



**UNIVERSIDADE FEDERAL DE SERGIPE**  
**CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA**  
**DEPARTAMENTO DE ESTATÍSTICA E CIÊNCIAS ATUARIAIS**



**JUAREZ DOS SANTOS ALVES**

**ANÁLISE DE SÉRIES TEMPORAIS APLICADA ÀS TAXAS DE  
DESEMPREGO NO BRASIL E EM SERGIPE NO PERÍODO DE 2012 A 2024**

**São Cristóvão – SE**

**2025**

**JUAREZ DOS SANTOS ALVES**

**ANÁLISE DE SÉRIES TEMPORAIS APLICADA ÀS TAXAS DE  
DESEMPREGO NO BRASIL E EM SERGIPE NO PERÍODO DE 2012 A 2024**

**Trabalho de Conclusão de Curso apresentado ao  
Departamento de Estatística e Ciências Atuariais da  
Universidade Federal de Sergipe, como parte dos  
requisitos para obtenção do grau de Bacharel em  
Ciências Atuariais.**

**Orientador: Prof. Dr. Cleber Martins Xavier**

**São Cristóvão – SE**

**2025**

**JUAREZ DOS SANTOS ALVES**

**ANÁLISE DE SERIES TEMPORAIS APLICADA ÀS TAXAS DE  
DESEMPREGO NO BRASIL E EM SERGIPE NO PERÍODO DE 2012 A 2024**

**Trabalho de Conclusão de Curso apresentado ao  
Departamento de Estatística e Ciências Atuariais da  
Universidade Federal de Sergipe, como um dos pré-  
requisitos para obtenção do grau de Bacharel em  
Ciências Atuariais.**

**Aprovado em 04 / 04 / 2025**

**Banca Examinadora:**

---

**Prof. Dr. Cleber Martins Xavier  
Orientador**

---

**Prof. Dr. Carlos Raphael Araújo Daniel**

---

**Prof. Dr. Jose Rodrigo Santos Silva**

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus, e segundo a minha família e amigos. A minha esposa Maria Aparecida Ramos Da Silva, que sempre me deu força durante todo o curso estando ao meu lado, obrigado por passar inúmeras noites tentando me esperar acordada. A minha mãe Maria Jose Dos Santos e meu pai Juarez De Deus Alves, que sempre me incentivaram a estudar, obrigado e desculpe a demora para entender a necessidade de cursar o nível superior. Aos meus amigos que conheci durante o curso, não vou citar o nome porque são muitos e acabaria gerando confusão, mas muito obrigado pessoal vocês são muito importantes para mim, cada dificuldade que passamos e cada momento de superação, todos eles estão gravados em minha memória.

Agradeço também a todos os docentes do Departamento de Estatística e Ciências Atuariais, por toda a dedicação e conhecimento passado durante todo o curso, em especial ao professor Cleber, meu muito obrigado pela orientação e pela paciência, obrigado por ter ido até o fim, apesar do trabalho que dei, meus mais sinceros agradecimentos.

E por fim dedico esse trabalho a memória do meu padrasto Paulo Sergio Dos Santos, o senhor foi um grande pai, obrigado por apoiar-me sempre, obrigado pelos conselhos, obrigado pelo carinho e afeto. O senhor sempre estará em minha memória.

## RESUMO

O presente trabalho aborda o estudo das séries temporais das taxas de desemprego no Brasil e em Sergipe no período de 2012 a 2024, fazendo uso de modelagem estatística pelo método Box e Jenkins, através da linguagem de programação estatística do R versão 4.4.2. Para selecionar o modelo adequado, foi testado a estacionariedade das series pelos testes ADF e KPSS, e utilizado o critério AIC. Para a série do Brasil foi ajustado um modelo ARIMA (0,2,3)(2,0,0)<sub>12</sub> e para a série de Sergipe um modelo ARIMA (0,1,1). Após a seleção do modelo foi realizada análise dos resíduos, em que para a série do Brasil o teste Ljung-Box apresentou p-valor de 0,0854 indicando que a série não apresentou correlações significativas nos resíduos e o teste Shapiro-Wilk com p-valor de 0,03148 indicando que o ajuste não segue distribuição normal; para a série de Sergipe o teste Ljung-Box apresentou um p-valor de 0,766 indicando que não há correlação significativa dos resíduos e o teste Shapiro-Wilk apresentou um p-valor de 0,05609 indicando que o ajuste do modelo segue uma distribuição normal. Por fim foi realizado as previsões, para a série do Brasil foram feitas previsões para seis e doze meses e para a série de Sergipe foi feita previsão para seis trimestres.

**Palavras-chave:** taxa de desemprego, serie temporal, previsões, estacionariedade, crise.

## ABSTRACT

This paper addresses the study of historical time series of unemployment rates in Brazil and Sergipe from 2012 to 2024, using statistical modeling by the Box and Jenkins method, through the statistical programming language R version 4.4.2. To select the appropriate model, the stationarity of the series was tested by the ADF and KPSS tests, and the AIC criterion was used. For the Brazil series, an ARIMA model  $(0,2,3)(2,0,0)_{12}$  was adjusted, and for the Sergipe series, an ARIMA model  $(0,1,1)$ . After selecting the model, an analysis of the residuals was performed, in which for the Brazil series, the Ljung Box test presented a p-value of 0.0854, indicating that the series did not present significant correlations in the residuals, and the Shapiro-Wilk test with a p-value of 0.03148, indicating that the adjustment does not follow a normal distribution; for the Sergipe series, the Ljung Box test presented a p-value of 0.766, indicating that there is no significant correlation of the residuals, and the Shapiro-Wilk test presented a p-value of 0.05609, indicating that the adjustment of the model follows a normal distribution. Finally, the forecasts were made; for the Brazil series, forecasts were made for six and twelve months, and for the Sergipe series, forecasts were made for six quarters.

**Keywords:** unemployment rate, time series, forecasts, stationarity, crisis.

## LISTA DE ILUSTRAÇÕES

Figura 1: Média móvel de desemprego IBGE x calculada.....	15
Figura 2: Série de desemprego no Brasil de março de 2012 a novembro de 2024.....	26
Figura 3: Série original do percentual de pessoas ocupadas no Brasil de 2012 a 2024.....	27
Figura 4: Gráfico de autocorrelação da série de desemprego no Brasil.....	29
Figura 5: Gráfico de autocorrelação parcial da série de desemprego no Brasil.....	30
Figura 6: Análise da primeira diferença aplicada à série de desemprego no Brasil...	31
Figura 7: Análise dos resíduos da série de desemprego no Brasil após aplicação da diferença de segunda ordem.....	34
Figura 8: Série original de desemprego no Brasil x Modelo ajustado.....	35
Figura 9: Previsão do desemprego no Brasil para 12 meses.....	36
Figura 10: Série original de desemprego em Sergipe.....	37
Figura 11: Série de desemprego Brasil x Sergipe.....	37
Figura 12: Análise da ACF e PACF da série de desemprego em Sergipe.....	39
Figura 13: Análise da ACF e PACF da série de desemprego em Sergipe após aplicação da primeira diferença.....	40
Figura 14: Análise dos resíduos do modelo aplicado a série de desemprego de Sergipe.....	42
Figura 15: Série original de desemprego em Sergipe x Modelo ajustado.....	43
Figura 16: Previsão do desemprego em Sergipe para 4 trimestres.....	44

## LISTA DE TABELAS

Tabela 1: Possíveis resultados combinados ADF test e KPSS test.....	21
Tabela 2: Teste de estacionariedade da série de desemprego no Brasil.....	28
Tabela 3: Modelos testados na série de desemprego no Brasil.....	32
Tabela 4: Resultado do teste dos modelos testados na série de desemprego no Brasil.....	33
Tabela 5: Resultado do teste de estacionariedade da série de desemprego em Sergipe.....	38
Tabela 6: Lista de modelos testados na série de desemprego de Sergipe com diferenciação.....	41
Tabela 7: Resultado do teste aplicado aos modelos da série de desemprego de Sergipe.....	41

## SUMARIO

<b>1 INTRODUÇÃO.....</b>	<b>09</b>
<b>2 OBJETIVOS.....</b>	<b>11</b>
2.1 Objetivo Geral.....	11
2.2 Objetivos Específicos.....	11
<b>3 JUSTIFICATIVA.....</b>	<b>12</b>
<b>4 REFERENCIAL TEÓRICO.....</b>	<b>13</b>
4.1 Definição de desemprego.....	13
4.2 Método de coleta, origem dos dados e cálculo das taxas.....	13
4.3 Histórico de desemprego no Brasil.....	15
4.4 Trabalhos correlatos.....	16
<b>5 METODOLOGIA.....</b>	<b>18</b>
5.1 Base de dados.....	18
5.2 Séries temporais.....	18
5.3 Modelos SARIMA.....	21
5.4 Suporte computacional.....	24
<b>6 RESULTADOS E DISCUSSÃO.....</b>	<b>26</b>
<b>6.1 Análise da série original de desemprego no Brasil.....</b>	<b>26</b>
6.1.1 Teste de estacionariedade da série original de desemprego no Brasil.....	28
6.1.2 Aplicando diferenças na série original de desemprego no Brasil.....	30
6.1.3 Verificação e aplicação do modelo a série original de desemprego no Brasil.....	31
6.1.4 Previsão de taxas de desemprego futuras no Brasil.....	35
<b>6.2 Análise da série de desemprego em Sergipe.....</b>	<b>36</b>
6.2.1 Teste de estacionariedade da série de desemprego em Sergipe.....	38
6.2.2 Aplicação e verificação do modelo a série de desemprego em Sergipe.....	40
6.2.3 Previsão de taxas futuras para Sergipe.....	44
<b>7 CONCLUSÃO.....</b>	<b>45</b>
<b>REFERÊNCIAS.....</b>	<b>46</b>
<b>APRÊNDICE A- Script no R.....</b>	<b>50</b>

## 1 INTRODUÇÃO

O desemprego é um problema que afeta todas as regiões do Brasil, principalmente em épocas de crises, tendo em vista que o número de postos de trabalho está diretamente ligado a economia do país. Segundo Barros, Camargo e Mendonça (1997), a qualidade das vagas de trabalho geradas no país é o principal fator do desemprego, e não a quantidade de vagas ofertadas, o autor também destaca que outra grande problemática é o grau de desigualdade existente no Brasil.

Os dados que mensuram o nível de desemprego a nível nacional, é calculado mensalmente. Segundo Moretto e Proni (2011), as estatísticas relacionadas ao desemprego no país, são todas baseadas em dados do IBGE (Instituto Brasileiro de Geografia e Estatística), que de modo geral, está atrelada ao número de pessoas procurando emprego, uma vez que, para ser considerado desempregado, o cidadão ou cidadã deve estar à procura de trabalho.

Como o mercado de trabalho está diretamente ligado à economia do país, ele está sujeito a grandes efeitos advindos de eventos que impactem na economia do país. Um dos eventos que impactam negativamente as taxas de desemprego são as crises econômicas, a exemplo a crise econômica internacional, que teve impacto na produção de países subdesenvolvidos e desenvolvidos, e como consequência, reduz drasticamente o número de postos de trabalho, acarretando o aumento das taxas de desemprego (MORETTO e PRONI, 2011).

A recessão de 2014 surtiu efeitos negativos em todo o Brasil, pois há 35 anos o Brasil não registrava taxas de desemprego tão elevadas. Em pesquisa realizada pela PME (Pesquisa Mensal de Emprego e Desemprego) realizada pelo IBGE, buscando demonstrar o comportamento da taxa de pessoas desocupadas em idade de trabalhar, onde foi registrado no mês de julho de 2015, um aumento de 41,5% nesse percentual, em relação ao mês de janeiro do mesmo ano; uma vez que em pesquisas realizadas nos anos de 2014, 2013 e 2012, para os mesmos meses, as taxas registradas sucessivamente, 2,1%, 3,7%, e 1,8% (POCHMANN, 2015).

A presença do desemprego em um país, é uma grande barreira para o setor econômico, sendo alvo de estudo de muitos economistas, na busca de soluções para tal cenário. Para Meneses e Cunha (2013), o valor do salário é um dos fatores mais impactantes, para que uma pessoa saia da condição de desempregado, tendo em

vista que para muitos desocupados o salário ofertado deve suprir seus gastos, caso não seja, torna-se inviável o aceite ao cargo.

Segundo Costa (2020), o Brasil enfrenta um grande desafio para elaborar estratégias que impulsionem a economia do país e alavanquem o mercado de trabalho, sendo evidente a insuficiência dos recursos emergenciais do país, uma vez que durante a recessão de 2014 foi comprovada a deficiência do Brasil, quando ficou à mercê da economia externa. Ainda segundo a autora, o setor de saúde é o que mais preocupou foi o cenário da pandemia, que tornou precária a situação dos empregados que tiveram seus salários reduzidos, com risco de demissão e desempregados, que tiveram uma drástica redução na pretensão por um cargo, com um futuro muito incerto.

Nesse estudo é abordado o comportamento das taxas mensais de desemprego no Brasil, com início em março de 2012 a novembro 2024, com 153 observações; para os mesmos anos também é estudado o comportamento das taxas trimestrais de desemprego de Sergipe, iniciando no primeiro trimestre de 2012 ao terceiro trimestre de 2024, com total de 53 observações. As variáveis observadas no estudo são a data da ocorrência e a taxa observada. Todos os dados da pesquisa foram coletados no IPEADATA (Instituto de Pesquisa Econômica Aplicada), e foram modelados e manipulados no Microsoft Excel e no Rstudio.

## **2 OBJETIVOS**

### **2.1 Objetivo Geral**

O presente trabalho, tem como principal objetivo analisar o comportamento do desemprego no Brasil e no estado de Sergipe, com o intuito de prever taxas futuras, fazendo uso dos modelos de análise de séries temporais.

### **2.2 Objetivos Específicos**

- Descrever o comportamento das taxas de desemprego mensais no Brasil no período de março de 2012 a novembro de 2024;
- Descrever o comportamento das taxas de desemprego trimestrais no estado de Sergipe, do primeiro trimestre de 2012 ao terceiro trimestre de 2024;
- Aplicar a metodologia de Box e Jenkins, para encontrar um modelo que melhor represente as séries de desemprego no Brasil e em Sergipe;
- Realizar previsões para a série de desemprego do Brasil, em um horizonte de 6 e 12 meses
- Realizar previsões para a série de desemprego de Sergipe, em um horizonte de 4 trimestres;

### 3 JUSTIFICATIVA

O desemprego no Brasil é um problema que perdura há anos, sendo fonte de estudo em vários artigos e trabalhos acadêmicos, a exemplo Quadros (2003), Silva (2014), Silveira (2022), Louredo et al (2021), tendo em vista os impactos negativos gerados no país como um todo e principalmente nos menores estados, a exemplo Sergipe; como o desemprego é uma variável sensível a mudanças, ele tende a elevar suas taxas sempre que ocorre mudanças na estrutura econômica da região ou no país e seus impactos trazem cenários para o estado sergipano e para o Brasil, como fechamento de empresas, taxas de juros elevadas, redução do PIB, dentre outros impactos. Uma vez que essa variável está em constante mudança, há uma necessidade de pesquisas atualizadas sobre o tema. Dessa forma o tema abordado nessa pesquisa é de suma importância, pois é possível avaliar o comportamento da série antes, durante e depois do período de crise em 2014; observando também seu comportamento em meio a pandemia da corona vírus; um outro ponto que torna o tema proposto relevante é o fato da abrangência de tempo e o nível da pesquisa, que além do país também está voltado para o estado sergipano, onde a coleta dos dados ocorre trimestralmente.

## **4 REFERENCIAL TEÓRICO**

### **4.1 Definição de desemprego**

Segundo Sauthier (2009), a definição de desemprego vem dos padrões capitalistas, originados entre o fim do século XIX e início do século XX, devido aos impactos causados pelos novos padrões mundiais, sendo originada a primeira definição de desemprego, pela Organização Internacional de Trabalho (OIT). Essa definição foi criada pela OIT, sendo termo “desemprego” visto como o contrário de trabalho, uma noção básica que evoluiu com o passar dos anos.

A categoria do desemprego é vista como um aspecto social. Segundo Gautié (1998), após o término da segunda guerra mundial, o tema desemprego tornou-se importante para a Organização para a Cooperação e o Desenvolvimento Econômico (OCDE), tornando-se um problema social para os governos, devido à evolução econômica na época, associando o desemprego à falta de trabalho ligada à relação salarial, onde o indivíduo passou a ter direito ao emprego formal, participando socialmente e juridicamente no ativo de produção das riquezas do país, dessa forma aquele que não assegurou uma vaga nesse novo mercado de trabalho, foi classificado como desempregado.

Segundo o Instituto Brasileiro de Geografia e Estatística IBGE, a pessoa desempregada é aquela que está à procura de trabalho em idade ativa, considerando como idade ativa, as pessoas com idade igual ou superior aos 14 anos. Dessa forma aquelas pessoas que estão em idade ativa, mas não estão à procura de trabalho, não se enquadram na categoria de desempregadas ou desocupadas.

### **4.2 Método de coleta, origem dos dados e cálculo das taxas**

Segundo a PUC-RIO (2011), o início da coleta dos dados mensais de emprego realizado pela Pesquisa Mensal de Emprego (PME), iniciou-se em 1980, realizada pelo IBGE de forma amostral nas regiões de São Paulo, Rio de Janeiro, Recife, Belo Horizonte, Porto Alegre e Salvador, sendo essa pesquisa realizada em domicílios e a principal forma de avaliar o cenário de desemprego no Brasil, onde o IBGE passou a fornecer os dados em formato de painéis, visto que a pesquisa visitava o mesmo domicílio várias vezes.

A coleta de dados para cálculo das taxas de desemprego é realizada mensalmente, contudo ela pode ser disponibilizada de forma mensal, trimestral ou anual. Segundo a Pesquisa de Emprego e Desemprego (PED), a coleta de dados teve início no Brasil em 1980, começando a coleta em São Paulo, graças a uma parceria com Sistema Estadual de Análise de Dados (SAEDE), que posteriormente seguiu para as demais regiões metropolitanas do país.

Segundo o IBGE (2002), a Pesquisa Mensal de Emprego PME, segue os padrões da Organização Internacional de Trabalho OIT, tendo início em 1980 e passando por revisões em 1982, 1988, 1993 e 2001; sendo que para o IBGE as principais mudanças são destacadas no início em 1980 sendo chamada de PME-antiga e em 2002 quando passou a ser chamada de PME-nova.

Para Ribas e Soares (2008), a PME é a maior série contínua e com maior extensão realizada em pesquisa no Brasil, que iniciou em 1980, e que em 1982 passou a integrar a Pesquisa Nacional por Amostra de Domicílios-PNAD realizada em seis regiões brasileiras, sendo essa a primeira reformulação de impacto real. Em 2002 com a chamada PME-nova foram realizadas alterações substanciais, seguindo recomendações da OIT.

Todos os dados referentes ao desemprego são coletados, tratados e fornecidos pelo IBGE, que possui setores responsáveis por esses dados, a exemplo da Pnad, que está disponível no site do Instituto de Pesquisa Econômica Aplicada-IPEA, de modo a melhorar o funcionamento das políticas públicas, uma vez que o IPEA atua em parceria com o Ministério do Planejamento e Orçamento, dessa forma facilita o acesso às informações, não só do desemprego, mas também de diversas outras áreas.

Segundo a DIEESE (1993), o método para cálculo das taxas é realizado por amostras em conglomerados, sorteado mensalmente um determinado número de domicílios em cada setor censitário, no desenho amostral da PED são sorteados 600 setores censitários no país, para uma pesquisa trimestral, sendo 200 setores a cada mês, onde são visitados 15 domicílios em cada setor.

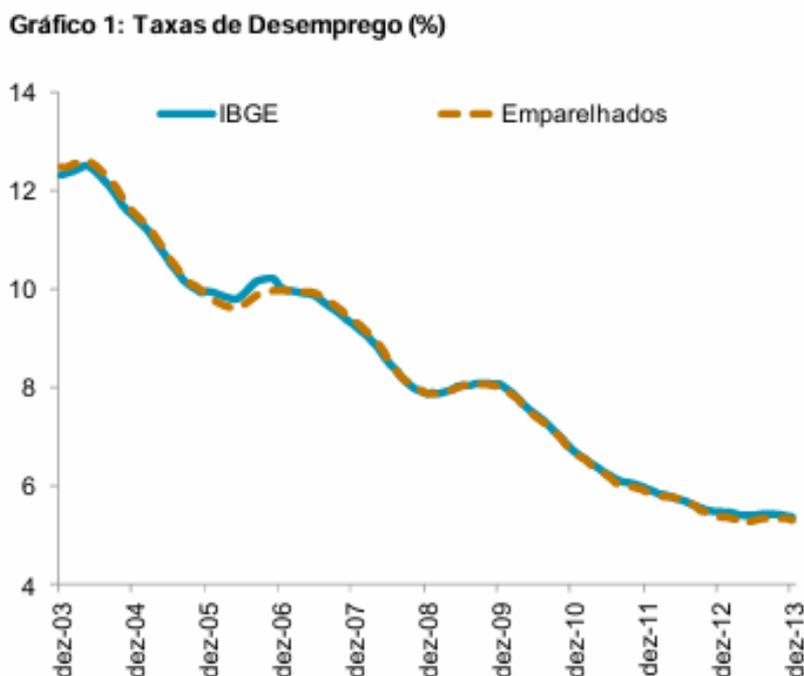
No ano de 2012 houve uma alteração no método de coleta, seguindo a nota técnica versão 1.18, onde foi definido que a partir desse ponto seriam coletadas amostras em 15096 unidades censitárias, com um total de 14 domicílios visitados em cada setor, e ao final de cada mês são visitados 70448 e ao final do trimestre são

contabilizados um total de 211344 domicílios visitados, a partir daí são calculadas as taxas mensais das amostras (IBGE, 2024).

### 4.3 Histórico de desemprego no Brasil

Segundo Silva e Pires (2014), o Brasil passou por uma considerável mudança no cenário de desemprego no país, tendo uma queda abrupta na média anual da taxa de desemprego no país, passando de 12,3% para 5,4% nos anos de 2003 a 2013, pesquisa essa realizada nas capitais metropolitanas pelo órgão de Pesquisa Mensal do Emprego (PME) do Instituto Brasileiro de Geografia e Estatística (IBGE), sendo essa considerada pelo órgão como a menor média histórica já registrada, conforme apresentado na Figura 1, onde a média móvel calculada pelo IBGE está destacada em azul.

Figura 1: Média móvel de desemprego IBGE x calculada



Fonte: Silva e Pires (2014)

No Brasil, a taxa de pessoas economicamente ativas que estão ocupadas sofreu poucas alterações no decorrer dos anos, contudo ao fazer essa análise por sexo, foi observado um significativo aumento no grupo feminino principalmente nas

grandes metrópoles, a exemplo a região de São Paulo nos anos 90, que aumentou de 50% para 55% (SANTOS, 2013).

Nos meses de dezembro dos anos de 2003 a 2013 a probabilidade de um desempregado se manter desocupado no mês seguinte, chegou a 56,2%, e a probabilidade de que esse mesmo grupo conseguisse se empregar chegou a 27,2%; analisando separadamente o intervalo de 2003 a 2005 é observado um aumento no percentual, chegando a marca de 57,3%, somente vindo a reduzir esse percentual, entre os anos de 2006 a 2008, chegando a 53,8%, o que apontou para uma possível melhora no cenário dessas pessoas, porém a crise de 2009 impactou diretamente no mercado de trabalho, revertendo a tendência de queda, com um aumento que elevou a taxa para 55,7%; contudo o impacto da crise ficou concentrado no ano 2009, pois nos anos subsequentes as taxas voltaram a cair, com o aumento do emprego formal e o maior ingresso das mulheres no mercado, (Silva e Pires, 2014).

Segundo Quadros (2003), no ano de 2001 o Brasil tinha mais de 10 milhões de habitantes desempregados, tendo uma maior concentração desse grupo nas áreas urbanas, aproximadamente 71%, totalizando mais de 7 milhões de pessoas desocupadas residindo nas zonas urbanas.

#### **4.4 Trabalhos correlatos**

Na pesquisa da série de desemprego da região de São Paulo, Marques e Fava (2011), utilizaram o modelo SARIMA (6,1,0)(1,0,1)<sub>12</sub> para modelar a série dos dados coletados pela PED, segundo os autores esse modelo apresentou melhores resultados, acompanhando o comportamento da série de maneira satisfatória, capturando bem o comportamento da autocorrelação e os movimentos sazonais, contudo o modelo SARFIMA (6;0,9156;0)(0;0,3642;0)<sub>12</sub> apresentou resultados melhores em intervalos de 1 a 20 anos. Analisando os dois modelos por meio da Função de Impulso-resposta Sazonal FIR-S, onde o primeiro modelo apresentou um valor de 1,3084 e o segundo 1,2675; indicando que apesar do SARFIMA realizar previsões com mais precisão a longo prazo, ambos apresentam boas previsões no curto prazo.

Na análise das taxas de desemprego de Porto Alegre, foi utilizado o modelo SARIMA (0,1,3)(2,1,0)<sub>12</sub>, sendo esse composto por um termo de diferença e três médias móveis em período sazonal igual a 12, sendo esse o modelo que melhor se

ajustou aos dados, acompanhando o comportamento da série original de desemprego em Porto Alegre de janeiro de 1993 a dezembro de 2009, ao analisar a autocorrelação dos resíduos da primeira diferença ele apresentou um comportamento semelhante a um ruído branco. Mas apesar do bom ajuste, o modelo não se mostrou favorável para realizar previsões (BECKER, 2010).

Na dissertação de mestrado, Silva (2015), realizou uma análise da série de desemprego no Brasil de 2002 a 2013, onde o mesmo fez uma análise de adequação da curva de Phillippe a série original, para tal o autor utilizou o software livre Acrônimo de Gnu Regression, Econometrics and Time-series Library (GRET), para modelagem e ajuste da série a curva Phillips, onde foi constatado que a série original é não-estacionária com 13 defasagens e um coeficiente sigma igual a  $-0,75$ ; posterior ao teste o autor verificou a co-integração da variável com outras séries, porém pelo fato da série de desemprego ser não-estacionária, os resultados obtidos negaram a co-integração.

A série de desemprego no Brasil de 2000 a 2021, sofre impactos de muitas variáveis, segundo Silveira (2022), as variáveis que mais impactam no desemprego são a dívida pública com uma correlação de 77,4%, o salário-mínimo com 35,2% e o Índice de Desenvolvimento Humano com 34,9%. Para ajustar o modelo a autora utilizou o Support Vector Machine, ajustando o modelo AR (1), o modelo Random Forecast, função SMV e o modelo de regressão linear LASSO, para análise das previsões, onde constatou que o Random Forecast apresentou melhor desempenho com um MAE de 0,4915 e erro quadrático médio de 0,7042.

## **5 METODOLOGIA**

### **5.1 Base de dados**

O estudo apresentado nesse trabalho tem como finalidade, fazer uma análise na série histórica de desemprego no Brasil, levando em consideração as taxas de ocorrências mensais no período de março de 2012 a novembro de 2024. A mesma análise será aplicada ao estado de Sergipe, tendo início no primeiro trimestre de 2012 e fechando o ciclo no terceiro trimestre de 2024.

Os dados utilizados foram extraídos no dia 27 de dezembro de 2024 do site IPEADATA (Instituto de Pesquisa econômica aplicada) no endereço eletrônico <http://www.ipeadata.gov.br/>. Para o estudo em questão foi levado em consideração as variáveis: Data – o mês e o ano de ocorrência, sendo considerado como mensal, trimestral ou anual; Taxa – o percentual de pessoas economicamente ativas, que estão à procura de emprego, levando em consideração as pessoas que estão trabalhando; Região – regiões brasileiras e suas respectivas taxas.

### **5.2 Séries temporais**

Segundo Morettin e Tolo (2006) série temporal pode ser definida como conjunto de dados quantitativos e sequenciados no decorrer do tempo, podendo a série ser discreta ou contínua, sendo que a primeira pode ser representada como uma amostragem em um intervalo de tempo definido de maneira igualitária de uma série contínua, enquanto a contínua é o conjunto de todas as informações geradas ao longo do tempo.

Uma série temporal pode ser classificada em dois grupos, dependendo da sua estrutura, podendo ser uma série curta, que ocorre quando são observados fenômenos ocorridos em curtos períodos, tendo essas alterações significativas em cada período observado; o outro grupo é a série longa, sendo caracterizada pela estrutura mantida pelos dados ao longo do tempo. Uma série possui três componentes, que são a tendência, a sazonalidade e aleatoriedade (Fonseca, Martins e Toledo, 1978).

É possível identificar uma série, observando suas características, como por exemplo, a sua recorrência dos dados e a dependência, sendo os principais fatores

implícitos para identificar e modelar esses dados, tendo em vista que o principal objetivo do estudo de séries temporais é modelar previsões futuras, com base na dependência dos dados históricos de um determinado conjunto em um intervalo de tempo iniciado em um tempo “x”, indo até “T” (MORETTIN E TOLOI, 2006).

O processo para modelar uma série consiste em uma análise detalhada de seus principais componentes por meio do processo chamado de decomposição da série, que consiste em realizar primeiramente a estimativa da tendência e a presença de sazonalidade, sequencialmente observa-se os resíduos, após observado esses componentes é possível ter um melhor embasamento do possível modelo que se ajusta melhor a série (MORETTIN E TOLOI, 2006).

Para Fonseca, Martins e Toledo (1978), a tendência é o comportamento que os dados apresentam ao apontar para uma direção determinada, podendo esse apresentar um comportamento decrescente ou crescente; já a denominação de uma série oscilatória pode ser observada por meio do movimento cíclico, sendo observada como variações em torno da tendência; e a aleatoriedade é justificada pelo comportamento irregular que a série apresenta, sendo visualizadas como flutuações.

O processo de modelagem de uma série consiste em uma análise dos seus dados, testando a dependência que os dados possuem entre si, por meio da autocorrelação ou pelo coeficiente de correlação. De acordo com Souza e Camargo (1996), a função de autocorrelação funciona como uma memória dos dados, sendo vista como uma autocovariância padronizada onde um valor  $t$ , possui uma dependência de outro valor no tempo  $t-k$ , desse modo a função de autocorrelação é simétrica em relação a origem de  $k=0$ , onde a autocorrelação de  $k$  é definida por:

$$p^k = \frac{y^k}{y^0} = \frac{Cov[Z_t, Z_{t+k}]}{\sqrt{Var(Z_t)Var(Z_{t+k})}}$$

Segundo Camargo e Souza (1996), a dependência entre os valores tem forte influência sobre a série, para remover essa influência é calculada a autocorrelação parcial, que calcula a dependência entre duas observações da série,  $Z_t$  e  $Z_{t+k}$ , dessa forma  $Z_{t+1}$ ,  $Z_{t+2}$ ,  $Z_{t+k-1}$  deixa de influenciar os demais dados, sendo definida pela função:

$$Cor(Z_t, Z_{t+k} | Z_{t+1}, \dots, Z_{t+k-1})$$

No processo de modelagem de uma série é necessário testar se ela é estacionária. Para Morettin e Tolo (2006), algumas séries temporais têm comportamento estacionário, sendo definida como estacionária, a série que apresenta uma média constante em qualquer tempo aleatório, ou seja, em uma estrutura de dados qualquer a média no tempo  $t$  vai ser a mesma no tempo  $t+1$ ,  $t+2$  ou  $t+n$ . Para Souza e Camargo (1996), a série pode ainda ser estacionária de três modos, estritamente estacionária, onde a série é distribuída de maneira igualitária, não sendo influenciada pela escolha do ponto de partida; estacionária de ordem finita, quando apenas partes isoladas da série são consideradas estacionárias; e estacionária de segunda ordem, que ocorre quando a covariância depende somente das diferenças e a função de médio valor é uma constante, dada por.

$$\mu_t = E[y_t]$$

Para testar a estacionariedade de uma série são aplicados dois testes: Dickey Fuller e o KPSS, onde o teste Dickey Fuller, verifica se uma série temporal possui raiz unitária, ou seja, se ela é não estacionária; e o teste KPSS faz o oposto do ADF, verificando se a série é estacionária em torno de uma tendência determinística; seguindo teste de hipótese (GUJARATI E POTER, 2011).

Hipóteses para o teste Dickey Fuller (ADF):

- $H_0$  (Hipótese Nula): (A série possui raiz unitária não é estacionária) p-valor  $< \alpha$
- $H_1$  (Hipótese Alternativa): (A série não possui raiz unitária é estacionária) p-valor  $> \alpha$

Se o p-valor for menor que 0,05, rejeitamos  $H_0$ , ou seja, a série é estacionária. Se o p-valor for maior que 0,05 não rejeitamos  $H_0$ , ou seja, a série não é estacionária (precisa de diferenciação).

Hipóteses para o teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS):

- $H_0$  (Hipótese Nula): (A série é estacionária) p-valor  $< \alpha$
- $H_1$  (Hipótese Alternativa): (A série não é estacionária) p-valor  $> \alpha$

Se o p-valor for menor que 0,05, rejeitamos  $H_0$ , ou seja, a série não é estacionária.

Se o p-valor for maior que 0,05, não rejeitamos  $H_0$ , ou seja, a série é estacionária.

Para os resultados dos testes, podem ocorrer conclusões de acordo com a Tabela 1, em caso de resultado inconclusivo com possível tendência ou quebra estrutural, são realizados testes de hipóteses ou utilizados modelos de decomposição.

Para a primeira conclusão segue o processo de modelagem da série, para os demais, faz se necessário a aplicação de técnicas que a tornem estacionária. Segundo Box; Jenkins e Reinsel (2008), uma das maneiras de tornar uma série estacionária, é a diferenciação, onde é calculado o número de diferenças, calculando a diferença do intervalo  $X_t$  a  $X_{t-1}$ , depois de  $X_{t-1}$  a  $X_{t-2}$ , e assim sucessivamente até encontrar o total de diferenças, sendo possível remover esses valores, desse modo é possível reduzir a sazonalidade e os sinais de tendência.

Tabela 1: Possíveis resultados combinados ADF test e KPSS test.

KPSS	ADF	
	Não Rejeita	Rejeita
Não Rejeita	Decisão inconclusiva (Informações insuficientes)	Decisão Conclusiva (Estacionária)
Rejeita	Decisão Conclusiva (Estacionária)	Decisão inconclusiva (Integração fracionária)

Fonte: Nusair (2003)

### 5.3 Modelos SARIMA

O processo para construção do modelo ARIMA(p,d,q) é baseado na identificação e ajuste de modelos autorregressivos, com médias móveis integradas; seguindo um ciclo com base nos dados, dessa forma é possível estruturar o modelo analisando quatro etapas, a primeira consiste em uma análise geral do modelo; na segunda etapa são analisados os resultados da funções de autocorrelação, autocorrelação parcial e demais funções, para definição de um modelo específico; na terceira etapa observa-se os parâmetros do modelo definido; na quarta e última etapa é feita uma verificação detalhada do ajuste do modelo, com base no resultado dos resíduos, para definir se o modelo é bom para realizar previsões (MORETTIN E TOLOI, 2006).

Nos modelos de autorregressão para séries estacionárias, é realizada uma combinação linear múltipla dos dados, onde a entrada são pontos  $p$  anteriores e a saída os pontos futuros, onde o  $p$  representa os atrasos na equação, de forma que a equação do modelo AR ( $p$ ) é utilizado para prever um passo à frente, dependendo apenas dos dados anteriores, sendo a equação escrita de modo geral dada por (MORETTIN E TOLOI, 2006).

$$x_t = c + \theta_1 x_{t-1} + \theta_2 x_{t-2} \cdots \theta_p x_{t-p} + w_t,$$

em que o termo  $c$  representa uma constante,  $\theta_p$  representa o termo autorregressivo da equação e  $w_t$  o termo aleatório da equação, ou seja, o erro no tempo  $t$ .

Para o modelo MA ( $q$ ) média móvel, aplicado a séries estacionárias, é definido o passo à frente a partir do cálculo de soma dos resíduos (MORETTIN E TOLOI, 2006), ou seja, a modelagem da série é feita por meio de combinação linear dos ruídos aleatórios  $q$  independentes, sendo esse modelo escrito de forma geral por.

$$x_t = c + \varphi_1 w_{t-1} + \varphi_2 w_{t-2} \cdots \varphi_j w_{t-j},$$

em que o termo  $c$  representa uma constante,  $\varphi_j$  representa o termo de média móvel e  $w_t$  o termo aleatório, ou seja, o erro no tempo  $t-j$ .

Nos modelos ARMA ( $p,q$ ), utilizados para modelar séries estacionárias, sendo definido como combinação do modelo AR (autorregressivo) com o modelo MA (médias móveis), esse modelo tem como característica a capacidade de regressar uma variável aos seus valores passados, ao mesmo tempo que modela o termo do erro como combinação linear, sendo escrito de modo geral por.

$$x_t = c + \theta_1 x_{t-1} + \theta_2 x_{t-2} \cdots \theta_p x_{t-p} + \varphi_1 w_{t-1} + \varphi_2 w_{t-2} \cdots \varphi_j w_{t-j} + w_t,$$

Segundo Morettin e Tolo (2006), os modelos ARIMA ( $p,d,q$ ), modelo utilizado em séries não estacionárias, característico por possuir o parâmetro  $d$ , que representa o número de diferenças responsável por tornar a serie estacionaria, onde ele remove os pontos que causam variação na série, reduzindo a tendência e tornando a série estacionária, dado por.

$$X'_t = c + \theta_1 x'_{t-1} + \theta_2 x'_{t-2} \cdots \theta_p x'_{t-p} + \varphi_1 w_{t-1} + \varphi_2 w_{t-2} \cdots \varphi_j w_{t-j} + w_t,$$

onde o operador de diferença  $d$  está representado na equação pelo termo  $x'$ .

Para realizar a aplicação do modelo SARIMA (p,d,q)(P,D,Q), característico por ser um modelo autorregressivo, com a particularidade de integração de médias móveis, ao mesmo tempo que capta a sazonalidade da série, de modo que esse modelo consegue generalizar os modelos ARIMA, permitindo modelar séries não estacionárias e sazonais, e séries não sazonais, o termo P representa o termo sazonal autorregressivo, o D representa a diferença sazonal e o Q, representa a média móvel sazonal, ou seja,

$$\varphi_p(B)\varphi(B^s)[(1-B)^d(1-B)^D - \mu] x_t = \theta(B)\theta(B^s) W_t$$

- B : “backward shift” (representa o deslocamento de uma unidade em cada tempo t em função de  $x_t$ );
- d : número de diferenças;
- D : número de diferenças sazonais;

Segundo Cryer e Chan (2008), o dispositivo backward shift é um termo notacional utilizado em séries temporais com defasamentos, para aplicar o deslocamento de dados um período para trás, para duas aplicações deslocam-se dois períodos; ainda segundo os autores a ordem de diferença é representada pelo termo  $\nabla^d$  e o termo que representa a diferença da sazonalidade é descrito  $\nabla_s^D$ , e o termo w representa os resíduos.

Após definir os possíveis modelos é realizada uma análise dos parâmetros, por meios que aferem a qualidade do conjunto de dados, após aplicação de um modelo estatístico. Para Box; Jenkins e Reinsel (2008), o método mais adequado para escolha do modelo que melhor se ajusta aos dados é o critério AIC, por esse método é definido como mais adequado aquele que possuir o menor critério AIC (Akaike Information Criterion), levando em consideração que nenhum modelo possui precisão exata, esse método faz a análise e contabiliza o número de informações perdidas,

consequentemente modelos com maior AIC, deixam escapar muitos dados, logo perdem a credibilidade.

No método AIC é multiplicado o total de parâmetros que o modelo estima por dois, subtraindo por duas vezes o logaritmo natural da função de verossimilhança, onde o parâmetro do modelo é representado pelo termo  $k$  e a função de verossimilhança é representada pelo termo  $L$  (TADDY, 2019), ou seja,

$$AIC = 2k - 2\ln(L)$$

Para avaliar a qualidade do ajuste do modelo é realizada a análise dos resíduos, tendo em vista que alguns modelos não captam bem o comportamento dos dados, deixando o comportamento da série, apresentando um comportamento irreal, dado a natureza sutil dos resíduos. Para Sáfadi e Barros (2004), um modelo com um bom ajuste possui características específicas, com resíduos apresentando o comportamento de um ruído branco, onde o resíduo é uma variável aleatória independente e distribuído de maneira igualitária.

Para Rossi (2019), deve-se analisar e testar a normalidade dos resíduos e verificar se são independentes, para que o modelo ajustado possa realizar previsões, onde é verificado a independência pelo Ljung-Box test; já para verificar a normalidade são realizados, o teste Durbin-Watson e o teste Shapiro-Wilk. No teste Durbin Watson é considerado positivo quando o p-valor for maior que o nível de significância, ou seja, p-valor  $>\alpha$  não rejeitando a hipótese nula; já no Shapiro-Wilk test, é considerado com distribuição normal, quando a estatística de teste for maior que o nível de significância de alpha igual a 0,05.

De modo geral o processo para realizar análise e previsão em séries temporais aplicando o modelo Box e Jenkins, consiste em três etapas, primeiramente realizar uma análise da série original, sequencialmente deve-se definir o modelo que melhor se ajusta a série, e por último fazer as previsões.

#### **5.4 Suporte computacional**

A análise de dados é um campo bastante estudado em todas as áreas, contudo com o desenvolvimento de softwares, essa análise torna-se mais rápida e prática. Uma das linguagens de programação, que colabora com esses estudos, e é muito

usual para dados estatísticos, é a linguagem de programação R, que possui pacotes que facilitam nos cálculos, na manipulação e modelagem dos dados e na criação de ilustrações gráficas (RSTUDIO, 2024).

Um dos pacotes do Rstudio que testa a estacionariedade da série por meio do teste Dickey-Fuller (ADF), é o `tseries`, onde é apresentada hipóteses, nas quais uma define a estacionariedade, ou seja, caso o p-valor apresentado for inferior a 0,05 rejeita-se a hipótese nula e confirma que a série é estacionária (TRAPLETTI; HORNIK; LEBARON, 2020).

Segundo Hyndman (2014), o `forecast` é um importante recurso, visto que ele permite integração entre os dados, fazendo combinações, de modo que prevê o último componente sazonal, dessa forma é possível realizar previsões dos componentes sazonais originais com os dados ajustados sazonalmente, para obter previsões com maior confiabilidade.

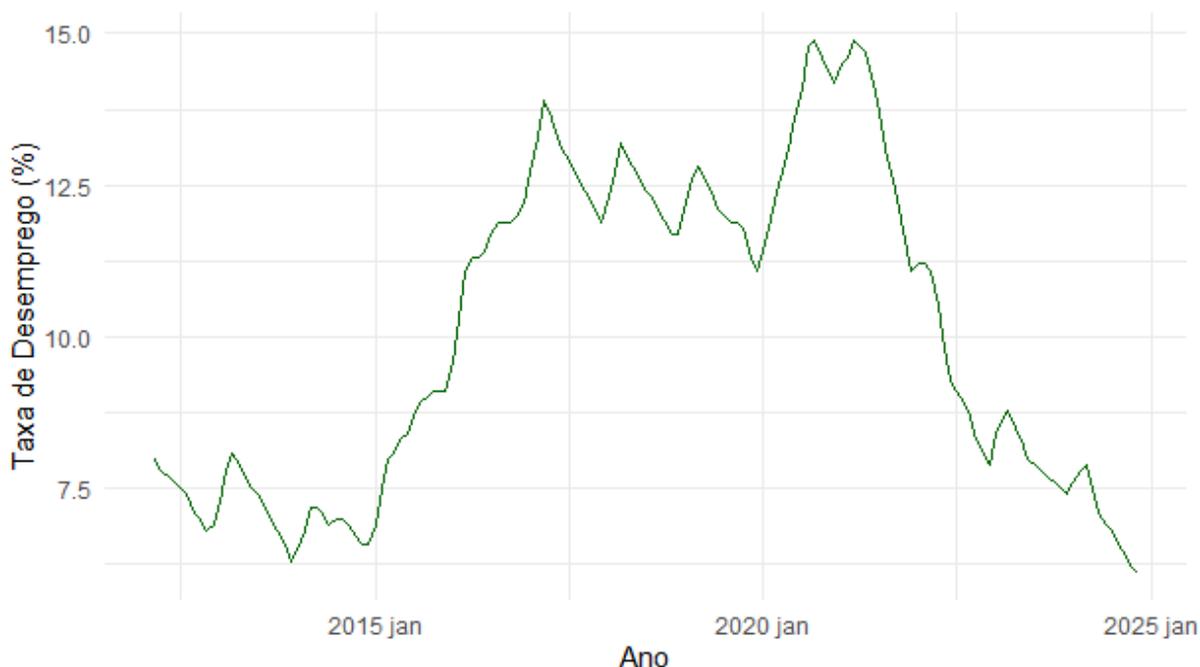
Um outro pacote essencial para modelar séries temporais é o `fpp3`, sem uma ferramenta crucial, pois ele integra outros pacotes como o `tsibble`, que por sua vez organiza a estrutura da série; um outro fator que o essencial é sua sintaxe intuitiva, o que facilita seu uso, desde o básico da importação dos dados, até as análises mais complexas, como modelagem, previsões e decomposição (HYNDMAN, 2014)

## 6 RESULTADOS E DISCUSSÃO

### 6.1 Análise da série original de desemprego no Brasil

Na Figura 2 está a evolução da taxa de desemprego no Brasil em forma de série temporal, no período de março de 2012 a novembro de 2024; observa-se uma tendência ao longo do tempo, com períodos de aumento e queda no indicador. Observando o período de 2012 a 2015, a taxa de desemprego manteve-se relativamente estável, variando entre 6% e 8%, com pequenas oscilações, mas a partir de 2015, nota-se um crescimento acentuado, refletindo a crise econômica que impactou o mercado de trabalho no Brasil, o que culminou em um pico de desemprego por volta de 2017, quando as taxas ultrapassaram os 13%. Segundo Balassiano (2018) esse comportamento está relacionado à recessão econômica de 2014 durante o governo Dilma, o que gerou um aumento das demissões no período.

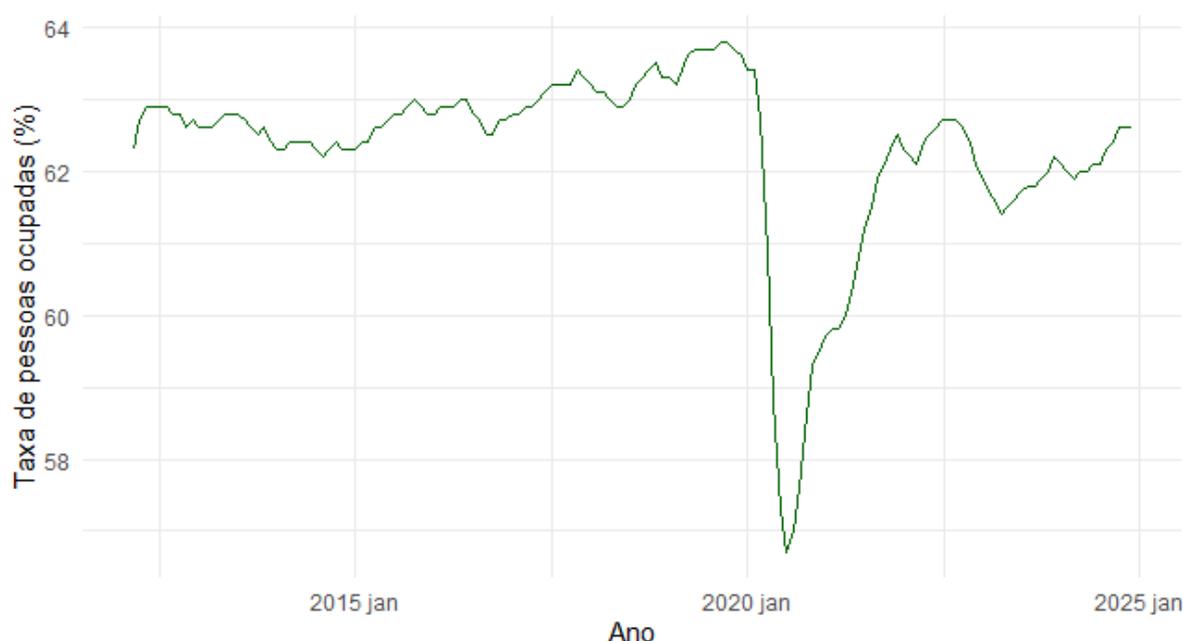
Figura 2: Série de desemprego no Brasil de março de 2012 a novembro de 2024



Fonte: Elaborado pelo autor

Após o pico de 2017, a taxa de desemprego apresentou certa oscilação, mantendo-se elevada até o final de 2021, quando o país chega próximo a 15%. Segundo Louredo et al (2022), esse aumento coincide com a crise provocada pela pandemia do vírus da Covid-19, que afetou todo o mundo, resultando no fechamento de empresas e na redução de postos de trabalho. Uma das medidas do governo foi a medida provisória, a MP 936, implementada no dia 1º de abril de 2020, com o objetivo de proteger o trabalho e a renda das famílias, de modo a evitar demissões definitivas. Conforme apresentado na Figura 3, pode-se observar uma queda abrupta no percentual de pessoas ocupadas no período da pandemia com aproximadamente 64% no final de 2019 e 57% durante o ano de 2020.

Figura 3: Série original do percentual de pessoas ocupadas no Brasil de 2012 a 2024



Fonte: Elaborado pelo autor

No entanto, somente em meados de 2021, o desemprego começa a apresentar uma tendência de queda impulsionada pela recuperação econômica gradual, pela reabertura das atividades econômicas e pela adaptação do mercado de trabalho às novas condições. Segundo dados do IBGE, com a chegada da pandemia, a modalidade de trabalho remoto ganhou força, chegando a dobrar o percentual de pessoas nessa modalidade que no ano de 2019 era aproximadamente 4,5% e no ano de 2022 chegou a 8,5%, sendo uma das principais medidas mitigadoras para evitar

que as taxas de desemprego continuassem a subir, isso aliado a medidas provisórias criada pelo governo.

Nos anos mais recentes, observa-se uma trajetória consistente de redução da taxa de desemprego, alcançando níveis próximos aos registrados antes da crise de 2015, esse movimento pode ser atribuído ao crescimento econômico, ao aumento da formalização do emprego e a políticas de incentivo à contratação (Velooso et al, 2022). Assim, a análise da série histórica sugere que, embora a taxa de desemprego esteja em queda, sua sustentabilidade dependerá de fatores macroeconômicos, políticas públicas e da dinâmica do mercado de trabalho nos próximos anos.

#### 6.1.1 Teste de estacionariedade da série original de desemprego no Brasil

Para analisar a estacionariedade da taxa de desemprego no Brasil, foi utilizado os testes Augmented Dickey-Fuller (ADF) e KPSS, conforme é mostrado na Tabela 2, nos quais indicaram que a série temporal não é estacionária; o teste ADF apresentou uma estatística de -0,90915 e um p-valor de 0,9496, muito acima do nível de significância usual de 0,05, o que leva à não rejeição da hipótese nula de raiz unitária, esse resultado também implica que a série apresenta uma tendência estocástica. Ou seja, não é estacionária; já o teste KPSS, que tem como hipótese nula a estacionariedade, apresentou uma estatística com valor 0,81876 e um p-valor inferior a 0,01, desse modo rejeitamos a hipótese nula e reforçamos a ideia de que a série não é estacionária.

Tabela 2: Teste de estacionariedade da série de desemprego no Brasil

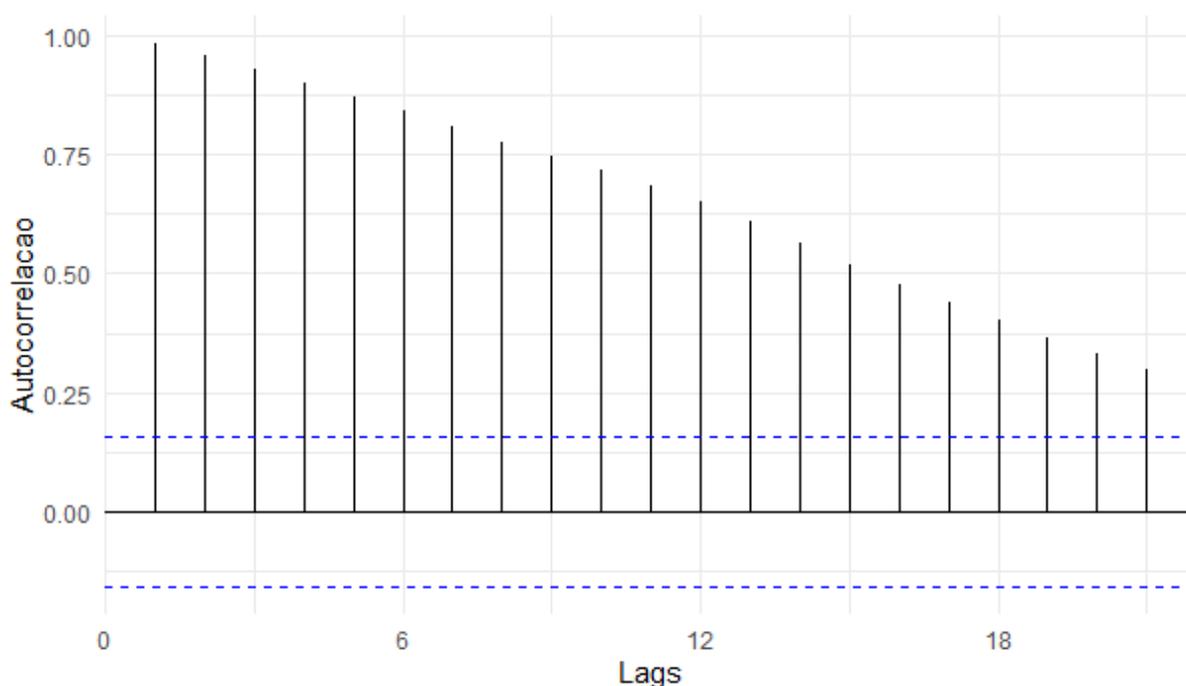
Teste	Estatística	p-valor	Hipótese Nula	Conclusão
<b>ADF</b>	-0.90915	0.9496	Série possui raiz unitária (não estacionária)	Não rejeita $H_0$ (série não estacionária)
<b>KPSS</b>	0.81876	0.01	Série é estacionária	Rejeita $H_0$ (série não estacionária)

Fonte: Elaborado pelo autor

Diante dos resultados obtidos, foram necessárias duas diferenciações para tornar a série estacionária, sugerindo que a série segue um processo integrado de ordem dois, ou seja, a taxa de desemprego tem uma forte dependência dos dados anteriores e a série pode ter efeitos de choques persistentes ao longo do tempo.

Na Figura 4 pode-se observar o gráfico de autocorrelação (ACF) da série da taxa de desemprego no Brasil que apresenta uma forte dependência temporal, com coeficientes elevados e decaindo lentamente ao longo das defasagens, o que indica que os valores da série estão fortemente correlacionados com seus valores passados, tal comportamento sugere a presença de uma tendência persistente, característica de séries não estacionárias, o que reforça os resultados dos testes ADF e KPSS.

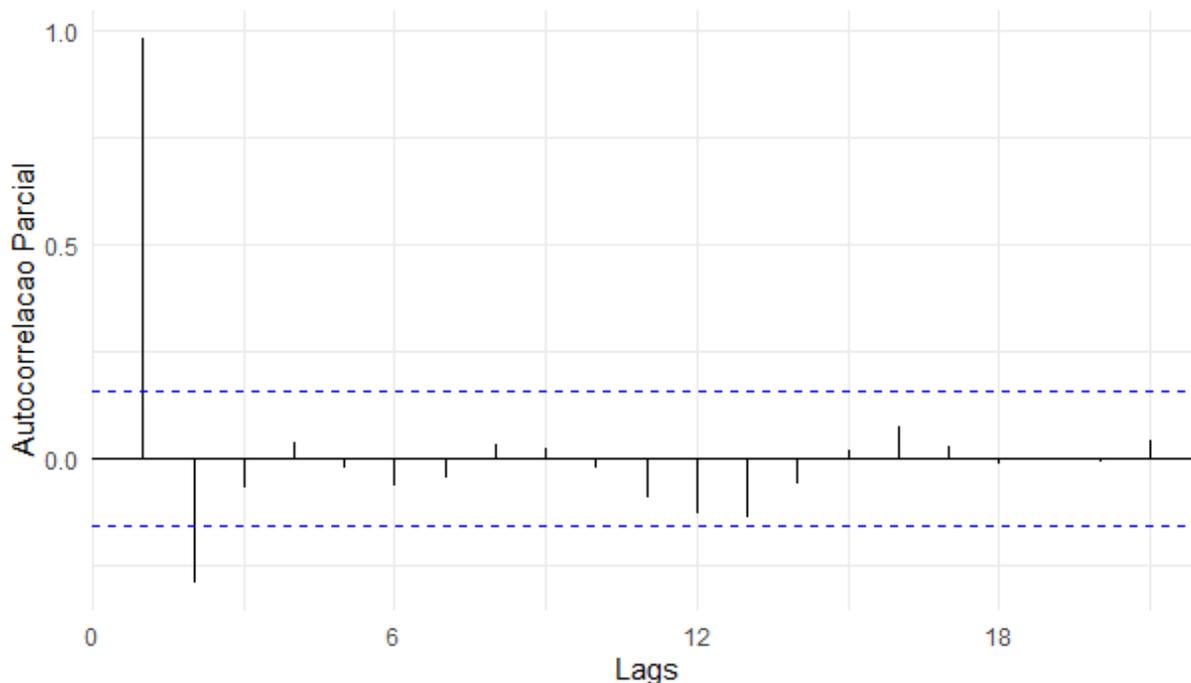
Figura 4: Gráfico de autocorrelação da série de desemprego no Brasil



Fonte: Elaborado pelo autor

O gráfico de autocorrelação parcial (PACF) da série da taxa de desemprego no Brasil apresentado na Figura 5, mostra um coeficiente significativo apenas na primeira defasagem, seguido por valores próximos de zero nos demais lags, esse padrão é característico de séries que apresentam um comportamento de passeio aleatório, a queda abrupta após o primeiro lag sugere que a série pode ser bem representada por um modelo autorregressivo de baixa ordem.

Figura 5: Gráfico de autocorrelação parcial da série de desemprego no Brasil

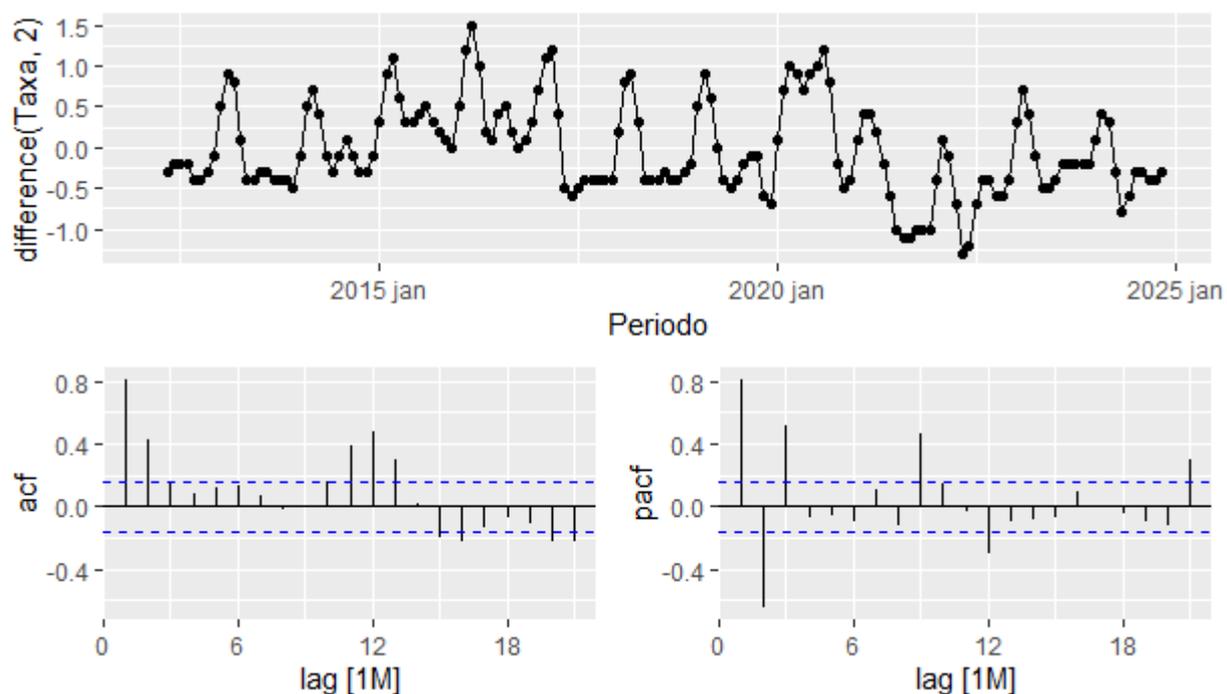


Fonte: Elaborado pelo autor

### 6.1.2 Aplicando diferenças na série original de desemprego no Brasil

A Figura 6 apresenta a série após a realização das duas diferenças para tornar a série estacionária, tornando a série mais adequada para modelagem, obtendo também uma maior estabilidade na variação ao longo do tempo, os gráficos da função de autocorrelação ACF e autocorrelação parcial PACF, os gráficos foram utilizados para identificar a estrutura da dependência temporal remanescente na série diferenciada, onde o gráfico ACF apresenta um decaimento rápido, o que sugere um comportamento autorregressivo; no gráfico PACF é visualizado um corte significativo após poucas defasagens. Um outro comportamento observado na série diferenciada é o comportamento da sazonalidade, onde é observado um aumento nas taxas sempre nos meses de janeiro, segundo Santone (2024), esses períodos são marcados por contratações temporárias, dado o aquecimento do mercado de vendas, contudo, essas contratações tendem a encerrar entre os meses de janeiro e fevereiro, explicando o comportamento sazonal da série.

Figura 6: Análise da segunda diferença aplicada a série de desemprego no Brasil



Fonte: Elaborado pelo autor

### 6.1.3 Verificação e aplicação do modelo a série original de desemprego no Brasil

Conforme a análises realizadas, foi definido alguns modelos ARIMA, levando em conta o ajuste de diferenciação de segunda ordem, para isso foi criado um quadro com os modelos que melhor se ajustaram a série, apresentado na Tabela 3.

Tabela 3: Modelos testados na série de desemprego no Brasil

<b>Modelos ARIMA testados para a serie</b>	
<b>Modelo</b>	<b>Descrição</b>
<b>ARIMA(2,2,3)</b>	Modelo ARIMA não sazonal com 2 termos autorregressivos, 2 diferenciações e 3 termos de média móvel. Adequado para séries temporais com tendência e ciclos curtos.
<b>ARIMA(3,2,2)</b>	Modelo ARIMA não sazonal com 3 termos autorregressivos, 2 diferenciações e 2 termos de média móvel. Similar ao anterior, mas com maior ênfase nos termos autorregressivos.
<b>ARIMA(0,2,1)(1,0,0)[12]</b>	Modelo ARIMA sazonal com 1 termo de média móvel e 2 diferenciações na parte não sazonal, e 1 termo autorregressivo na parte sazonal. Adequado para dados mensais com sazonalidade e tendência.
<b>ARIMA(0,2,3)(2,0,0)[12]</b>	Modelo ARIMA sazonal com 3 termos de média móvel e 2 diferenciações na parte não sazonal, e 2 termos autorregressivos na parte sazonal. Similar ao anterior, mas com maior complexidade na parte sazonal.

Fonte: Elaborado pelo autor

Para comparar os modelos ARIMA ajustados à série, foi realizada uma comparação com base nos critérios estatísticos AICc (Akaike Information Criterion corrected) e BIC (Bayesian Information Criterion), no qual o modelo identificado como (ARIMA(0,2,3)(2,0,0)[12]) apresentou o menor AICc igual a -112 e um BIC igual a -94, conforme é mostrado na Tabela 4, esse modelo combina diferenciação de segunda ordem  $d=2$  com três termos de média móvel e dois termos sazonais autorregressivos, o que sugere que a série requer ajustes para capturar tanto a estrutura de tendência quanto possíveis padrões sazonais.

Tabela 4: Resultado do teste dos modelos testados na série de desemprego no Brasil

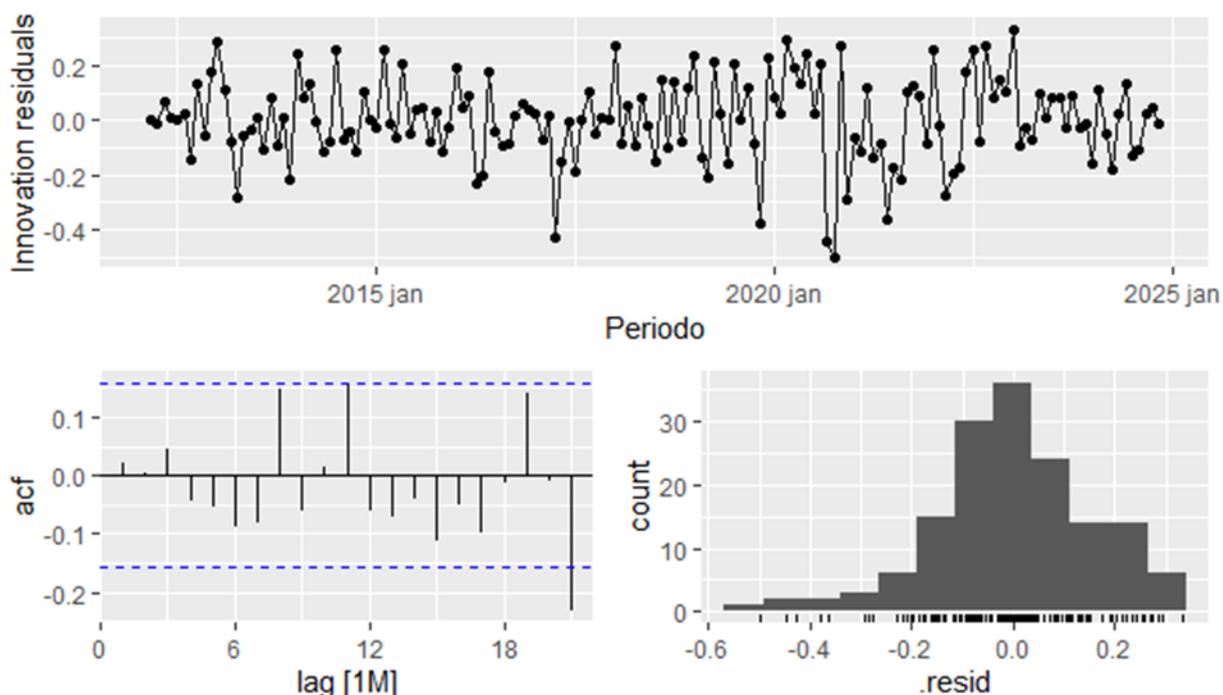
Teste dos modelos					
Modelo ARIMA	sigma2	log_lik	AIC	AICc	BIC
<b>ARIMA(0,2,3)(2,0,0)[12]</b>	250	62.0	-112.0	-112.0	-94.0
<b>ARIMA(0,2,1)(1,0,0)[12]</b>	338	39.1	-72.3	-72.1	-63.2
<b>ARIMA(2,2,3)</b>	388	32.4	-52.8	-52.2	-34.6
<b>ARIMA(3,2,2)</b>	445	22.7	-33.5	-32.9	-15.4

Fonte: Elaborado pelo autor

A Figura 7 apresenta a análise dos resíduos, os quais parecem oscilar em torno de zero sem apresentar padrões claros ou tendências sistemáticas, sugerindo um bom ajuste do modelo, a função de autocorrelação dos resíduos, está com a maioria das defasagens dentro dos limites de significância, sugerindo que os resíduos se comportam como um ruído branco; já o histograma dos resíduos segue um comportamento aproximadamente normal, com uma distribuição centrada em torno de zero e leve assimetria, mas sem desvios extremos que comprometam a validade do modelo.

Após a escolha do modelo foi testado a correlação dos resíduos, onde o Ljung-Box apresentou um p-valor de 0,0854, não rejeitando a hipótese nula de que não há correlação significativa entre os resíduos; no teste Shapiro-Wilk foi encontrado um p-valor de 0,03148, indicando que os dados não seguem distribuição normal, apesar de não haver autocorrelação entre os resíduos, esse resultado impacta na confiabilidade dos intervalos de previsão.

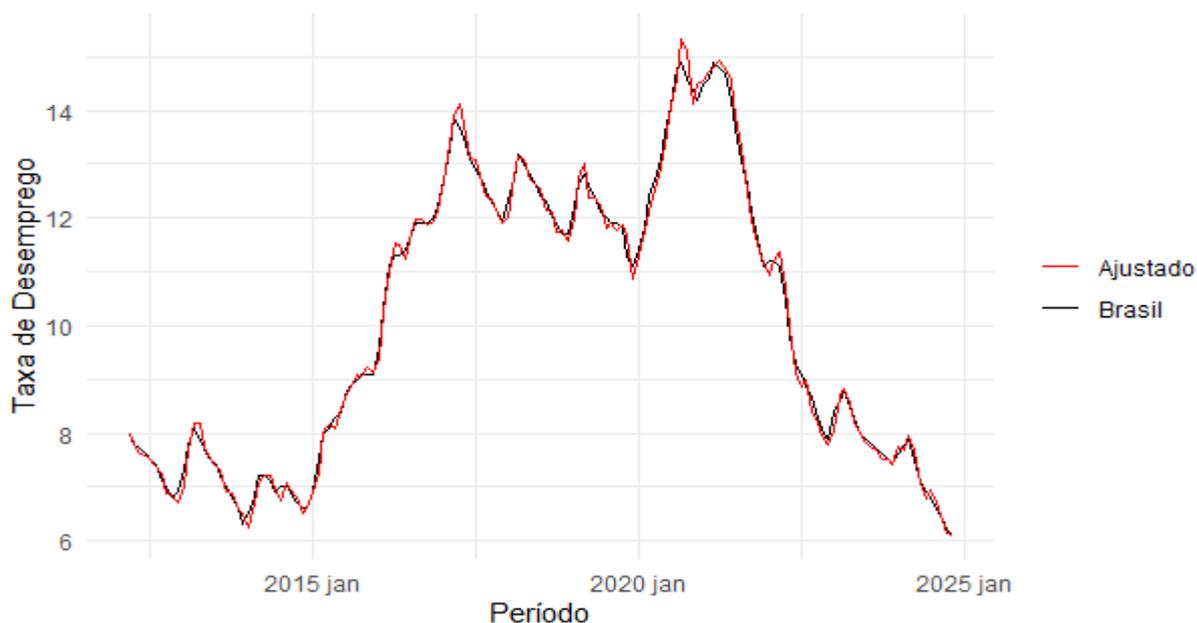
Figura 7: Análise dos resíduos da série de desemprego no Brasil após aplicação da diferença de segunda ordem



Fonte: Elaborado pelo autor

Após aplicado o modelo, foi realizada uma comparação conforme a Figura 8, comparando a série original da taxa de desemprego no Brasil com a série ajustada pelo modelo estatístico, observa-se que as duas séries se sobrepõem quase perfeitamente, o que indica que o modelo conseguiu capturar a dinâmica dos dados de maneira eficiente, seguindo a trajetória da série tanto nos períodos de crescimento, como exemplo entre 2015 e 2020, quanto as fases de queda, especialmente após 2021, esse ajuste adequado reforça a validade das transformações aplicadas anteriormente, como a diferenciação para garantir estacionariedade, e a escolha de um modelo apropriado para capturar a dependência temporal da série.

Figura 8: Série original de desemprego no Brasil x Modelo ajustado



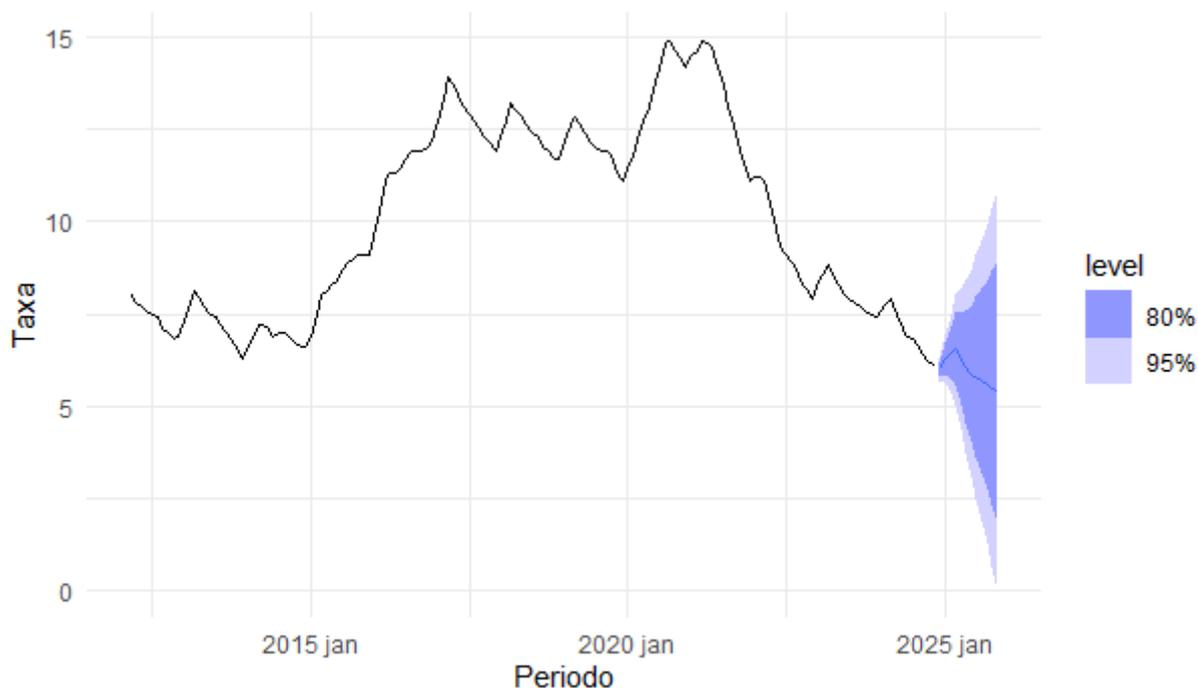
Fonte: Elaborado pelo autor

#### 6.1.4 Previsão de taxas de desemprego futuras no Brasil

A previsão da taxa de desemprego no Brasil foi estimada para os próximos doze meses, com intervalos de confiança de 80% e 95%, analisando os primeiros seis meses dessa projeção, é observada uma tendência de queda seguindo o comportamento da série histórica recente, o que indica uma possível continuidade da redução do desemprego, de modo, que essa previsão pode ser útil para políticas públicas e decisões econômicas, conforme apresentado na Figura 9.

Ainda analisando a Figura 9, nota-se um aumento da incerteza nas estimativas de previsão para os doze meses, uma vez que os intervalos de confiança tornam-se mais amplos, refletindo a dificuldade crescente de prever eventos futuros à medida que o horizonte se expande, ainda assim, o modelo sugere que a tendência de queda do desemprego pode se manter ao longo do próximo ano, embora com alguma possibilidade de estabilização ou menor ritmo de redução, portanto, embora essa projeção ainda forneça direções úteis, sua interpretação deve ser acompanhada de uma análise mais ampla dos cenários econômicos.

Figura 09: Previsão do desemprego no Brasil para 12 meses

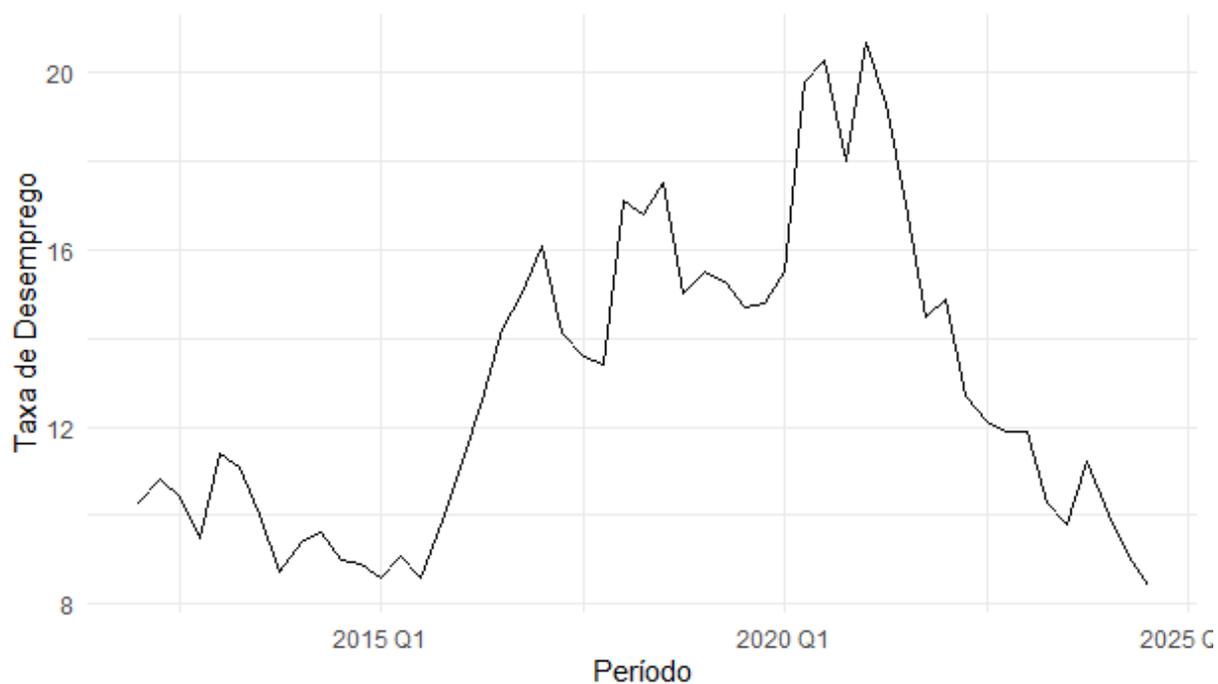


Fonte: Elaborado pelo autor

## 6.2 Análise da série de desemprego em Sergipe

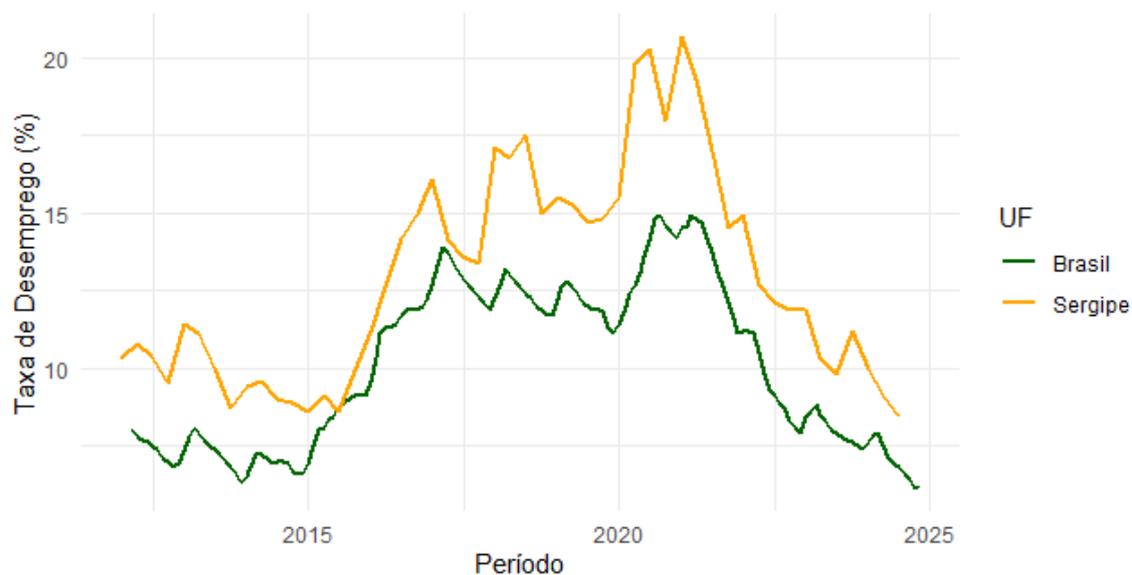
Analisando a série original da taxa de desemprego em Sergipe, apresentada na Figura 10, em períodos trimestrais de 2012 a 2024, nota-se que ela exibe uma dinâmica semelhante à observada na série nacional, contudo a mesma apresenta algumas particularidades regionais, ao observar o período de 2012 a 2015 onde a taxa de desemprego mantém-se em patamares relativamente baixos, apresentando pequenas oscilações e com uma leve tendência de queda; mas a partir de 2015, ocorre um forte crescimento, alcançando um pico entre 2020 e 2021, o que é justificado, pela epidemia do corona vírus, seguindo o mesmo comportamento observado na série nacional, o que evidencia a influência de fatores macroeconômicos comuns, mas com possíveis variações de intensidade no nível estadual.

Figura 10: Série original de desemprego em Sergipe



Fonte: Elaborado pelo autor

Figura 11: Série de desemprego Brasil x Sergipe



Fonte: Elaborado pelo autor

A partir de 2021 a taxa de desemprego em Sergipe inicia um movimento de queda contínua, indicando uma recuperação progressiva do mercado de trabalho, da

mesma forma que ocorreu a nível nacional, no entanto, observa-se que a taxa de desemprego permanece em um nível ainda elevado em comparação com a série do Brasil. Apesar da boa recuperação, Sergipe demonstrou uma maior sensibilidade aos eventos ocorridos, durante a pandemia, Sergipe apresentou picos mais acentuados, chegando a taxas superiores a 20%, quase 6% acima do pico da série nacional, conforme é apresentado na Figura 11. Essa diferença pode ser atribuída à informalidade, que tem um impacto maior nos estados do nordeste, sendo refletido nas taxas de desemprego (ABDALA, 2023).

### 6.2.1 Teste de estacionariedade da série de desemprego em Sergipe

Os resultados dos testes de estacionariedade aplicados à série apresentaram resultados conflitantes, o Dickey-Fuller test ADF, apresentou um p-valor de 0.9602, não rejeitando a hipótese nula de presença de raiz unitária, indicando que a série não é estacionária, porém ao realizar o teste KPSS, foi encontrado um p-valor igual a 0.0675, que também não rejeitou a hipótese nula, definindo a série como estacionária em nível, tal discrepância entre os testes pode sinalizar uma complexidade na estrutura da série temporal, conforme é apresentado na Tabela 5.

Tabela 5: Resultado do teste de estacionariedade da série de desemprego em Sergipe

Teste	Estatística do Teste	Ordem da Defasagem	p-valor	Resultado	Explicação
<b>ADF</b> (Augmented Dickey-Fuller)	-0,75127	3	0,9602	Não Rejeita H0	A série temporal provavelmente não é estacionária.
<b>KPSS</b> (Kwiatkowski-Phillips-Schmidt-Shin)	0,42241	3	0,0675	Não Rejeita H0	A série temporal provavelmente é estacionária em nível.

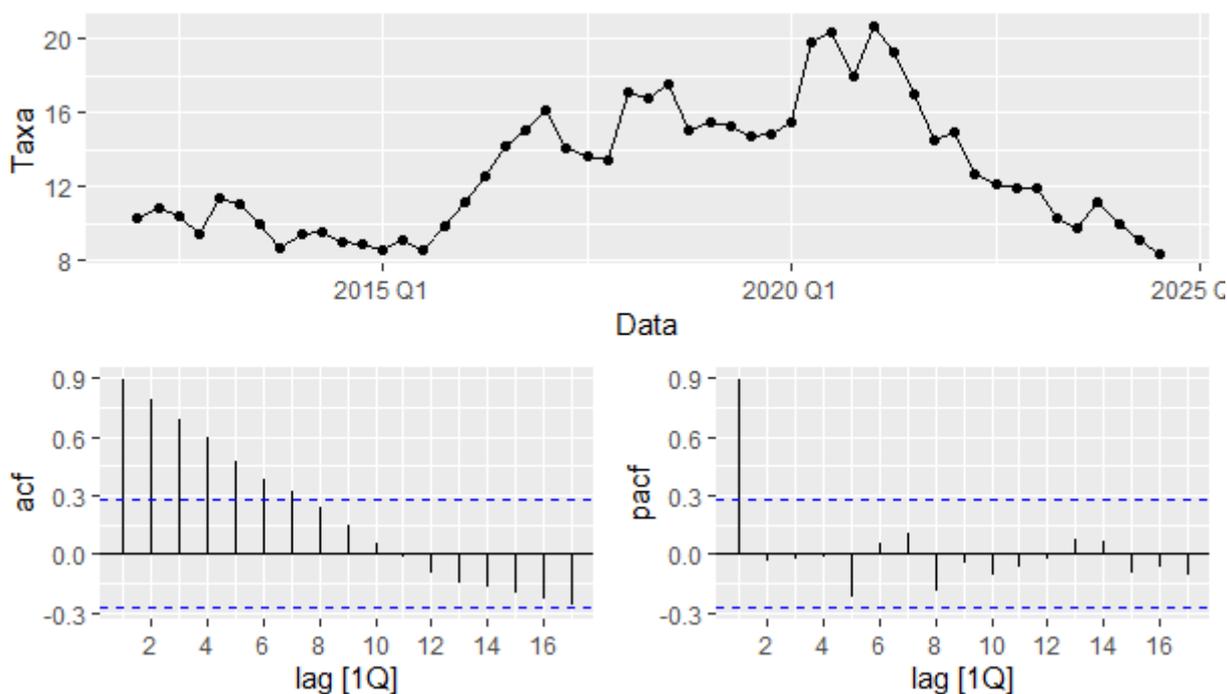
Fonte: Elaborado pelo autor

A estacionariedade é uma propriedade crucial para a aplicação de modelos de séries temporais, como o ARIMA, tal inconsistência nos resultados dos testes ADF e KPSS demanda cautela na escolha do modelo, uma vez que a presença de uma raiz unitária, indicada pelo Dickey-Fuller test ADF, sugere a necessidade de diferenciação da série para torná-la estacionária, por outro lado, a estacionariedade em nível,

apontada pelo teste KPSS, indica que a série pode ser modelada sem diferenciação; diante dessa ambiguidade, optou-se pela aplicação de diferenciação.

Na Figura 12 está apresentado o conjunto de gráficos da série temporal original da taxa de desemprego em Sergipe e suas respectivas funções de autocorrelação (ACF) e autocorrelação parcial (PACF), a série exibe uma tendência clara, caracterizada pelo aumento das taxas entre 2014 e 2020 e posterior queda até 2024, essa forte tendência indica que a série não é estacionária, conforme indicado pelo Dickey-Fuller test ADF, a função ACF mostra uma correlação significativa para diversos lags, evidenciando uma dependência de longo prazo entre as observações, já a PACF apresenta um decaimento progressivo, sugerindo a necessidade de um modelo autorregressivo ou a aplicação de uma diferenciação para eliminar a tendência.

Figura 12: Análise da ACF e PACF da série de desemprego em Sergipe

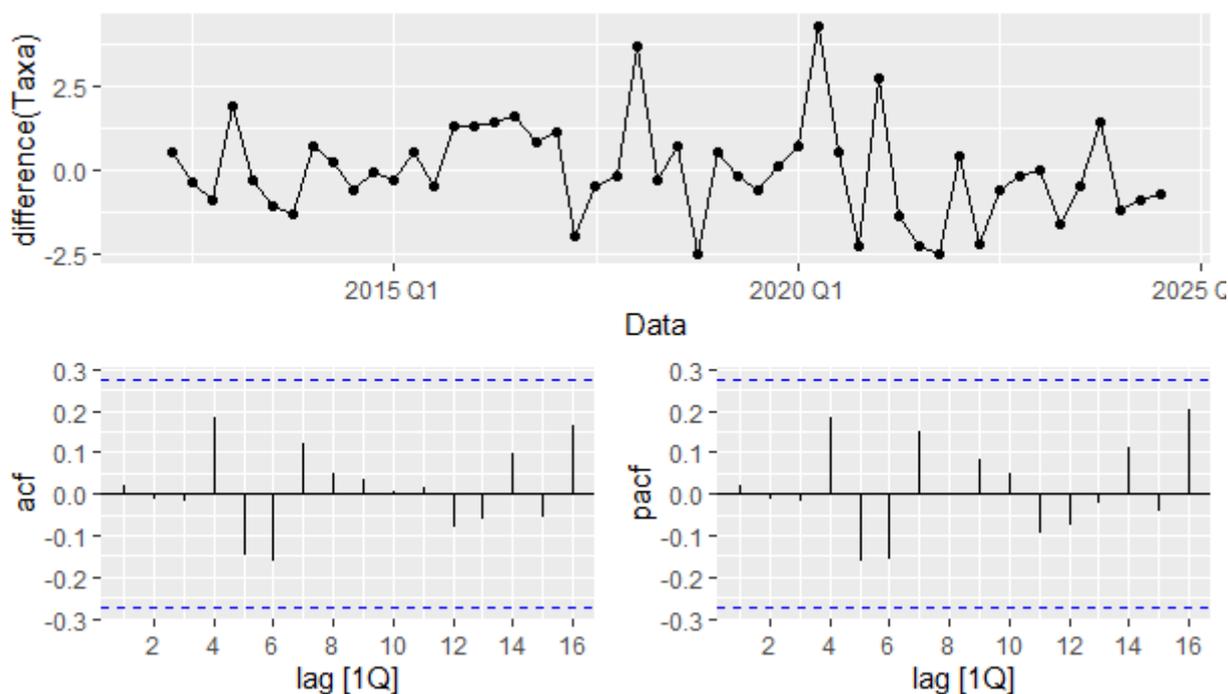


Fonte: Elaborado pelo autor

Na Figura 13 é apresentado a série após a aplicação da primeira diferenciação, onde evidencia que os valores oscilam de forma mais equilibrada em torno de uma média constante, eliminando a tendência de longo prazo observada na série original, essa estabilização indica que a série diferenciada se aproxima de um comportamento estacionário, essencial para a modelagem com processos estocásticos, a função ACF

agora mostra correlações significativamente menores, com picos apenas nos primeiros lags, indicando que a dependência temporal foi reduzida, no gráfico PACF é observado a presença de poucos lags significativos.

Figura 13: Análise da ACF e PACF da série de desemprego em Sergipe após aplicação da primeira diferença



Fonte: Elaborado pelo autor

### 6.2.2 Aplicação e verificação do modelo a série de desemprego em Sergipe

Para escolha do melhor modelo, foi criado a Tabela 6, onde está descrito um breve resumo de cada modelo, e suas aplicações a série; o ARIMA(1,1,0), o ARIMA(0,1,1) e o ARIMA(1,0,0), sendo o primeiro termo de cada modelo o parâmetro autorregressivo, o segundo o número de diferenças e o terceiro a média móvel.

Tabela 6: Lista de modelos testados na série de desemprego de Sergipe com diferenciação

<b>Modelo</b>	<b>Resumo</b>
<b>ARIMA(1,1,0)</b>	Modelo ARIMA com 1 termo autorregressivo (AR), 1 diferenciação e 0 termos de média móvel (MA). Adequado para séries temporais com tendência e dependência de valores passados.
<b>ARIMA(0,1,1)</b>	Modelo ARIMA com 0 termos AR, 1 diferenciação e 1 termo MA. Adequado para séries temporais com tendência e dependência de erros de previsão passados.
<b>ARIMA(1,0,0) com média</b>	Modelo ARIMA com 1 termo AR, 0 diferenciações e 0 termos MA, incluindo uma média. Adequado para séries estacionárias com dependência de valores passados.

Fonte: Elaborado pelo autor

Analisando os resultados da Tabela 7, observa-se que o modelo ARIMA(0,1,1) e o modelo ARIMA(1,1,0) apresentam os menores valores de AIC, AICc e BIC, indicando um melhor ajuste aos dados em comparação com o modelo ARIMA(1,0,0), especificamente, ambos apresentam os mesmos valores de  $\sigma^2$ , log\_lik, AIC, AICc e BIC, o que significa que, estatisticamente, eles têm o mesmo desempenho, entretanto, foi testado ambos, mas o modelo ARIMA(0,1,1) apresentou um melhor ajuste aos dados.

Tabela 7: Resultado do teste aplicado aos modelos da série de desemprego de Sergipe

<b>Modelos</b>	<b>sigma2</b>	<b>log_lik</b>	<b>AIC</b>	<b>AICc</b>	<b>BIC</b>
<b>ARIMA(0,1,1)</b>	2.10	-88.9	182.0	182.0	186.0
<b>ARIMA(1,1,0)</b>	2.10	-88.9	182.0	182.0	186.0
<b>ARIMA(1,0,0)</b>	2.04	-90.4	187.0	187.0	193.0

Fonte: Elaborado pelo autor

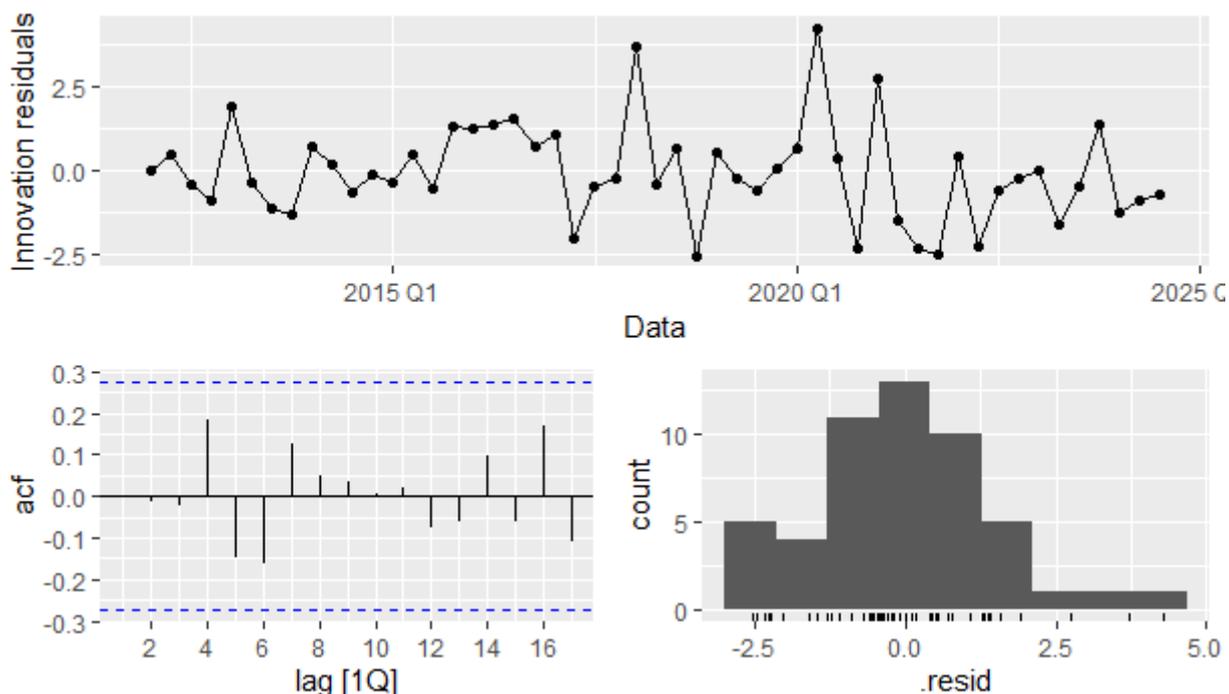
O conjunto de resíduos apresentado na Figura 14, possui um comportamento estacionário sem tendência aparente ao longo do tempo, a análise da autocorrelação (ACF) indica que os valores dos resíduos não apresentam correlações estatisticamente significativas para a maioria das defasagens, um indicativo de

comportamento próximo ao de um ruído branco, e observando a distribuição dos resíduos no histograma, onde há uma aproximação da distribuição normal, com uma leve assimetria.

O teste Ljung-Box apresentou um p-valor de 0,766, indicando que não há presença de autocorrelação nos resíduos, de modo que não há evidências para rejeitar a hipótese nula de que os resíduos são não correlacionados; também foi testada a normalidade, onde o teste Shapiro-Wilk encontrou um p-valor de 0,05609, desse modo não rejeitamos a hipótese nula de que os dados seguem distribuição normal.

Na ideia de ajuste para o modelo, obtivemos resíduos relativamente bem ajustados, dado a semelhança de um ruído branco, que pode ser identificado pela menor presença de padrões estruturais na autocorrelação e pela distribuição mais simétrica dos resíduos em torno da média, apesar de conter pequenas discrepâncias, o modelo de modo geral atende a necessidade do intuito abordado nessa pesquisa.

Figura 14: Análise dos resíduos do modelo aplicado a série de desemprego de Sergipe

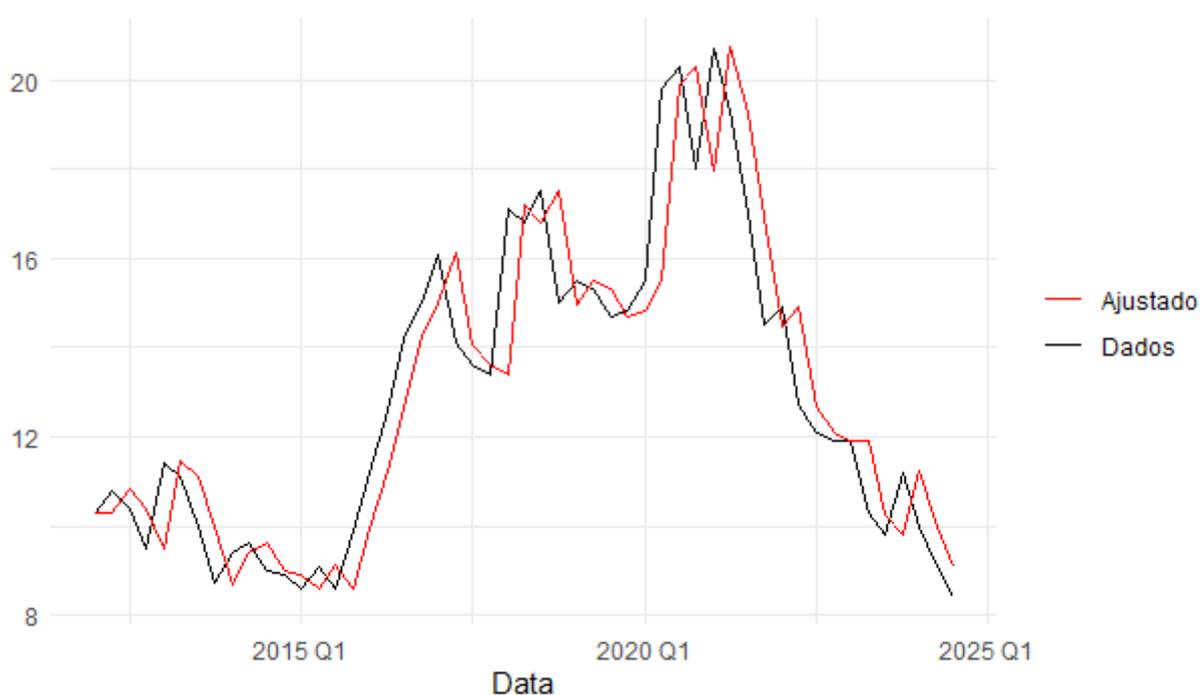


Fonte: Elaborado pelo autor

Ao realizar o comparativo da série original com o modelo ajustado, apresentado na Figura 15, ficou evidenciado que o ajuste demonstra uma boa capacidade de

capturar a tendência e os principais padrões da série temporal, uma vez que o ajuste acompanha de perto a evolução da taxa de desemprego ao longo do período analisado, refletindo tanto os momentos de crescimento acentuado quanto os períodos de queda, a suavização promovida pelo modelo contribui para reduzir pequenas variações aleatórias, esse comportamento é especialmente visível nos períodos de alta entre 2015 e 2021, onde o modelo segue fielmente os picos e vales observados nos dados reais, onde a proximidade entre as linhas preta dados reais e vermelha modelo ajustado, conseguiram capturar bem a dinâmica subjacente.

Figura 15: Série original de desemprego em Sergipe x Modelo ajustado



Fonte: Elaborado pelo autor

Apesar do modelo possuir uma defasagem acentuada em relação a série original, ele tende a se destacar por sua capacidade de replicar o comportamento geral da série sem extrapolar os ajustes dos ruídos da amostra, observando as variações sazonais de curto prazo, nota-se que a suavização, foi bem modelada, permitindo uma melhor interpretação da tendência predominante, a escolha do modelo  $ARIMA(0,1,1)$  sem parâmetros autorregressivos complexos foi a mais adequada, para um ajuste mais parcimonioso e generalizável, tendo em vista, que o modelo mostrou um desempenho satisfatório na fase mais recente da série, onde as

projeções seguem de perto a queda no desemprego, indicando que a estrutura do modelo é adequada para capturar a evolução da variável no curto prazo.

### 6.2.3 Previsão de taxas futuras para Sergipe

Ao realizar as previsões para a taxa de desemprego em Sergipe em um horizonte de seis trimestres, foi observado uma tendência de queda, condizendo com o comportamento da série nos últimos períodos, contudo, há uma faixa de incerteza crescente à medida que o horizonte de previsão se estende, as áreas sombreadas em azul representam os intervalos de confiança de 80% e 95%, indicando a variação esperada das projeções, a faixa mais clara sugere uma maior incerteza nos valores extremos, enquanto a mais escura concentra as previsões mais prováveis, sabendo que esse comportamento é comum em modelos de séries temporais, dado que a precisão diminui conforme a previsão se distancia dos dados observados, mas apesar da incerteza, o modelo sugere que a taxa de desemprego permanecerá em níveis mais baixos em comparação aos períodos de pico anteriores, conforme é visto na Figura 16.



Fonte: Elaborado pelo autor

## 7 CONCLUSÃO

O presente trabalho objetivou realizar uma análise das séries históricas de desemprego no Brasil e no estado de Sergipe, no período de 2012 a 2024, fazendo uso da linguagem de programação estatística R, aplicando modelos da família SARIMA, e fazendo uso do método Box e Jenkins, para obter modelos capazes de acompanhar o comportamento das séries em questão e posteriormente realizar previsões, de modo a entender as variações apresentadas por ambas as séries.

A análise voltada à série nacional demonstrou possuir um bom ajuste, captando os eventos ocorridos ao longo do tempo, como a tendência de redução das taxas no início da série, e principalmente nos picos observados durante a recessão de 2014 e a pandemia causada pela covid-19 em 2019, onde ambos foram responsáveis por dois dos piores cenários já enfrentados no Brasil. Ao realizar as previsões foi observada uma limitação no horizonte de previsões, ficando claro que o modelo aumenta muito a abrangência do nível de confiança, a medida que estende-se o horizonte de previsão, apesar disso, as previsões indicam um possível cenário favorável com redução das taxas de desemprego.

Na análise da série voltada a Sergipe ficou evidente a necessidade de mais dados, uma vez que apesar do modelo ajustado ter conseguido capturar o comportamento geral da série, ele se mostrou relativamente deficiente, pois foi insuficiente, por não conseguir acompanhar a dinâmica dos dados, e os períodos de com alta volatilidade não foram captados precisamente. Nas previsões realizadas o modelo demonstrou que as taxas tendem a se estabilizar nos próximos 4 trimestres, mas devido a defasagem aparente do modelo, esse resultado torna-se imprevisível.

A análise das duas séries apresentou resultados conclusivos sobre a sensibilidade do mercado de trabalho a eventos ocorridos a nível nacional, uma vez que o mercado de trabalho apresentou aumento nas taxas de desemprego em períodos distintos, como a pandemia da covid-19. Um outro ponto observado é falta de dados para uma melhor análise no estado de Sergipe, tendo em vista que ele se mostrou mais sensível aos eventos ocorridos, e possui uma recuperação mais lenta pós-eventos.

## REFERÊNCIAS

ABDALA, Vitor. **Desocupação cresce em 16 unidades da Federação no primeiro trimestre**. Agência Brasil 18 de maio de 2023. Disponível em: [Desocupação cresce em 16 unidades da Federação no primeiro trimestre | Agência Brasil](#).

BALASSIANO, Marcel Grillo. **Recessão Brasileira (2014-2016): Uma Análise por Meio do Método do Controle Sintético do PIB, PIB per capita, Taxa de Investimento e Taxa de Desemprego**. Macroeconomia, Economia Monetária e Finanças, artigo homônimo selecionado para apresentação no 46º Encontro Nacional de Economia – ANPEC 2018.

BARROS, Ricardo Paes de; CAMARGO, José Marcio e MENDONÇA, Rosane. **TEXTO PARA DISCUSSÃO Nº 478, A ESTRUTURA DO DESEMPREGO NO BRASIL**. Maio de 1997.

BECKER, Marcel Henrique. **Modelos para Previsão em Séries Temporais: uma Aplicação para a Taxa de Desemprego na Região Metropolitana de Porto Alegre**. Monografia apresentada para obtenção de grau de Bacharel em Estatística, Porto Alegre, 15 de dezembro de 2010.

BOX, George E. P.; JENKINS, Gwilym M. and REINSEL, Gregory C. **Time series analysis: forecasting and control**, 4º ed., San Francisco: Holden-Day, 2008.

COSTA, Simone da Silva. **Pandemia e desemprego no Brasil**. Universidade Federal do Rio Grande do Norte / Departamento de Arquitetura e Urbanismo, Natal / RN – Brasil, agosto de 2020.

CRYER, JD; CHAN, KS. **Time series analysis with applications in R**. 2. Ed, Nova Iorque: Springer, 2008.

DIEESE. Sistema PED – Metodologia utilizada. Disponível Acesso 10/03/2025

FONSECA, Jairo Simon da; MARTINS, Gilberto de Andrade e TOLEDO, Geraldo Luciano. **Estatística Aplicada 1º EDIÇÃO**. Editora Atlas S. A. 1978.

GAUTIÉ, Jérôme. **DA INVENÇÃO DO DESEMPREGO SUA DESCONSTRUÇÃO**. MANA 4(2):67-83, 1998

GUJARATI, Domador N. e POTER, Dawn C. **ECONOMETRIA BASICA Quinta edição**. AMGH Editora Ltda. 2011.

HYNDMAN, Rob J. **Forecasting: Principles & Practice**. 23-25 September 2014 University of Western Australia.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílios Contínua Notas técnicas Versão 1.18**. Rio de Janeiro, 2024.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. [Busca | IBGE](#). Acessado em 22 de fevereiro de 2025.

IBGE. Notas metodológicas da PNAD Contínua. Disponível em [IBGE | Portal do IBGE | IBGE](#). Acesso em: 10/03/2025.

LOUREDO, Fábio Moita; PEREIRA, Pedro Henrique Fonseca; JÚNIOR, Péricles Dias De Souza e SANTOS, Renan André Barbosa. **PANDEMIA E DESEMPREGO NO BRASIL: IMPACTOS E CONSEQUÊNCIAS PARA OS JOVENS NO MERCADO DE TRABALHO**. Revista Eletrônica Do Instituto De Humanidades, fevereiro de 2022.

MARQUES, Guilherme de O. L. e FAVA, Vera Lúcia. **PERSISTÊNCIA E MEMÓRIA LONGA SAZONAL NA SÉRIE DE DESEMPREGO DA REGIÃO METROPOLITANA DE SÃO PAULO**. Economia Aplicada, v. 15, n. 2, 2011, pp. 177-198.

MENEZES, Andrey Ivale e CUNHA, Marina Silva da. **Uma análise da duração do desemprego no Brasil (2002 - 2011)**. R. Bras. Eco. de Emp. 2013.

MORETTIN, Pedro; TOLOI, Clélia. **Análises de Séries Temporais**, 2. ed. São Paulo: Blucher, 2006.

MORETTO, Amilton J. e PRONI, Marcelo Weishaupt. **O DESEMPREGO NO BRASIL: ANÁLISE DA TRAJETÓRIA RECENTE**. Economia e Desenvolvimento, Recife (PE), v. 10, n. 2, 2011.

NUSAIR, Salah A. **TESTING THE VALIDITY OF PURCHASING POWER PARITY FOR ASIAN COUNTRIES DURING THE CURRENT FLOAT**. JOURNAL OF ECONOMIC DEVELOPMENT. Volume 28, Number 2, December 2003.

POCHMANN, Marcio. **DESEMPREGO: Ajuste econômico de desemprego recente no Brasil metropolitano**. ESTUDOS AVANÇADOS 29 (85), 2015.

PUC-RIO/ Portal de Microdados Brasileiro 2011. Disponível em: [Data Zoom](#). Acessado em 09 de março de 2025.

QUADROS, Waldir Jose de. **Classes sociais e desemprego no Brasil dos anos 1990**. Economia e Sociedade, Campinas, v. 12, n. 1 (20), p. 109-135, jan./jun. 2003.

RIBAS, Rafael Perez e SOARES, Sergei Suarez Dillon. **TEXTO PARA DISCUSSÃO Nº 1348, SOBRE O PAINEL DA PESQUISA MENSAL DE EMPREGO (PME) DO**

**IBGE**. Rio de Janeiro, agosto de 2008. Disponível em: [td\\_1348.pdf](#). acessado em 09/03/2025.

ROSSI, Alfredo. **Diagnóstico em Regressão**, LAMFO, 2019. Disponível em: <https://lamfounb.github.io/2019/04/13/Diagnostico-emRegressao/#:~:text=Os%20res%C3%ADduos%20indicam%20a%20varia%C3%A7%C3%A3o,e%20n%C3%A3o%20confi%C3%A1veis%20para%20infer%C3%AAnca>. Acesso em: janeiro de 2025

RSTUDIO, <https://www.rstudio.com/products/rstudio> . Acesso: janeiro de 2025.

SÁFADI, T. **Uso de séries temporais na análise de vazão de água na represa de Furnas**. Ciência e Agrotecnologia, Lavras, v. 28, n. 1, 2004.

SANTOS, Fernando Siqueira dos. **Ascensão e Queda do Desemprego no Brasil: 1998-2012**. Área ANPEC: 13 – Economia do Trabalho, 2013.

SANTONE, Paula Corina. **Festas de fim de ano e trabalho temporário: Pontos de atenção**. MIGALHAS DE PESO, 13 de novembro de 2024. Disponível em:

SAUTHIER, Ingrid Liebeskind. **Histoire de la définition du chômage. Courier des Statistiques, e en.127**. mai-a-oût2009. Disponível em: [http://www.insee.fr/fr/ffc/docs\\_ffc/cs127b](http://www.insee.fr/fr/ffc/docs_ffc/cs127b) pdf. Acesso em: 10 mar. 2025.

SILVEIRA, Raiane Padilha. **TAXA DE DESEMPREGO NO BRASIL: VARIÁVEIS DETERMINANTES PARA PREVER O PATAMAR**. Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Economia, Porto Alegre, 2022.

SILVA, Bruno dos Santos. **CURVA DE PHILLIPS: UMA CONSTRUÇÃO PARA O BRASIL (2002 – 2013)**. DISSERTAÇÃO DE MESTRADO, São Cristóvão - SE 2015.

SILVA, Fábio José Ferreira e PIRES, Leandro Siani. **Evolução do Desemprego no Brasil no Período 2003-2013: análise através das probabilidades de transição**. Fevereiro, 2014.

SOUZA, Reinaldo Castro e CAMARGO, Maria Emilia. **Análise e Previsão de Séries Temporais: Modelos ARIMA**. Ijuí: SEDIGRAF, 1996.

CN, Congresso Nacional. **Medida Provisória nº 936, de 2020**. Disponível em: [MPV 936/2020 - Congresso Nacional](#). Acessado em 23/01/2025.

TADDY, Matt. **Ciência de dados de negócios: combinando aprendizado de máquina e economia para otimizar, automatizar e acelerar decisões de negócios**. Nova York: McGraw-Hill. 2019.

TRAPLETTI, A.; HORNIK, K.; LEBARON, B. tseries: Time Series Analysis and Computational Finance. R package version 0.10-47, [2023]. Disponível em: <https://CRAN.R-project.org/package=tseries>. Acesso em: 11 março 2025.

VELOSO, Fernando; FILHO, Fernando de Holanda Barbosa; PERUCHETTI, Paulo e FEIJÓ, Janaína. **Mercado de Trabalho no Brasil: Evolução, Efeitos da Pandemia, Perspectivas e Propostas**. FGV IBRE, 07 de abril de 2022.

## APRÊNDICE A- Script no R

```

## Pacotes
library(dplyr)
library(xts)
library(fpp2)
library(tseries)
library(lubridate)
library(readr)
library(readxl)
library(fpp3)
library(ggplot2)
library(forecast)
## CARREGANDO A BASE DO BRASIL
dados<-read_excel(file.choose())
brasil<-dados %>%
  add_column(Periodo=yearmonth("2012 Mar") + 0:152)%>%
  as_tsibble(index = Periodo)
autoplot(brasil)
# Exibir a série temporal
#plot(dados.ts, main = "Série Temporal da Taxa de Desemprego", ylab = "Taxa (%)",
xlab = "Ano")
autoplot(brasil) +
  geom_line(color = "darkgreen") +
  xlab("Ano") + ylab("Taxa de Desemprego (%)") +
  theme_minimal()
## Testando a estacionariedade da série -----
adf.test(brasil$Taxa)
kpss.test(brasil$Taxa)
## Número de diferenças
unitroot_ndiffs(brasil$Taxa)
## Autocorrelacao (ACF) e Autocorrelacao Parcial (PACF)
brasil%>% ACF(Taxa) %>%
  autoplot()+

```

```

labs(x="Lags",
      y="Autocorrelacao")+
theme_minimal()
brasil%>% PACF(Taxa) %>%
autoplot()+
labs(x="Lags",
      y="Autocorrelacao Parcial")+
theme_minimal()
brasil%>%gg_tsdisplay(Taxa,
                      plot_type='partial', lag=36) +
labs(title="Serie diferenciada", y="")
## Autocorrelação serie diferenciada
brasil %>%
  gg_tsdisplay(difference(Taxa,2), plot_type='partial')
brasil_fit <-Taxa
brasil_fit <- brasil %>%
  model(arima223 = ARIMA(Taxa ~ pdq(2,2,3)+PDQ(0,0,0)),
        arima322 = ARIMA(Taxa ~ pdq(3,2,2)+PDQ(0,0,0)),
        stepwise = ARIMA(Taxa),
        search = ARIMA(Taxa, stepwise=FALSE))
brasil_fit
### Modelos menor AICc
glance(brasil_fit) |> arrange(AICc) |> select(.model:BIC)
brasil_fit |>
  select(search) |>
  gg_tsresiduals()
# Teste de Ljung-Box para série mensal
ljung_box_result <- augment(brasil_fit) %>%
  filter(.model == "search") %>%
  features(.resid, ljung_box, lag = 24)
print(ljung_box_result)
# Teste de normalidade de Shapiro-Wilk para resíduos
arimasearch_res<-augment(brasil_fit) %>%
  filter(.model == "search") %>%
  select(.innov)

```

```

shapiro.test(arimasearch_res$.innov)
hist(arimasearch_res$.innov)
## Ajuste x original
augment(brasil_fit) %>%
  filter(.model == "search") %>%
  ggplot(aes(x = Período)) +
  geom_line(aes(y = Taxa, colour = "Brasil")) +
  geom_line(aes(y = .fitted, colour = "Ajustado")) +
  labs(y = "Taxa de Desemprego", x = "Período") +
  scale_colour_manual(values = c("Brasil" = "black", "Ajustado" = "red")) +
  guides(colour = guide_legend(title = NULL)) +
  theme_minimal()
#Previsões para 12 meses
brasil_fit |>
  forecast(h=12) |>
  filter(.model=='search') |>
  autoplot(brasil) + theme_minimal()
#Importando dados do percentual de pessoas ocupadas
dados1<-read_excel(file.choose())
ocupados<-dados1 %>%
  add_column(Período=yearmonth("2012 Mar") + 0:153)%>%
  as_tsibble(index = Período)
autoplot(ocupados)
## Carregando a base de dados de Sergipe
dados2<-read_excel(file.choose())
## Filtrando sergipe
sergipe<-dados2 %>% filter(Estado=="Sergipe") %>%
  pivot_longer(.,
    cols = `2012 T1`:`2024 T3`,
    values_to = "Taxa"
  )
## Adicionando a data
sergipe<- sergipe %>%
  add_column(Data=yearquarter("2012 Q1") + 0:50)

```

```

## Criando a serie temporal
sergipe_ts<- sergipe %>% tsibble(index = Data)
autoplot(sergipe_ts, vars(Taxa)) +
  labs(y = "Taxa de Desemprego", x = "Período") +
  theme_minimal()
## Testando estacionariedade
tseries::adf.test(sergipe$Taxa) # não rejeita
tseries::kps.test(sergipe$Taxa) # não rejeita
## Resultado do teste inconclusivo
## Autocorrelação da serie original
sergipe_ts %>%
  gg_tsdisplay(Taxa, plot_type='partial')
## Autocorrelação serie diferenciada
sergipe_ts %>%
  gg_tsdisplay(difference(Taxa), plot_type='partial')
sergipe_fit <- sergipe_ts %>%
  model(arima110 = ARIMA(Taxa ~ pdq(1,1,0)),
        arima011 = ARIMA(Taxa ~ pdq(0,1,1)),
        stepwise = ARIMA(Taxa),
        search = ARIMA(Taxa, stepwise=FALSE))
sergipe_fit
#### Modelos menor AICc
glance(sergipe_fit) |> arrange(AICc) |> select(.model:BIC)
sergipe_fit |>
  select(arima011) |>
  gg_tsresiduals()
# Teste de Ljung-Box para série trimestral de Sergipe
ljung_box_result <- augment(sergipe_fit) %>%
  filter(.model == "arima011") %>%
  features(.resid, ljung_box, lag = 12, dof = 2) # dof = p + q
print(ljung_box_result)
# Teste de normalidade de Shapiro-Wilk para resíduos
sarima_res <- augment(sergipe_fit) %>%
  filter(.model == "arima011") %>%

```

```

select(.resid)
shapiro.test(sarima_res$.resid)
### Ajuste x original
augment(sergipe_fit) %>%
  filter(.model=="arima011")%>%
  ggplot(aes(x = Data)) +
  geom_line(aes(y = Taxa, colour = "Dados")) +
  geom_line(aes(y = .fitted, colour = "Ajustado")) +
  labs(y = NULL) +
  scale_colour_manual(values=c(Dados="black",
                               Ajustado="red")) +
  guides(colour = guide_legend(title = NULL))+
  theme_minimal()
## Previsões para 4 trimestres
sergipe_fit |>
  forecast(h=4) |>
  filter(.model=='arima011') |>
  autoplot(sergipe_ts) + theme_minimal()
sergipe_fit |>
  forecast(h=12) |>
  filter(.model=='arima011') |>
  autoplot(sergipe_ts) + theme_minimal()
# Extraí dados do ajuste e converte para tibble com datas no formato Date
brasil_data <- augment(brasil_fit) %>%
  filter(.model == "search") %>%
  as_tibble() %>%
  select(Periodo, Taxa) %>%
  mutate(
    Periodo = as_date(Periodo), # converte yearmonth para Date
    UF = "Brasil"
  )
sergipe_data <- augment(sergipe_fit) %>%
  filter(.model == "arima011") %>%
  as_tibble() %>%

```

```
select(Data, Taxa) %>%
rename(Periodo = Data) %>%
mutate(
  Periodo = as_date(Periodo), # converte yearquarter para Date
  UF = "Sergipe"
)
# Agora a junção funciona sem erro
series_comparadas <- bind_rows(brasil_data, sergipe_data)
# Gráfico comparando as séries originais
ggplot(series_comparadas, aes(x = Periodo, y = Taxa, color = UF)) +
  geom_line(size = 1) +
  labs(
    title = "",
    x = "Período",
    y = "Taxa de Desemprego (%)"
  ) +
  scale_color_manual(values = c("Brasil" = "darkgreen", "Sergipe" = "orange")) +
  theme_minimal()
```