos

como:

**Definição 2.1. (Processo estocástico)** *Seja  $T$  um conjunto arbitrário. Um processo estocástico é uma família  $Z = \{Z(t), t \in T\}$ , tal que, para cada  $t \in T$ ,  $Z(t)$  é uma variável aleatória.*

Conforme Definição 2.1, observe que  $Z(t)$  é uma função de dois argumentos, ou seja,  $Z(t, \omega)$ ,  $t \in T$  e  $\omega \in \Omega$ . A Figura 2, ilustra a interpretação de um processo estocástico.

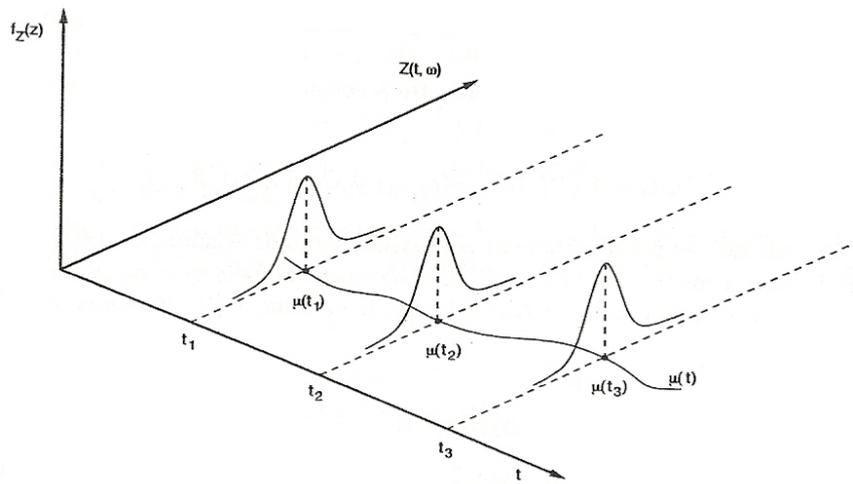


Figura 2: Um processo estocástico interpretado como uma família de variáveis aleatórias. (MORETTIN; TOLOI, 2004)

Na Figura 2 pode-se observar a interpretação do processo estocástico, no qual, para cada  $t \in T$ ,  $Z(t)$  tem uma variável aleatória  $Z(t, \omega)$  com uma distribuição de probabilidade, porém, é possível que as função de densidade de probabilidade sejam diferentes entre os períodos  $t_1$  e  $t_2$  por exemplo. Ademais, espera-se que a função de densidade de probabilidade de  $Z(t, \omega)$  seja a mesma para todo  $t \in T$ . Nesse contexto, para cada  $\omega \in \Omega$  fixado, obtém-se uma função de  $t$ , ou seja, uma série temporal do tipo  $Z(t)$ .

Segundo Morettin e Toloí (2004), há situações que pretende-se utilizar modelos para descrever séries de tempo, para isso, faz-se necessário seguir algumas suposições:

- *i.* Processos estacionários ou não-estacionários, de acordo com a independência ou não relativamente à origem dos tempos;
- *ii.* Processos normais (Gaussianos) ou não-normais, de acordo com as funções de densidade de probabilidade (f.d.p) que caracterizam os processos;
- *iii.* Processos Markovianos ou não-Markovianos, de acordo com a independência dos valores do processo, em dado instante, de seus valores em instantes precedentes.

## 2.1.2 Processo estacionário

**Definição 2.2. (Processo estacionário)** *Um processo estocástico  $Z = \{Z(t), t \in T\}$  diz-se estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo.*

$$F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n),$$

para quaisquer  $t_1, \dots, t_n, \tau$  de  $T$ .

Logo, todas as distribuições unidimensionais são invariantes sob translação do tempo, então, tem-se média  $\mu(t)$  e variância  $\sigma^2(t)$  constante, ou seja,

$$E[Z(t)] = \mu(t) = \mu \quad e \quad Var[Z(t)] = \sigma^2(t) = \sigma^2.$$

Os processos estocásticos podem ser classificados como fracamente estacionário, de acordo com a Definição 2.3 a seguir.

**Definição 2.3. (Processo fracamente estacionário)** *Um processo estocástico  $Z = \{Z(t), t \in T\}$  diz-se fracamente estacionário ou estacionário de segunda ordem se e somente se*

$$E\{Z(t)\} = \mu(t) = \mu, \quad \text{constante, para todo } t \in T;$$

$$E\{Z^2(t)\} < \infty, \quad \text{para todo } t \in T;$$

$$\gamma(t_1, t_2) = Cov\{Z(t_1), Z(t_2)\} \text{ é uma função de } |t_1 - t_2|.$$

De acordo com Morettin e Tolo (2004), a maioria das séries encontradas, apresentam alguma forma de não estacionariedade e tendências. Logo, para vários procedimentos e análises de séries temporais faz-se necessário que as mesmas sejam estacionárias. Nesse sentido, é necessário a utilização de transformações para surgir uma série estacionária e prosseguir com as aplicações. São várias as formas de transformações para tornar uma série estável, considerando-se  $Z(t)$  uma série temporal não estacionária, aplica-se a primeira diferença para tornar a série estável, de acordo com o seguinte processo

$$\Delta Z(t) = Z(t) - Z(t - 1). \tag{2.1}$$

A segunda diferença será:

$$\begin{aligned}
\Delta^2 Z(t) &= \Delta[\Delta Z(t)] \\
&= \Delta[Z(t) - Z(t-1)] \\
&= \Delta Z(t) - \Delta[Z(t-1)] \\
&= Z(t) - Z(t-1) - [Z(t-1) - Z(t-2)] \\
&= Z(t) - Z(t-1) - Z(t-1) + Z(t-2) \\
&= Z(t) - 2Z(t-1) + Z(t-2).
\end{aligned}$$

Tomando-se a  $n$ -ésima diferença tem-se

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)]. \quad (2.2)$$

Assim como tomar  $n$  diferenças com a finalidade de resultar numa série estacionária, pode-se aplicar o logaritmo nos dados e também aplicar sucessivas diferenças na série. Contudo, ao aplicar o logaritmo na série faz-se necessário que nas análises finais o pesquisador volte a escala original para finalizações e conclusões sobre o estudo da série.

### 2.1.3 Função de autocorrelação (FAC)

Segundo GREENE (apud LEROY, 2009, p. 46): “autocorrelação é a correlação entre membros de uma série de observações ordenadas no tempo”. Logo, a Função de autocorrelação (FAC) é derivada da função de autocovariância podendo ser definida como:

$$\rho_\tau = \frac{\text{cov}(Y_t, Y_{t-\tau})}{\text{Var}(Y_t)}, \quad (2.3)$$

que pode ser estimada por:

$$r_\tau = \hat{\rho}_\tau = \frac{\sum_{t=1}^{n-\tau} (Y_t - \bar{Y})(Y_{t-\tau} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}. \quad (2.4)$$

Considere  $\tau$  como a defasagem da série. Tendo-se uma série estacionária então a covariância entre  $Y_t$  e  $Y_{t-\tau}$  não dependerá de  $t$ , bem como as variâncias de  $Y_t$  e  $Y_{t-\tau}$  são constantes, logo, a equação poderá ser escrita como:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}. \quad (2.5)$$

## 2.2 Modelos para séries temporais

Os modelos de Séries Temporais podem-se classificar quanto ao número de parâmetros, alguns deles utilizam um número finito de parâmetros, ou seja, modelos paramétricos, enquanto outros, não-paramétricos, envolvem um número infinito de parâmetros. Dos modelos paramétricos, os mais conhecidos são os modelos de Erro, os Modelos Autoregressivos (ARMA, ARIMA) e os modelos não-lineares (ARCH, GARCH), onde, em todos estes modelos, a análise da série encontra-se no domínio do tempo.

### 2.2.1 Passeio aleatório

Ao estudar modelos autoregressivos é importante mencionar a ideia do Passeio aleatório em que os  $\epsilon_t$  são independentes. Conforme a definição de Ehlers (2007) tem-se que:

**Definição 2.4. (Passeio aleatório):** *Considere  $\{\epsilon_t\}$  um processo discreto, aleatório com média  $\mu$  e variância  $\sigma_\epsilon^2$ . Um processo  $Y_t$  é chamado de passeio aleatório se:*

$$Y_t = Y_{t-1} + \epsilon_t. \quad (2.6)$$

Por meio de várias substituições, tem-se

$$\begin{aligned} Y_t &= Y_{t-2} + \epsilon_{t-1} + \epsilon_t \\ Y_t &= Y_{t-3} + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t \\ &\vdots \\ Y_t &= Y_0 + \sum_{j=1}^t \epsilon_j. \end{aligned}$$

Iniciando-se o processo  $Y_0 = 0$ , obtém-se as seguintes expressões:

$$E(Y_t) = \sum_{j=1}^t E(\epsilon_j) = t\mu \quad (2.7)$$

$$Var(Y_t) = \sum_{j=1}^t Var(\epsilon_j) = t\sigma_\epsilon^2. \quad (2.8)$$

Com função de covariância definida por:

$$Cov(Y_t, Y_{t-k}) = Cov(\epsilon_1 + \dots + \epsilon_{t-k} + \dots + \epsilon_t, \epsilon_1 + \dots + \epsilon_{t-k}) = (t-k)\sigma_\epsilon^2. \quad (2.9)$$

Logo, pelas Equações (2.3), (2.8) e (2.9) a Equação de correlação será:

$$\rho_t(k) = \frac{t-k}{t}.$$

Com média, variância e autocovariâncias dependendo de  $t$  o processo será não estacionário, porém, tomando-se a primeira diferença, ou seja,

$$\Delta Y_t = Y_t - Y_{t-1} = \varepsilon_t,$$

o passeio aleatório será estacionário.

## 2.2.2 Modelo autoregressivo (AR)

Considere que  $\{\varepsilon_t\}$  seja um processo aleatório com média zero e variância  $\sigma_\varepsilon^2$ . Um processo  $Y_t$  é denominado processo autoregressivo de ordem  $p$ , ou  $AR(p)$  se

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \varepsilon_t. \quad (2.10)$$

Segundo Ehlers (2009), processos AR podem ser usados como modelos se for razoável assumir que o valor atual de uma série temporal depende do seu passado imediato mais um erro aleatório. Por meio do operador  $L$  como translação para o passado, tem-se

$$\begin{aligned} LY_t &= Y_{t-1} \\ L^m Y_t &= Y_{t-m} \\ \Delta Y_t &= Y_t - Y_{t-1} = (1 - L)Y_t \\ \Delta &= 1 - L, \end{aligned}$$

pode-se reescrever a Equação (2.10) como:

$$\begin{aligned} Y_t - \beta_1 Y_{t-1} - \beta_2 Y_{t-2} - \dots - \beta_p Y_{t-p} &= \varepsilon_t \\ Y_t - \beta_1 LY_t - \beta_2 L^2 Y_t - \dots - \beta_p L^p Y_t &= \varepsilon_t \\ Y_t(1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_p L^p) &= \varepsilon_t. \end{aligned} \quad (2.11)$$

Dessa forma tem-se um polinômio de ordem  $p$  multiplicando  $Y_t$ , com isso, pode-se escrever o modelo  $AR(p)$  de forma reduzida

$$B_p(L)Y_t = \varepsilon_t, \quad (2.12)$$

em que  $B_p(L)$  representa um polinômio de  $L$  como observado na Equação (2.11).

### 2.2.3 Função de autocorrelação parcial (FACP)

De acordo com Ehlers (2009) uma função de autocorrelação parcial pode ser identificada com o  $j$ -ésimo coeficiente de uma série AR( $p$ ), dessa forma tem-se por exemplo processos AR( $p$ ):

$$AR(1) : Y_t = \beta_1 Y_{t-1} + \epsilon_t$$

$$AR(2) : Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \epsilon_t$$

$$AR(p) : Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t.$$

Dessa forma,  $\beta_1, \beta_2, \beta_p$  são coeficientes de correlação parcial, considerando-se  $k$  e  $\rho$ , tem-se Equações de Yule-Waker, seguindo a ideia de Morettin e Tolo (2004), tem-se que:

$$\rho_1 = \beta_1 + \beta_2 \rho_1 + \cdots + \beta_p \rho_{p-1}$$

$$\rho_2 = \beta_1 \rho_1 + \beta_2 + \cdots + \beta_p \rho_{p-2}$$

$$\rho_p = \beta_1 \rho_{p-1} + \beta_2 \rho_{p-2} + \cdots + \beta_p.$$

Logo, pode-se calcular os coeficientes de autocorrelação, por meio do valor  $p$  e das autocorrelações  $\rho_1, \rho_2, \dots, \rho_k$  utilizando-se a seguinte forma matricial:

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \beta_{k1} \\ \beta_{k2} \\ \vdots \\ \beta_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (2.13)$$

Segundo Morettin e Tolo (2004), as Equações de Yule-Walker podem ser resolvidas por meio de um procedimento recursivo conhecido como *algoritmo de Durbin-Levinson*.

### 2.2.4 Processo de média móvel (MA)

No processo de médias móveis a variável em questão  $Y_t$  é uma combinação entre os  $\epsilon_t$  ocorridos no tempo ( $t$ ) e ( $t-1$ ), formando-se um processo conhecido como médias móveis e nesse exemplo tem-se uma MA(1). Segundo Ehlers (2009) um processo MA( $q$ ) pode ser descrito como

$$Y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q}, \quad (2.14)$$

utilizando-se o operador  $L^n \epsilon_t = \epsilon_{t-n}$ , tem-se que

$$Y_t = \epsilon_t(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q), \quad (2.15)$$

resumidamente:

$$Y_t = \Theta_q(L)\epsilon_t, \quad (2.16)$$

em que  $\epsilon_t$  representa uma sequência *iid*<sup>1</sup> com média zero e variância  $\sigma_\epsilon^2$  e  $\Theta_q(L)$  representa um polinômio de L por meio da Equação (2.15).

No modelo  $Y_t = \epsilon_t - \theta_1 \epsilon_{t-1}$ , observe que para qualquer valor dado  $Y_t$  na série, ele é diretamente proporcional ao erro aleatório  $\epsilon_t$  do período anterior mais um erro aleatório corrente. Em (2.14),  $\theta_1, \theta_2, \dots, \theta_q$  são parâmetros da séries MA(q). Nesse contexto, qualquer valor da série  $Y_t$  é expresso como uma combinação aritmética dos  $q$  erros aleatórios passados mais um erro aleatório corrente.

## 2.2.5 Modelos mistos (ARMA)

Segundo Ehlers (2007) o modelo ARMA(p,q) pode ser representado como combinação dos modelos  $AR(p)$  e  $MA(q)$ , com isso, pode-se obter uma representação adequada para o modelo com um número menor de parâmetros. Processos autoregressivos de médias móveis (ARMA) formam um classe de modelos muito úteis e parcimoniosos para descrever dados de séries temporais. O modelo ARMA(p, q) será descrito como:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}, \quad (2.17)$$

ou, de acordo com as expressões (2.12) e (2.16),

$$B_p(L)Y_t = \Theta_q(L)\epsilon_t. \quad (2.18)$$

Observe que modelos AR(p) ou MA(q) podem ser obtidos quando  $p = 0$  ou  $q = 0$ . Seguindo a ideia de Fischer (1982), a justificativa da utilização de modelos ARMA reside no fato de que em muitas séries há necessidade de ambos os termos, tanto autoregressivo quanto de médias móveis, por isso, com a combinação de ambos os termos, tem-se modelos com número de parâmetros reduzido, sendo mais satisfatório do que a utilização de apenas modelos autoregressivos ou de médias móveis. A Figura 3, apresenta uma ideia mais ampla no processo de seleção do modelo. No primeiro momento parece um pouco repetitivo, porém, com o auxílio dos *softwares* e suas rotinas computacionais há um ganho

---

<sup>1</sup>Independente e igualmente distribuída

significativo em relação ao tempo gasto na identificação do modelo por outros métodos.

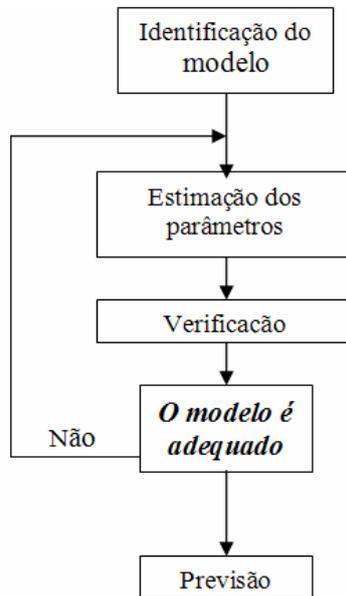


Figura 3: Ciclo iterativo da metodologia de Box e Jenkins.

## 2.2.6 Modelos mistos integrados (ARIMA)

Segundo (EHLERS, 2007), ao considerar um processo estocástico não estacionário, se esse, apresenta características onde sua estacionariedade é do tipo homogênea, pode-se por meio da transformação com um operador diferença tornar a série estacionária, posteriormente, com a utilização de um processo misto autoregressivo de médias móvel, será chamado de processo autoregressivo de médias móveis integrado de ordem  $d$  ou ARIMA( $p,d,q$ ). Em casos que a variável não é estacionária  $Y_t$  (não estacionária), define-se uma função  $Z(t)$ , sendo a primeira diferença de  $Y_t$ , ou seja,

$$Z(t) = Y_t - Y_{t-1} = \Delta Y_t,$$

nesse caso, considerando-se  $Z(t)$  estacionária, então a variável  $Y_t$  é integrada de ordem 1. Tomando  $d$  diferenças, pode-se reescrever a equação da seguinte forma:

$$Z(t) = \Delta^d Y_t.$$

Logo, a nova variável  $Z(t)$  segue um processo ARIMA( $p, d, q$ ), ou seja:

$$Z(t) = \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \dots + \beta_p Z_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}. \quad (2.19)$$

Finalmente  $Y_t$ , segue um processo ARIMA( $p, d, q$ ), ou seja,  $Y_t$  é uma variável integrada de ordem  $d$  e sua  $d$ -ésima diferença segue um processo autoregressivo de ordem  $p$  e de médias móveis de ordem  $q$ , podendo ser descrito como:

$$\Delta^d Y_t = \beta_1 \Delta^d Y_{t-1} + \beta_2 \Delta^d Y_{t-2} + \cdots + \beta_p \Delta^d Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q}. \quad (2.20)$$

Na prática valores pequenos são em geral especificados para  $d$ , sendo  $d = 1$  o valor mais frequentemente utilizado e excepcionalmente  $d = 2$ . Note também que o passeio aleatório pode ser considerado um processo ARIMA(0,1,0) (EHLERS, 2007).

### 2.2.7 Modelos Mistos Integrados Sazonais (SARIMA)

Conforme Ehlers (2009), os modelos de séries que apresentam autocorreção sazonal são geralmente conhecidos como modelos SARIMA, este modelo é composto por uma parte não sazonal com parâmetros  $(p, d, q)$ , e uma parte sazonal com parâmetros  $(P, D, Q)$ . Ao trabalhar com dados sazonais, pode-se observar que há uma periodicidade e essa periodicidade se repete a cada  $s$  observações. Logo, ao utilizar dados mensais ( $s = 12$ ), espera-se que  $Z_t$  dependa de  $Z_t - 12$  e sucessivamente a  $Z_t - 24$ .

O modelo SARIMA poderá ser representado por meio dos seguintes passos (EHLERS, 2009):

Ao utilizar a primeira diferença sazonal com  $s = 12$ , tem-se:

$$\nabla_{12} Y_t = (1 - L^{12}) Y_t = Y_t - Y_{t-12}.$$

Logo, pode-se observar que a diferença sazonal é denotada por  $\nabla_s$ , em que  $s$  é o período sazonal. De forma genérica  $\nabla_s^D$  é a  $D$ -ésima diferença sazonal, porém, em alguns modelos sazonais é necessário trabalhar com ambos operadores, dessa forma tem-se  $\nabla_s^D \nabla^d$  como combinação das duas diferenças.

Tomando uma diferença simples e uma diferença sazonal em uma série mensal encontra-se:

$$\nabla \nabla_{12} Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}.$$

Segundo Box e Jenkins (apud EHLERS, 2009, p. 44) o modelo ARIMA para série com sazonalidade pode ser definido como ARIMA sazonal multiplicativo. Observando-se a

notação desse trabalho tem-se as seguintes expressões:

$$B(L)b(L^s)W_t = \Theta(L)\phi(L^s)\varepsilon_t,$$

em que

$$\begin{aligned} B(L) &= (1 - \beta_1 L - \dots - \beta_p L^p) \\ b(L^s) &= (1 - B_s L^s - \dots - B_P L^{Ps}) \\ W_t &= \nabla^d \nabla_s^D Y_t \\ \Theta(L) &= (1 - \theta_1 L - \dots - \theta_q L^q) \\ \phi(L^s) &= (1 - \Theta_s L^s - \dots - \Theta_Q L^{Qs}). \end{aligned}$$

Logo, o modelo SARIMA multiplicativo será denotado por SARIMA  $(p, d, q) \times (P, D, Q)_s$  onde geralmente os valores de  $d$  e  $D$  não são superiores a 1.

## 2.2.8 Método de previsão

De acordo com Morettin e Toloi (2004), designando-se  $\hat{Z}_t(h)$ , a previsão de origem  $t$  e horizonte  $h$ ,  $\hat{Z}_t(h)$  poderá ser expressado de três maneiras:

*i.* Previsão utilizando-se a equação de diferenças

Tomando-se a esperança condicional em uma equação de diferenças, tem-se

$$\hat{Z}_t(h) = \varphi_1 [Z_{t+h-1}] + \dots + \varphi_{p+d} [Z_{t+h-p-d}] - \theta_1 [a_{t+h-1}] - \dots - \theta_q [a_{t+h-q}] + [a_{t+h}] \quad (2.21)$$

para  $h \geq 1$ . Aqui deve-se utilizar as seguintes propriedades

$$\begin{aligned} [Z_{t+k}] &= \hat{Z}_t(h), \quad k > 0 \\ [Z_{t+k}] &= Z_{t+k}, \quad k \leq 0 \\ [a_{t+k}] &= 0, \quad k > 0 \\ [a_{t+k}] &= a_{t+k}, \quad k \leq 0 \end{aligned}$$

*ii.* Previsão utilizando-se a forma de choques aleatórios

Utilizando-se a equação de choques aleatórios, tem-se

$$\hat{Z}_t(h) = \psi_1 [a_{t+h-1}] + \dots + \psi_{h-1} [a_{t+1}] + \psi_h [a_t] + \dots + [a_{t+h}]. \quad (2.22)$$

*iii.* Previsão utilizando-se a forma invertida

Obtida a partir da equação da forma invertida

$$\hat{Z}_t(h) = \sum_{j=1}^{\infty} \pi_j [Z_{t+h-j}] + [a_{t+h}]. \quad (2.23)$$

A equação de previsão, considerada como uma função de  $h$ , com origem  $t$  fixa, satisfaz a equação de diferenças

$$\sum_{i=1}^{p+d} \varphi_i \hat{Z}_t(h-1), h > q$$

ou

$$\varphi(B) \hat{Z}_t(h) = (1-B)^d \phi(B) \hat{Z}_t(h) = 0, h > q$$

com  $\varphi(B)$  operando sobre  $h$ .

## 2.3 Metodologia de Box e Jenkins

A estratégia para a construção do modelo será baseada em um ciclo iterativo, ao qual a escolha da estrutura do modelo é baseada nos próprios dados. Os estágios do ciclo iterativos proposto por Morettin e Toloí (2004) são:

- *i.* Uma classe geral de modelos é considerada para a análise e especificação;
- *ii.* Há identificação de um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- *iii.* A seguir vem a fase de estimação, na qual os parâmetros do modelo identificado são estimados;
- *iv.* Finalmente, há verificação ou diagnóstico do modelo ajustado, por meio de uma análise de resíduos, para se saber se este é adequado para os fins em vista (previsão, por exemplo).

Seguindo esses passos e o modelo não sendo adequado, o ciclo deverá ser repetido voltando a fase inicial, ou seja, volta-se à fase de identificação do modelo, porém, muitos modelos poderão ser estimados em uma mesma série, portanto, se a finalidade da

estimação é a previsão, busca-se o modelo que tiver o menor erro quadrático médio de previsão (MORETTIN; TOLOI, 2004).

### 2.3.1 Identificação do modelo

Considerada uma das etapas mais difíceis, a identificação do modelo poderá ser realizada por meio da análise de autocorrelação e autocorrelação parcial amostrais, ou através de um critério de seleção de modelos. Na primeira estratégia, faz-se necessário identificar os modelos AR, MA e ARMA observando-se as características da FAC e FACP como apresentado anteriormente, nesse contexto, utiliza-se a Tabela 1, a seguir, para identificá-los.

Tabela 1: Identificação dos modelos ARIMA.

| Tipo de processo | FAC           | FAC Parcial   |
|------------------|---------------|---------------|
| AR(p)            | declinante    | truncada em p |
| MA(q)            | truncada em q | declinante    |
| ARMA(p;q)        | declinante    | declinante    |

Um processo será AR( $p$ ) quando apresentar FAC declinante e FAC Parcial truncada em  $p$ , será MA( $q$ ) ao apresentar FAC truncada em  $q$  e FAC Parcial declinante, ou ainda será um ARMA( $p, q$ ) quando apresentar FAC e FAC Parcial declinantes (MORAIS, 2010). Além da verificação dos valores de  $p$  e  $q$ , será necessário identificar o grau das diferenças  $d$  para modelos do tipo ARIMA( $p, d, q$ ).

Segundo Melo (2001), existem outras formas de identificar o modelo, ou seja, a utilização do AIC (*Akaike information criterion*) e o BIC (*Bayesian information criterion*), ambos são critérios que incorporam um termo de penalidade para o aumento do número de parâmetros no modelo. Portanto, com a visualização dos gráficos da FAC e FACP, torna-se difícil identificar o modelo com aspectos de crescimento e declínio das funções, por isso, para resolver esse “problema”, faz-se o uso do AIC.

O critério de informação de Akaike será expresso por

$$AIC = -2 \log \hat{L} + 2(p + q), \quad (2.24)$$

em que,  $\hat{L}$  é o log de verossimilhança maximizado, e o melhor modelo será aquele que apresentar o menor valor de AIC.

A presença de  $p$  e  $q$ , na fórmula do critério AIC, tem por objetivo “penalizar” os modelos com muitos parâmetros, tendo em vista que modelos mais parcimoniosos de-

vem ser privilegiados, por apresentarem menor número de parâmetros a serem estimados (KIRCHNER, 2006).

A comparação de modelos também poderá ser realizada com o Critério de Informação Bayesiano (BIC) como aproximações do fator de Bayes. O BIC resolve este problema por meio da introdução de um termo de penalidade para o número de parâmetros do modelo.

$$BIC = -2 \log \hat{L} + (p + q) \log n, \quad (2.25)$$

em que  $n$  é o número de observações ou tamanho da amostra,  $\log \hat{L}$  é o valor maximizado da função de verossimilhança para o modelo estimado. Assim como o AIC, o melhor modelo será aquele que tiver o menor valor BIC.

Além do AIC e o BIC, existem outros testes que auxiliaram na aplicação da análise, por exemplo, tem-se a estatística  $Q$  calculada pelos testes de Box-Pierce ou Ljung-Box que avaliam a qualidade do modelo, verificando se o conjunto de autocorrelações residuais é estatisticamente diferente de zero. A estatística Box-Pierce é definida por

$$Q = n \sum_{k=1}^m \rho_k^2. \quad (2.26)$$

Na prática o número  $m$  de autocorrelações amostrais é tipicamente escolhido entre 15 e 30 (EHLERS, 2007). Se o modelo for apropriado,  $Q$  terá uma distribuição aproximadamente qui-quadrado com  $m - p - q$  graus de liberdade. Assim, seguindo a ideia de Ehlers (2007), nota-se que valores grandes de  $Q$  fornecem indicação contra a hipótese de que as autocorrelações são todas nulas, em favor da hipótese que pelo menos uma difere de zero, isto é,

$$\begin{aligned} H_0 &= \rho_1 = \rho_2 = \dots = \rho_k = 0 \\ H_1 &= \rho_k \neq 0. \end{aligned} \quad (2.27)$$

Contudo, o teste Box-Pierce não tem bom desempenho em amostras pequenas pois a distribuição afasta-se da qui-quadrado, então aplica-se o teste Ljung-Box, que é aproximadamente qui-quadrado com  $m - p - q$  graus de liberdade. A estatística Ljung-Box é definida por

$$Q = n(n + 2) \sum_{k=1}^m \frac{\rho_k^2}{n - k}. \quad (2.28)$$

Em síntese, pode-se explicar que a estatística  $Q$ , calculada por meio do teste Box-Pierce ou pelo teste de Ljung-Box é utilizada para verificar conjunto de autocorrelações de resíduos sendo estatisticamente diferente ou não de zero. Contudo, ambos os testes

possuem distribuição  $\chi^2$  (qui-quadrado), teste esse, sendo realizado da seguinte forma: se  $Q_{calculado} > Q_{tabelado}$ , então rejeita-se a hipótese nula ( $H_0$ ) de que os resíduos não são correlacionados, ou seja, pelo menos uma autocorrelação difere estatisticamente de zero, com isso, o modelo deverá ser rejeitado.

É importante ressaltar que, ao estimar-se um modelo, deseja-se que o erro produzido por ele possua característica de ruído branco<sup>2</sup>, isto é, este venha ser independente e identicamente distribuído (LEROY, 2009).

### 2.3.2 Estimação

Segundo Morettin e Tolo (2004), para estimação do modelo utiliza-se estimadores de máxima verossimilhança, porém, há casos que será necessário utilizar procedimentos iterativos de estimação não-linear de mínimos quadrados e as estimativas encontradas serão usadas como valores iniciais neste procedimento. Atualmente os programas computacionais incorporam esses valores iniciais.

### 2.3.3 Verificação e diagnóstico

Na busca de um modelo parcimonioso para realizar possíveis previsões, faz-se necessário verificar os pressupostos para validação do modelo. Nesse sentido, pode-se por meio da visualização dos gráficos FAC e FACP dos resíduos, observa se ambos estão dentro do intervalo de confiança, em caso afirmativo, há presença do ruído branco, ou seja, temos uma série de variáveis aleatórias independentes identicamente distribuídas (*iid*) com média zero e variância constante (PEDROSO *et al.*, 2007).

O resíduo nada mais é que a diferença entre a observação e os valores ajustados. Assim, como em outros modelos espera-se que os resíduos se distribuam aleatoriamente em torno de zero com variância constante e sejam não correlacionados, porém, há casos que a variância dos resíduos apresenta-se de forma crescente, então, é aconselhável utilizar transformação logarítmica nos dados para tornar a série estacionária e estabilizar a variância.

O fenômeno de não “constância” na variância é denominado de volatilidade na literatura de séries temporais e pode ser tratado por meio de transformações nos dados (EHLERS, 2009). Seguindo a ideia de Ehlers (2009), ao invés de visualizar as autocorrelações

---

<sup>2</sup>Ruído branco corresponde a um processo estacionário, com média zero e que não apresenta autocorrelação.

residuais individualmente pode-se testar um grupo de autocorrelações para averiguar se os resíduos são significativamente diferentes de zero por meio das estatísticas  $Q$ , visto em (2.26) e (2.28).

### 3 Aplicação

Uma extensão dos modelos  $ARIMA(p,d,q)$  para séries com sazonalidade e tendência é o modelo  $SARIMA(p, d, q) \times (P, D, Q)$ . Contudo, nessa aplicação será utilizado uma série de dados já conhecida e bastante trabalhada por Ehlers (2009), os dados em questão, são totais mensais de passageiros em linhas aéreas internacionais nos EUA entre 1949 e 1960, disponível no pacote *datasets* do *Software R* versão 2.14.2.

Motivado em prever totais mensais de passageiros em linhas aéreas internacionais, procura-se modelar uma série temporal que possa prever a curto prazo o quantitativo de passageiros num certo período de tempo. Com auxílio do *Software R* versão 2.14.2, será utilizado alguns pacotes e funções para estimar e prever os resultados.

No primeiro momento, será utilizado apenas os dados da série “AirPassengers” no período de janeiro de 1949 a dezembro de 1959, esse conjunto de dados será exemplificado nesse trabalho e posteriormente com as previsões de 12 meses a frente, compará-las com os dados reais da série no período de janeiro de 1960 a dezembro de 1960. Esse método de comparação é comum para verificar a performance do modelo. A visualização gráfica faz-se necessária para identificação do modelo, porém, também é necessário que o pesquisador tenha bastante cautela na identificação por meio de gráficos da função de autocorrelação dada por (2.3) e da função de autocorrelação parcial identificada em (2.13). Outra maneira de identificar o modelo será com base nos critérios AIC (*Akaike information criterion*) e o BIC (*Bayesian information criterion*), ambos são critérios que incorporam um termo de penalidade para o aumento do número de parâmetros no modelo, visto em (2.24) e (2.25). Nesse contexto, com a finalidade de reduzir possíveis distorções na identificação do modelo e maximização do erro, será utilizado com auxílio do pacote *forecast* a função *auto.arima* que retorna o melhor modelo ARIMA de acordo com qualquer valor AICC, AIC ou BIC. A função realiza uma pesquisa sobre os possíveis modelos, dentro das limitações das ordens fornecidas, além de utilizar o algoritmo Stepwise delineado no Hyndman e Khandakar (2008). Nesse caso, será utilizado o critério BIC visto na Seção (2.3.1) para identificação do modelo. Por meio dos pacotes *Hmisc*, *graphics* e outros, serão gerados os gráficos e

as tabelas exportadas em formato Latex.

Observa-se por meio da Figura 4 a série dos dados “AirPassengers” com certa tendência ao longo do tempo, na sequência segue os gráficos de autocorreção e autocorrelação parcial do mesmo. Considerando-se a não estacionariedade da série na Figura 4, tenta-se por meio de várias transformações tornar a série estacionária, pode-se observar por meio da Figura 5 a série com uma diferença simples e seus respectivos gráficos de autocorreção e autocorrelação parcial, é notável a não estacionariedade na série “AirPassengers”.

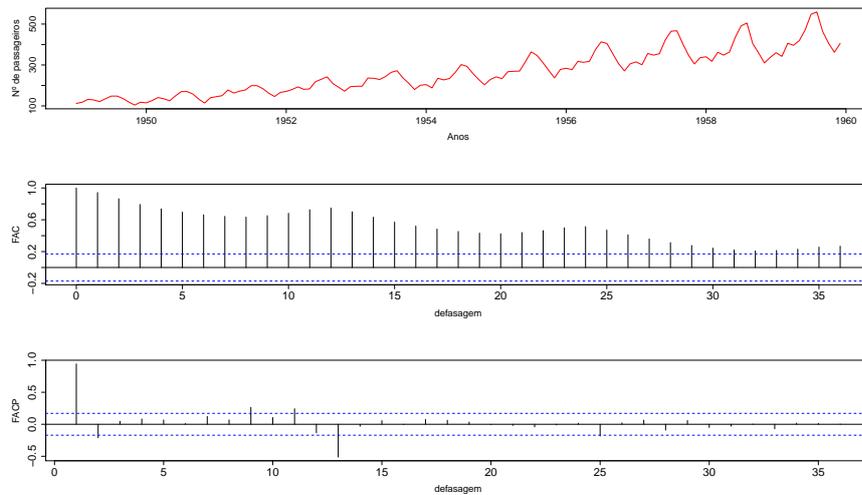


Figura 4: Gráficos da série AirPassengers

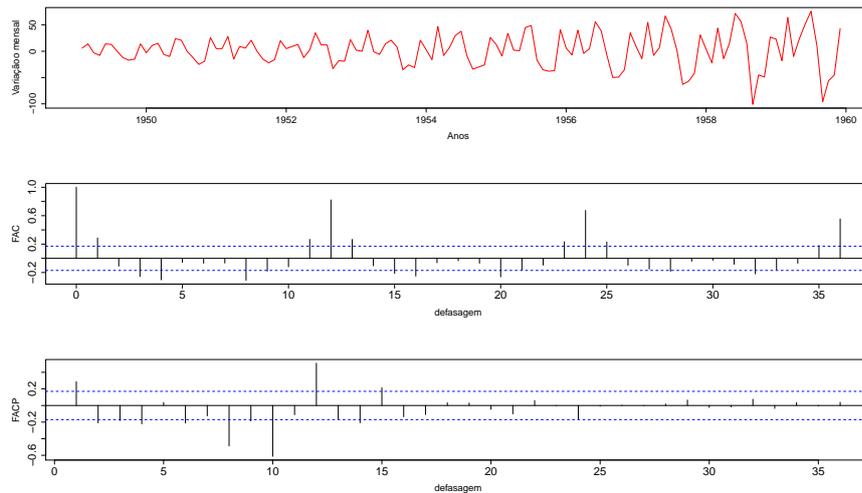


Figura 5: Gráficos da série AirPassengers com uma diferença mensal.

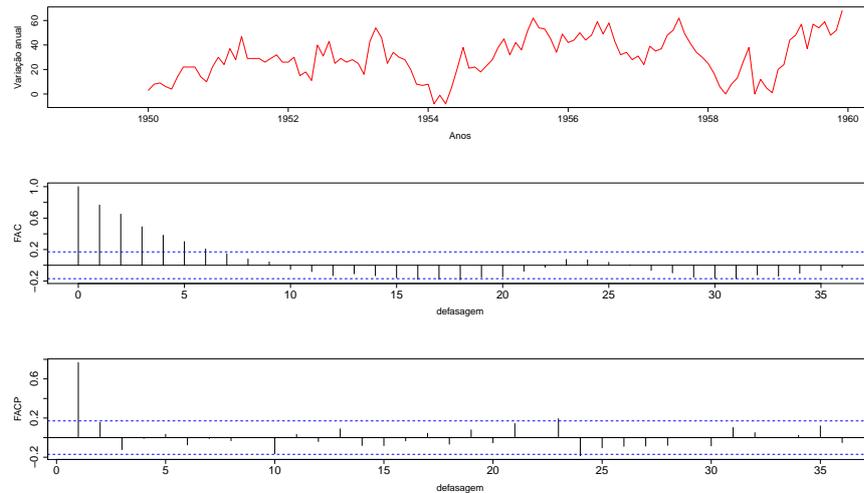


Figura 6: Gráficos da série AirPassengers com diferença sazonal

Diante deste problema, tenta-se aplicar uma diferença sazonal nas observações da série e por meio dessa transformação pode-se visualizar, conforme Figura 6, uma série com altos e baixos, ou seja, com certa variação sazonal, porém nota-se uma certa estabilidade nos gráficos de autocorrelação FAC e FACP. Nesse contexto, ao aplicar uma diferença simples e uma diferença sazonal, conforme Figura 7, será possível obter o número de parâmetros significativos ao modelo e satisfatório para previsão, uma vez, que a série seja estacionária.

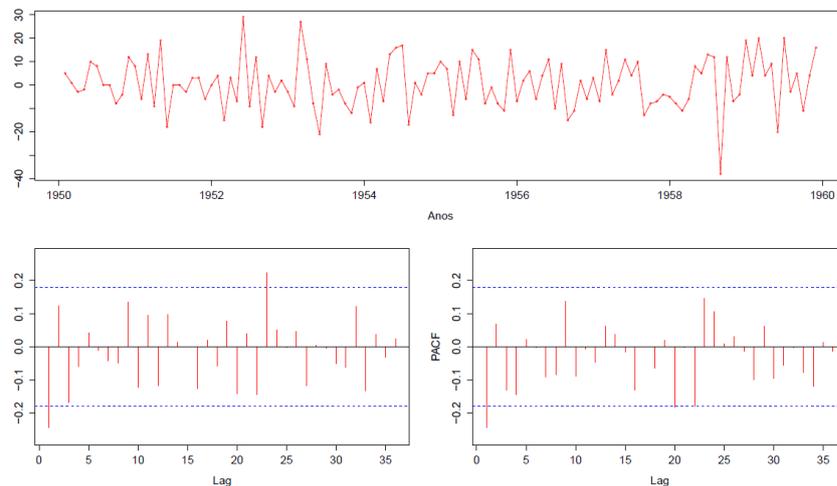


Figura 7: Gráfico da série AirPassengers com uma diferença simples e uma diferença sazonal.

Dando sequência as análises, pode-se observar por meio da Figura 7 o gráfico da série “AirPassengers” com uma diferença simples e uma diferença sazonal, em alguns casos a transformação logarítmica é usual, porém, nesse trabalho não será utilizada tal transformação.

Ao utilizar a função *auto.arima* do pacote *forecast* do *software* R versão 2.14.2, seleciona-se de forma automática o melhor modelo de acordo com o critério estabelecido, nesse trabalho, será utilizado o critério BIC como critério de seleção Bayesiano (2.25). Com auxílio dessa função será selecionado o modelo de forma rápida sem que para isso necessite de uma análise mais específica nos gráficos da FAC e da FACP de cada modelo candidato, ou seja, de forma prática e ágil seleciona-se o modelo com menor valor do BIC, critério esse visto anteriormente. Logo, o modelo selecionado fora o modelo SARIMA(1, 1, 0)  $\times$  (0, 1, 0), ou seja, um modelo com um parâmetro autoregressivo com uma diferença simples e uma diferença sazonal. De acordo a Tabela 2 pode-se observar o valor do parâmetro com o respectivo desvio-padrão e o valor do critério de seleção BIC.

Tabela 2: Tabela do modelo ajustado para séries AirPassengers.

| AR(1)   | s.e    | BIC    |
|---------|--------|--------|
| -0,2431 | 0,0894 | 905,46 |

Por meio da Tabela 3, apresenta-se os valores reais e os valores previstos com o modelo SARIMA(1, 1, 0)  $\times$  (0, 1, 0).

Tabela 3: Valores observados e previsto para a séries AirPassengers.

| mês/ano  | valor real | predição | erro     |
|----------|------------|----------|----------|
| jan 1960 | 417        | 424,1099 | -7,1099  |
| fev 1960 | 391        | 407,0557 | -16,0557 |
| mar 1960 | 419        | 470,8258 | -51,8258 |
| abr 1960 | 461        | 460,8817 | 0,1183   |
| mai 1960 | 472        | 484,8681 | -12,8681 |
| jun 1960 | 535        | 536,8714 | -1,8714  |
| jul 1960 | 622        | 612,8706 | 9,1294   |
| ago 1960 | 606        | 623,8708 | -17,8708 |
| set 1960 | 508        | 527,8707 | -19,8707 |
| out 1960 | 461        | 471,8707 | -10,8707 |
| nov 1960 | 390        | 426,8707 | -36,8707 |
| dez 1960 | 432        | 469,8707 | -37,8707 |

A Figura 8 apresenta os gráficos de diagnóstico para o modelo SARIMA(1, 1, 0)  $\times$  (0, 1, 0) onde pode-se observar o gráfico dos resíduos estudentizados e o gráfico de autocorrelação - FAC, com uma visão mais crítica, poder-se inferir algumas suposições quanto a correlação residual do modelo, portanto faz-se necessário analisar as hipóteses em (2.27) do teste de Ljung-Box, visto em (2.28) e exposto no último gráfico da Figura 8.

Por meio do gráfico Q-Q Plot apresentado na Figura 8 é possível visualizar que os resíduos parecem seguir uma distribuição normal.

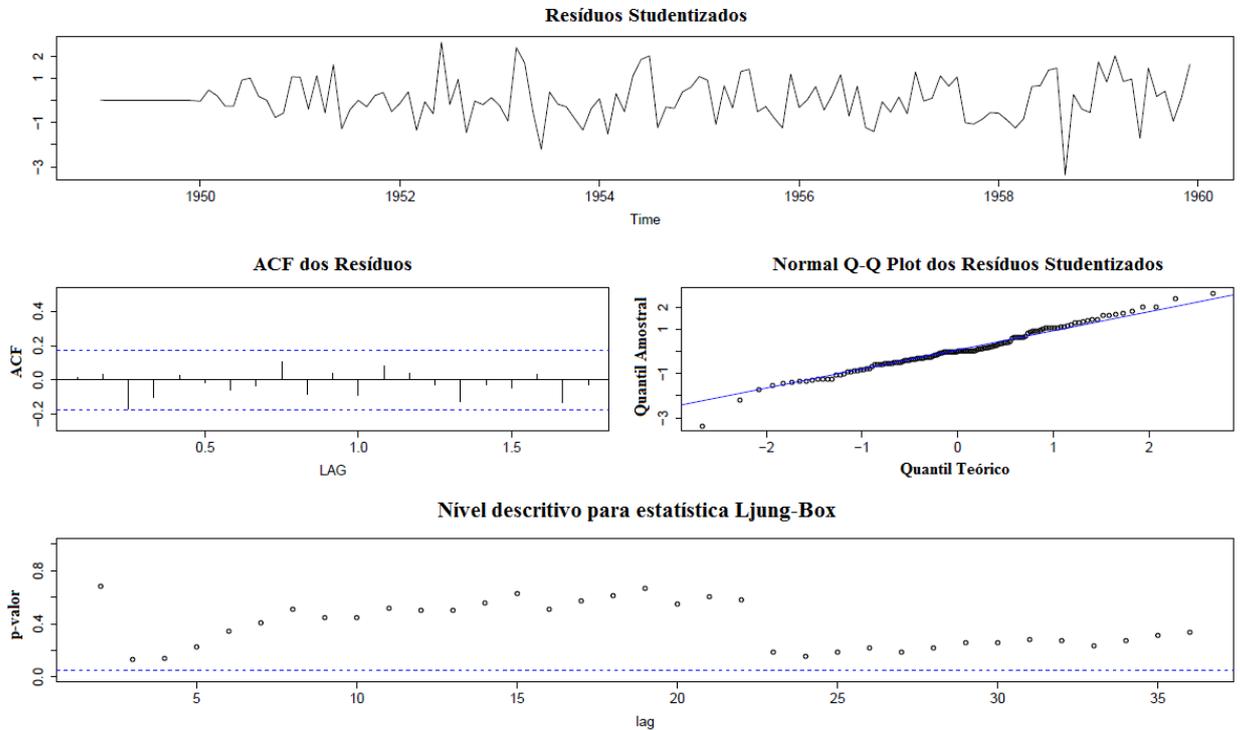


Figura 8: Gráfico para validação do modelo.

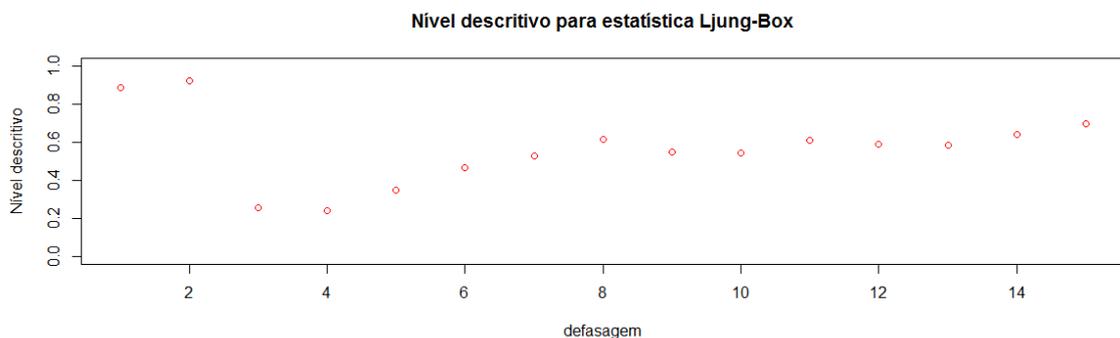


Figura 9: Gráfico do nível descritivo para o teste Ljung-Box.

Para a validação do modelo, será imprescindível a utilização do teste Ljung-Box, por isso, na Figura 9 tem-se uma visualização ampliada dos níveis descritivos. Segundo Moret-tin e Toloí (2004), basta utilizar as 10 ou 15 primeiras autocorrelações para a verificação da estatística  $Q$ . Então, espera-se que os níveis descritivos sejam superiores a significância de 0.05 para que não haja evidências de rejeição da hipótese nula dada em (2.3.1). Logo, não há indícios pra rejeitar a hipótese nula, que os resíduos não são correlacionados, portanto os mesmos parecem estar distribuídos aleatoriamente.

Na Figura 8 pode-se observar os pressupostos para validação do modelo, porém, na verificação da normalidade é utilizado o teste Shapiro Wilks e através do gráfico da Figura 10 e Figura 11 tem-se o p-valor observado sendo superior ao nível de significância  $\alpha = 0.05$ , com isso, há indícios de aceitar a hipótese nula que os resíduos seguem uma distribuição normal.

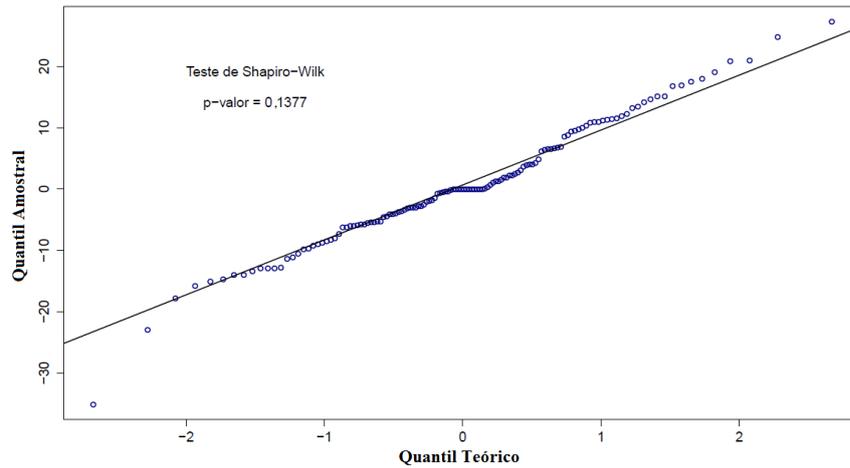


Figura 10: Gráfico da normalidade dos resíduos com p-valor para o teste Shapiro Wilks.

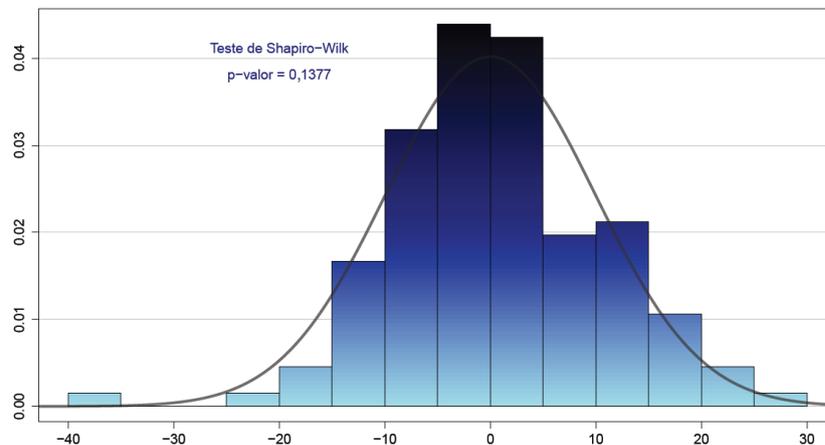


Figura 11: Histograma dos resíduos com teste Shapiro Wilks.

Após a validação dos pressupostos do modelo segue a etapa de previsão, nessa etapa, como descrito no início da aplicação, será previsto 12 meses a frente. Na Tabela 4 pode-se observar cada período predito com o respectivos intervalos de confiança nas colunas, para Li80 e Ls80 temos 80% de confiança, para Li95 e Ls95 tem-se 95% de confiança.

Tabela 4: Previsões e Limites inferior (Li) e Limite Superior (Ls) dos intervalos de previsões para os totais mensais de passageiros em linhas aéreas internacionais no EUA no período de janeiro de 1960 a dezembro de 1960.

| Mês/ano  | predição | Li 80    | Ls 80    | Li 95    | Ls 95    |
|----------|----------|----------|----------|----------|----------|
| jan 1960 | 424,1099 | 410,7378 | 437,4820 | 403,6590 | 444,5608 |
| fev 1960 | 407,0557 | 390,2853 | 423,8261 | 381,4076 | 432,7039 |
| mar 1960 | 470,8258 | 450,8181 | 490,8334 | 440,2267 | 501,4248 |
| abr 1960 | 460,8817 | 438,1835 | 483,5798 | 426,1678 | 495,5955 |
| mai 1960 | 484,8681 | 459,7461 | 509,9900 | 446,4473 | 523,2888 |
| jun 1960 | 536,8714 | 509,5442 | 564,1985 | 495,0781 | 578,6647 |
| jul 1960 | 612,8706 | 583,5023 | 642,2388 | 567,9557 | 657,7854 |
| ago 1960 | 623,8708 | 592,5946 | 655,1470 | 576,0380 | 671,7036 |
| set 1960 | 527,8707 | 494,7964 | 560,9450 | 477,2879 | 578,4535 |
| out 1960 | 471,8707 | 437,0912 | 506,6503 | 418,6800 | 525,0615 |
| nov 1960 | 426,8707 | 390,4657 | 463,2758 | 371,1940 | 482,5474 |
| dez 1960 | 469,8707 | 431,9097 | 507,8317 | 411,8144 | 527,9270 |

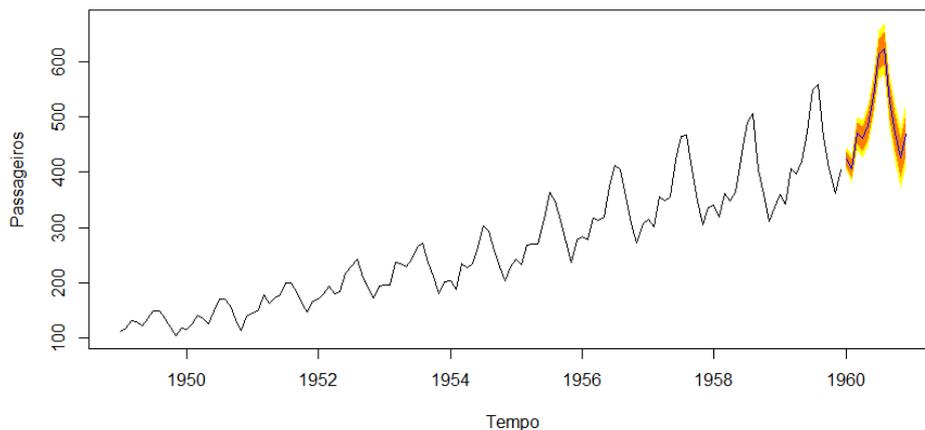


Figura 12: Gráfico dos dados preditos no período de janeiro a dezembro de 1960.

Por meio da Figura 12 tem-se o gráfico dos dados preditos no período de janeiro a dezembro de 1960, demarcado com o respectivo intervalo de confiança e acrescido na série original.

## 4 Conclusão

Os resultados encontrados na aplicação da série “AirPassengers”, indicaram bom ajuste para a série temporal, contendo um parâmetro autoregressivo, uma diferença simples e uma diferença sazonal, ou seja, o modelo  $SARIMA(1, 1, 0) \times (0, 1, 0)$ .

Além da visualização dos gráficos de autocorrelação e autocorrelação parcial, para encontrar um modelo parcimonioso, utilizou-se processos iterativos computacionais a exemplo da função *auto.arima* do pacote *forecast*, que agilizaram o processo de geração da série até a verificação do modelo, identificando assim, períodos sazonais e melhores modelos através do critério de seleção BIC. Contudo, as previsões encontradas para totais mensais de passageiros em linhas aéreas internacionais nos EUA é um conjunto de observação bastante conhecido na área de séries temporais, porém, trabalhos de previsão dessa magnitude são de grande importância para o desenvolvimento científico e para possíveis políticas públicas, em destaque ao tráfego aéreo de passageiros em determinado país ou região, de forma que, em casos especiais, a elaboração de um planejamento adequado contribuirá para prever antecipadamente o volume de tráfego aéreo em um determinado local e através das políticas públicas tornar possível identificar as prioridades e minimizar transtornos ou catástrofes aéreas por falta de um pré-planejamento.

De forma objetiva, esse trabalho teve resultados satisfatórios a contribuir positivamente, na tentativa de expor análises de séries temporais como ferramenta aos órgãos públicos no desenvolver de suas estratégias e no planejamento de ações futuras na análise de dados dos respectivos órgãos.

# Referências

- BOX, G. E. P.; JENKINS, G. M. Time series analysis, forecasting and control. 1970.
- EHLERS, R. S. *Análise de Séries Temporais*. 2007. Departamento de Estatística, UFPR. Disponível em: <<http://www.each.usp.br/rvicente/AnaliseDeSeriesTemporais.pdf>>.
- EHLERS, R. S. *Análise de Séries Temporais*. 2009. Disponível em: <<http://www.icmc.usp.br/~ehlers/stemp/stemp.pdf>>.
- FISCHER, S. *Séries univariantes de tempo - metodologia de Box & Jenkins*. Porto Alegre: Fundação de Economia Estatística, 1982.
- GREENE, W. H. Econometric analysis. *New Jersey: Prentice Hall*, v. 5 edição, 2003.
- KIRCHNER, R. M. *MODELAGEM E PREVISÃO POR MEIO DE METODOLOGIA BOX E JENKINS: UMA FERRAMENTA DE GESTÃO*. Dissertação (Mestrado) — Universidade Federal de Santa Maria, 2006.
- LEROY, F. L. D. Projeção do preço futuro de uma ação da usiminas: uma abordagem econométrica. *REVISTA DA FACULDADE DE CIÊNCIAS ADMINISTRATIVAS DE CURVELO - MG*, Vol. 5, p. 37–70, 2009. Disponível em: <<http://www.fac.br/multiplosarquivos/Vol5Parte03.pdf>>.
- MELO, B. S. V. de.  
*MODELO DE PREVISÃO PARA ARRECADAÇÃO TRIBUTÁRIA*  
 — Escola de Administração Fazendária, Brasília-DF, 2001. Disponível em: <<http://www3.tesouro.fazenda.gov.br/PremioTN/VIPremio/sistemas/1siafpVIPTN/MELOBruno.pdf>>.
- MORAIS, E. F. *Análise de Regressão versus Séries Temporais: Uma previsão para o Índice Ibovespa*. Monografia (Graduação) — Universidade Estadual da Paraíba, Campina Grande, 2010.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. São Paulo: Edgard Blücher, 2004.
- PEDROSO, T. R.; SOUZA, A. M.; JACOBI, L. F. *Previsão e monitoramento do preço das ações do Bradesco*. 2007. Universidade Federal de Santa Maria - Santa Maria, RS. Disponível em: <<http://www.sumarios.org/sites/default/files/pdfs/171-532-1-pb.pdf>>.