



UNIVERSIDADE ESTADUAL DA PARAÍBA  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

Nayara Fernanda da Silva Lima

# **Modelagem e previsão da taxa de desemprego na Região Metropolitana de São Paulo**

Campina Grande - PB

Dezembro de 2017

Nayara Fernanda da Silva Lima

## **Modelagem e previsão da taxa de desemprego na Região Metropolitana de São Paulo**

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: Prof. Dr. Sílvio Fernando Alves Xavier Júnior

Campina Grande - PB

Dezembro de 2017

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

L732m Lima, Nayara Fernanda da Silva.  
Modelagem e previsão da taxa de desemprego na Região Metropolitana de São Paulo [manuscrito] : / Nayara Fernanda da Silva Lima. - 2017.  
52 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2017.

"Orientação : Prof. Dr. Sílvio Fernando Alves Xavier Júnior, Coordenação do Curso de Estatística - CCT."

1. Séries temporais. 2. Desemprego. 3. Processos estocásticos. 4. Modelo SARIMA.

21. ed. CDD 519.5

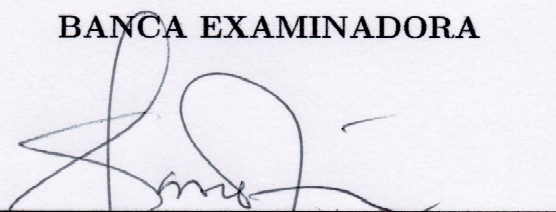
Nayara Fernanda da Silva Lima

## Modelagem e previsão da taxa de desemprego na Região Metropolitana de São Paulo

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.

Trabalho aprovado em 07 de Dezembro de 2017.

### BANCA EXAMINADORA



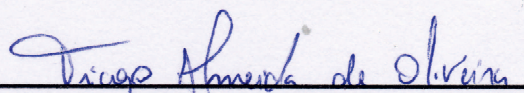
---

Prof. Dr. Sílvio Fernando Alves Xavier  
Júnior  
Universidade Estadual da Paraíba

*DMEstes*

---

Prof. Dra. Divanilda Maia Esteves  
Universidade Estadual da Paraíba



---

Prof. Dr. Tiago Almeida de Oliveira  
Universidade Estadual da Paraíba

*Dedico este trabalho aos meus pais, **Edilson** e **Maria**, que foram os maiores exemplos de dedicação e superação. Sem vocês eu não seria metade do que sou hoje, nem teria chegado até onde cheguei. Eu amo vocês!*

# Agradecimentos

Agradeço primeiramente a Deus, por ter iluminado minha jornada e ter renovado as minhas forças todos os dias através da minha fé. Aos meus pais, vovó e meus irmãos, por todo o apoio, por terem me aguentado e me motivado a nunca desistir, sempre mostrando que através do meu esforço e estudo eu chegaria longe.

Agradeço aos meus amigos, que estiveram me incentivando a buscar sempre o melhor de mim. Obrigada por nunca me deixarem desistir, por todo o apoio (seja com palavras de conforto, ou na ajuda para digitar as tabelas, rs) e reconhecimento, o meu MUITO OBRIGADO!

Ao meu orientador, o professor Sílvio, por todo o apoio e paciência na elaboração deste trabalho. E a todos os professores que fizeram parte da minha jornada acadêmica, que foram exemplos ímpares de que um professor vai muito além do que simplesmente ministrar aulas.

Não poderia esquecer dos "nerds de ferro", obrigada por fazerem das manhãs na universidade momentos descontraídos, por todo o apoio e cumplicidade nas disciplinas.

Por fim, agradeço a todos que contribuíram direta ou indiretamente para a minha jornada acadêmica.

*“A conquista de si mesmo está ao alcance  
do querer para ser, do esforçar-se para triunfar,  
do viver para jamais morrer..”  
(Joanna de Ângelis)*

# Resumo

O desemprego tem se tornado um dos grandes problemas deste século, principalmente, devido a crise econômica que assola não só o Brasil, mas o mundo. A região metropolitana de São Paulo compõe-se de 39 municípios, incluindo a principal metrópole do Brasil (a cidade São Paulo), esta região é considerada como o maior polo de riqueza nacional. Este trabalho tem por objetivo fazer previsões baseando-se em metodologias utilizadas na análise de séries temporais, tendo como finalidade prever a taxa de desemprego em São Paulo até o primeiro trimestre de 2018, fundamentando-se através dos dados no período de janeiro de 2002 à setembro de 2017, provenientes da Pesquisa Mensal de Emprego. Os métodos utilizados para estas previsões serão os algoritmos de Holt-Winters (aditivo e multiplicativo) e os modelos SARIMA e SARIMAX, focando nas suas respectivas previsões. Para o caso do modelo SARIMAX, aplicou-se uma variável dummy, com valores 1 para o meses a partir de Janeiro/15, adotando como referência o começo do segundo mandato da ex presidente Dilma Rousseff, e 0 para os meses anteriores. Através destas previsões, observou-se que as medidas adotadas pelo governo, em 2015, no intuito de fugir da crise, acabou tendo efeito contrário, contribuindo para o crescimento da taxa de desemprego, pois, os resultados adquiridos propõem que o modelo SARIMAX denota uma melhor capacidade de previsão entre os outros modelos sugeridos.

**Palavras-chaves:** Séries temporais. Desemprego. Holt-Winters. Modelos SARIMA.



# Abstract

Unemployment has become one of the great problems of this century, due to the economic crisis that plagues not only Brazil, but the world. The metropolitan region of São Paulo consists of 39 municipalities, including the main metropolis of Brazil (the city of São Paulo), which is considered to be the largest national wealth center. The purpose of this study is to make predictions based on methodologies used in the analysis of time series, with the purpose of predicting the unemployment rate in São Paulo until the first quarter of 2018, based on data from January 2002 to September 2017, from the Monthly Employment Survey. The methods used for these predictions will be the Holt-Winters (additive and multiplicative) algorithms and the SARIMAX and SARIMAX models, focusing on their respective forecasts. For the case of the SARIMAX model, dummy variable was applied, with values 1 for the months from January/15, adopting as reference the beginning of the second term of former president Dilma Rousseff, and 0 for the previous months. Through these forecasts, it was observed that the measures adopted by the government in 2015 in order to escape the crisis, ended up having the opposite effect, contributing to the growth of the unemployment rate, since, the acquired results propose that the SARIMAX a better predictability among the other suggested models.

**Key-words:** Time series. Unemployment. Holt-Winters. SARIMA models.

# Lista de ilustrações

Figura 1 – Taxa de Desemprego em São Paulo (de janeiro de 2012 à setembro de 2017). . . . .	32
Figura 2 – Funções de autocorrelação amostral e autocorrelação parcial da série. .	33
Figura 3 – Funções de autocorrelação amostral da primeira diferença e autocorrelação parcial da primeira diferença da série. . . . .	33
Figura 4 – Gráfico da série (linha preta) e seu ajuste de Holt-Winters (linha vermelha) aditivo e multiplicativo (1 passo à frente). . . . .	35
Figura 5 – Gráfico de previsão de Holt-Winters aditivo e multiplicativo da série Taxa de Desemprego em SP (7 passos à frente). . . . .	36

# Lista de tabelas

Tabela 1 – Testes da raiz unitária de Dickey-Fuller . . . . .	27
Tabela 2 – Valores otimizados das constantes de suavização para Holt-Winters aditivo e multiplicativo. . . . .	34
Tabela 3 – Previsões intervalares do algoritmo Holt-Winters aditivo. . . . .	35
Tabela 4 – Valores dos erros de previsão de acordo com a série. . . . .	36
Tabela 5 – Valores referentes ao critério de informação dos modelos da série e erros relativos, 1 passo à frente. . . . .	38
Tabela 6 – Previsões dos modelos SARIMA e SARIMAX estudados, 7 passos à frente, de Setembro/17 à Dezembro/17. . . . .	38
Tabela 7 – Previsões dos modelos SARIMA e SARIMAX estudados, 7 passos à frente. . . . .	39
Tabela 8 – Previsões intervalares do algoritmo Holt-Winters aditivo. . . . .	44
Tabela 9 – Previsões intervalares do algoritmo Holt-Winters multiplicativo. . . . .	45
Tabela 10 – Taxa de desemprego em São Paulo, entre 2002 e 2017, em porcentagem	52

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>14</b>
<b>2.1</b>	<b>Conceitos Básicos</b>	<b>14</b>
<b>2.2</b>	<b>Processos Estocásticos</b>	<b>15</b>
2.2.1	Processos Estacionários	15
2.2.2	Ruído Branco	16
<b>2.3</b>	<b>Funções de Autocovariância e Autocorrelação</b>	<b>17</b>
<b>2.4</b>	<b>Modelos de Suavização Exponencial</b>	<b>17</b>
2.4.1	Alisamento Exponencial Simples (AES)	18
2.4.2	Alisamento Exponencial de Holt	19
2.4.3	Alisamento Exponencial de Holt-Winters	19
<b>2.5</b>	<b>Modelagem Box-Jenkins</b>	<b>21</b>
2.5.1	Modelo Autorregressivo (AR)	21
2.5.2	Modelo de Médias Móveis (MA)	22
2.5.3	Modelo autorregressivo de médias móveis (ARMA)	23
2.5.4	Modelo autorregressivo integrado de médias móveis (ARIMA)	23
2.5.5	Modelo autorregressivo de médias móveis sazonal (SARIMA)	24
2.5.6	Modelo autoregressivo de médias móveis sazonal com variáveis exógenas (SARIMAX)	24
2.5.7	Função de autocorrelação (FAC)	25
2.5.8	Função de autocorrelação parcial (FACP)	25
<b>2.6</b>	<b>Etapas do modelo SARIMA</b>	<b>26</b>
2.6.1	Identificação	26
2.6.2	Teste de Estacionariedade de Dickey-Fuller	26
2.6.2.1	Análise Gráfica	28
2.6.3	Estimação	28
2.6.4	Validação	29
<b>2.7</b>	<b>Avaliação de Previsão</b>	<b>30</b>
<b>3</b>	<b>APLICAÇÃO</b>	<b>32</b>
<b>3.1</b>	<b>Análise exploratória dos dados</b>	<b>32</b>
<b>3.2</b>	<b>Modelagem</b>	<b>34</b>
3.2.1	Alisamento Exponencial de Holt-Winters	34
3.2.2	Modelagem Box-Jenkins - SARIMA	36

<b>4</b>	<b>CONCLUSÃO</b> . . . . .	<b>40</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>41</b>
	<b>APÊNDICES</b>	<b>43</b>
	<b>APÊNDICE A – PREVISÕES INTERVALARES DE HOLT-WINTERS</b>	<b>44</b>
	<b>APÊNDICE B – ROTINA DO R</b> . . . . .	<b>46</b>
	<b>ANEXOS</b>	<b>51</b>
	<b>ANEXO A – CONJUNTO DE DADOS</b> . . . . .	<b>52</b>

# 1 Introdução

A questão do desemprego ocupa um papel central na discussão das políticas tanto macroeconômica quanto de compensações sociais (PICCHETTI; CHAHAD, 2003). A crise econômica enfrentada pelo Brasil, reflete diretamente para o aumento do desemprego, onde, de acordo com o relatório da Organização Internacional do Trabalho (OIT), o Brasil apresenta uma quantidade de desempregados inferior apenas à da China e Índia, apresentando o maior crescimento da taxa de desemprego entre as 20 maiores economias do mundo. De forma geral, o desempregado caracteriza-se como a não possibilidade de trabalho assalariado. O que, de acordo com Garraty (1978), desemprego significa

"A condição da pessoa sem algum meio aceitável de ganhar a vida e os desempregados são pessoas capazes de trabalhar para satisfazer suas necessidades, mas ociosas, independentemente de sua boa vontade para trabalhar ou do que elas possam fazer para atender as necessidades da sociedade".

Diversas causas estão relacionadas para o surgimento deste problema, cita-se como principais: o desenvolvimento da tecnologia, a terceirização da mão-de-obra, o excesso de concentração da renda por empresários, baixa qualificação do trabalhador, a globalização, entre outros diversos fatores. Entretanto, o aumento do desemprego estrutural no Brasil dá-se, principalmente, devido à baixa qualificação na mão-de-obra resultante da educação defasada, baixo crescimento econômico e a legislação inflexível (KATO; PONCHIROLLI, 2017).

A Região Metropolitana de São Paulo é composta por 39 municípios e é o maior polo de riqueza nacional, onde abriga a principal metrópole nacional: São Paulo. Além disso, é considerada o centro das decisões políticas do Estado, concentrando serviços especializados e diversificados, sendo considerado também, como o polo de turismo de negócios da América Latina, e, ainda, abriga sedes de empresas transnacionais. Por isso, torna-se um mercado atrativo, visto como um meio de conseguir melhoria de vida, considerada como o destino principal para a busca de condições mais abrangentes de emprego. Dito isto, surge a necessidade de um estudo temporal acerca da taxa de desemprego nessa região (??).

Os dados utilizados neste trabalho são provenientes da **PME** (Pesquisa Mensal de Emprego), referente à região metropolitana de São Paulo. Segundo o **DIEESE** (Departamento de Estatística e Estudos Econômicos), a taxa de desemprego refere-se a proporção da população economicamente ativa (PEA) e a população desempregada. Esta taxa é calculada da seguinte maneira:

$$\text{Taxa de Desemprego} = \frac{N^{\circ} \text{ de desempregados}}{PEA} \times 100$$

O presente trabalho tem por objetivo utilizar a metodologia de Holt-Winters e Box-Jenkins para analisar a taxa de desemprego, observando, primeiramente, o comportamento temporal da taxa de desemprego na região metropolitana de São Paulo, bem como, as tendências de crescimento e decrescimento. Subsequente, através do Alisamento Exponencial de Holt-Winters, aditivo e multiplicativo, busca-se previsões 7 passos à frente, ou seja, prever como essa taxa de desemprego se comportará até o primeiro trimestre de 2018 (previsões de Outubro/2017 à Março/2018). E, analogamente, utiliza-se da modelagem de Box-Jenkins, com os modelos SARIMA e SARIMAX, onde busca-se um melhor modelo através dos dados e utiliza um modelo base (BECKER, 2010), para obter uma melhor previsão para a taxa de desemprego.

O trabalho está dividido da seguinte maneira: Capítulo 1 - Introdução, Capítulo 2 - Fundamentação teórica, Capítulo 3 - Aplicação, Capítulo 4 - Conclusão, seguido das referências bibliográficas, anexos e apêndices.

## 2 Fundamentação Teórica

### 2.1 Conceitos Básicos

Considera-se Séries Temporais como um conjunto de observações geradas sequencialmente em um determinado tempo  $t$ , de modo que para o tempo  $t$  esteja associada uma variável  $Y_t$ . Estas observações podem ser consideradas como *discretas*, quando estas acontecem em um tempo específico, como  $(T = t_1, \dots, t_n)$ , ou como *contínuas*, quando as observações acontecem continuamente, sendo  $(T = \{t: t_1 < t < t_2\})$ , (MORETTIN; TOLOI, 2006). Como exemplos de séries temporais, temos:

- valores diários da temperatura do Rio de Janeiro;
- quantidade semanal de chuva no Maranhão;
- valores diários de vendas de automóveis em Pernambuco.

Por outro lado, as séries temporais podem ser *determinísticas*, quando a série temporal é exatamente previsível; *não-determinística* ou *estocásticas*, caso a série seja determinada de modo probabilístico; *univariada* e *multivariada* quando uma variável é observada e quando duas ou mais variáveis são observadas, respectivamente.

De acordo com Box et al. (2015), a série temporal possui uma característica típica, onde as suas observações mais próximas possuem uma dependência, tornando-a bem atrativa na busca em encontrar uma certa dependência entre as observações utilizando suas metodologias. Claro que, de certa forma, para termos uma maior percepção de seu comportamento, é necessário saber que a série  $Y_t$  pode ser gerada da combinação de todos os seus seus característicos componentes, sendo eles:

- **Tendência:** o comportamento a longo prazo apresentado pela série, que reflete as mudanças, como diminuição ou aumento, que ocorreram nos dados de maneira consistente;
- **Sazonalidade:** movimentos padrões que se repetem em períodos idênticos de tempo;
- **Ciclo:** equivale aos movimentos de elevação e queda, que ocorrem repetidamente, em torno da tendência;
- **Erro aleatório:** são as variações que não são identificadas pelos demais componentes.



Todavia, é importante salientar que processo de séries temporais originou-se de *processos estocásticos*, onde para um melhor entendimento acerca de séries temporais, se faz necessário uma explanação sobre o mesmo.

## 2.2 Processos Estocásticos

Um processo estocástico é uma família de variáveis aleatórias indexadas por elementos  $t$  que pertencem a um determinado intervalo temporal. De modo geral, os processos estocásticos são processos aleatórios que dependem do tempo.

Assim, matematicamente, define-se o processo estocástico como uma família  $Z = \{Z_t, t \in T\}$  tal que, para cada  $t \in T$ , temos uma variável aleatória  $Z_t$  (MORETTIN; TOLOI, 2006), evoluindo ao longo do tempo, tendo as leis probabilísticas como base. Tratando-se de séries temporais, fazer mais de uma observação para cada tempo é utópico. Não obstante, efetua-se o processo estocástico apenas uma vez.

Normalmente, os processos estocásticos são relatados por meio de suas funções de média, variância e autocovariância. Assim,

- **Função da média** ( $Z_t$ ) :  $E[Z_t] = \mu_t$ ;
- **Variância**:  $\text{Var}[Z_t] = \sigma_t^2$ .
- **Função de autocovariância**:  $\gamma_{t,s} = \text{Cov}[Z_t, Z_s] = E[(Z_t - \mu_t)(Z_s - \mu_t)]$ , lembrando que  $t$  e  $s$  são tempos distintos;

Existe uma classe de tamanha importância nos processos estocásticos denominada de processos estacionários, onde convém um maior aprofundamento acerca do assunto.

### 2.2.1 Processos Estacionários

Uma das suposições mais importantes para um processo estocástico é a *estacionariedade*. Basicamente, isto significa que as leis de probabilidade que agem no processo não mudam com o tempo, ou seja, o processo mantém o equilíbrio estatístico. Portanto, uma série temporal é dita como estritamente estacionária quando a distribuição de probabilidade conjunta de  $X(t_1), \dots, X(t_k)$  é a mesma de  $X(t_1 + \tau), \dots, X(t_k + \tau)$ . Isto é, a mudança da origem dos tempos por uma quantidade  $\tau$  não obtém efeito sob a distribuição conjunta, dependendo assim, apenas dos intervalos entre  $t_1, \dots, t_k$ .

Existem dois tipos de estacionariedade que são, estritamente (fortemente) estacionária e de segunda ordem (fracamente estacionária). Segundo Morettin e Tolo (2006) tem-se as seguintes definições:

- **Estacionariedade Forte:** Um processo estocástico  $Z = \{Z_t, t \in T\}$  diz-se estritamente estacionário se todas as distribuições finito-dimensionais, ou seja,  $F(z_1, \dots, z_n; t_1, \dots, t_n) = P\{Z_{t_1} \leq z_1, \dots, Z_{t_n} \leq z_n\}$ , permanecem invariantes no tempo.

Assim, temos que sua média e variância são constantes nas translações do tempo, ou seja:

$$E(t) = \mu \text{ e } Var(t) = \sigma^2, \forall t \in T.$$

- **Estacionariedade Fraca:** Um processo estocástico  $Z = \{Z_t, t \in T\}$  diz-se fracamente estacionário, se e somente se:
  1.  $E[Z_t] = \mu$ , constante  $\forall t \in T$ ;
  2.  $E^2[Z_t] < \infty$ ,  $t \in T$ ;
  3.  $\gamma_{t,s} = Cov[Z_t, Z_s]$  é uma função de t-s, chamado de defasagem.

Um exemplo bastante importante de processo estacionário, é o ruído branco. Onde, a partir dele, pode-se construir diversos processos.

### 2.2.2 Ruído Branco

Segundo Box et al. (2015), ruído branco é um processo estacionário com uma sequência de variáveis aleatórias i.i.d., denotado por  $u_1, \dots, u_t$ , que assumem média 0 e variância  $\sigma_u^2$ , sendo um processo estritamente estacionário. Em resumo, o ruído branco é um processo estocástico não-correlacionado puramente aleatório.

O ruído branco possui as seguintes propriedades:

$$\begin{aligned} E(\mu_t) &= 0 \\ Var(\mu_t) &= \sigma^2 \\ Cov(\mu_t, \mu_j) &= 0, \forall t \neq j \end{aligned}$$

Como os  $u_t$  são não-correlacionados, por conta de sua independência, então sua função de autocovariância é dada por:

$$\gamma_k = E[u_t u_{t+k}] = \begin{cases} \sigma_u^2 & , k = 0; \\ 0 & , k \neq 0. \end{cases}$$

O termo *ruído branco* vem do fato de podermos mostrar que todas as frequências são iguais através de uma análise de frequência do modelo. Todavia, uma sequência de variáveis aleatórias pode estar concentrada em um processo de segunda ordem e, mesmo obtendo as mesmas características da média e variância, também será considerado um ruído branco, porém, terá uma estacionariedade fraca.

## 2.3 Funções de Autocovariância e Autocorrelação

O conhecimento acerca da dependência (relação) entre as observações, recentes e passadas, é fundamental para uma análise estruturada utilizando a metodologia de séries temporais. Para isto, utiliza-se a *função de autocovariância* para a análise da dependência na estrutura estocástica da série.

De acordo com Brockwell e Davis (2016), uma série temporal  $Y_t$  estacionária, com  $E[Z_t^2] < \infty$ , apresenta uma função de autocovariância (FACV), dada por,

$$\gamma(\tau) = Cov[Y_t, Y_{t+\tau}]$$

Esta função, apresenta as seguintes propriedades,

1.  $\gamma(0) > 0$ ;
2.  $\gamma(-\tau) = \gamma(\tau), \forall t \in T$ ;
3.  $|\gamma(\tau)| \leq \gamma(0), \forall t \in T$ .

A FAC (função de autocorrelação) é originada através da desvantagem na dependência da unidade de medição de  $\gamma$ . Onde, temos que a FAC é dada por,

$$\rho(\tau) = \gamma(\tau)/\gamma(0), \forall t \in T.$$

e apresenta propriedades análogas a  $\gamma(\tau)$ , menos para  $\rho(0) = 1$ .

## 2.4 Modelos de Suavização Exponencial

Durante muito tempo, a análise de regressão teve uma razoável aceitação como método de ajustar os modelos autorregressivos, cujo o objetivo é calcular previsões de séries temporais. No entanto, não é apropriado que a análise seja feita quando o número de observações é bem pequeno, pois, normalmente, temos que a hipótese básica de independência dos resíduos é violada, o que resulta em estimadores inconsistentes, impossibilitando testar hipóteses e estabelecer intervalos de confiança para os parâmetros (MORETTIN; TOLOI, 2006).

O propósito dos métodos é distinguir o padrão de qualquer ruído que possa estar contido nas observações e, só então, utilizar este padrão para prever os futuros valores das séries (MORETTIN; TOLOI, 2006). Para verificar estes acontecimentos nas previsões, torna-se indispensável utilizar o método de *alisamento* (suavização). É importante salientar que os modelos de suavização exponencial são uma classe de algoritmos de previsão *ad hoc*, onde temos métodos simples de previsões adaptativas, ajustando os dados automaticamente para serem incluídos na série.

Considerando-se  $Z_1, Z_2, \dots, Z_t$  uma série temporal, suas previsões serão construídas para  $Z_{t+1}, Z_{t+2}, (\dots)$ , ou seja, através dos erros encontrados em previsões passadas é que obtemos as previsões dos modelos de suavização exponencial.

### 2.4.1 Alisamento Exponencial Simples (AES)

O **método AES** é utilizado para séries temporais onde não existe a presença de tendência, nem sazonalidade, de acordo com Farias (2017), isto é, são séries compostas apenas por seu nível e ruído aleatório. Podemos decompor a série da seguinte forma:

$$Z_t = N_t + \epsilon_t, \text{ com } t = 1, 2, \dots, n.$$

em que,  $\epsilon_t \sim (0, \sigma_\epsilon^2)$  e  $N_t$  representam o nível da série no instante  $t$ .

Estima-se o parâmetro  $N_t$  através da média de observações passadas, onde os pesos crescem exponencialmente ao longo do tempo, ou seja, os pesos relativamente maiores são atribuídos as observações mais recentes. Segundo Chatfield (2016), temos:

$$N_t = \alpha Z_t + \alpha(1 - \alpha)Z_{t-1} + (1 - \alpha)^2 Z_{t-2} + \alpha(1 - \alpha)^3 Z_{t-3} + \dots$$

na qual,  $Z_t$  são as  $n$  observações da série e  $\alpha$  é a constante de suavização com  $0 < \alpha < 1$ .

Utiliza-se, frequentemente, a forma recursiva do algoritmo, devido ao uso da observação do uso do nível no instante anterior e pela observação no instante atual, que é dada por:

$$N_t = \alpha Z_t + (1 - \alpha)N_{t-1}$$

Logo, observa-se que para a estimativa do nível, temos uma medida ponderada entre  $N_{t-1}$ , que é a estimativa anterior do nível e  $Z_t$  (onde temos a observação atual da série).

A previsão de valores futuros com  $h$  passos a frente para qualquer  $h = 1, 2, \dots$  é dada pelo último valor do nível no momento  $t$ , isto é:

$$\hat{Z}_t(h) = N_t, \text{ com } h = 1, 2, \dots$$

em que, geralmente o algoritmo se inicializa pela primeira observação ( $N_1 = Z_1$ ).

Segundo Morettin e Tolo (2006), deve-se escolher a constante de suavização  $\alpha$ , para que tal valor forneça uma melhor previsão, de modo que minimize a soma dos quadrados dos erros de previsão.

O Alisamento Exponencial Simples é um método bastante utilizado por ser de fácil entendimento, por ter uma grande flexibilidade permitida pela variação da constante de alisamento ( $\alpha$ ), por possuir uma aplicação não dispendiosa e, por ter apenas a necessidade de armazenar somente  $Z_t, \bar{Z}$  e  $\alpha$ . Por outro lado, a principal desvantagem é a dificuldade na

determinação do valor mais apropriado da constante de alisamento que pode ser superada através da utilização do alisamento exponencial adaptativo de Trigg-Leach (MORETTIN; TOLOI, 2006).

### 2.4.2 Alisamento Exponencial de Holt

O método de Holt utiliza dois coeficientes diferentes de alisamento: o nível (também presente no AES) e a tendência linear (crescente ou decrescente) (MONTGOMERY; JENNINGS; KULAHCI, 2015). Utiliza-se do método de Alisamento Exponencial de Holt quando uma série apresenta tendência, pois previsões futuras pelo AES induziriam a um erro sistemático de previsão devido ao atraso no acompanhamento da tendência. De modo geral, este método permite a obtenção das estimativas de nível e da tendência da série (MARTÍNEZ; ZAMPROGNO, 2003).

Os coeficientes são estimados da seguinte maneira:

$$\begin{aligned} N_t &= \alpha Z_t + (1 - \alpha)(N_{t-1} + T_{t-1}) \\ T_t &= \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1}, \end{aligned}$$

com  $0 < \alpha < 1$  e  $0 < \beta < 1$ , em que  $\alpha$  e  $\beta$  são as constantes de suavização .

Para minimizar a soma dos quadrados dos erros de previsão, utilizando os coeficientes de suavização  $\alpha$  e  $\beta$ , utiliza-se um processo idêntico ao do AES, porém o algoritmo se inicia pela segunda observação, ou seja,  $N_2 = Z_2$  e  $T_2 = Z_2 - Z_1$  (MORETTIN; TOLOI, 2006).

Logo, temos que a previsão é dada por:

$$\hat{Z}_t(h) = N_t + hT_t, \text{ com } h = 1, 2, \dots$$

Para a correção dos erros neste algoritmo, utilizamos uma maneira análoga ao que utiliza-se no AES, ou seja,

$$\begin{aligned} N_t &= N_{t-1} + T_{t-1} + \alpha\epsilon_t \\ T_t &= T_{t-1} + \beta\epsilon_t. \end{aligned}$$

### 2.4.3 Alisamento Exponencial de Holt-Winters

O *Alisamento Exponencial de Holt-Winters* é utilizado em séries onde existe a presença de tendência e sazonalidade, considerado um padrão de comportamento para a série mais complexo. É importante salientar que, este algoritmo é dividido em: *modelos aditivos* e *multiplicativos*, e, que a utilização destes algoritmos baseia-se nos seguintes componentes: nível, tendências e sazonalidade (MORETTIN; TOLOI, 2006).

- *Série sazonal multiplicativa*: Considerando uma série sazonal que possui período  $s$ , que pode ser para dados mensais, anuais, por exemplo. Então, o fator sazonal  $F_t$  é multiplicativo, enquanto a tendência é aditiva (MORETTIN; TOLOI, 2006). A forma de recorrência deste algoritmo, é dada por:

$$\begin{aligned} N_t &= \frac{\alpha Y_t}{F_{t-s}} + (1 - \alpha)(N_{t-1} + T_{t-1}) \\ T_t &= \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1} \\ F_t &= \frac{\gamma Y_t}{N_t} + (1 - \alpha)F_{t-s}, \end{aligned}$$

com  $0 < \alpha < 1$ ,  $0 < \beta < 1$ ,  $0 < \gamma < 1$ , na qual  $\alpha, \beta$  e  $\gamma$  são as constantes de suavização do algoritmo.

As previsões dos valores futuros são obtidas através das expressões:

$$\begin{aligned} \hat{Y}_t(h) &= (N_t + hT_t)F_{t+h-s}, \text{ com } h = 1, 2, \dots, s \\ \hat{Y}_t(h) &= (N_t + hT_t)F_{t+h-2s}, \text{ com } h = s+1, s+2, \dots, 2s. \end{aligned}$$

A partir destas, utilizando algumas manipulações matemáticas, obtém-se a seguinte forma de correção de erros:

$$\begin{aligned} N_t &= N_{t-1} + T_{t-1} + \frac{\alpha \epsilon_t}{F_{t-s}} \\ T_t &= T_{t-1} + \alpha \beta \frac{\epsilon_t}{F_{t-s}} \\ F_t &= F_{t-s} + \gamma(1 - \alpha) \frac{\epsilon_t}{F_{t-s}} \end{aligned}$$

- *Série sazonal aditiva*: O procedimento exposto acima, pode passar por modificações em casos onde o fator sazonal passa a ser aditivo. Segundo Morettin e Tolo (2006), as equações de atualização são dadas por:

$$\begin{aligned} N_t &= \alpha(Y_t - F_{ts}) + (1 - \alpha)(N_{t-1} + T_{t-1}) \\ T_t &= \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1} \\ F_t &= \gamma(Y_t - N_t) + (1 - \gamma)F_{t-s} \end{aligned}$$

com  $0 < \alpha < 1$ ,  $0 < \beta < 1$ ,  $0 < \gamma < 1$ , sendo  $\alpha, \beta$  e  $\gamma$  são as constantes de suavização do algoritmo.

Para as previsões futuras, com a previsão  $h$  passos à frente, temos, a partir de  $Y_t$  que:

$$\begin{aligned} \hat{Y}_t(h) &= N_t + hT_t + F_{t+h-s}, \text{ para } h = 1, 2, \dots, s \\ \hat{Y}_t(h) &= N_t + hT_t + F_{t+h-2s}, \text{ para } h = s+1, s+2, \dots, 2s. \end{aligned}$$

A partir disto, utilizando algumas manipulações matemáticas, obtem-se a seguinte forma de correção de erros:

$$N_t = N_{t-1} + T_{t-1} + \alpha\epsilon_t$$

$$T_t = T_{t-1} + \alpha\beta\epsilon_t$$

$$F_t = F_{t-s} + \gamma(1 - \alpha)\epsilon_t.$$

## 2.5 Modelagem Box-Jenkins

Uma metodologia de previsão clássica é a modelagem de Box-Jenkins, que fundamenta-se na estruturação de funções que baseiam-se num determinado ciclo iterativo, utilizando dos próprios dados contidos na série para a obtenção de uma estrutura favorável para as previsões.

Esta abordagem pode ser dividida em quatro passos:

1. *Especificação*: uma classe geral de modelos é considerada (especificação);
2. *Identificação*: identificação do modelo, baseando-se na análise de autocorrelações, autocorrelações parciais e critérios secundários;
3. *Estimação*: estima-se os parâmetros do modelo;
4. *Verificação ou diagnóstico*: Através da análise de resíduos, verifica-se o modelo ajustado, analisando se este é o modelo adequado para as previsões.

Estes quatro passos são repetidos, caso o modelo escolhido não seja adequado, voltando para a fase de *identificação*. Vale ressaltar que é indicado a análise de mais de um modelo, afinal, quando a finalidade é fazer previsões, é indicado utilizar o modelo que apresente o menor erro quadrático médio de previsão. Normalmente, considera-se melhor os modelos mais simples e com menos parâmetros, pois tendem a proporcionar previsões mais precisas.

### 2.5.1 Modelo Autorregressivo (AR)

Quando os parâmetros possuem quantidades fixas e não podem assumir variações ao longo do tempo, utiliza-se os modelos autorregressivos. Porém, estes são muito utilizados para diversas aplicações, principalmente na análise clássica.

Denotado por  $AR(p)$ , a série  $Y_t$  é explicada baseada na média ponderada de seus valores prévios e pelo *ruído branco* (um erro aleatório) (BOX et al., 2015). De modo geral,

descreve-se a série  $Y_t$  como a combinação linear de  $p$  valores passados de  $Y_t$  e pelo ruído branco no tempo  $t$ . Logo, tem-se que o modelo  $AR(p)$  é dado por:

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + u_t,$$

em que,  $\phi_i, \forall_i = 1, 2, \dots, p$  são parâmetros do modelo e o ruído branco é o  $u_t$  (no tempo  $t$ ).

Para o modelo autorregressivo, a variância e a autocovariância são definidas pelas equações

- Variância:

$$\gamma_0 = \phi_1 \gamma_1 - \phi_2 \gamma_2 - \dots - \phi_p \gamma_p + \sigma_u^2$$

na qual  $\sigma_u^2$  é a variância do ruído branco  $u_t$ .

- Autocovariância:

$$\gamma_k = \phi_1 \gamma_{k-1} - \phi_2 \gamma_{k-2} - \dots - \phi_p \gamma_{k-p} + \sigma_u^2, k \geq 1.$$

Se utilizarmos a equação de autocovariância, e fizermos  $k = 1, 2, \dots, p$  e utilizarmos  $\gamma_0 = 1$  e  $\gamma_{-k} = \gamma_k$ , conseguimos obter as equações de autocorrelações, que são conhecidas como as equações de Yule-Walker (BOX et al., 2015).

## 2.5.2 Modelo de Médias Móveis (MA)

Utiliza-se o modelo de médias móveis na análise de séries temporais para a modelagem de modelos das séries univariadas no tempo. Conseqüentemente, como a série  $Y_t$  é construída com base na combinação linear de seus erros aleatórios passados para gerar o valor mais recente da série (BOX et al., 2015).

A nomenclatura  $MA(q)$ , refere-se a um modelo cujo  $\mu$  é a média,  $u_t \sim RB(0, \sigma^2)$ ,  $\theta_1, \theta_2, \dots, \theta_q$  são os parâmetros e  $q$  é a ordem do modelo. De acordo com Morettin e Tolo (2006), o modelo de médias móveis, denotado por  $MA(q)$ , é representado por:

$$Y_t = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}.$$

Assim, tem-se que a variância deste modelo é dada por,

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_u^2.$$

Quando temos que  $k > q$ , suas autocovariâncias são nulas, isto é,  $\gamma_k = 0$ . Mas, para os valores de  $k < q$ , as autocovariâncias são dadas por,

$$\left\{ \begin{array}{ll} 0 & , \text{ se } k > q; \\ \gamma_k = (-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_q \theta_{q-k}) \sigma_u^2 & , \text{ em que } k = 0, 1, \dots, q = 0; \\ \gamma(-k) & , \text{ se } k < 0. \end{array} \right.$$



### 2.5.3 Modelo autorregressivo de médias móveis (ARMA)

Resultando da combinação dos modelos autorregressivos e de médias móveis, o modelo  $ARMA(p, q)$  é uma combinação linear dos valores passados ( $p$ ) de  $Y_t$  com  $u_t \sim RB(0, \sigma^2)$ ,  $\forall t$  (ruído branco), no tempo presente e nos  $q$  tempos passados (BOX et al., 2015). Este modelo procura utilizar de uma quantidade menor de parâmetros na representação da série (MORETTIN; TOLOI, 2006).

O modelo autorregressivo de médias móveis,  $ARMA(p, q)$ , é descrito por,

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q},$$

sendo os  $\phi$ 's e os  $\theta$ 's são parâmetros autorregressivos e parâmetros das médias móveis, respectivamente.

Reescrevendo esta equação de modo que temos os polinômios do  $AR(p)$  e  $MA(q)$ , que são  $\phi(B)$  e  $\theta(B)$ , respectivamente, temos,

$$\phi(B)Z_t = \alpha + \theta(B)u_t$$

Temos que as equações de variância e autocovariâncias, são expressas, respectivamente, por,

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} - \theta_1 \gamma_{yu}(k-1) - \dots - \theta_q \gamma_{yu}(k-q), k \leq q,$$

em que  $\gamma_{yu}(k) = E[u_t Y_{t-k}]$ .

e,

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}, k > q.$$

### 2.5.4 Modelo autorregressivo integrado de médias móveis (ARIMA)

O modelo autorregressivo integrado de médias móveis (ARIMA) é uma generalização do modelo ARMA. Este, como foi visto anteriormente, é utilizado quando as séries são estacionárias, porém, raramente tem-se a ocorrência disto, pois, normalmente as séries não apresentam estacionariedade. Então, quando esta hipótese não é satisfeita, e pode tornar-se estacionária a partir da aplicação de  $d$  diferenças, utiliza-se o modelo ARIMA, no qual, podem ser aplicadas, no máximo, duas diferenças (BOX et al., 2015).

O modelo é denotado por  $ARIMA(p, d, q)$ , onde  $p$  é o operador autorregressivo,  $d$  é o operador diferença e  $q$  é o operador de médias móveis, cuja equação é,

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}$$

em que,  $w_t = \Delta^d Y_t$  (diferença).

E, pode-se ser escrita por:

$$\phi(B)((1 - B)^d Y_t - \alpha) = \theta(B)u_t.$$

De modo geral, define-se a diferença de ordem 1 por:

$$w_t = \Delta^1 Y_t = Y_t - Y_{t-1}.$$

Consequentemente, para as diferenças de ordem ( $d > 1$ ), onde  $\Delta^d$  é o operador de defasagem, tem-se:

$$w_t = \Delta^d Y_t = \Delta^{d-1}(\Delta^1 Y_t),$$

Quando define-se  $\gamma_0 = E[w_t^2]$ ,  $\gamma_k = E[w_t w_{t-k}]$  e  $\gamma_w u(k) = E[u_t w_{t-k}]$ , a variância e as autovariâncias são iguais, dadas por:

$$\phi(B)((1 - B)^d Y_t - \alpha) = \theta(B)u_t$$

### 2.5.5 Modelo autorregressivo de médias móveis sazonal (SARIMA)

Trata-se de um modelo para séries sazonais, onde tem-se o  $s$  como o período de sazonalidade. O modelo SARIMA é uma extensão do modelo ARIMA( $p, q, d$ ), mas, aqui, leva-se em conta a sazonalidade estocástica dos dados (MORETTIN; TOLOI, 2006).

Possui notação SARIMA ( $p, d, q$ ) x ( $P, D, Q$ ), em que  $P$  equivale ao número de coeficientes sazonais regressivos (referente a estacionariedade), o  $D$  ao número de diferenças sazonais ( $(1 - B^s)Z_t = Z_t - Z_{t-s}$ ) que tornam a série estacionária levando em conta o número de períodos ( $s$ ) e o  $Q$  como o número de coeficientes sazonais de médias móveis relacionados a invertibilidade. Sendo assim, o modelo é dado por:

$$\phi(B)\Phi(B)[(1 - B)^d(1 - B^s)^D - \alpha]Y_t = \theta(B)\Theta(B)u_t.$$

### 2.5.6 Modelo autoregressivo de médias móveis sazonal com variáveis exógenas (SARIMAX)

De uma adaptação feita a partir do modelo SARIMA, para adicionar uma estrutura de regressão, tem-se o SARIMAX. Este, é denotado por SARIMAX ( $p, d, q, s$ ) x ( $P, D, Q$ ) e a equação do modelo é dada por:

$$\phi(B)\Phi(B)(1 - B)^d(1 - B^s)^D Z_t = \alpha + X\beta + \theta(B)\Theta(B^s)u_t,$$

onde  $X$  uma matriz  $n \times k$  de regressores não aleatórios e  $\beta$  um vetor de parâmetros  $k \times 1$ .

### 2.5.7 Função de autocorrelação (FAC)

A função de autocorrelação é bastante utilizada na modelagem Box-Jekins, fornecendo uma descrição parcial do processo que está modelando, bem como, mede a correlação, em períodos distintos, entre valores de duas variáveis  $w_t$  e  $w_{t-k}$  (ENDERS, 2008). Então, de uma série temporal que é estacionária, onde  $w_t$  e  $w_{t-k}$  são duas observações, tem-se que sua distribuição de probabilidade conjunta é a mesma para qualquer período de tempo,  $t$  e  $t+k$ , sendo assim, separados por um mesmo intervalo, chamado de *lag*( $k$ ).

Portanto, tal equação, para estimativas populacionais, é definida através de:

$$\rho_k = \frac{Cov[w_t, w_{t-k}]}{Var[w_t]} = \frac{\gamma_k}{\gamma_0}$$

No entanto, sabe-se que raramente trabalha-se com estimativas populacionais. Dessa forma, a FAC para estimativas amostrais é dada por:

$$r_k = \frac{\sum_{t=k+1}^n (w_t - \bar{w})(w_{t-k} - \bar{w})}{\sum_{t=1}^n (w_t - \bar{w})^2}, \text{ onde } n \text{ é o número de observações.}$$

É necessário descobrir qual a distribuição de probabilidade de  $r_k$  para que seja feita uma análise satisfatória. Quando o  $n$  é suficientemente grande, a distribuição de  $r_k \sim N(0, \frac{1}{n})$ , portanto, a partir da distribuição encontrada, pode-se encontrar seus respectivos intervalos de confiança e utilizar testes de hipóteses tendo em vista verificar a hipótese em que as correlações são nulas (LÚCIA, 2000).

Para testar as hipóteses em que os primeiros coeficientes de autocorrelação  $k$  são conjuntamente iguais a zero, utiliza-se a seguinte equação:

$$Q(K) = n(n+2) \sum_{k=1}^K \frac{r_k^2}{n-k},$$

$Q(k) \sim \chi^2$ , com  $k$  graus de liberdade.

### 2.5.8 Função de autocorrelação parcial (FACP)

A função de autocorrelação parcial, que é denotada por  $\phi_{kk}$ , é utilizada para analisar a correlação entre pontos da série,  $w_t$  e  $w_{t-k}$ , reduzindo a influência que  $w_{t-1}, \dots, w_{t-k+1}$  exerce sobre  $w_t$ . Todavia, a FACP também é usada para a identificação de modelos para determinada série temporal, auxiliando na identificação da estacionariedade da série.

$$\left\{ \begin{array}{l} w_t = \phi_{11}w_{t-1} + u_t, \\ w_t = \phi_{11}w_{t-1} + \phi_{22}w_{t-2} + u_t, \\ \vdots \\ w_t = \phi_{k1}w_{t-1} + \phi_{k2}w_{t-2} + \dots + \phi_{kk}w_{t-k} + u_t. \end{array} \right.$$

Uma maneira alternativa utilizada para estimar os valores de  $\phi_{kk}$  é utilizando o sistema de equações de *Yule-Walker*.

$$\begin{cases} r_1 = \phi_{11} + \phi_{22}r_1 + \dots + \phi_{kk}r_{k-1}, \\ r_2 = \phi_{11}r_1 + \phi_{22} + \dots + \phi_{kk}r_{k-2}, \\ \vdots \\ r_k = \phi_{11}r_{k-1} + \phi_{22}r_{k-2} + \dots + \phi_{kk}. \end{cases}$$

Lembrando que, no caso do  $n$  ser suficiente grande, tem-se que  $\phi_{kk}$  apresenta distribuição  $N(0, \frac{1}{n})$ , para  $k > p$ .

## 2.6 Etapas do modelo SARIMA

### 2.6.1 Identificação

A fase de identificação é crucial, pois, aqui, está presente a dificuldade na determinação dos parâmetros do modelo ( $p, d$  e  $q$ ), afinal, se estes forem determinados de maneiras errôneas, o modelo será inadequado, comprometendo diretamente os resultados. Vale ressaltar que esta fase engloba conjuntos de procedimentos que proporcionam a ideia aproximada acerca da estrutura do modelo.

Todavia, sabe-se que o mais simples de encontrar é o  $d$ , que é o parâmetro de integração, pois aplica-se algum teste de estacionariedade com o objetivo de encontrar a ordem de integração. Este parâmetro pode assumir, normalmente, os valores  $d = 0$ , quando a série é estacionária e  $d = 1$ , quando a série não é estacionária, possuindo uma raiz unitária, mas, raramente o valor  $d = 2$ . Existem meios para testar se a série  $Y_t$  tem raiz unitária, para mais detalhes consultar (FULLER, 2009) e para a ordem de integração  $d > 1$  consultar (DICKY; PANTULA, 1987).

### 2.6.2 Teste de Estacionariedade de Dickey-Fuller

Considerando o modelo  $Y_t = \rho Y_t + \epsilon_t$ , em que  $\epsilon_t$  é o ruído branco. A série  $Y_t$  é estacionária e descrita por um AR(1), se  $|\rho| < 1$ . Mas, se  $\rho = 1$ , a série  $Y_t$  é não estacionária. Dickey e Pantula (1987) propuseram o teste sobre a presença da raiz unitária na série  $Y_t$ , quando o processo gerador da série é expresso por um destes modelos:

$$\Delta Y_t = \alpha + \beta_t + \lambda_3 Y_{t-1} + \epsilon_t$$

$$\Delta Y_t = \alpha + \lambda_2 Y_{t-1} + \epsilon_t$$

$$\Delta Y_t = \lambda_1 Y_{t-1} + \epsilon_t$$

em que  $\lambda_i = \rho - 1$ ,  $\forall_i = 1, 2, 3$ ,  $\alpha$  e  $\beta$  são constantes para serem estimadas.

Para o teste da raiz unitária de Dickey-Fuller, estima-se a seguinte autorregressão:

$$\begin{aligned}\Delta Y_t &= (\rho - 1)Y_{t-1} + \varepsilon_t \\ \Delta Y_t &= \gamma Y_{t-1} + \varepsilon_t\end{aligned}$$

na qual  $\Delta = (Y_t - Y_{t-1})$  é o operador de diferença e  $\gamma = \rho - 1$ .

As hipóteses a serem testadas são apresentadas de acordo com a Tabela 1, onde  $H_0 : \rho = 1$  equivale a  $H_0 : \lambda = 0$ .

Tabela 1 – Testes da raiz unitária de Dickey-Fuller

Modelos	$H_0$	Regras de decisão
1	$\lambda_3 = 0$	$\tau_3 > \text{valor crítico} \Rightarrow H_0$ não é rejeitada
	$(\alpha, \beta, \lambda_3) = (0, 0, 0)$	$\delta_2 < \text{valor crítico} \Rightarrow H_0$ não é rejeitada
	$(\alpha, \beta, \lambda_3) = (\alpha, 0, 0)$	$\delta_3 < \text{valor crítico} \Rightarrow H_0$ não é rejeitada
2	$\lambda_2 = 0$	$\tau_2 > \text{valor crítico} \Rightarrow H_0$ não é rejeitada
	$(\alpha, \lambda_2) = (0, 0)$	$\delta_1 < \text{valor crítico} \Rightarrow H_0$ não é rejeitada
3	$\lambda_1 = 0$	$\tau_1 > \text{valor crítico} \Rightarrow H_0$ não é rejeitada

Fonte: Adaptada de Dickey e Fuller (1979)

Finalmente, se ao menos uma das hipóteses apresentadas na Tabela 1, não for rejeitada, a série é não estacionária, possuindo pelo menos uma raiz unitária.

Neste caso, testaremos as seguintes hipóteses:

- $H_0 : \gamma = 0$ , existe pelo menos uma raiz unitária, e sua variável não é estacionária;
- $H_1 : \gamma < 0$ , não existe raiz unitária, e sua variável é fracamente estacionária.

As estatísticas  $\tau_1, \tau_2, \tau_3, \delta_1, \delta_2$  e  $\delta_3$  que estão presentes na Tabela 1, são encontradas de acordo com:

$$\begin{aligned}\tau_3 &= \frac{\lambda_3}{\sigma_{\lambda_3}} \\ \delta_3 &= \frac{SQR(1) - SQR(1)}{\frac{3SQR(1)}{n}} \\ \tau_2 &= \frac{\lambda_2}{\sigma_{\lambda_2}} \\ \delta_2 &= \frac{SQR(2) - SQR(2)}{\frac{3SQR(2)}{n}} \\ \tau_1 &= \frac{\lambda_1}{\sigma_{\lambda_1}}\end{aligned}$$

$$\delta_1 = \frac{SQR(3) - SQR(2)}{\frac{3SQR(3)}{n}}$$

em que  $SQR(1)$  é a soma de quadrado do resíduo do modelo  $\Delta Z_t = \alpha + \beta_t + \lambda_3 Z_{t-1} + \varepsilon_t$ ,  $SQR(2)$  é a soma de quadrado do resíduo do modelo  $\Delta Z_t = \alpha + \lambda_2 Z_{t-1} + \varepsilon_t$ ,  $SQR(3)$  é a soma de quadrado do resíduo do modelo  $\Delta Z_t = \lambda_1 Z_{t-1} + \varepsilon_t$ , já  $\sigma_{\lambda_1}$  é a variância de  $\lambda_1$ ,  $\sigma_{\lambda_2}$  é a variância de  $\lambda_2$  e  $\sigma_{\lambda_3}$  é a variância de  $\lambda_3$ .

### 2.6.2.1 Análise Gráfica

Com a série estacionária definida, bem como o valor do parâmetro  $d$ , o próximo passo é a determinação dos parâmetros  $p$  e  $q$ , o que nos leva a um  $ARIMA(p, q)$  da série estacionária. Para isto, usualmente, utiliza-se a metodologia que baseia-se na análise dos gráficos estabelecidos através das funções de autocorrelação (FAC) e autocorrelação parcial (FACP).

Os valores obtidos através do modelo  $AR(p)$  tendem a zero aceleradamente à medida que  $k$  cresce, isto é, apresenta uma FAC declinante. Já a FACP é truncada em  $p$ , onde a FACP dos  $p$  primeiros  $k$  são diferentes de zero e para os demais  $k$ , a FACP é igual a zero. O modelo  $MA(q)$  possui um comportamento inverso ao supracitado, onde a FAC é truncada em  $q$  da FACP é declinante.

Quando a série estacionária é descrita a partir da combinação do AR e MA, tornando-se um modelo  $ARMA(p, q)$ , as FAC e FACP apresentarão comportamentos similares, declinando com o aumento de  $k$ . Todavia, a determinação dos valores de  $p$  e  $q$  torna-se mais difícil. Uma forma de resolver este problema, é estimar diversas combinações destes dois parâmetros e após verificar a qualidade dos modelos obtidos, selecionar o mais adequado.

### 2.6.3 Estimação

Identificado um modelo provisório para a série, o passo seguinte é estimar seus parâmetros, onde se pode usar o procedimento iterativo de Mínimos Quadrados Condicionais ou o Método de Máxima Verossimilhança (MORETTIN; TOLOI, 2006).

Neste estudo será utilizado o método de Máxima Verossimilhança pelo fato dele maximizar a verossimilhança de um conjunto de observações, e também devido a aplicabilidade do software usado apresentar cálculos bem implementados das estimativas dos parâmetros.

Considerando um modelo  $ARIMA(p, d, q)$  e colocando seus  $p + q + 1$  parâmetros no vetor  $\xi = (\phi_i, \theta_j, \sigma_\xi^2)$ , onde  $\phi_i = (\phi_1, \dots, \phi_p)$ ,  $\theta_j = (\theta_1, \dots, \theta_q)$ . Para estimar  $\xi$ , será usado o método de Máxima Verossimilhança: dadas as  $N$  observações  $Z_1, \dots, Z_N$ , considera-se a função de verossimilhança  $L = (\xi | Z_1, \dots, Z_N)$  como função de  $\xi$ . Além disso, deve-se supor

que  $\xi$  (ruído branco) tenha distribuição normal (MORETTIN; TOLOI, 2006). Porém, se a modelagem é do tipo SARIMAX, então considera-se que o ruído branco tem distribuição normal, e  $\mathbf{w} = (w_1, w_2, \dots, w_n)$  apresentará distribuição normal conjunta com o vetor de médias nulo e matriz de covariância  $\Sigma$ .

A função adaptada de máxima verossimilhança é dada por Brockwell e Davis (2016)

$$L(\phi_i, \theta_j, \sigma_\xi^2/w) = (2\pi\sigma_\xi^2)^{-\frac{n}{2}} \left| \Sigma \right|^{-\frac{1}{2}} e^{-\frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{w}} / \sigma_\xi^2$$

Ao maximizar esta função obtemos as estimativas dos parâmetros  $\phi_i$ ,  $\theta_j$  e  $\sigma_\xi^2$ .

## 2.6.4 Validação

Com o modelo estimado, a próxima etapa consiste em uma análise de resíduos, para a avaliação se a quantidade de parâmetros é conveniente. De qualquer forma, a análise de resíduos ( $\epsilon_t$ ) corrobora para verificar indícios de ruído branco, onde, a partir do teste de *Ljung-Box*, verifica-se que os coeficientes de autocorrelação dos resíduos são estatisticamente iguais a zero. Todavia, se faz necessário a verificação menos específica empregue aos resíduos do modelo que foi ajustado (BOX et al., 2015). Aqui, pode-se utilizar gráficos de resíduos, com  $(t, e_t)$ , a FAC do erro e o gráfico de Ljung-Box  $(K, Q(K))$ , ou, alternativamente, testar a normalidade dos resíduos através do teste de Kolmogorov-Smirnov.

Os modelos não parcimoniosos podem englobar coeficientes que estejam correlacionados (ou não-significativos). Para resolver tal problema, não existe uma solução específica, mas, pode-se através dos softwares estatísticos fazer uma análise mais rebuscada acerca da estimação e utilizar várias combinações de  $p$  e  $q$ , descartando as combinações que correspondem a modelos não significativos. Para esta seleção, utiliza-se os seguintes modelos como critérios:

- **FPE** - Erro de Previsão Final

$$FPE = \tilde{\sigma}^2 x \frac{M+p}{M-p}$$

- **AIC** - Critério de Informação de Akaike

$$AIC = 2\log\hat{L} + 2(p+q)$$

em que  $\hat{L}$  é o valor maximizado da verossimilhança.

- **AICC** - Critério de Informação de Akaike Corrigido

$$AICC = -2\log\hat{L} + \frac{2(p+q)M}{M-p-q-1}$$

- **BIC** - Critério de Informação Bayesiano

$$BIC = -2\log\hat{L} + (p + q)\log M$$

Vale salientar que por tenderem a selecionar modelos superparametrizados assintoticamente, os critérios de *AIC* e *AICC* não são consistentes. Contrário a isto, tem-se o *BIC* que é um critério bastante rigoroso, consequentemente consistente para a inclusão de parâmetros do modelo.

## 2.7 Avaliação de Previsão

Um dos principais motivos para a realização da análise de séries temporais é a previsões de valores futuros. No entanto, se faz necessário a avaliação das previsões futuras originadas do modelo escolhido. Assim, considera-se uma série temporal com  $\mathbf{M}$  observações, tiramos desta série as  $n$  últimas observações. Utilizando a nova série, sem as  $\mathbf{n}$  observações, gera  $\mathbf{n}$  previsões a partir do modelo que foi estimados para obter os erros de previsão, a partir de,

$$\epsilon_j = Y_j - [\bar{Y}]_j$$

onde  $[\bar{Y}]_j$  é a previsão da  $Y_j$ .

Para avaliar a capacidade de previsão do modelo que foi estimado, as estatísticas que avaliam os erros serão utilizadas a seguir. Lembrando que,  $k = M - n$ , então:

1. **MPE** - Erro Percentual Médio

$$MPE = \left( \frac{1}{n} \sum_{j=k+1}^M \frac{\epsilon_j}{Z_j} \right) \times 100$$

2. **MSE** - Erro Quadrático Médio

$$MSE = \frac{1}{n} \sum_{j=k+1}^M (\epsilon_j)^2$$

3. **MAE** - Erro Absoluto Médio

$$MAE = \frac{1}{n} \sum_{j=k+1}^M |\epsilon_j|$$

4. **TE** - Erro Total

$$TE = \sum_{j=k+1}^M \epsilon_j$$

5. **TAE** - Erro Absoluto Total

$$TAE = \sum_{j=k+1}^M |\epsilon_j|$$



## 6. TPE - Erro Percentual Total

$$TPE = \left( \frac{\sum_{j=k+1}^M \epsilon_j}{\sum_{j=k+1}^M Z_j} \right) \times 100$$

Por fim, sabe-se que a utilização de qualquer desses erros pode ser efetivada de acordo com cada tipo de análise. Porém, para o caso de séries temporais, o mais indicado a ser utilizado é o **TPE - Erro Percentual Total** (FARIAS, 2017).

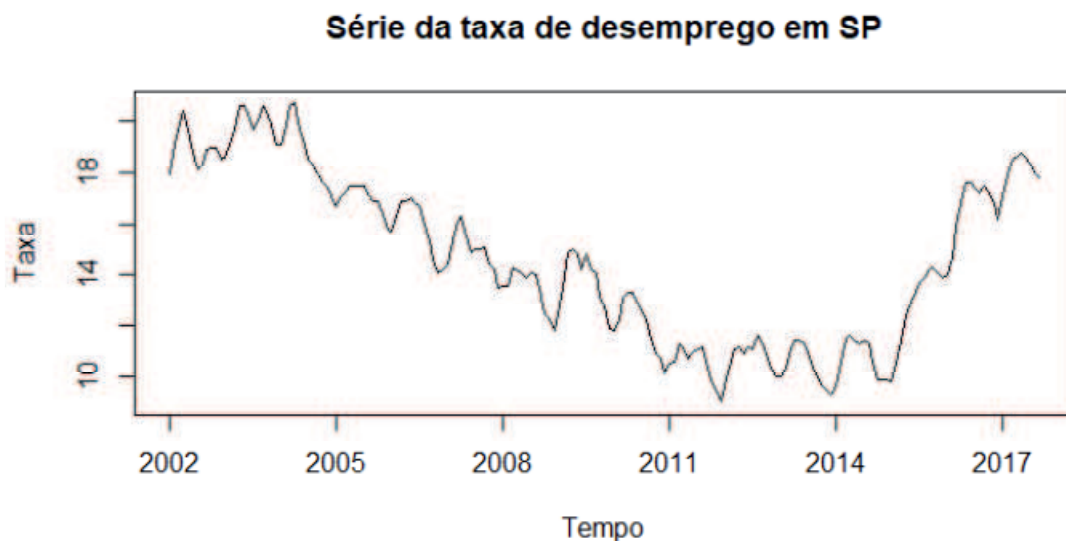
## 3 Aplicação

Para tal análise, optou-se pelo uso do software R (R Core Team, 2017), um software estatístico gratuito que permite análises estatísticas e oferece recursos para a construção de gráficos.

### 3.1 Análise exploratória dos dados

A série temporal da taxa de desemprego na Região Metropolitana de São Paulo, utilizada neste trabalho, compõe-se de 189 observações mensais que vão de janeiro de 2002 à setembro de 2017. A maior taxa de desemprego desta série é de 20,7%, equivalente ao mês de abril de 2004, e a menor taxa é de 9,0%, referente ao mês de dezembro de 2011. A média amostral da série da taxa de desemprego é 11,02% com um desvio padrão de 3,32%.

Figura 1 – Taxa de Desemprego em São Paulo (de janeiro de 2012 à setembro de 2017).



De acordo com a Figura 1, observa-se o gráfico da série, onde apresenta o seu comportamento ao longo do tempo. Por meio deste, percebe-se indícios de tendência, sazonalidade e não estacionariedade.

No gráfico de autocorrelação amostral da série (*ACF*), presente da Figura 2, nota-se um decaimento lento para zero, comprovando a não estacionariedade da série, e, a partir do lag 4, aproximadamente, a correlação se torna negativa. No que refere-se a função de autocorrelação parcial da série (*PACF*), observa-se picos que ultrapassam o limite, indicando que existe uma forte correlação entre esses meses.

Figura 2 – Funções de autocorrelação amostral e autocorrelação parcial da série.

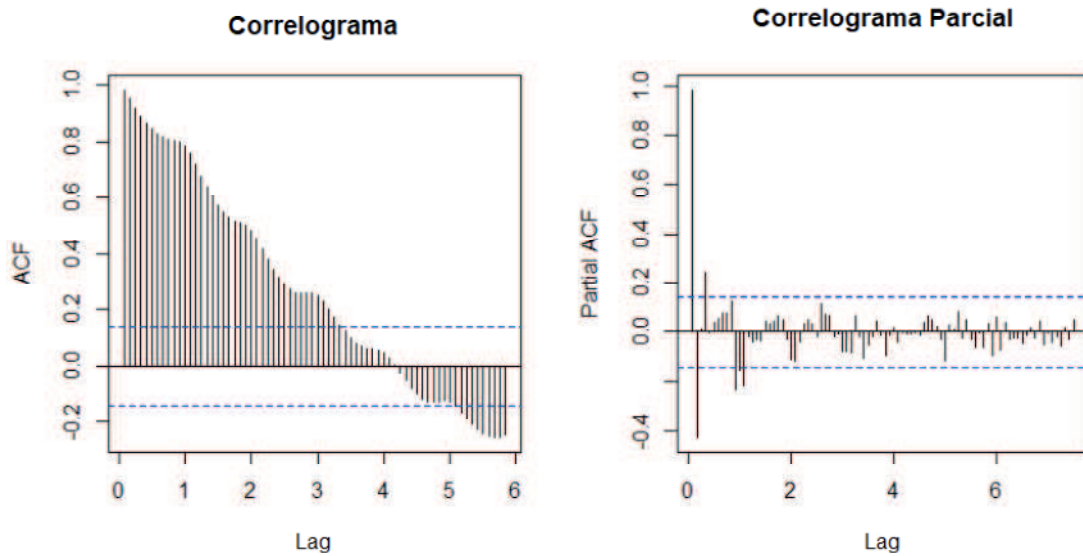
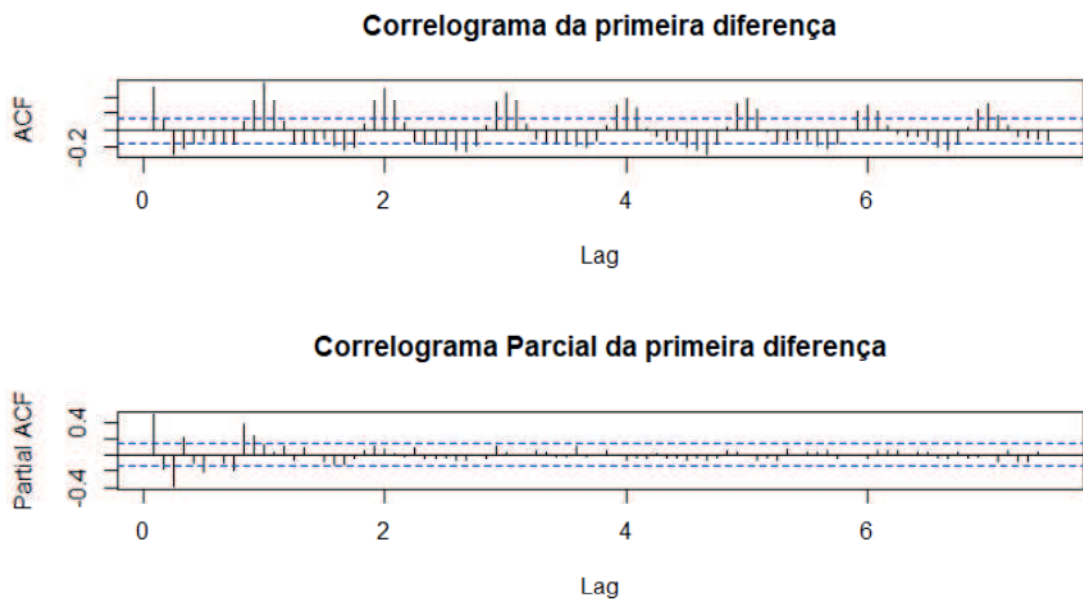


Figura 3 – Funções de autocorrelação amostral da primeira diferença e autocorrelação parcial da primeira diferença da série.



Para verificar a estacionariedade da série, novamente, aplicou-se a primeira diferença ( $d = 1$ ). A Figura 3, mostra que para o **ACF** e **PACF**, há picos que ultrapassam os limites propostos. Porém, no correlograma da primeira diferença da série, nota-se que há picos de correlação que se repetem, o que sugere uma forte sazonalidade na série.

Com base no objetivo do trabalho, que é a previsão, faz-se a análise de possíveis modelos que sejam capazes de representar adequadamente a série da taxa de desemprego em São Paulo.

## 3.2 Modelagem

A partir daqui, investigaremos modelos que sejam capazes de representar a série da taxa de desemprego em São Paulo, que possui valores que abrangem desde janeiro de 2002 à setembro de 2017. Para isto, será considerada previsões 1 e 7 passos à frente, em que, foram utilizados os métodos de Holt-Winters e Box-Jenkins, o primeiro na forma do modelo aditivo e multiplicativo, o segundo, nas formas SARIMA e SARIMAX.

### 3.2.1 Alisamento Exponencial de Holt-Winters

Como observou-se a presença de sazonalidade na série, então opta-se pelo uso dos algoritmos de suavização aditivo e multiplicativo de Holt-Winters. Vale ressaltar que foi retirado o último valor da série original, a fim de verificar o erro na previsão de um passo à frente. Posteriormente, considerará a previsão de 7 passos à frente, ou seja, até Março de 2018 (primeiro trimestre). Todavia, se faz necessário um maior cuidado na escolha dos parâmetros de suavização, afinal, estes contribuem para a minimização da soma dos quadrados dos erros de previsão, um passo à frente.

Inicialmente, se faz importante a escolha dos parâmetros de suavização adequados para, buscar ao máximo, minimizar a soma dos quadrados dos erros de previsão um passo à frente. Estes, estão dispostos na Tabela 2:

Tabela 2 – Valores otimizados das constantes de suavização para Holt-Winters aditivo e multiplicativo.

MÉTODO	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
HW Aditivo	0,6510948	0,04867425	1
HW Multiplicativo	0,8383338	0,02834135	1

Observa-se que os valores equivalentes a  $\beta$  obtidos pelo método sazonal aditivo e multiplicativo são muito próximos de 0, indicando que ao longo do tempo existe um decréscimo lento das previsões da taxa de desemprego. Através dos algoritmos de Holt-Winters, um passo à frente, foi encontrado os valores de previsão 17,95% e 17,47%, para o aditivo e o multiplicativo, respectivamente. Essas previsões trouxeram valores abaixo e acima do valor esperado, isto é, no caso do aditivo temos um valor superestimado e no multiplicativo, subestimado.

Através da Figura 4, observa-se que os dados previstos comportam-se de maneira satisfatória em relação aos dados reais, graficamente mostrando pouca diferença.

Após verificar que o método de alisamento exponencial de Holt-Winters obteve um comportamento satisfatório, realizou-se uma previsão da série com 7 passos à frente, no intuito de prever o comportamento da taxa de desemprego em São Paulo até o primeiro trimestre de 2018 (março/2018), através do método aditivo e multiplicativo.

Figura 4 – Gráfico da série (linha preta) e seu ajuste de Holt-Winters (linha vermelha) aditivo e multiplicativo (1 passo à frente).

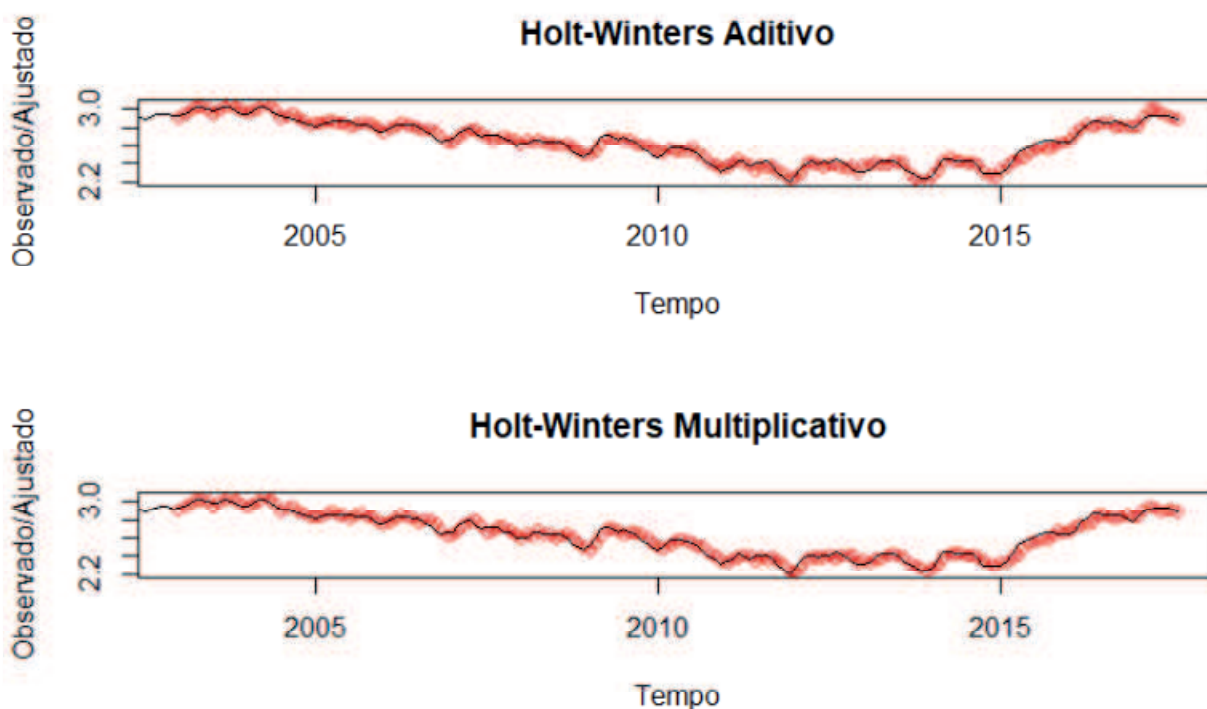


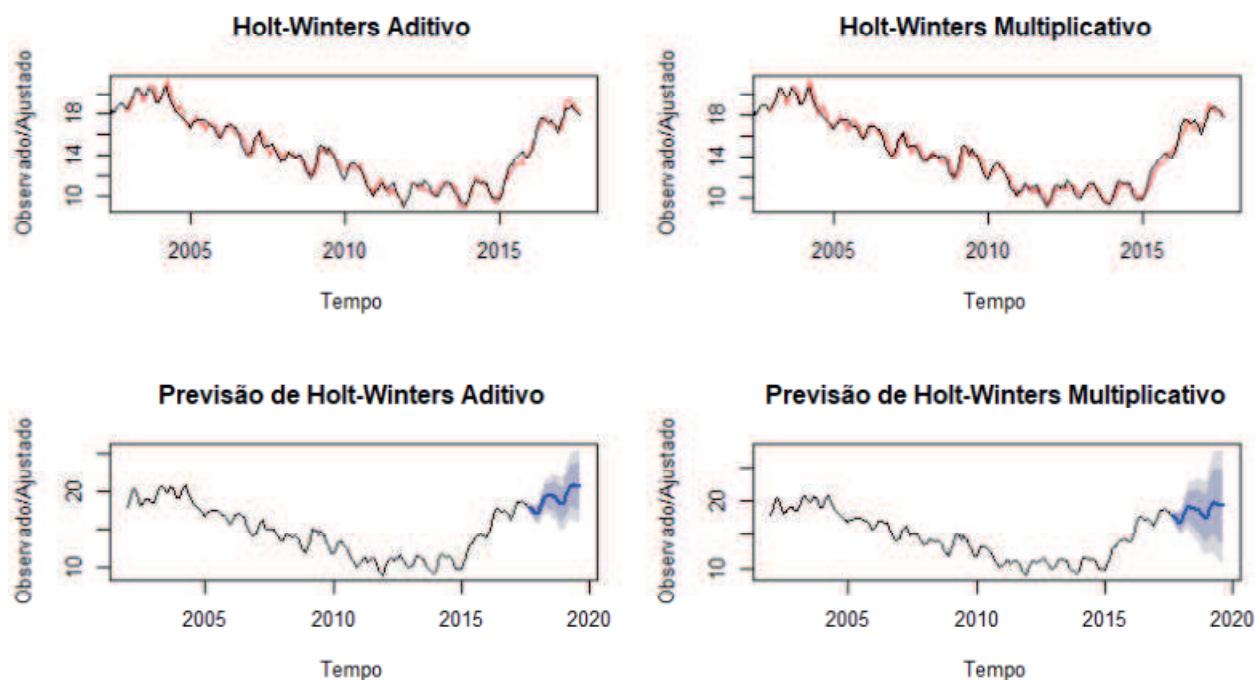
Tabela 3 – Previsões intervalares do algoritmo Holt-Winters aditivo.

Mês	HW Aditivo	HW Multiplicativo
Setembro 2017	17,95396	17,46555
Outubro 2017	17,36332	16,91619
Novembro 2017	16,85944	16,89931
Dezembro 2017	16,45919	16,96663
Janeiro 2018	16,93800	17,77501
Fevereiro 2018	17,63974	18,94592
Março 2018	18,57591	20,10309

Na Tabela 3, pode-se perceber que em novembro e dezembro, a previsão das taxas de desemprego apresentaram os menores valores, isto pode ser devido ao grande número de emprego que é ofertado nos últimos meses do ano (meses de festas). Observa-se também, que nos dois primeiros meses, o método de alisamento aditivo apresenta valores superiores ao multiplicativo, porém, a partir do terceiro mês, o método de alisamento multiplicativo mostra valores superiores ao aditivo.

A Figura 5 apresenta uma melhor explicação acerca da previsão de Holt-Winters (aditivo e multiplicativo) graficamente, onde pode-se observar a série e suas previsões com o intervalo de confiança.

Figura 5 – Gráfico de previsão de Holt-Winters aditivo e multiplicativo da série Taxa de Desemprego em SP (7 passos à frente).



De acordo com o pacote *forecast*, no software R (R Core Team, 2017), pode-se utilizar a função *accuracy* que nos fornece o teste de precisão da previsão ajustada com base nos dados da série (esta função só produz estas medidas com base na série), definidas em (HYNDMAN; KHANDAKAR, 2007). Através da Tabela 4, observa-se que o teste de previsão resultou em erros baixos, indicando uma boa capacidade de previsão para o modelo.

Tabela 4 – Valores dos erros de previsão de acordo com a série.

ME	RMSE	MAE	MPE	MAPE	MASE
-0,007106991	0,3343627	0,2694853	-0,03625529	1,904181	0,1971298

### 3.2.2 Modelagem Box-Jenkins - SARIMA

Utilizou-se da modelagem Box-Jenkins, para a verificação da melhor previsão através dos modelos SARIMA e SARIMAX. O SARIMA é composto de uma parte não sazonal, cujo parâmetros são  $(p, d, q)$  e uma sazonal com parâmetros  $(P, D, Q)$ . Ressalta-se que, aqui, também observa-se o comportamento da FAC e da FACP, porém, observam-se as defasagens sazonais, onde para as séries mensais, por exemplo, temos defasagens de 12, 24, 36, entre outras.

Para este trabalho, a modelagem SARIMA foi utilizada com o propósito de encontrar uma previsão adequada para a taxa de desemprego na região metropolitana de São Paulo de Outubro/2017 à Março/2018, como foi feito no tópico anterior. Através de Hyndman e Khandakar (2007), encontra-se o embasamento para utilização da função *auto.arima*, que faz parte do pacote *forecast*, no software R, para encontrar um modelo SARIMA mais adequado para o nosso tipo de situação, com a melhor capacidade preditiva.

Com o intuito de analisar a capacidade preditiva, retiramos o último valor da série e alocamos a um objeto separado, para modelarmos e observamos o erro desta predição comparando o valor encontrado com o observado. Após a utilização da função citada acima, foi encontrado o modelo  $SARIMA(2, 1, 2) \times (1, 0, 0)_{12}$ , onde encontramos uma previsão de, aproximadamente, 18,1% para a taxa de desemprego na região metropolitana de São Paulo, em que o valor real é de 17,8%, com um erro de previsão de  $-1,69\%$ , tem-se um valor predito acima do esperado, porém, aproximado.

Analisando pesquisas anteriores para o mesmo tipo de modelagem, foi encontrada uma pesquisa referente a taxa de desemprego na região metropolitana de Porto Alegre, em que foi utilizado um modelo  $SARIMA(0, 1, 3) \times (2, 1, 0)_{12}$ , proposto por (BECKER, 2010). Para uma melhor análise, utilizamos este modelo base como objeto para comparação entre as previsões. Através deste, obteve-se uma previsão de 18,05%, com um erro de previsão de  $-1,39\%$ , que se aproxima bastante ao que encontramos para nossos dados.

Sabe-se que com a prolongação da crise econômica mundial, o governo precisou adotar medidas com o intuito de estimular a produção e o consumo, como por exemplo, a redução dos impostos, liberação de crédito subsidiado e desonerações fiscais. A partir destas medidas, o governo federal passou a gastar cada vez mais e a arrecadação diminuiu, o que contribuiu diretamente para desequilibrar as contas públicas. Com o aumento da dívida, a ex presidente Dilma Rousseff começou, em 2015, o seu segundo mandato adotando o ajuste fiscal, visando equilibrar o orçamento do governo, o que, de modo geral, é o corte no orçamento, restrição de benefícios e aumento dos impostos e tributos. Este, causou um impacto amplo na economia, bem como, no seu governo, onde suas medidas acabaram gerando um efeito contrário, afinal, quando empresas fecham ou diminuem suas produções e vendas, menor será a contribuição para a receita federal. É importante salientar, que mesmo após o impeachment da ex presidente, ainda não foi apresentada uma solução acerca deste problema.

Sendo assim, aplicou-se uma variável dummy, com valores de 1 a partir de janeiro de 2015 até 2017, e 0 para os meses anteriores, a fim de prever diferenciando os acontecimentos marcantes na economia brasileira. Agora, avaliando novamente o modelo  $SARIMA(0, 1, 3) \times (2, 1, 0)_{12}$  e o *modelo base*, com o acréscimo da variável dummy, que resultará em modelos SARIMAX, calculou-se novamente as previsões.

Para o caso do  $SARIMAX(2, 1, 2) \times (1, 0, 0)_{12}$ , obteve-se uma previsão de 18,12%

e um erro de previsão de  $-1,82\%$ . Já para o  $SARIMA(0, 1, 3) \times (2, 1, 0)_{12}$  (*modelo base*), observou-se uma previsão de  $18,10\%$  com um erro de previsão de  $-1,69\%$ . Comparando os dois modelos, nota-se que existe uma diferença mínima entre eles. Para uma melhor análise, calculou-se os critérios de informação **AIC** (*Critério de Informação Akaike*) e o **BIC** (*Critério de Informação Bayesiano*), que se baseiam no **SQE** (*Soma dos Quadrados dos Erros*). Vale lembrar que, como EMILIANO et al. (2010) se posiciona:

"Ao selecionarmos modelos é preciso ter em mente que não existem modelos verdadeiros. Há apenas modelos aproximados da realidade que, causam perda de informações. Deste modo, é necessário fazer a seleção do "melhor" modelo, dentre aqueles que foram ajustados, para explicar o fenômeno sob estudo."

A partir da Tabela 5, observa-se os valores referentes aos critérios AIC e BIC, bem como, os erros relativos dos modelos:

Tabela 5 – Valores referentes ao critério de informação dos modelos da série e erros relativos, 1 passo à frente.

ÍNDICE	MODELO	AIC	BIC	Erro de Previsão (%)
1	$SARIMA(2, 1, 2) \times (1, 0, 0)_{12}$	129,19	148,58	-1,69
2	$SARIMA(0, 1, 3) \times (2, 1, 0)_{12}$ ( <i>base</i> )	123,09	142,08	-1,39
3	$SARIMAX(2, 1, 2) \times (1, 0, 0)_{12}$	123,57	146,19	-1,82
4	$SARIMAX(0, 1, 3) \times (2, 1, 0)_{12}$ ( <i>base</i> )	119,41	141,57	-1,69

Para escolhermos com base nos critérios de informação, o modelo que apresenta o melhor ajuste é o que apresenta os menores valores do **AIC** e **BIC**. Através da Tabela 5, nota-se que o último modelo (o modelo 4) apresenta um melhor ajuste, pois o mesmo apresenta os menores valores para os critérios de informação.

Após esta análise, iremos refazer todas as análises, agora com o intuito de prever com 7 passos à frente, ou seja, de Setembro/2017 até Março/2018. De forma análoga, utilizaremos o modelo previsto no software R, através da função *auto.arima*, bem como, o modelo base, decorrente de estudos anteriores, na busca do modelo que forneça uma melhor previsão.

Tabela 6 – Previsões dos modelos SARIMA e SARIMAX estudados, 7 passos à frente, de Setembro/17 à Dezembro/17.

Modelos	Setembro/17	Outubro/17	Novembro/17	Dezembro/17
$SARIMA(2, 1, 2) \times (1, 0, 0)_{12}$	18,10	17,87	17,68	17,37
$SARIMA(0, 1, 3) \times (2, 1, 0)_{12}$ ( <i>base</i> )	18,05	17,79	17,64	17,34
$SARIMAX(2, 1, 2) \times (1, 0, 0)_{12}$	18,12	17,90	17,69	17,38
$SARIMAX(0, 1, 3) \times (2, 1, 0)_{12}$ ( <i>base</i> )	18,10	17,89	17,60	17,29

Sabendo dos valores referentes aos critérios de informação, dados na Tabela 4, agora temos os valores das previsões, com 7 passos à frente, dispostos na Tabela 7. Com



Tabela 7 – Previsões dos modelos SARIMA e SARIMAX estudados, 7 passos à frente.

Modelos	Janeiro/18	Fevereiro/18	Março/18
$SARIMA(2, 1, 2) \times (1, 0, 0)_{12}$	17,84	18,26	18,58
$SARIMA(0, 1, 3) \times (2, 1, 0)_{12}$ ( <i>base</i> )	17,71	18,45	19,34
$SARIMAX(2, 1, 2) \times (1, 0, 0)_{12}$	17,86	18,29	18,62
$SARIMAX(0, 1, 3) \times (2, 1, 0)_{12}$ ( <i>base</i> )	17,68	18,42	19,31

esta tabela, pode-se observar de uma maneira mais ampla o que foi descrito acima, para as previsões do mês de Setembro/17, onde os valores preditos estão bem próximos. Através dos critérios de informação, os modelos com a variável dummy, apresentaram um melhor comportamento, o que já era esperado, dado que no próprio gráfico da série temporal nota-se uma quebra estrutural a partir de, aproximadamente, 2015, onde observa-se um aumento da taxa, visto na Figura 1.

## 4 Conclusão

Para a tomada de decisões das políticas públicas acerca do mercado de trabalho, é necessário um estudo aprofundado sobre a taxa de desemprego, para que sejam definidos meios que tragam uma melhoria na vida da população. Este estudo veio com o propósito de buscar modelos que possibilitem uma boa previsão futura para a taxa de desemprego, especialmente na maior metrópole nacional, afinal, sabe-se que este problema abrange não só a área econômica, mas a social.

Para tais previsões, foram utilizados os algoritmos de Holt-Winters (aditivo e multiplicativo), os modelos SARIMA e SARIMAX, onde, neste último, foi acrescentado uma variável dummy, que assumiu valores 1 para Janeiro de 2015 (referente ao segundo mandato da ex presidente Dilma) ao último mês da série, e 0 para os valores antecedentes. Com base em estudos anteriores, foi acrescentado um modelo base (com e sem a variável dummy), para fazer a comparação e buscar uma melhor previsão (BECKER, 2010).

De modo geral, os modelos apresentaram previsões aproximadas. Através dos erros, verificou-se que o Holt-Winters Aditivo, o SARIMA (base) e o SARIMAX (base), apresentaram previsões um pouco melhores. Porém, de acordo com os critérios de informação (AIC e BIC), o modelo  $SARIMAX (0, 1, 3) \times (2, 1, 0)_{12}$  (base) obteve um melhor ajuste, indicando que as medidas tomadas pelo governo da ex presidente Dilma, em sua reeleição no ano de 2015, afetaram a taxa de desemprego na região metropolitana de São Paulo.

Como proposta futura, através do modelo proposto por Becker (2010), podemos abranger este estudo acerca da taxa de desemprego para diversas regiões brasileiras, possibilitando estabelecer um comparativo que permita averiguar o comportamento desta taxa, bem como, onde situa-se as maiores taxas de desemprego no Brasil.

# Referências

- BECKER, M. H. Modelos para previsão em séries temporais: uma aplicação para a taxa de desemprego na Região Metropolitana de Porto Alegre. 2010. Citado 3 vezes nas páginas 13, 37 e 40.
- BOX, G. E. P. et al. *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015. ISSN 1118674928. Citado 6 vezes nas páginas 14, 16, 21, 22, 23 e 29.
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to time series and forecasting*. [S.l.]: springer, 2016. ISSN 3319298542. Citado 2 vezes nas páginas 17 e 29.
- CHATFIELD, C. *The analysis of time series: an introduction*. [S.l.]: CRC press, 2016. ISSN 0203491688. Citado na página 18.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, Taylor & Francis, v. 74, n. 366a, p. 427–431, 1979. ISSN 0162-1459. Citado na página 27.
- DICKEY, D. A.; PANTULA, S. G. Determining the order of differencing in autoregressive processes. *Journal of Business & Economic Statistics*, Taylor & Francis Group, v. 5, n. 4, p. 455–461, 1987. ISSN 0735-0015. Citado na página 26.
- EMILIANO, P. C. et al. Critérios de informação de akaike versus bayesiano: análise comparativa. *19º Simpósio Nacional de Probabilidade e Estatística*, 2010. Citado na página 38.
- ENDERS, W. *Applied econometric time series*. [S.l.]: John Wiley & Sons, 2008. ISSN 8126515643. Citado na página 25.
- FARIAS, J. F. Uma aplicação de séries temporais na arrecadação do ICMS do estado da Paraíba (TCC). 2017. Citado 2 vezes nas páginas 18 e 31.
- FULLER, W. A. *Introduction to statistical time series*. [S.l.]: John Wiley & Sons, 2009. v. 428. ISSN 0470317752. Citado na página 26.
- GARRATY, J. A. *Unemployment in history: Economic thought and public policy*. [S.l.]: New York: Harper & Row, 1978. ISSN 0060114576. Citado na página 12.
- HYNDMAN, R. J.; KHANDAKAR, Y. *Automatic time series for forecasting: the forecast package for R*. [S.l.]: Monash University, Department of Econometrics and Business Statistics, 2007. Citado 2 vezes nas páginas 36 e 37.
- KATO, J. M.; PONCHIROLLI, O. O desemprego no brasil e os seus desafios éticos. *Revista da FAE*, v. 5, n. 3, 2017. ISSN 2447-2735. Citado na página 12.
- LÚCIA, F. A. V. A. V. Manual de econometria. *Vasconcelos, MAS; Alves, D. São Paulo: Editora Atlas*, 2000. Citado na página 25.
- MARTÍNEZ, R. O.; ZAMPROGNO, B. Comparação de algumas técnicas de previsão em análise de séries temporais. *Revista Colombiana de Estadística*, Universidad Nacional de Colombia, v. 26, n. 2, p. 129–157, 2003. ISSN 0120-1751. Citado na página 19.

---

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. *Introduction to time series analysis and forecasting*. [S.l.]: John Wiley & Sons, 2015. ISSN 1118745116. Citado na página 19.

MORETTIN, P. A.; TOLOI, C. *Análise de séries temporais*. [S.l.]: Blucher, 2006. ISSN 8521203896. Citado 11 vezes nas páginas 14, 15, 17, 18, 19, 20, 22, 23, 24, 28 e 29.

PICCHETTI, P.; CHAHAD, J. P. A evolução da taxa de desemprego estrutural no brasil: uma análise entre regiões e características dos trabalhadores. In: . [S.l.]: ANPEC - Associação Nacional dos Centros de Pósgraduação em Economia [Brazilian Association of Graduate Programs in Economics], 2003. Citado na página 12.

/

# Apêndices

# APÊNDICE A – Previsões Intervalares de Holt-Winters

Tabela 8 – Previsões intervalares do algoritmo Holt-Winters aditivo.

mês/ano	Ponto previsão	Inferior 80%	Superior 80%	Inferior 95%	Superior 95%
Setembro 2017	17.80737	17.36909	18.24565	17.13708	18.47766
Outubro 2017	17.42502	16.79915	18.05088	16.46784	18.38220
Novembro 2017	17.24222	16.46237	18.02207	16.04955	18.43489
Dezembro 2017	16.97343	16.05575	17.89110	15.56996	18.37689
Janeiro 2018	17.21757	16.17148	18.26365	15.61772	18.81741
Fevereiro 2018	17.95217	16.78370	19.12065	16.16514	19.73920
Março 2018	18.84170	17.55486	20.12853	16.87365	20.80974
Abril 2018	19.30567	17.90325	20.70810	17.16085	21.45049
Mai 2018	19.33779	17.82169	20.85389	17.01911	21.65646
Junho 2018	19.26624	17.63777	20.89472	16.77571	21.75678
Julho 2018	19.28590	17.54590	21.02589	16.62480	21.94699
Agosto 2018	19.26642	17.41542	21.11743	16.43556	22.09729
Setembro 2018	19.17374	17.20868	21.13880	16.16845	22.17904
Outubro 2018	18.79139	16.71581	20.86698	15.61706	21.96573
Novembro 2018	18.60859	16.42235	20.79484	15.26502	21.95217
Dezembro 2018	18.33980	16.04263	20.63697	14.82658	21.85302
Janeiro 2019	18.58394	16.17548	20.99240	14.90052	22.26736
Fevereiro 2019	19.31855	16.79835	21.83875	15.46423	23.17286
Março 2019	20.20807	17.57560	22.84054	16.18206	24.23408
Abril 2019	20.67205	17.92673	23.41737	16.47344	24.87065
Mai 2019	20.70416	17.84535	23.56297	16.33199	25.07633
Junho 2019	20.63262	17.65965	23.60559	16.08585	25.17938
Julho 2019	20.65227	17.56443	23.74012	15.92982	25.37472
Agosto 2019	20.63280	17.42933	23.83626	15.73352	25.53207

Tabela 9 – Previsões intervalares do algoritmo Holt-Winters multiplicativo.

mês/ano	Ponto previsão	Inferior 80%	Superior 80%	Inferior 95%	Superior 95%
Setembro 2017	17.78284	17.19004	18.37565	16.87623	18.68946
Outubro 2017	17.34951	16.51471	18.18430	16.07280	18.62622
Novembro 2017	17.18407	16.13860	18.22954	15.58516	18.78298
Dezembro 2017	16.83596	15.61285	18.05908	14.96537	18.70656
Janeiro 2018	17.12415	15.68617	18.56213	14.92495	19.32335
Fevereiro 2018	17.92448	16.22128	19.62768	15.31966	20.52931
Março 2018	18.88192	16.88222	20.88161	15.82364	21.94019
Abril 2018	19.19149	16.95194	21.43103	15.76640	22.61658
Mai 2018	19.02220	16.59816	21.44623	15.31495	22.72944
Junho 2018	18.77748	16.18347	21.37149	14.81028	22.74468
Julho 2018	18.73949	15.95009	21.52890	14.47347	23.00552
Agosto 2018	18.62571	15.65372	21.59770	14.08044	23.17097
Setembro 2018	18.48709	15.33097	21.64321	13.66023	23.31396
Outubro 2018	18.02037	14.75075	21.28999	13.01992	23.02082
Novembro 2018	17.83290	14.40568	21.26013	12.59141	23.07439
Dezembro 2018	17.45674	13.91377	20.99971	12.03823	22.87525
Janeiro 2019	17.74080	13.94860	21.53300	11.94113	23.54047
Fevereiro 2019	18.55492	14.38783	22.72201	12.18190	24.92794
Março 2019	19.53061	14.93245	24.12878	12.49833	26.56290
Abril 2019	19.83556	14.94984	24.72129	12.36349	27.30764
Mai 2019	19.64587	14.59267	24.69906	11.91768	27.37406
Junho 2019	19.37897	14.18269	24.57525	11.43195	27.32599
Julho 2019	19.32600	13.93234	24.71967	11.07711	27.57490
Agosto 2019	19.19533	13.62757	24.76309	10.68018	27.71048

# APÊNDICE B – Rotina do R

```
#####
#      AUTORA: Nayara Fernanda da Silva Lima      #
#      FUNÇÃO: Modelar e fazer previsões para a   #
#              taxa de desemprego em SP, 1 e 7    #
#              passos à frente.                  #
#              Trabalho de Conclusão de Curso    #
#              Orientador: Prof. Dr. Sílvio Fernando #
#####

# Pacotes
require(TSA)
require(tseries)
require(urca)
require(stats)
require(forecast)
require(nortest)

# Leitura dos dados
taxa <- read.table("C:/Users/Nayara/Desktop/Series Temporais/desemprego2.txt")
# Verificar o tamanho da série
lengths(taxa)

# Formatação dos dados para uma série temporal
desemprego <- ts(taxa,start = c(2002,1),frequency=12)

# Gráfico dos dados
par(mfrow = c(1,1))
plot.ts(desemprego, ylab="Taxa",xlab="Tempo",
main="Série da taxa de desemprego em SP",xaxp=c(2002,2017,5))

par(mfrow = c(1,2))
# Correlograma
acf(desemprego,main="Correlograma",lag.max = 70)
# Correlograma parcial
pacf(desemprego,main="Correlograma Parcial",lag.max = 90)

par(mfrow = c(1,2))
# Correlograma da primeira diferença
acf(diff(desemprego),main="Correlograma da primeira diferença",lag.max = 90)
# Correlograma parcial da primeira diferença
pacf(diff(desemprego),main="Correlograma Parcial da primeira diferença",lag.max = 90)

summary(desemprego)
```



```
sqrt(var(desemprego))
summary(diff(log(desemprego))*100)

# Aplicação da transformação na série
desemprego1 = log(desemprego)

# Teste de Dickey-Fuller - Verificar Estacionariedade
# p-value > 0.05 -> não rejeita H0: Existe pelo menos uma raiz dentro do círculo unitário
adf.test(desemprego1, alternative = "stationary")

#####
# Alisamento Exponencial Holt-Winters Aditivo e Multiplicativo #
#####

# Previsão 1 passo à frente (sem a última observação)
taxa1p <- read.table("C:/Users/Nayara/Desktop/Series Temporais/desemprego21p.txt")
desemprego1p <- ts(taxa1p,start=c(2002,1),frequency = 12)

# medidas de precisão das previsões
etsfit1 <- ets(desemprego1p)
accuracy(etsfit1)

# Previsão 7 passos à frente (Aditivo e Multiplicativo)
taxa7pp <- read.table("C:/Users/Nayara/Desktop/Series Temporais/desemprego21p.txt")
desemprego7pp <- ts(taxa7pp,start=c(2002,1),frequency = 12)

# Previsão do Holt-Winters Aditivo
hwa7p2 <- HoltWinters(desemprego7pp, seasonal = "additive")
predict(hwa7p2,7)

#Previsão do Holt-Winters Multiplicativo
hwm7p2 <- HoltWinters(desemprego7pp, seasonal = "multiplicative")
predict(hwm7p2,7)

# dados sem tendência
hwa7p2p <- hw(desemprego7pp)
# dados com tendência
hwm7p2p <- hw(desemprego7pp, damped = TRUE,seasonal = "multiplicative")

# Gráficos
par(mfrow=c(2,2))
plot(hwa7p2, xlab = "Tempo", ylab = "Observado/Ajustado",
main = "Holt-Winters Aditivo")
plot(hwm7p2, xlab = "Tempo", ylab = "Observado/Ajustado",
main = "Holt-Winters Multiplicativo")
plot(hwa7p2p, xlab = "Tempo", ylab = "Observado/Ajustado",
main = "Previsão de Holt-Winters Aditivo")
```

```

plot(hwm7p2p, xlab = "Tempo", ylab = "Observado/Ajustado",
main = "Previsão de Holt-Winters Multiplicativo")

#####
##          Modelagem SARIMA          ##
##          1 passo à frente          ##
#####

# 1 passo à frente
auto.arima(desemprego1p)
arimafit1 <- arima(desemprego1p, order = c(2,1,2), seasonal = list(order=c(1,0,0)))
predict(arimafit1,1)$pred
observado <- 17.8
(observado - (predict(arimafit1,1)$pred))/observado*100

AIC(arimafit1)
BIC(arimafit1)
tsdiag(arimafit1)

# Comparação com o modelo base (0,1,3)(2,1,0)
arimafit1c <- arima(desemprego1p, order = c(0,1,3), seasonal = list(order=c(2,1,0)))
predict(arimafit1c,1)$pred
(observado - (predict(arimafit1c,1)$pred))/observado*100

AIC(arimafit1c)
BIC(arimafit1c)
tsdiag(arimafit1c)

#####
##          Modelagem SARIMAX          ##
##          1 passo à frente          ##
#####

# Acrescentando a variável Dummy (com valores 1 a partir de janeiro de 2015)
dummy1p <- rep(0,length(desemprego1p))
dummy1p[(length(desemprego1p)-33):length(desemprego1p)]=1

# modelo sarimax
arimaxfit1 <- arima(desemprego1p, order = c(2,1,2),
seasonal = list(order=c(1,0,0)), xreg=dummy1p)
predict(arimaxfit1, 1, newxreg = 1)$pred
(observado - predict(arimaxfit1, 1, newxreg = 1)$pred)/observado*100

AIC(arimaxfit1)
BIC(arimaxfit1)

# modelo sarimax base

```

```

arimaxfit2 <- arima(desemprego1p, order = c(0,1,3),
seasonal = list(order=c(2,1,0)), xreg=dummy1p)
predict(arimaxfit2, 1, newxreg = 1)$pred
(observado - predict(arimaxfit2, 1, newxreg = 1)$pred)/observado*100

AIC(arimaxfit2)
BIC(arimaxfit2)

#####
##          Modelagem SARIMA          ##
##          7 passos à frente          ##
#####

# 7 passos à frente
arimafit2 <- arima(desemprego1p, order = c(2,1,2), seasonal = list(order=c(1,0,0)))
predict(arimafit2,7)$pred

AIC(arimafit2)
BIC(arimafit2)
tsdiag(arimafit2)

# Comparação com o modelo base (0,1,3)(2,1,0)
arimafit2c <- arima(desemprego1p, order = c(0,1,3), seasonal = list(order=c(2,1,0)))
predict(arimafit2c,7)$pred

AIC(arimafit2c)
BIC(arimafit2c)
tsdiag(arimafit2c)

#####
##          Modelagem SARIMAX          ##
##          7 passos à frente          ##
#####

# Acrescentando a variável Dummy (com valores 1 a partir de janeiro de 2015)
dummy2p <- rep(0,length(desemprego1p))
dummy2p[(length(desemprego1p)-33):length(desemprego1p)]=1

# modelo sarimax
arimaxfit2 <- arima(desemprego1p, order = c(2,1,2),
seasonal = list(order=c(1,0,0)), xreg=dummy2p)
predict(arimaxfit2, 7, newxreg = 1)$pred

AIC(arimaxfit2)
BIC(arimaxfit2)
tsdiag(arimaxfit2)

```

---

```
# modelo sarimax base
arimaxfit2c <- arima(desemprego1p, order = c(0,1,3),
seasonal = list(order=c(2,1,0)), xreg=dummy2p)
predict(arimaxfit2c, 7, newxreg = 1)$pred

AIC(arimaxfit2c)
BIC(arimaxfit2c)
tsdiag(arimaxfit2c)
```

# Anexos

# ANEXO A – Conjunto de Dados

Tabela 10 – Taxa de desemprego em São Paulo, entre 2002 e 2017, em porcentagem

Mês	Anos															
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Janeiro	17,9	18,6	19,1	16,7	15,7	14,4	13,6	12,5	11,8	10,5	9,6	10,0	9,6	9,8	14,0	17,1
Fevereiro	19,1	19,1	19,8	17,1	16,3	15,3	13,6	13,5	12,2	10,6	10,4	10,3	10,6	10,5	14,7	17,9
Março	19,9	19,7	20,6	17,3	16,9	15,9	14,3	14,9	13,1	11,3	11,1	10,9	11,5	11,4	15,9	18,5
Abril	20,4	20,6	20,7	17,5	16,9	16,3	14,2	15,0	13,3	11,2	11,2	11,4	11,6	12,4	16,8	18,6
Mai	19,7	20,6	19,7	17,5	17,0	15,5	14,1	14,8	13,3	10,7	10,9	11,4	11,4	12,9	17,6	18,8
Junho	18,8	20,3	19,1	17,5	16,8	14,9	13,9	14,2	12,9	11,0	11,2	11,3	11,3	13,2	17,6	18,6
Julho	18,1	19,7	18,5	17,5	16,7	15,0	14,1	14,8	12,6	11,1	11,1	11,0	11,4	13,7	17,4	18,3
Agosto	18,3	20,0	18,3	17,1	16,0	15,0	14,0	14,2	12,3	11,2	11,6	10,4	11,3	13,9	17,2	17,9
Setembro	18,9	20,6	17,9	16,9	15,3	15,1	13,5	14,1	11,5	10,6	11,3	10,0	10,6	14,2	17,5	17,8
Outubro	19,0	20,4	17,6	16,9	14,6	14,4	12,5	13,2	10,9	9,9	10,9	9,6	9,87	14,3	17,2	–
Novembro	19,0	19,9	17,4	16,4	14,1	14,2	12,3	12,8	10,7	9,5	10,3	9,4	9,85	14,1	16,8	–
Dezembro	18,5	19,1	17,1	15,8	14,2	13,5	11,8	11,9	10,1	9,0	10,0	9,3	9,9	13,9	16,2	–

Fonte: Fundação Sistema Estadual de Análise de Dados, Pesquisa de Emprego e Desemprego (Seade/PED)