



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE CIÊNCIAS EXATAS E DA TERRA  
DEPARTAMENTO DE ESTATÍSTICA  
BACHARELADO EM ESTATÍSTICA

Lucas Medeiros dos Santos

**Modelagem de Séries Temporais para Previsão da  
Quantidade de Micro e Pequenas Empresas atendidas  
pelo Sebrae/RN**

Natal - RN

16 de agosto de 2024

**Lucas Medeiros dos Santos**

**Modelagem de Séries Temporais para Previsão da  
Quantidade de Micro e Pequenas Empresas atendidas  
pelo Sebrae/RN**

Monografia de Graduação apresentada ao  
Departamento de Estatística do Centro de  
Ciências Exatas e da Terra da Universidade  
Federal do Rio Grande do Norte como re-  
quisito parcial para a obtenção do grau de  
Bacharel em Estatística.

Universidade Federal do Rio Grande do Norte  
Centro de Ciências Exatas e da Terra  
Departamento de Estatística

Orientadora:

Profa. Dra. Andressa Nunes Siroky

**Natal - RN**

16 de agosto de 2024

Universidade Federal do Rio Grande do Norte - UFRN  
Sistema de Bibliotecas - SISBI  
Catalogação de Publicação na Fonte. UFRN - Biblioteca Setorial Prof. Ronaldo Xavier de Arruda - CCET

Santos, Lucas Medeiros dos.

Modelagem de séries temporais para previsão da quantidade de micro e pequenas empresas atendidas pelo Sebrae/RN / Lucas Medeiros dos Santos. - 2024.

60 f.: il.

Monografia (graduação) - Universidade Federal do Rio Grande do Norte, Centro de Ciências Exatas e da Terra, Curso de Estatística. Natal, RN, 2024.

Orientação: Profa. Dra. Andressa Nunes Siroky.

1. Séries temporais - Monografia. 2. Pequenos negócios - Monografia. 3. Regressão - Monografia. 4. SARIMA - Monografia. 5. Sebrae - Monografia. I. Siroky, Andressa Nunes. II. Título.

RN/UF/CCET

CDU 519.246.8



**MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO  
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE CIÊNCIAS EXATAS E DA TERRA  
CURSO DE GRADUAÇÃO EM ESTATÍSTICA**

**ATA DE DEFESA DE MONOGRAFIA  
BACHARELADO EM ESTATÍSTICA**

Às 14h00min do dia 18 de junho de 2024, compareceu para defesa da monografia do curso de graduação em Estatística o aluno *Lucas Medeiros dos Santos* tendo como Título da Monografia “Modelagem de Séries Temporais para Previsão da Quantidade de Micro e Pequenas Empresas atendidas pelo Sebrae/RN”. Constituíram a Banca Examinadora os professores Andressa Nunes Siroky (presidente), Luz Milena Zea Fernández (examinadora) e Marcus Alexandre Nunes (examinador). Após a apresentação e as observações dos membros da banca avaliadora, ficou definido que o trabalho foi considerado aprovado.

---

Prof. Dra. Andressa Nunes Siroky (DEST/UFRN)  
Orientadora

---

Prof. Dra. Luz Milena Zea Fernández (DEST/UFRN)  
Examinador

---

Prof. Dr. Marcus Alexandre Nunes (DEST/UFRN)  
Examinador



---

*ATA Nº 3/2024 - EST/CCET (12.02)*

*(Nº do Protocolo: NÃO PROTOCOLADO)*

*(Assinado digitalmente em 19/06/2024 11:50 )*

ANDRESSA NUNES SIROKY  
PROFESSOR DO MAGISTERIO SUPERIOR  
EST/CCET (12.02)  
Matrícula: ###623#6

*(Assinado digitalmente em 19/06/2024 12:11 )*

LUZ MILENA ZEA FERNANDEZ  
PROFESSOR DO MAGISTERIO SUPERIOR  
EST/CCET (12.02)  
Matrícula: ###835#3

*(Assinado digitalmente em 19/06/2024 09:57 )*

MARCUS ALEXANDRE NUNES  
PROFESSOR DO MAGISTERIO SUPERIOR  
EST/CCET (12.02)  
Matrícula: ###663#8

Visualize o documento original em <https://sipac.ufrn.br/documentos/> informando seu número: 3, ano: 2024, tipo:  
ATA, data de emissão: 18/06/2024 e o código de verificação: 50efa74ea8

# Agradecimentos

Gostaria de expressar minha profunda gratidão a todos aqueles que me apoiaram ao longo desta jornada acadêmica. Concluir este curso representa um marco significativo em minha vida, uma conquista que demandou dedicação e determinação, e que só foi possível graças ao apoio inestimável de muitas pessoas incríveis.

Em primeiro lugar gostaria de agradecer a minha família. Seu apoio incondicional, encorajamento e compreensão foram fundamentais para que eu pudesse alcançar este momento tão importante. A eles expresso minha mais sincera gratidão.

Aos meus amigos e colegas de estágio e faculdade, agradeço por estarem ao meu lado, compartilhando desafios, vitórias e aprendizados. Cada um de vocês contribuiu de maneira única para o meu crescimento pessoal e acadêmico.

Aos meus professores, sou imensamente grato pela orientação, conhecimento compartilhado e disponibilidade para esclarecer todas as minhas dúvidas. Em especial, gostaria de expressar minha profunda gratidão a minha orientadora, cuja paciência, orientação e incentivo foram fundamentais para o desenvolvimento deste trabalho.

A todos que de alguma forma contribuíram para esta jornada, o meu mais sincero obrigado. Cada palavra de incentivo, cada gesto de apoio, foi fundamental para que eu pudesse alcançar este momento tão especial. Que nossa jornada juntos possa continuar sendo marcada por aprendizado, crescimento e sucesso. Muito obrigado por tudo.

*“Be the best version of yourself in anything  
you do. You don’t have to live anybody  
else’s story”*

*Stephen Curry*

# Resumo

Este estudo investiga a aplicação da análise de séries temporais aos dados mensais sobre o número de Microempreendedores Individuais (MEI), Microempresas (ME) e Empresas de Pequeno Porte (EPP) atendidos pelo Sebrae/RN, no período de janeiro de 2014 a dezembro de 2023, com o objetivo de prever a quantidade futura de atendimentos distintos dessas categorias empresariais. O principal objetivo é ajustar um modelo de regressão dinâmica com erros SARIMA (SARIMAE), utilizando a quantidade de atendimentos realizados pelo Sebrae como variável regressora. O estudo seleciona o modelo mais adequado para cada categoria de interesse, realiza previsões para 24 passos à frente e compara as previsões dos primeiros 12 passos com os dados observados, utilizando a raiz do erro quadrático médio (REQM) como critério de avaliação, além de realizar previsões para o ano de 2024 nos últimos 12 passos. Para a análise, o conjunto de dados foi dividido em treino e teste, abrangendo o período de janeiro de 2014 a dezembro de 2022 para treino e o ano de 2023 para teste. O processo de modelagem incluiu a identificação do modelo, avaliação dos parâmetros, diagnóstico dos modelos selecionados e a previsão. Os resultados indicam um aumento gradual no número de atendimentos distintos ao longo do tempo, com um aumento significativo observado em 2023 para todas as variáveis estudadas. Entre os modelos analisados, o SARIMAE(2,1,1)(0,1,3) mostrou-se mais adequado para prever o número de MEIs atendidos com a quantidade de atendimentos aos MEIs como variável regressora, o SARIMAE(3,1,1)(0,1,1) para o número de MEs atendidos com a quantidade de atendimentos a MEs como variável regressora, e o SARIMA(0,0,2)(0,1,1) para o número de EPPs atendidos.

**Palavras-chave:** Séries Temporais, Pequenos negócios, Regressão, SARIMA, Sebrae.



# Abstract

This study investigates the application of time series analysis to monthly data on the number of Individual Microentrepreneurs (MEI), Microenterprises (ME), and Small Businesses (EPP) served by Sebrae/RN from January 2014 to December 2023, aiming to forecast the future number of distinct attendances for these business categories. The main objective is to fit a dynamic regression model with SARIMA errors (SARIMAE), using the number of attendances performed by Sebrae as the regressor variable. The study selects the most appropriate model for each category of interest, makes forecasts for 24 steps ahead, and compares the forecasts for the first 12 steps with the observed data, using the root mean squared error (RMSE) as the evaluation criterion, in addition to making forecasts for the year 2024 for the last 12 steps. For the analysis, the dataset was divided into training and testing sets, covering the period from January 2014 to December 2022 for training and the year 2023 for testing. The modeling process included model identification, parameter evaluation, diagnostic of the selected models, and forecasting. The results indicate a gradual increase in the number of distinct attendances over time, with a significant increase observed in 2023 for all studied variables. Among the analyzed models, SARIMAE(2,1,1)(0,1,3) proved to be the most suitable for forecasting the number of MEIs served with the number of MEI attendances as the regressor variable, SARIMAE(3,1,1)(0,1,1) for the number of MEs served with the number of ME attendances as the regressor variable, and SARIMA(0,0,2)(0,1,1) for the number of EPPs served.

**Keywords:** Time Series, Small Businesses, Regression, SARIMA, Sebrae.

# Lista de Figuras

2.1	Exemplo de séries temporais não estacionárias e estacionárias. (a) Preço de fechamento do Google, (b) Diferença no preço de fechamento do Google. Fonte: Hyndman, R.J., & Athanasopoulos, G.(2021) . . . . .	6
2.2	Exemplo de série temporal sazonal, $s = 12$ . Gráfico do Número total de passageiros por mês pelo tempo; Fonte: Box, George E. P.(2016) . . . . .	6
3.1	Ciclo Iterativo, Metodologia de Box Jenkins. Fonte: Elaboração Própria . .	12
4.1	Série temporal de atendimentos a MEI e o número de MEI atendidos de Janeiro de 2014 a Dezembro de 2023. . . . .	18
4.2	Série temporal de atendimentos a ME e o número de ME atendidos de Janeiro de 2014 a Dezembro de 2023. . . . .	18
4.3	Série temporal de atendimentos a EPP e número de EPP atendidos de Janeiro de 2014 a Dezembro de 2023. . . . .	19
4.4	Série temporal de atendimentos a MEI, ME e EPP de Janeiro de 2014 a Dezembro de 2023. . . . .	19
4.5	Série temporal de atendimentos a MEI, ME e EPP por ano de Janeiro de 2014 a Dezembro de 2023. . . . .	20
4.6	Série temporal de MEI, ME e EPP atendidos de Janeiro de 2014 a Dezembro de 2023. . . . .	21
4.7	Série temporal de MEI, ME e EPP atendidos por ano de Janeiro de 2014 a Dezembro de 2023. . . . .	21
4.8	Gráficos da diferença da diferença sazonal do log série de atendimentos ao MEI, da autocorrelação e da autocorrelação parcial . . . . .	24
4.9	Análise diagnóstica do modelo de atendimentos ao MEI . . . . .	25

4.10	Gráficos da diferença da diferença sazonal da série de MEI atendidos, da autocorrelação e da autocorrelação parcial . . . . .	26
4.11	Análise diagnóstica do modelo MEI atendidos . . . . .	27
4.12	Gráficos da diferença sazonal da série de atendimentos as MEs, da autocorrelação e da autocorrelação parcial . . . . .	28
4.13	Análise diagnóstica do modelo de atendimentos a ME . . . . .	29
4.14	Gráficos da diferença da diferença sazonal da série de ME atendidos, da autocorrelação e da autocorrelação parcial . . . . .	30
4.15	Análise diagnóstica do modelo ME atendidos . . . . .	31
4.16	Gráficos da diferença sazonal da série de atendimentos as EPPs, da autocorrelação e da autocorrelação parcial . . . . .	32
4.17	Análise diagnóstica do modelo de atendimentos a EPP . . . . .	33
4.18	Gráficos do log da diferença da diferença sazonal da série de EPP atendidas, da autocorrelação e da autocorrelação parcial . . . . .	34
4.19	Análise diagnóstica do modelo EPP atendidas . . . . .	35
4.20	Gráfico da previsão de de MEIs atendidos no ano de 2024 . . . . .	37
4.21	Gráfico da previsão de de MEIs atendidos no ano de 2024 . . . . .	37
4.22	Gráfico da previsão de de MEIs atendidos no ano de 2024 . . . . .	39
4.23	Gráfico da previsão de de MEIs atendidos no ano de 2024 . . . . .	39
4.24	Gráfico da previsão de de EPPs atendidas no ano de 2024 . . . . .	41
4.25	Gráfico da previsão de de EPPs atendidos no ano de 2024 . . . . .	41
4.26	Gráfico da previsão de empresas atendidas no ano de 2024 . . . . .	43
4.27	Gráfico da previsão de empresas atendidas no ano de 2024 . . . . .	43

# Lista de Tabelas

4.1	Resultados obtidos no Teste de Dickey-Fuller para a estacionaridade das séries analisada. . . . .	22
4.2	Modelos propostos . . . . .	24
4.3	Modelos propostos para os MEIs atendidos. . . . .	26
4.4	Modelos para os atendimentos as MEs. . . . .	29
4.5	Modelos propostos para as MEs atendidos. . . . .	30
4.6	Modelos para os atendimentos as EPPs. . . . .	32
4.7	Modelos propostos para as EPPs atendidas. . . . .	34
4.8	Previsão de MEIs atendidos e Atendimentos aos MEIs . . . . .	36
4.9	Previsão de MEs atendidos e Atendimentos as MEs . . . . .	38
4.10	Previsão de EPPs atendidas e Atendimentos as EPPs . . . . .	40
4.11	Empresas atendidas e Atendimentos as empresas . . . . .	42

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
	<b>Introdução</b>	<b>1</b>
1.1	Objetivo Geral . . . . .	2
1.2	Objetivos Específicos . . . . .	2
<b>2</b>	<b>Revisão Teórica</b>	<b>4</b>
2.1	Revisão Teórica . . . . .	4
2.2	Série Temporal . . . . .	4
2.3	Estacionaridade . . . . .	5
2.4	Sazonalidade . . . . .	5
2.5	Transformações e Diferenciação . . . . .	6
2.5.1	Diferença sazonal . . . . .	7
2.5.2	Notação de <i>Backshift</i> . . . . .	7
2.6	Função de autocorrelação e autocorrelação parcial . . . . .	7
2.7	Modelos de séries temporais . . . . .	8
2.7.1	Modelo Autorregressivos . . . . .	8
2.7.2	Modelos de Médias Móveis . . . . .	9
2.7.3	Modelo ARMA . . . . .	9
2.7.4	Modelos ARIMA . . . . .	9
2.7.5	Modelos SARIMA . . . . .	10
2.7.6	Modelos de Regressão dinâmica com erros SARIMA . . . . .	10
<b>3</b>	<b>Modelagem</b>	<b>12</b>
3.1	Estimação dos parâmetros . . . . .	14
3.2	Análise diagnóstica . . . . .	14

3.3	Previsão . . . . .	16
<b>4</b>	<b>Resultados</b>	<b>17</b>
4.1	Análise da série . . . . .	17
4.2	Ajuste dos modelos . . . . .	22
4.2.1	Ajuste do modelo para o MEI . . . . .	23
4.2.2	Ajuste do modelo para a ME . . . . .	28
4.2.3	Ajuste do modelo para a EPP . . . . .	31
4.3	Previsão . . . . .	36
4.3.1	Previsão para o MEI . . . . .	36
4.3.2	Previsão para a ME . . . . .	38
4.3.3	Previsão para a EPP . . . . .	39
4.3.4	Previsão total . . . . .	41
<b>5</b>	<b>Considerações Finais</b>	<b>44</b>

# Capítulo 1

## Introdução

O Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (Sebrae) é uma entidade privada que promove a competitividade e o desenvolvimento sustentável dos empreendimentos de micro e pequenas empresas, dentre as quais tem-se: Microempreendedor Individual (MEI) que possui no máximo 1 funcionário, Microempresa (ME) que possui até 19 funcionários e Empresa de Pequeno Porte (EPP) que possui de 20 a 99 funcionários.

O Sebrae tem como foco atuar no fortalecimento do empreendedorismo e na aceleração do processo de formalização da economia por meio de parcerias com os setores público e privado, programas de capacitação, acesso ao crédito e à inovação, estímulo ao associativismo, incentivo a educação empreendedora na educação formal, feiras e rodadas de negócios. As soluções desenvolvidas pelo Sebrae atendem desde o empreendedor que pretende abrir seu primeiro negócio até as pequenas empresas que já estão consolidadas e buscam um novo posicionamento no mercado.

Os serviços desenvolvidos pelo Sebrae para o seu público-alvo são: consultorias, orientações técnicas, capacitação empresarial, eventos, palestras e entre outros, sendo fundamental a captação e fidelização do cliente.

Deste modo, é importante para o Sebrae prever o quantitativo de empresas que serão atendidas para uma melhor organização dos seus serviços. Todavia, no cenário atual, a previsão das métricas de quantidade de empresas atendidas ainda carece de estudos aprofundados. Com isso, iremos abordar técnicas preditivas utilizando uma modelagem de séries temporais.

Apesar da ausência de métricas preditivas estabelecidas para o quantitativo de Empresas atendidas, o Sebrae já utiliza técnicas de modelagem de séries temporais, como

os modelos ARIMA (*AutoRegressive Integrated Moving Average*) e ETS (*Error-Trend-Seasonality*), para prever o cenário das empresas em cada estado do Brasil, esses artigos não podem ser apresentados conforme Lei nº 9.610, Esta Lei regula os direitos autorais, entendendo-se sob esta denominação os direitos de autor e os que lhes são conexos. O Sebrae/RN tem como uma de suas metas atender um determinado número de empresas distintas. Portanto, para alcançar esse objetivo, a análise a ser realizada deve estar alinhada às expectativas esperadas em relação ao número de empresas atendidas por porte empresarial.

Para a realização dessa previsão, utilizaremos modelos de regressão dinâmica com erros SARIMA (*Seasonal AutoRegressive Integrated Moving Average*), na qual denominaremos de SARIMAE. Estes modelos permitem prever valores futuros a partir de valores passados, além de incorporarem sazonalidade e variáveis regressoras, possibilitando uma compreensão mais completa e precisa da dinâmica empresarial.

Com base nesses estudos, iremos desenvolver um modelo de série temporal para prever e antecipar o número de Microempreendedores Individuais (MEI), Microempresas (ME) e Empresas de Pequeno Porte (EPP) atendidos mensalmente pelo Sebrae/RN, com a maior precisão possível e relacionando com o número de atendimentos aos mesmos, sem incorrer em sobreajuste nos modelos. Esses modelos poderão contribuir significativamente para o aprimoramento das estratégias de atendimento e suporte oferecidas pelo Sebrae/RN aos empreendedores locais.

## 1.1 Objetivo Geral

O objetivo geral deste trabalho é ajustar modelos de regressão dinâmica com erros SARIMA (SARIMAE) para a quantidade de empresas por porte (MEI, ME, EPP) atendidos pelo Sebrae/RN utilizando como variável regressora a quantidade de atendimentos realizados por porte empresarial. Os dados de Janeiro de 2014 a Dezembro de 2022 formarão a base de treino e o ano de 2023 será a base de teste.

## 1.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:



- Ajustar um modelo SARIMAE para cada série temporal do número de MEI, ME e EPP atendidos pelo Sebrae/RN, utilizando como variável regressora a quantidade de atendimentos de MEI, ME e EPP;
- Realizar a previsão de 24 meses a frente de cada série MEI, ME e EPP;
- Comparar os 12 primeiros meses previstos com os valores observados em 2023 para selecionarmos o modelo com a menor Ráiz do Erro Quadrático Médio (REQM);
- Aplicar o modelo escolhido para prever o ano de 2024.

Para a estrutura deste trabalho, optamos por seguir uma abordagem sequencial que visa fornecer uma compreensão clara e lógica do conteúdo. Inicialmente, no Capítulo 2, iremos revisar a literatura relevante, onde serão explorados os fundamentos essenciais relacionados à análise e modelagem de regressão dinâmica com erros SARIMA para séries temporais.

Em seguida, no Capítulo 3, será detalhada a metodologia empregada para ajustar a série temporal, fornecendo uma explicação abrangente de como os dados foram tratados e os modelos foram selecionados.

No Capítulo 4, os resultados das análises realizadas serão minuciosamente discutidos, apresentando insights valiosos obtidos a partir da aplicação dos modelos.

Por fim, no Capítulo 5, serão apresentadas as conclusões finais, juntamente com reflexões sobre as implicações dos resultados e possíveis direções para pesquisas futuras.

# Capítulo 2

## Revisão Teórica

### 2.1 Revisão Teórica

Neste capítulo faremos uma revisão teórica a cerca dos modelos ARIMA e SARIMA, analisando principalmente os trabalhos realizados por DE PINHO, A. L. S. & SIROKY, A. N.(2023), Box, George E. P.(2016), Hyndman, R.J., & Athanasopoulos, G.(2021), Harris, R., & Sollis, R.(2003), G. M. Ljung & G. E. P. Box(1978), Andrew V. Metcalfe & Paul S.P. Cowpertwait(2008), Jonathan D. Cryer & Kung-Sik Chan(2008) e Robert H. Shumway & David S. Stoffer(2017) que tiveram um empenho para demonstrar e desenvolver um modelo temporal que consiga prever diversos resultados de estudos específicos.

### 2.2 Série Temporal

Uma série temporal  $y_t$  é uma sequência de observações ou medidas coletadas em intervalos de tempo regulares. Geralmente, essas observações são registradas em uma ordem cronológica, permitindo que padrões e tendências ao longo do tempo sejam identificados e analisados. As séries temporais são comumente utilizadas em diversas áreas, como economia, finanças, meteorologia, ciências sociais e engenharia, para modelar e prever o comportamento de fenômenos que variam ao longo do tempo. Como exemplo:

- Número de acidentes aéreos por mês;
- Índice diário da bolsa de valores;

- Números de vendas diário de uma loja;
- Temperatura de uma cidade diariamente;
- Índice pluviométrico por semana;
- Número de sismos por mês;
- Número de transações realizadas por um cliente de um banco por dia.

De acordo com Box, George E.P.(2016) uma das principais características de uma série temporal é que, tipicamente, as observações são dependentes. A natureza dessa dependência entre observações de uma série temporal é de considerável interesse prático. A análise de séries temporais preocupa-se com técnicas para a análise dessa dependência. Isso requer o desenvolvimento de modelos estocásticos e dinâmicos para dados de séries temporais e o uso desses modelos é fundamental em diversas áreas de aplicação.

## 2.3 Estacionaridade

Dizemos que uma série é fortemente estacionária quando a distribuição da variável resposta não depende do tempo em que ela foi observada, ou seja,  $\{y_t\}$  é uma série temporal estacionária, se para todo  $s$ , a distribuição de  $(y_t, \dots, y_{t+s})$  não depende de  $t$ .

Em geral, uma série temporal estacionária não apresenta padrões previsíveis ao longo do tempo. Os gráficos de séries temporais estacionárias tendem a exibir a série com as suas observações oscilando em torno de uma média constante sem apresentar tendências evidentes, embora algum comportamento cíclico possa estar presente, a variância permanece constante ao longo do tempo.

Conforme os exemplos na Figura 2.1, podemos evidenciar o comportamento de duas séries temporais não-estacionária e outra estacionária.

## 2.4 Sazonalidade

Dizemos que uma série temporal tem sazonalidade quando ela apresenta um padrão recorrente ao longo do tempo, sendo a sazonalidade denotada por  $s$  podendo ocorrer em diferentes períodos de  $s$ . Quando falamos de um padrão recorrente anual temos então  $s = 12$ , se for trimestral  $s = 4$ , e assim por diante.

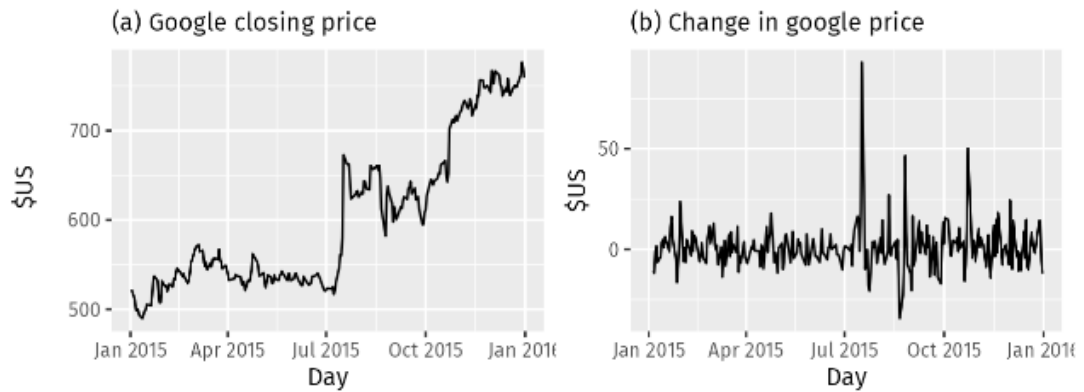


Figura 2.1: Exemplo de séries temporais não estacionárias e estacionárias.  
 (a) Preço de fechamento do Google, (b) Diferença no preço de fechamento do Google.  
 Fonte: Hyndman, R.J., & Athanasopoulos, G.(2021)

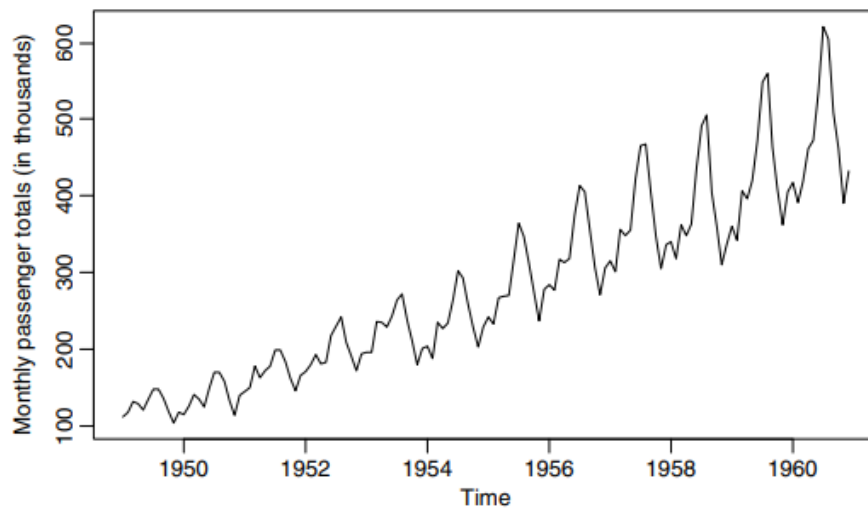


Figura 2.2: Exemplo de série temporal sazonal,  $s = 12$ .  
 Gráfico do Número total de passageiros por mês pelo tempo;  
 Fonte: Box, George E. P.(2016)

## 2.5 Transformações e Diferenciação

As transformações matemáticas são ferramentas essenciais para estabilizar a variância de séries temporais. Entre as várias transformações utilizadas, podemos citar a transformação logarítmica, a transformação Box-Cox e a transformação de raiz quadrada, entre outras. A transformação logarítmica, por exemplo, é particularmente apropriada quando a série temporal exibe um padrão exponencial.

A diferenciação é um dos métodos utilizados para tornar uma série temporal não-estacionária em uma série temporal estacionária. Podemos mostrar isso com a seguinte fórmula:

$$\Delta_1^d y_t = y'_t = y_t - y_{t-1} \quad (2.1)$$

Os dados diferenciados podem não parecer estacionários, o que pode indicar a necessidade de diferenciar os dados novamente para obter uma série estacionária:  $\Delta_1^2 y'_t = y''_t = y'_t - y'_{t-1}$ . Utilizaremos a notação  $\Delta_1^d$  para denotar uma série com diferença simples  $d$  vezes.

Para o ajuste dos modelos ARIMA, que usaremos neste trabalho, é necessário que a série temporal seja estacionária e tenha a variância constante.

### 2.5.1 Diferença sazonal

A diferença sazonal é um tipo de diferenciação aplicada na sazonalidade, na equação 2.1 substituímos o  $t - 1$  por  $t - m$ , ou seja se a sazonalidade é anual ( $s = 12$ ) o  $m$  será então 12, com isso teremos a seguinte equação geral:

$$\Delta_m^d y_t = y'_t = y_t - y_{t-m}$$

A diferença sazonal auxilia na remoção de tendência e sazonalidade dos dados.

### 2.5.2 Notação de *Backshift*

Utilizaremos o conceito de “*Backshift*” para mostrar a defasagem de  $y_t$ , sendo sua notação denotada por  $B$ , com isso temos a seguinte expressão:

$$y_{t-m} = B^m y_t,$$

sendo o  $m$  o tempo defasado de  $y_t$ . Esta notação será útil para a obtenção de uma expressão compacta para os modelos ARIMA.

## 2.6 Função de autocorrelação e autocorrelação parcial

Assim como a correlação mede a extensão de uma relação linear entre duas variáveis, a autocorrelação mede a relação linear entre valores defasados de uma série temporal.

Existem diversos coeficientes de autocorrelação, correspondendo a cada painel no gráfico de defasagem. Por exemplo,  $r_1$  mede a relação entre  $y_t$  e  $y_{t-1}$ ,  $r_2$  mede a relação entre  $y_t$  e  $y_{t-2}$ , e assim por diante.

Iremos analisar os gráficos da FAC (Função de Autocorrelação), ou em inglês "ACF" (*AutoCorrelation Function*), e da FACP (Função de Autocorrelação Parcial), ou em inglês "PACF" (*Partial AutoCorrelation Function*), que representam, respectivamente, a correlação entre  $y_t$  e seus valores defasados, e a correlação entre  $y_t$  e seus valores defasados após a remoção dos efeitos das demais defasagens. A expressão para a correlação pode ser denotada pela seguinte fórmula:

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

onde  $T$  é o tamanho da série temporal.

Os gráficos dessas funções serão utilizados para a escolha da ordem dos modelos que apresentaremos a seguir.

## 2.7 Modelos de séries temporais

Nesta seção iremos apresentar diversos modelos da classe ARIMA para uma análise de séries temporais.

### 2.7.1 Modelo Autorregressivos

O modelo autorregressivo de ordem  $p$ , que sua representação pode ser denotada por  $AR(p)$ , pode ser escrito da seguinte forma:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t,$$

em que  $\{\varepsilon_t\}$  é um ruído branco, ou seja,  $\varepsilon_t = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T)$  são variáveis aleatórias não correlacionadas, com esperança 0 e variância constante. Temos também que  $\phi_1, \dots, \phi_p$  são os coeficientes autorregressivos e  $c$  é uma constante real. Podemos perceber que o modelo autorregressivo é similar a uma regressão múltipla, em que  $y_t$  é a variável resposta e as variáveis explicativas são os valores passados de  $y_t$ . Pode-se mostrar que para processo seja autorregressivo  $E(y_t) = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$ .

### 2.7.2 Modelos de Médias Móveis

O modelo de médias móveis de ordem  $q$ , denotado por  $MA(q)$ , na qual pode ser expresso da seguinte maneira:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_p \varepsilon_{t-q},$$

em que  $\{\varepsilon_t\}$  é um ruído branco. Conseguimos visualizar que no modelo de médias móveis o valor de  $y_t$  está relacionado com erros passados. Pode-se mostrar que para processo seja de médias móveis  $E(y_t) = c$ .

### 2.7.3 Modelo ARMA

Ao juntarmos o modelo autorregressivo com o modelo de médias móveis, iremos obter um modelo autorregressivo de médias móveis, denominado de  $ARMA(p,q)$ . O qual é expresso por:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_p \varepsilon_{t-q},$$

em que  $\phi_1, \dots, \phi_p$  são os coeficientes autorregressivos,  $\theta_1, \dots, \theta_q$  os coeficientes de médias móveis e  $\{\varepsilon_t\}$  é um ruído branco.

### 2.7.4 Modelos ARIMA

Ao combinarmos a diferenciação, apresentada na Seção 2.5, com o modelo autorregressivo e o modelo de médias móveis, obtemos o modelo *AutoRegressive Integrated Moving Average* (ARIMA) com parâmetros  $p, d, q$ . Seja  $w_t$  a variável  $y_t$  diferenciada  $d$  vezes:

$$\Delta_1^d y_t = w_t$$

Vamos escrever o modelo ARMA em função de  $w_t$ , obtendo então o modelo ARIMA, com isso temos a seguinte expressão:

$$w_t = c + \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_p \varepsilon_{t-p} + \varepsilon_t.$$

Podemos, também, utilizar a notação de  $B$  para expressar o modelo ARIMA, com isso temos o seguinte modelo:

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) (1 - B^d) y_t = \left(1 + \sum_{i=1}^q \theta_i B^i\right) \varepsilon_t$$

em que  $p$  representa a ordem da parte AR,  $d$  é a ordem de diferenciação e  $q$  a ordem da parte MA.

### 2.7.5 Modelos SARIMA

Nas séries temporais podemos ter sazonalidade como vemos na seção 2.4, a sazonalidade pode ser adicionada ao modelo ARIMA, obtendo assim então o modelo ARIMA com sazonalidade ou também conhecido como SARIMA, denotado por:

$$\text{SARIMA}(p, d, q)(P, D, Q)_m$$

em que as letras minúsculas,  $(p, d, q)$ , fazem referência à parte não-sazonal da série e as letras maiúsculas,  $(P, D, Q)$ , fazendo alusão à parte sazonal com  $m$  sendo o período da sazonalidade.

Com isso em mente, temos que um modelo  $\text{SARIMA}(p, d, q)(P, D, Q)_m$  é expresso por:

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) \left(1 - \sum_{i=1}^P \Phi_i B^{m \times i}\right) (1 - B^d)(1 - B^{D \cdot m}) y_t = \left(1 + \sum_{i=1}^q \theta_i B^i\right) \left(1 + \sum_{i=1}^Q \Theta_i B^{m \times i}\right) \varepsilon_t$$

### 2.7.6 Modelos de Regressão dinâmica com erros SARIMA

Os modelos de regressão dinâmica com erros SARIMA, que denotaremos como SARIMAE, relacionam duas ou mais séries temporais por meio de uma equação de regressão e modelam o erro da regressão utilizando um modelo SARIMA. Essa formulação permite incluir informações provenientes das observações passadas da série, bem como outras informações que possam ser relevantes. Considere o seguinte modelo de regressão:



$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \eta_t,$$

em que  $y_t$  é o valor da variável resposta no tempo  $t$ ,  $x_{k,t}$  é o valor da  $k$ -ésima variável regressora no tempo  $t$  e os erros da regressão são correlacionados, pois  $y_t$  e  $x_{k,t}$  são séries temporais. Finalmente, assumimos que os erros  $\eta_t$  seguem um modelo ARIMA ou SARIMA.

Se  $\eta_t$  segue um modelo SARIMA( $p, d, q$ )( $P, D, Q$ ) $_m$ , podemos expressar da seguinte forma:

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) \left(1 - \sum_{i=1}^P \Phi_i B^{m \times i}\right) (1 - B^d)(1 - B^{D \cdot m}) \eta_t = \left(1 + \sum_{i=1}^q \theta_i B^i\right) \left(1 + \sum_{i=1}^Q \Theta_i B^{m \times i}\right) \varepsilon_t$$

na qual  $\{\varepsilon_t\}$  é um ruído branco.

# Capítulo 3

## Modelagem

Neste capítulo iremos explorar a metodologia realizada Box e Jenkins (1970), esta metodologia tem como estratégia encontrar e estimar o melhor modelo e estimar os seus parâmetros (SARIMA, ARIMA) com base em um ciclo iterativo que podemos visualizar na Figura 3.1, tal abordagem consiste em:

- Identificar os modelos adequados;
- Estimar os parâmetros do modelo;
- Realizar o diagnóstico do modelo;
- Uso do modelo para previsão de valores.

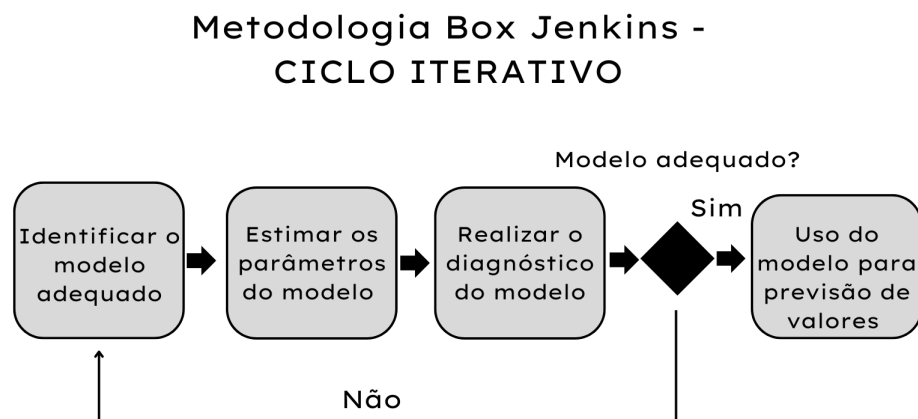


Figura 3.1: Ciclo Iterativo, Metodologia de Box Jenkins.  
Fonte: Elaboração Própria

Primeiramente, devemos definir o modelo com o qual iremos trabalhar, podendo ser o modelo SARIMA (2.7.5) ou o modelo de regressão dinâmica (2.7.6).

Após essa definição, iniciaremos os ciclos iterativos da metodologia citada para identificar e ajustar o modelo selecionado. Para isso, iremos verificar se a série possui uma variância constante, caso não tenha aplicaremos alguma transformação, posteriormente verificaremos se a série é estacionária, se a mesma não for realizaremos diferenciações na série tanto na parte sazonal ou/e na parte não sazonal e por fim iremos analisar os gráficos da auto correlação e da auto correlação parcial estimados para cada série.

Ao iniciarmos esta etapa de identificação devemos visualizar o gráfico da série que serão trabalhadas, que seriam o número de MEI atendidos pelo Sebrae considerando como variável regressora o número de Atendimentos pelo Sebrae as MEIs, o número de ME atendidos pelo Sebrae considerando como variável regressora o número de Atendimentos pelo Sebrae as MEs e o número de EPP atendidas pelo Sebrae considerando como variável regressora o número de Atendimentos pelo Sebrae as EPPs. Com as séries plotadas devemos observar as características que foram apresentadas no parágrafo anterior.

Utilizaremos a metodologia de Box Jenkins para analisarmos os possíveis modelos que foram identificados a partir da análise das autocorrelações e auto correlaçõesparcias amostrais das séries, juntamente com a função ARIMA do pacote *fable* do software R Core Team (2022) que gera automaticamente um modelo sem a necessidade da interpretação gráfica da ACF e da PACF.

O critério adotado para a escolha do modelo foi com base na raiz do erro quadrático médio (*REQM*), que serve como uma medida de acurácia de um modelo de previsão, sendo denotado por:

$$REQM = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

sendo  $\hat{y}_i$  os valores preditos pelo modelo,  $y_i$  os valores observados e  $n$  o número de observações. No nosso contexto, os dados foram divididos em conjuntos de treino e teste. O conjunto de treino, que abrange de janeiro de 2014 até dezembro de 2022, será utilizado para treinar os possíveis modelos. Em seguida, compararemos os valores previstos pelos modelos ( $\hat{y}_i$ ) para o ano de 2023 com os dados do conjunto de teste, ou seja, os valores observados ( $y_i$ ) no ano de 2023. A partir dessas comparações, calcularemos o *REQM*.

### 3.1 Estimação dos parâmetros

A partir da metodologia iterativa descrita, selecionaremos um modelo. Como utilizaremos um modelo de regressão dinâmica (2.7.6), para estimar seus parâmetros podemos empregar os métodos de mínimos quadrados e de máxima verossimilhança.

Para estimação utilizaremos o pacote *fable* que utiliza o método da máxima verossimilhança para obtenção das estimativas dos parâmetros do modelo.

### 3.2 Análise diagnóstica

Uma vez o selecionado o modelo e estimados seus parâmetros, iremos verificar se o modelo está bem ajustado, isto é, os resíduos são ruído branco. Caso o modelo selecionado não esteja de acordo com as suposições, então poderemos selecionar um outro modelo se adequar.

Primeiramente verificaremos se os resíduos são ruído branco através da inspeção visual do gráfico de dispersão, da FAC e de alguns testes adicionais. Em seguida, iremos verificar a distribuição dos resíduos se eles seguem uma distribuição normal, para isso iremos realizar dois testes sendo o primeiro o Teste de Shapiro-Wilk cujo objetivo é testar se os resíduos possuem uma distribuição normal. A estatística  $W$  do teste de Shapiro-Wilk é calculada da seguinte forma:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

em que:

- $x_{(i)}$  é o  $i$ -ésimo valor ordenado da amostra,
- $\bar{x}$  é a média da amostra,
- $a_i$  são coeficientes calculados a partir da matriz de covariância de uma amostra normal ordenada,
- $n$  é o tamanho da amostra.

e o segundo teste a ser realizado é o Teste de Kolmogorov-Sminov, em que é utilizado para verificar se os resíduos possuem alguma distribuição, podendo ser normal

ou outra distribuição teórica. A estatística  $D$  do teste de Kolmogorov-Smirnov é definida como:

$$D = \sup_x |F_n(x) - F(x)|$$

onde:

- $D$  é a estatística do teste K-S,
- $\sup_x$  representa o supremo (ou o máximo) sobre todas as observações  $x$ ,
- $F_n(x)$  é a função de distribuição acumulada empírica da amostra,
- $F(x)$  é a função de distribuição acumulada da distribuição teórica (neste caso, a distribuição normal).

em ambas as estatísticas testaremos a seguintes hipóteses:

- $H_0$  : Os resíduos possuem distribuição Normal;
- $H_1$  : Os resíduos não possuem distribuição Normal.

E por fim, iremos realizar o teste de Ljung-Box em que nele temos avaliamos se os resíduos são correlacionados ou não. A estatística  $Q$  do teste de Ljung-Box é dada por:

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k}$$

onde:

- $Q$  é a estatística do teste Ljung-Box,
- $n$  é o número de observações na série temporal,
- $m$  é o número de lags a serem testados,
- $\hat{\rho}_k$  é a autocorrelação amostral no lag  $k$ .

na qual testaremos as seguintes hipóteses:

- $H_0$  : Os resíduos são não correlacionados;
- $H_1$  : Os resíduos são correlacionados.

### 3.3 Previsão

Após concluirmos identificação do modelo, a estimação dos parâmetros e a análise diagnóstica do modelo, então, com o modelo já bem avaliado por essas etapas iremos por fim prever valores de  $y_{t+h}$  sendo  $h \geq 1$  em que  $h$  representa o quão a frente do tempo iremos prever.

Para fazermos a realização desta previsão iremos utilizar o comando *forecast* que estará contido no pacote *forecast* do software que será utilizado, no nosso caso é o R core team(2022).

# Capítulo 4

## Resultados

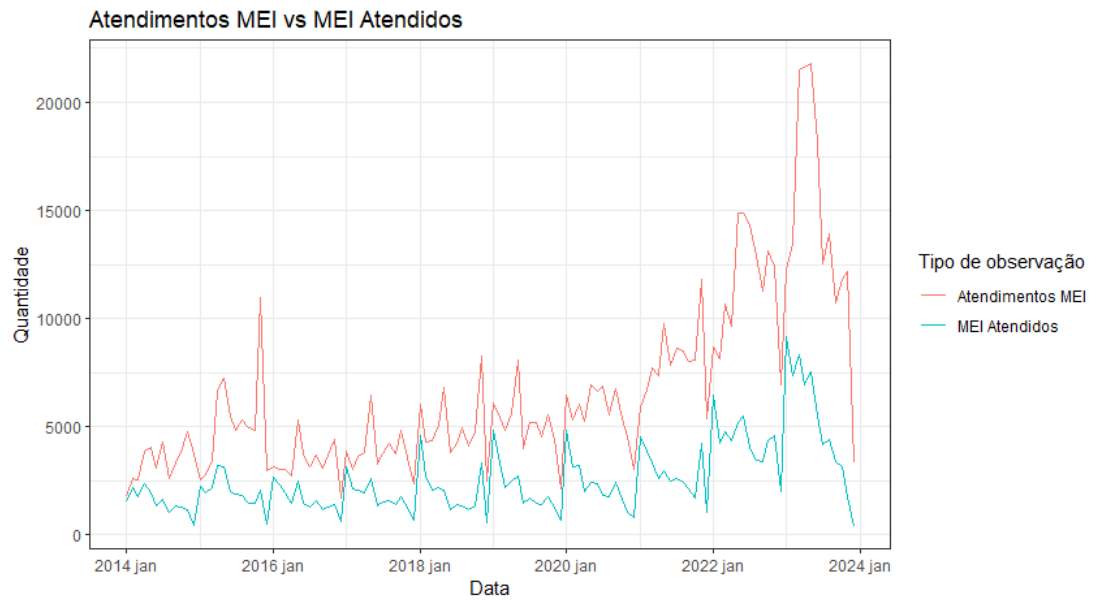
Neste capítulo, abordaremos os resultados dos modelos ajustados. Analisaremos duas séries: a série de atendimentos pelo Sebrae as MEIs, MEs e EPPs, que será a variável regressora, e a série do número de MEIs, MEs e EPPs atendidos, que é o nosso objetivo. Os dados abrangem o período de janeiro de 2014 a dezembro de 2023.

Os modelos serão selecionados com base no critério *REQM*, citado anteriormente no Capítulo 3.

### 4.1 Análise da série

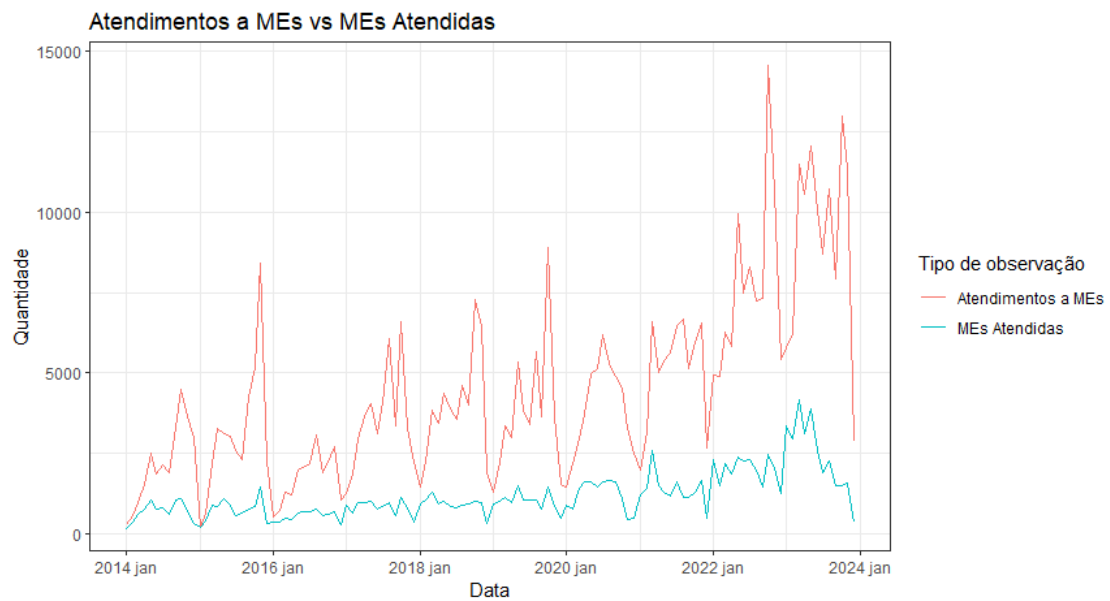
As séries apresentadas a seguir são provenientes de dados disponibilizados pelo Sebrae/RN, os gráficos das séries de MEI, ME e EPP atendidos com os seus atendimentos, são apresetnados nas Figuras 4.1, 4.2 e 4.3, respectivamente.

Figura 4.1: Série temporal de atendimentos a MEI e o número de MEI atendidos de Janeiro de 2014 a Dezembro de 2023.



Fonte: Autoria própria.

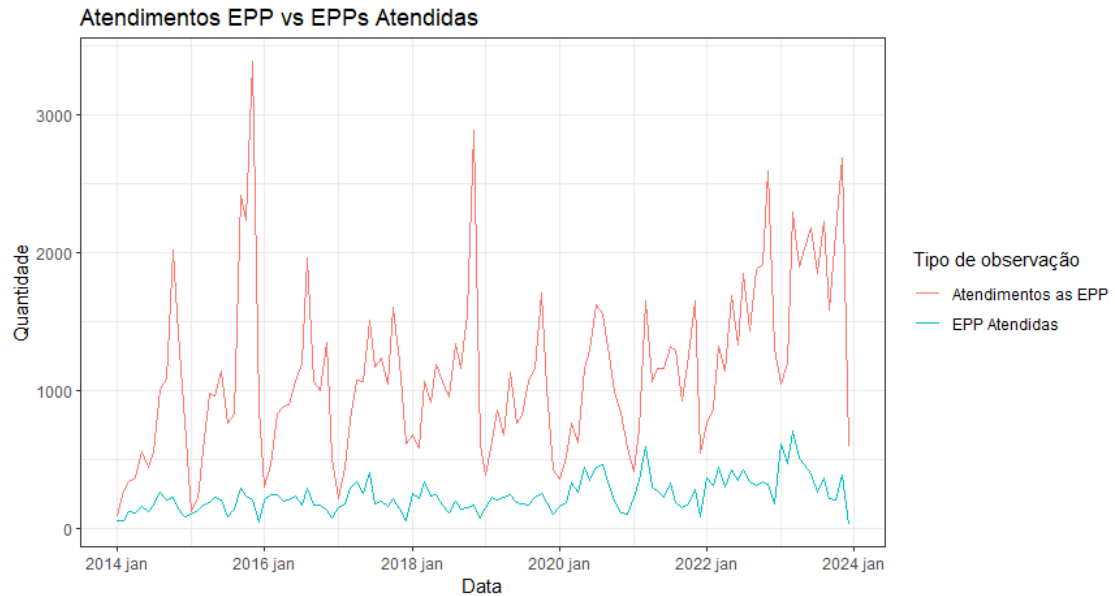
Figura 4.2: Série temporal de atendimentos a ME e o número de ME atendidos de Janeiro de 2014 a Dezembro de 2023.



Fonte: Autoria própria.



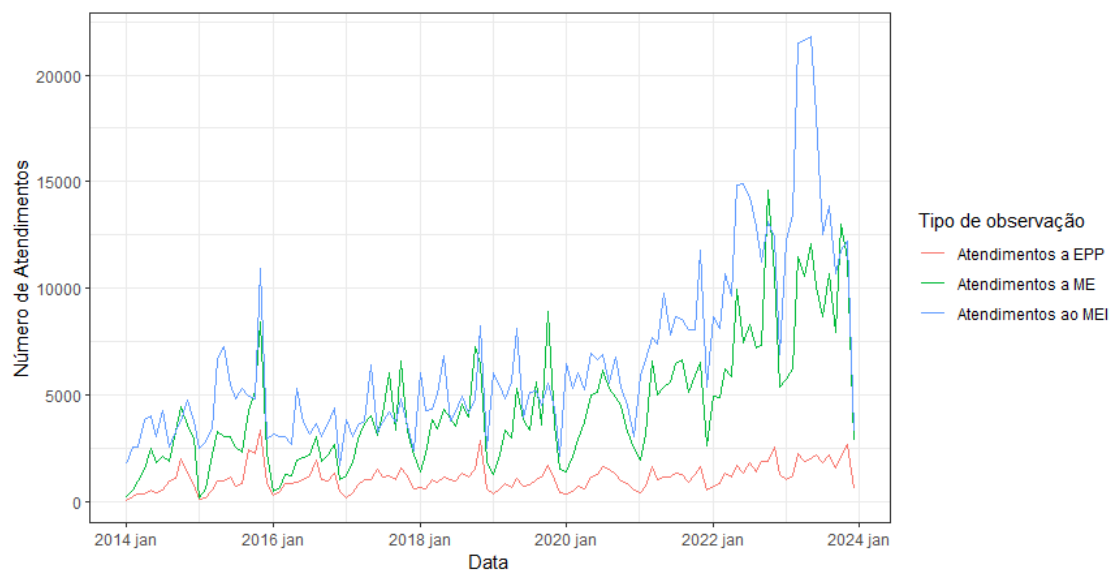
Figura 4.3: Série temporal de atendimentos a EPP e número de EPP atendidos de Janeiro de 2014 a Dezembro de 2023.



Fonte: Autoria própria.

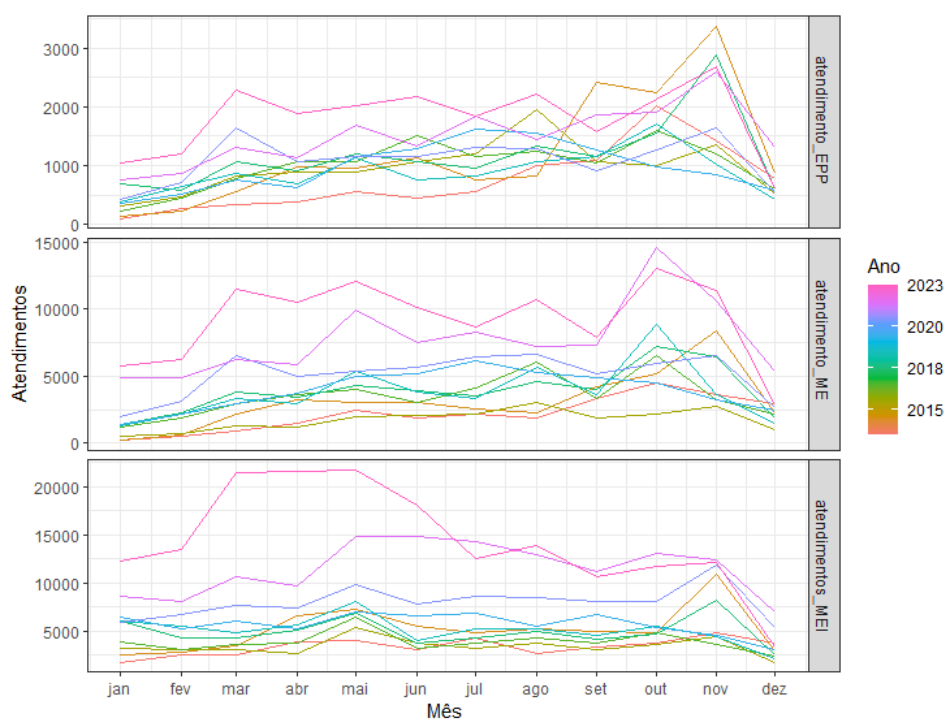
Agora, apresentamos as séries temporais dos atendimentos a MEI, ME e EPP nas Figuras 4.4 e 4.5. Podemos notar a sazonalidade em cada variável, em que no final de cada ano o número de atendimentos decai e no início do ano seguinte aumenta significativamente, podemos perceber também que o MEI e a ME possuem uma tendência sugerindo então que precisamos fazer diferenciação na série para tornar ela estacionária.

Figura 4.4: Série temporal de atendimentos a MEI, ME e EPP de Janeiro de 2014 a Dezembro de 2023.



Fonte: Autoria própria.

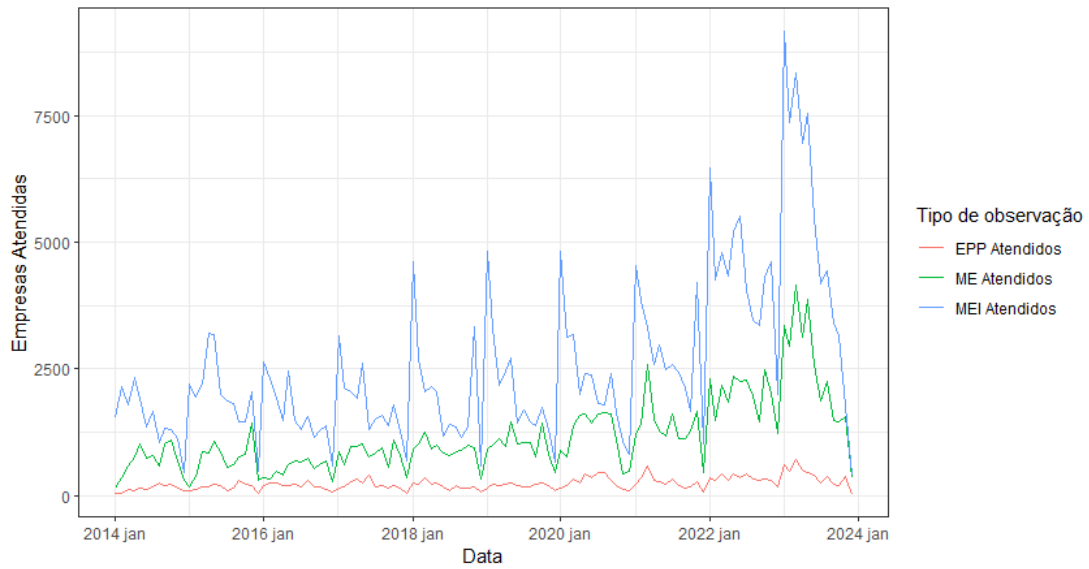
Figura 4.5: Série temporal de atendimentos a MEI, ME e EPP por ano de Janeiro de 2014 a Dezembro de 2023.



Fonte: Autoria própria.

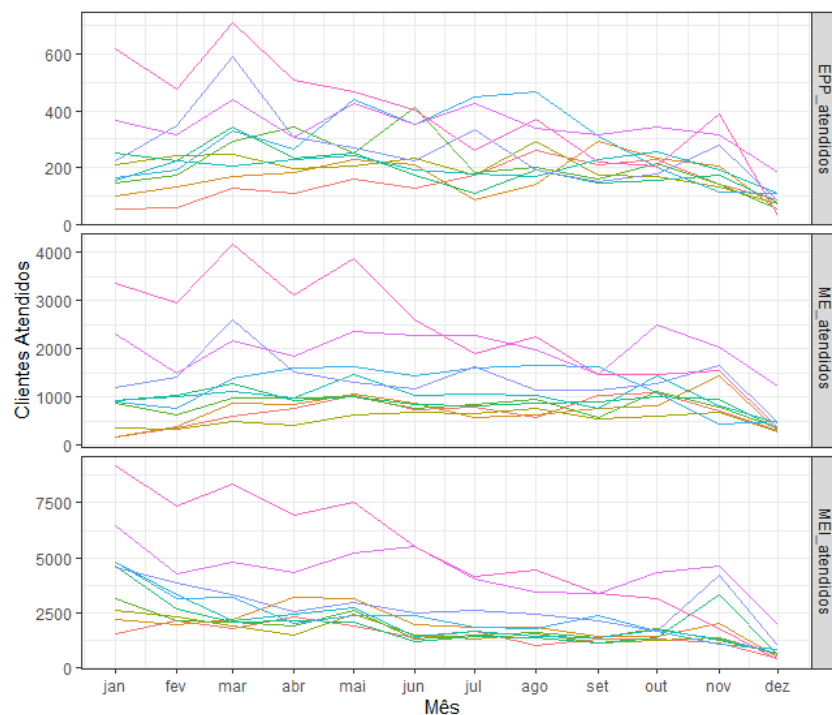
Em seguida visualizaremos o comportamento da série temporal de MEI, ME e EPP atendidos nas Figura 4.6 e 4.7 , podemos perceber um padrão similar ao padrão de atendimentos a MEI, ME e EPP, em que o MEI e a ME possuem alguma tendência sugerindo uma diferenciação na série e notamos também uma sazonalidades para todos os tipos de empresas.

Figura 4.6: Série temporal de MEI, ME e EPP atendidos de Janeiro de 2014 a Dezembro de 2023.



Fonte: Autoria própria

Figura 4.7: Série temporal de MEI, ME e EPP atendidos por ano de Janeiro de 2014 a Dezembro de 2023.



Fonte: Autoria própria.

A fim de verificar a estacionaridade de cada série iremos realizar o Teste de Dickey-Fuller, que tem as seguintes hipóteses:

- $H_0$  : Série não estacionária

- $H_1$  : Série estacionária

Portanto nosso objetivo é rejeitar a hipótese nula ( $H_0$ ) para concluirmos que a série é estacionária. Com isso vamos apresentar o resultados do teste realizado:

Tabela 4.1: Resultados obtidos no Teste de Dickey-Fuller para a estacionaridade das séries analisada.

Atendimentos a empresas	P-valor
MEI	0.122
ME	0.086
EPP	0.010
Empresas atendidas	
MEI	0.014
ME	0.034
EPP	0.010

Com os resultados apresentados na Tabela 4.1, ao nível de significância de 99%, devemos rejeitar a hipótese nula apenas para a EPP. Ou seja, a série temporal de EPP não precisa de diferenciação para se tornar estacionária, ao contrário das demais, que precisam ser diferenciadas para alcançar a estacionaridade. Portanto, realizaremos o mesmo teste nas séries temporais diferenciadas. Após essa etapa, observamos que os  $p$ -valores são todos menores que 0,01, o que indica que devemos diferenciar a série temporal de atendimento ao MEI e à ME. O mesmo é válido para a série temporal de MEI e ME atendidos.

## 4.2 Ajuste dos modelos

Neste capítulo, apresentaremos os modelos selecionados para modelar tanto o número de atendimentos às MEIs, MEs e EPPs quanto o número de MEIs, MEs e EPPs atendidos. Primeiro, ajustamos um modelo para a variável regressora e, em seguida, ajustamos um modelo para a variável de interesse. Para esse ajuste, devemos analisar os gráficos da função de autocorrelação (FAC ou ACF) e da função de autocorrelação parcial (FACP ou PACF), a fim de identificar o modelo adequado para cada série. Em seguida, analisaremos os lags significativos da ACF e da PACF para determinar o modelo SARIMA  $(p, d, q)(P, D, Q)_{12}$ .

Para selecionar as ordens “ $p$ ” e “ $P$ ”, analisamos os lags significativos da PACF, e para selecionar “ $q$ ” e “ $Q$ ”, analisamos os lags significativos da ACF, onde as letras maiúsculas correspondem aos lags sazonais, ou seja, os lags 12, 24 e 36. Também utilizaremos a função do pacote *fable* chamada ARIMA, sem definir os parâmetros, para

que a função gere automaticamente uma ARIMA com o menor Critério de Informação de Akaike (AIC). O AIC fornece uma maneira de avaliar a qualidade relativa do modelo e pode ser calculado pela fórmula:

$$\text{AIC} = -2 \ln(L) + 2k$$

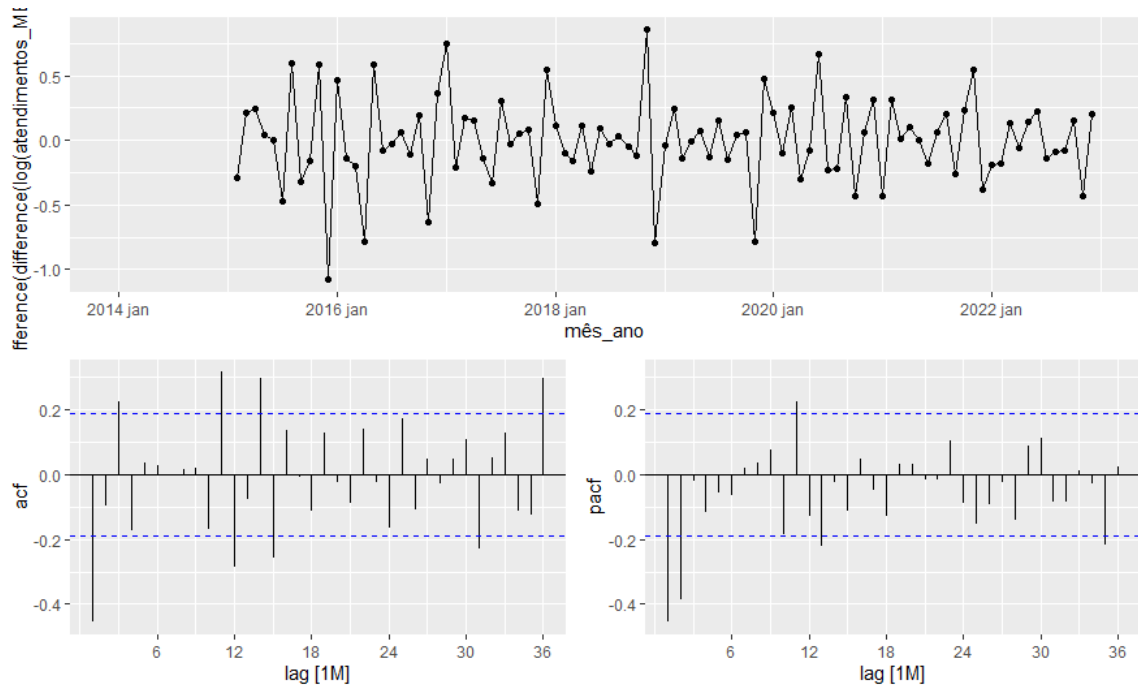
em que:

- $L$  é o valor da função de verossimilhança para o modelo ajustado,
- $k$  é o número de parâmetros estimados no modelo.

#### 4.2.1 Ajuste do modelo para o MEI

Para o atendimento ao MEI, realizamos uma transformação logarítmica na série para estabilizar a variância. Na Figura 4.8, apresentamos a diferença da diferença sazonal do logaritmo da série de atendimentos ao MEI, juntamente com a ACF e a PACF. Note que as 13 primeiras observações não aparecem devido à diferença da diferença sazonal realizada.

Figura 4.8: Gráficos da diferença da diferença sazonal do log série de atendimentos ao MEI, da autocorrelação e da autocorrelação parcial



Fonte: Autoria própria.

Após a análise dos lags significativos da PACF e da ACF, pode-se sugerir um modelo  $SARIMA(5,1,5)(0,1,2)$ . O modelo sugerido pela função ARIMA foi um  $SARIMA(1,0,1)(2,1,0)$ .

Portanto, selecionaremos o modelo conforme citado no Capítulo 3, a partir do REQM do modelo com os valores observados no conjunto de teste (ano de 2023). Com isso, temos os seguintes modelos:

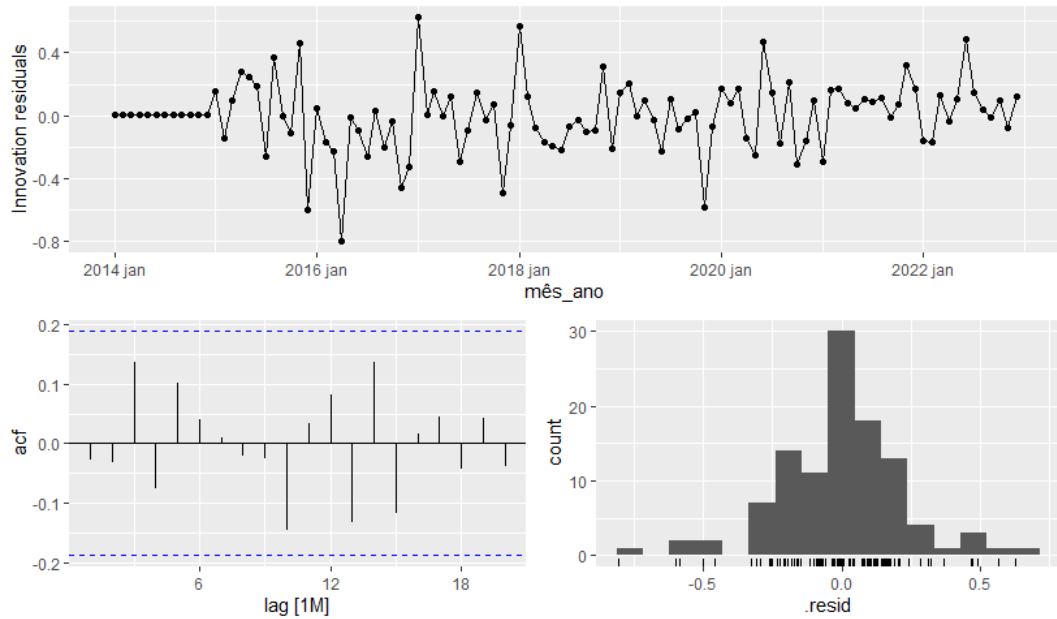
Tabela 4.2: Modelos propostos

Modelo	<i>REQM</i>
$SARIMA(1,0,1)(2,1,0)$	4194
$SARIMA(1,0,1)(1,1,1)$	4522
$SARIMA(1,0,2)(2,1,1)$	4533
$SARIMA(2,1,2)(0,1,2)$	4619

Percebe-se que o modelo sugerido não aparece entre os modelos propostos, pois possui muitos parâmetros. Portanto, faremos agora a análise diagnóstica do modelo, conforme visualizado na Figura 4.9. O teste de Ljung-Box, para o lag 24, obteve um p-valor de 0,643, e o teste de Kolmogorov-Smirnov para a distribuição normal apresentou um p-valor de 0,208. Assim, conclui-se que os resíduos são não correlacionados e possuem uma distribuição normal. Concluímos, então, que os atendimentos ao MEI serão modelados

por um SARIMA(1,0,1)(2,1,0).

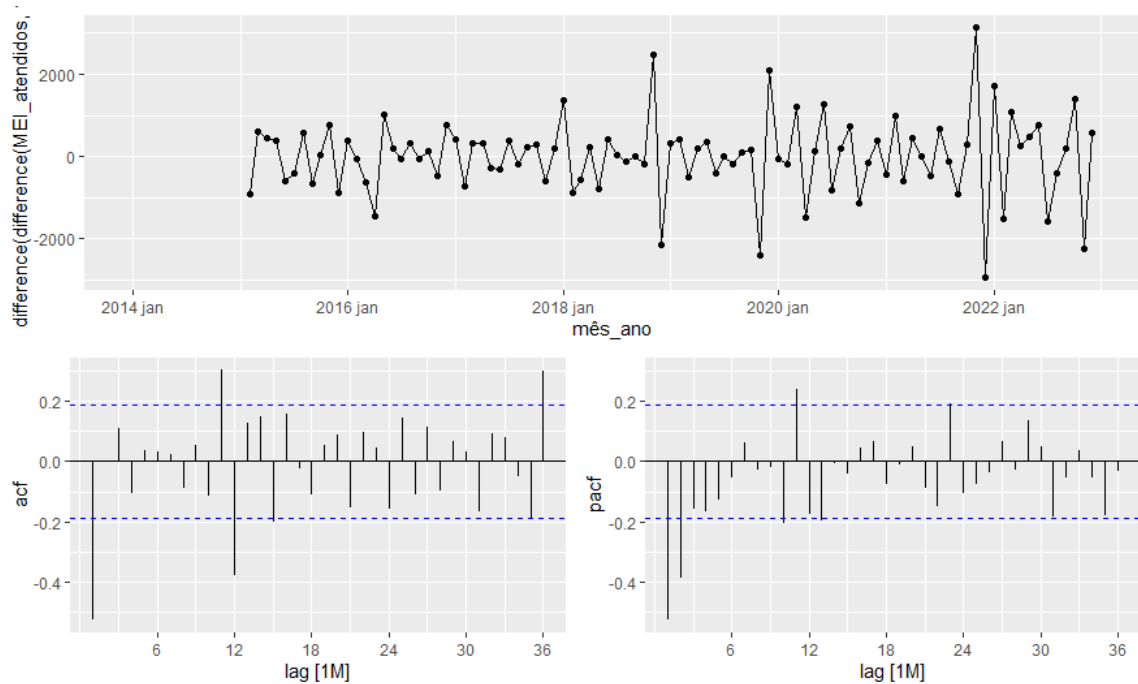
Figura 4.9: Análise diagnóstica do modelo de atendimentos ao MEI



Fonte: Autoria própria.

Agora, iremos ajustar o modelo para o número de MEIs atendidos. Para isso, utilizaremos a mesma metodologia citada anteriormente, mas com o acréscimo da variável regressora de atendimentos ao MEI. Portanto, analisaremos a Figura 4.10, onde são apresentados os gráficos da ACF e da PACF.

Figura 4.10: Gráficos da diferença da diferença sazonal da série de MEI atendidos, da autocorrelação e da autocorrelação parcial



Fonte: Autoria própria.

Ao analisar os gráficos da ACF e da PACF, podemos sugerir o modelo com os valores de  $(p, d, q)(P, D, Q)$  sendo  $(2, 1, 2)(0, 1, 3)$ . Para selecionar nosso modelo, compararemos o REQM dos modelos com os valores observados de MEIs atendidos na parte de teste (ano de 2023). Temos, então, os seguintes modelos:

Tabela 4.3: Modelos propostos para os MEIs atendidos.

Modelo	<i>REQM</i>
SARIMAE(2,1,1)(0,1,3)	1598
SARIMAE(2,1,1)(0,1,2)	1637
SARIMAE(2,0,2)(0,1,1)	1664
SARIMA(0,1,1)(0,1,1)	1963

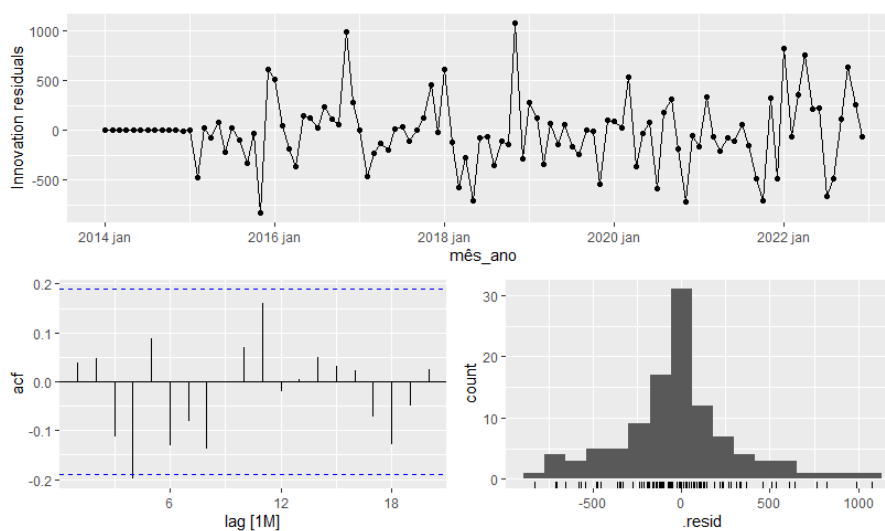
Para melhor compreensão, denotaremos os modelos de regressão dinâmica com erros SARIMA como SARIMAE. Observa-se que o modelo sugerido não está entre os propostos devido ao grande número de parâmetros. Assim, o modelo escolhido é aquele com o menor REQM, ou seja, o modelo SARIMAE(2,1,1)(0,1,3). Também foi considerado o modelo sugerido pela função **ARIMA**, que é o SARIMAE(2,0,2)(0,1,1), além de uma análise do modelo sem a variável regressora, que é o SARIMA(0,1,1)(0,1,1).

Para verificar se as suposições do modelo são atendidas, realizamos a análise diagnóstica do modelo, cuja representação pode ser vista na Figura 4.11. O teste de Ljung-



Box, para o lag 24, obteve um p-valor de 0,141, e o teste de Kolmogorov-Smirnov para a distribuição normal teve um p-valor de 0,05643. Portanto, ao nível de significância de 5%, constatamos que os resíduos são não correlacionados e possuem uma distribuição normal. Concluimos, então, que o modelo para os MEIs atendidos é o SARIMAE(2,1,1)(0,1,3).

Figura 4.11: Análise diagnóstica do modelo MEI atendidos

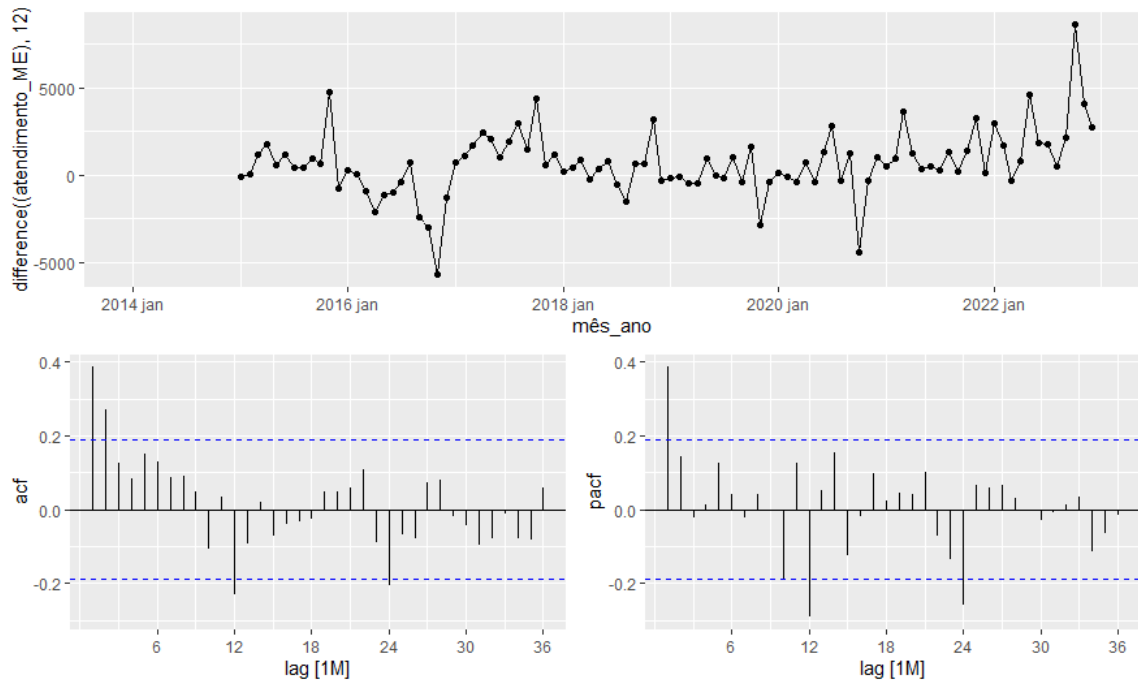


Fonte: Autoria própria.

## 4.2.2 Ajuste do modelo para a ME

Iremos ajustar o modelo para a série de atendimentos às MEs. Na Figura 4.12, apresentamos a diferença sazonal da série de atendimentos às MEs, bem como as autocorrelações (ACF) e as autocorrelações parciais (PACF).

Figura 4.12: Gráficos da diferença sazonal da série de atendimentos as MEs, da autocorrelação e da autocorrelação parcial



Fonte: Autoria própria.

Após a análise dos lags significativos, observamos que na PACF temos 1 lag significativo na parte não sazonal e 2 lags significativos na parte sazonal. Na ACF, identificamos 2 lags significativos na parte sazonal e 2 lags significativos na parte não sazonal, resultando em um modelo SARIMA(1,0,2)(2,1,2). O modelo sugerido pela função ARIMA foi um SARIMA(0,1,1)(0,1,1).

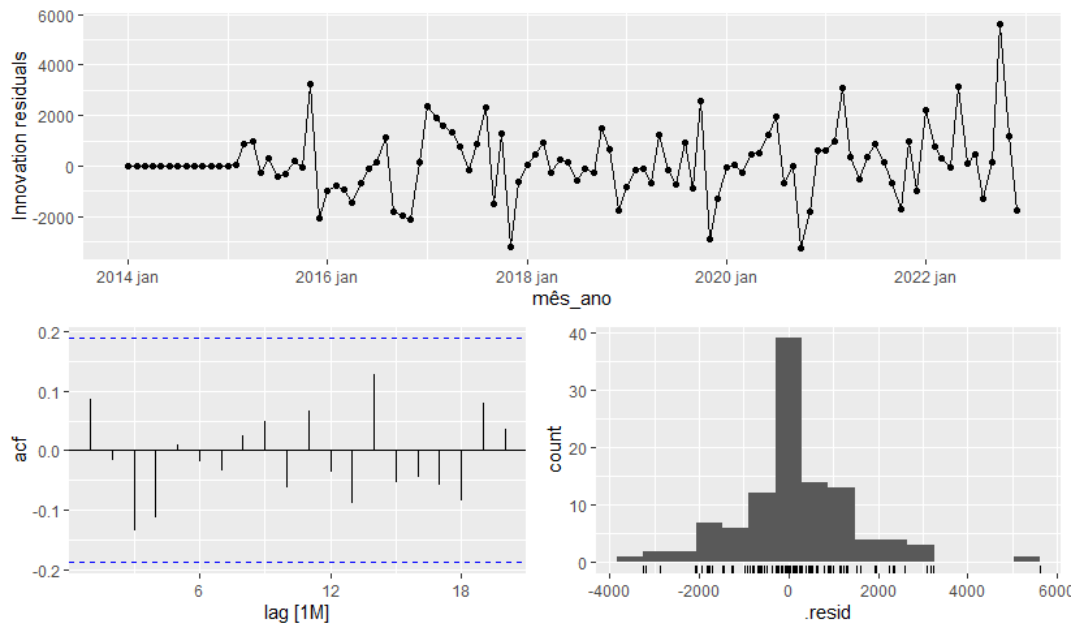
Agora, selecionaremos o modelo com base no *REQM*, utilizando os valores observados na parte de teste (ano de 2023) dos atendimentos às MEs. Com isso, temos os seguintes modelos:

Tabela 4.4: Modelos para os atendimentos as MEs.

Modelo	$REQM$
SARIMA(0,1,1)(0,1,1)	1885
SARIMA(1,1,1)(0,1,1)	2001
SARIMA(1,1,1)(1,1,1)	2005
SARIMA(1,0,2)(2,1,2)	2380

Iremos agora realizar a análise diagnóstica do modelo escolhido, que é o SARIMA(0,1,1)(0,1,1), o qual apresenta o menor  $REQM$ . A visualização dessa análise pode ser encontrada na Figura 4.13. Realizamos o teste de Ljung-Box para o lag 24, que obteve um p-valor de 0,504, e o teste de Kolmogorov-Smirnov para a distribuição normal apresentou um p-valor de 0,1739. Assim, ao nível de significância de 5%, concluímos que os resíduos são não correlacionados e possuem uma distribuição normal. Portanto, a série de atendimentos às MEs será modelada por um SARIMA(0,1,1)(0,1,1).

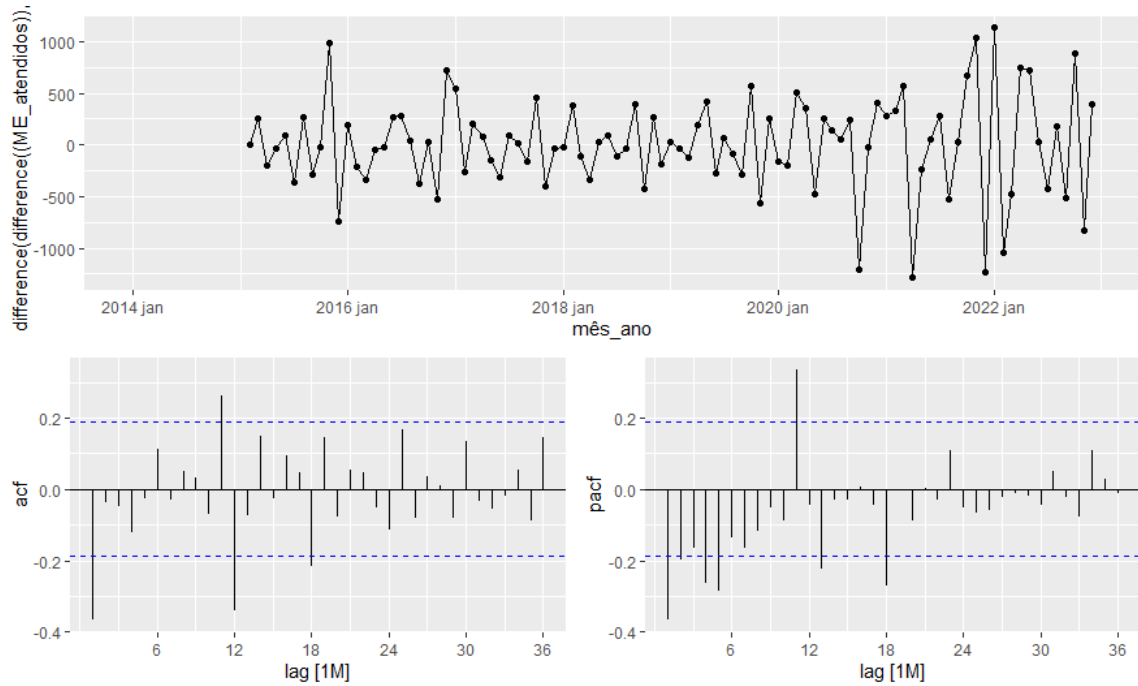
Figura 4.13: Análise diagnóstica do modelo de atendimentos a ME



Fonte: Autoria própria.

Agora, iremos ajustar o modelo para o número de MEs atendidos. Para isso, utilizaremos a mesma metodologia citada anteriormente, mas com o acréscimo da variável regressora de atendimentos a MEs. Portanto, analisaremos a Figura 4.14, que apresenta os gráficos da ACF e da PACF.

Figura 4.14: Gráficos da diferença da diferença sazonal da série de ME atendidos, da autocorrelação e da autocorrelação parcial



Fonte: Autoria própria.

Analisando os gráficos da ACF e da PACF, sugerimos o modelo com os valores de  $(p, d, q)(P, D, Q)$  sendo  $(3, 1, 2)(0, 1, 1)$ . Para selecionar o modelo, compararemos o *REQM* dos modelos com os valores observados de MEs atendidos na parte de teste (ano de 2023). Temos, então, os seguintes modelos:

Tabela 4.5: Modelos propostos para as MEs atendidos.

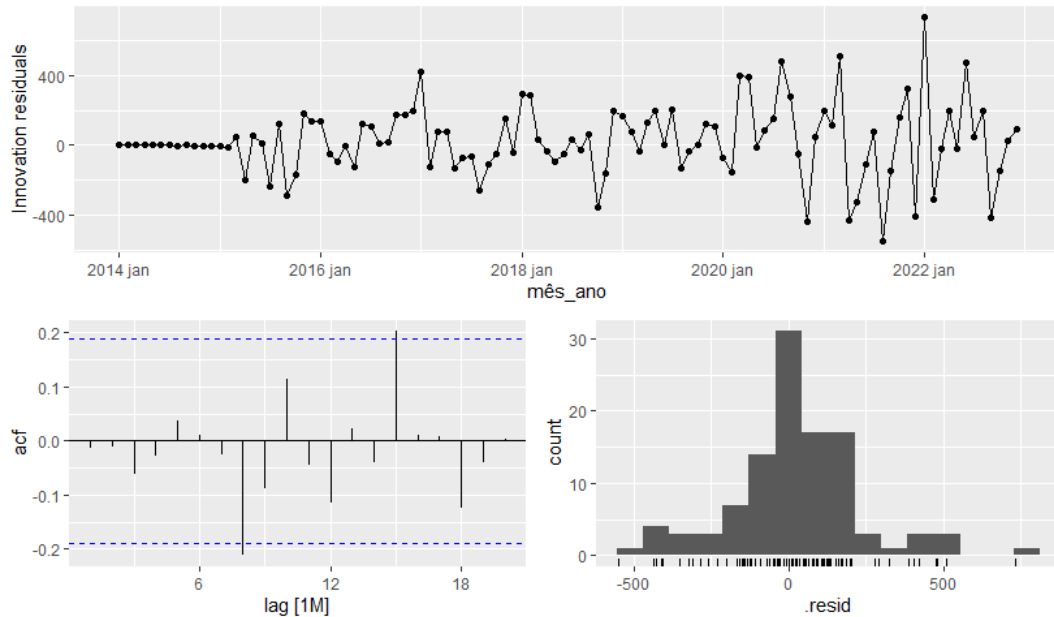
Modelo	<i>REQM</i>
SARIMAE(2,1,0)(1,1,0)	890
SARIMAE(3,1,1)(0,1,1)	910
SARIMAE(3,1,2)(0,1,1)	912
SARIMA(0,1,3)(2,0,0)	1040

Primeiramente, escolheremos o modelo com o menor *REQM*, ou seja, o modelo SARIMAE(2,1,0)(1,1,0). Contudo, ao analisarmos o gráfico de autocorrelação dos resíduos e o teste de Ljung-Box para o lag 24, observamos que os resíduos são autocorrelacionados. Portanto, devemos escolher outro modelo. O modelo final que satisfaz essas condições foi o SARIMAE(3,1,1)(0,1,1). Além disso, também consideramos o modelo sem a variável regressora, que foi um SARIMA(0,1,3)(2,0,0).

Como o primeiro modelo escolhido foi rejeitado devido à autocorrelação dos resíduos, faremos uma análise diagnóstica do modelo final escolhido, a qual pode ser visualizada

na Figura 4.15. O teste de Ljung-Box para o lag 24 obteve um p-valor de 0,339, e o teste de Kolmogorov-Smirnov para a distribuição normal teve um p-valor de 0,2604. Assim, ao nível de significância de 5%, concluímos que os resíduos são não correlacionados e possuem uma distribuição normal. Portanto, o modelo para a quantidade de MEs atendidas será o SARIMAE(3,1,1)(0,1,1).

Figura 4.15: Análise diagnóstica do modelo ME atendidos

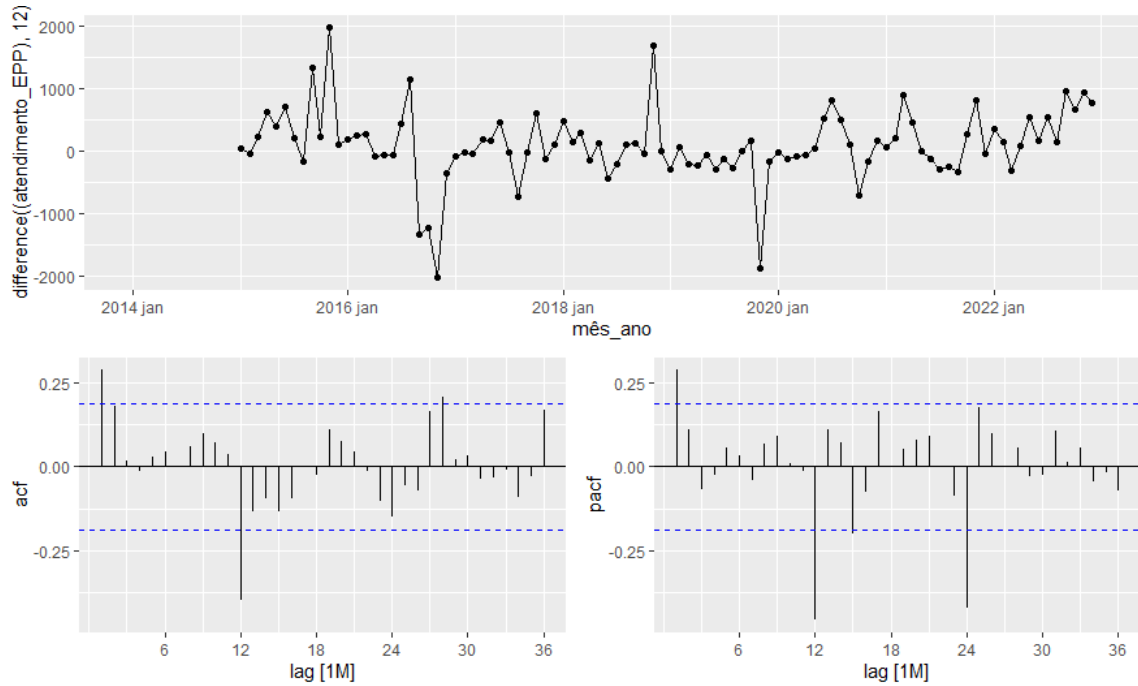


Fonte: Autoria própria.

### 4.2.3 Ajuste do modelo para a EPP

Por fim, realizaremos o ajuste do modelo para a série de atendimentos às EPPs. Na Figura 4.16, apresentamos a diferença sazonal da série de atendimentos às EPPs, bem como as ACF e as PACF.

Figura 4.16: Gráficos da diferença sazonal da série de atendimentos as EPPs, da autocorrelação e da autocorrelação parcial



Fonte: Autoria própria.

Ao analisar os lags significativos dos gráficos da ACF e PACF, podemos sugerir o modelo  $SARIMA(1,0,2)(2,1,1)$ . O modelo sugerido pela função `ARIMA` foi um  $SARIMA(1,0,0)(2,0,0)$ .

Selecionaremos o modelo com base no *REQM* utilizando os valores observados na parte de teste (ano de 2023) dos atendimentos às EPPs. Com isso, temos os seguintes modelos:

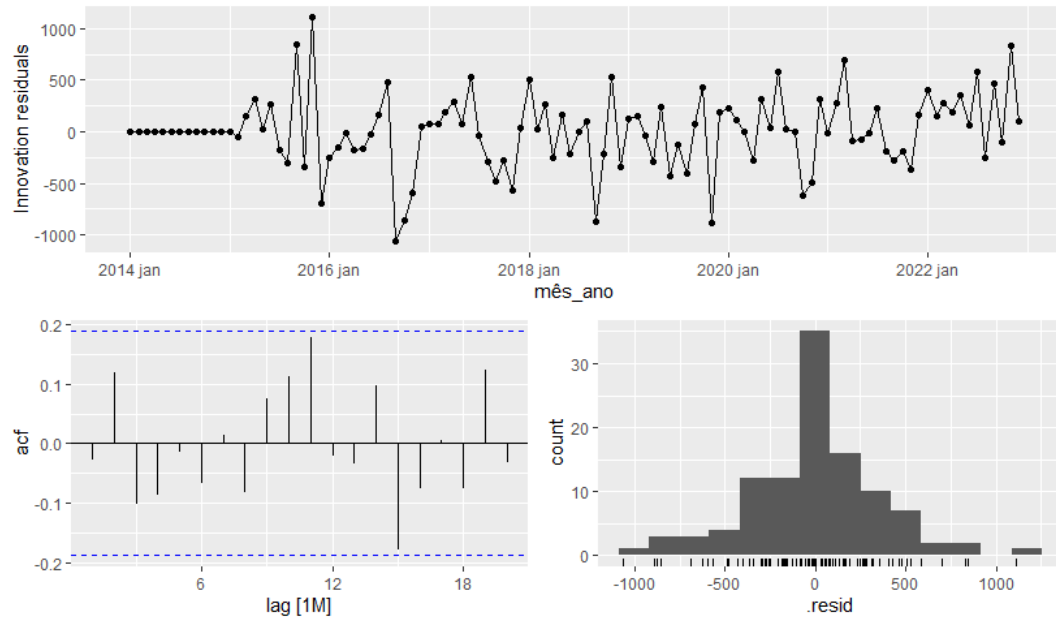
Tabela 4.6: Modelos para os atendimentos as EPPs.

Modelo	<i>REQM</i>
$SARIMA(1,1,1)(2,1,1)$	474
$SARIMA(1,1,1)(1,1,1)$	481
$SARIMA(1,0,0)(2,0,0)$	724
$SARIMA(1,0,2)(2,1,1)$	869

Agora, realizaremos a análise diagnóstica do modelo escolhido, que pode ser visualizada na Figura 4.17. O modelo com o menor *REQM*, escolhido, é o  $SARIMA(1,1,1)(2,1,1)$ . Também realizaremos o teste de Ljung-Box para o lag 24, que obteve um p-valor de 0,269, e o teste de Kolmogorov-Smirnov para a distribuição normal, que teve um p-valor de 0,1256. Assim, ao nível de significância de 5%, concluímos que os resíduos são não correlacionados e possuem uma distribuição normal. Portanto, a série de atendimentos às EPPs será

modelada por um SARIMA(1,1,1)(2,1,1).

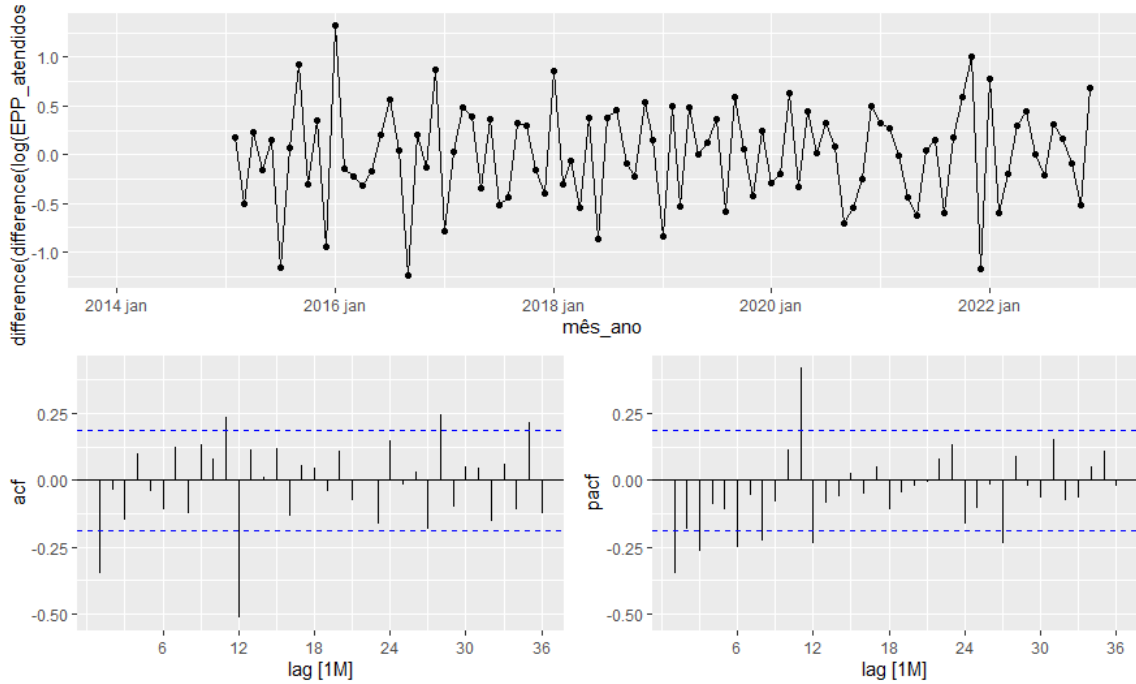
Figura 4.17: Análise diagnóstica do modelo de atendimentos a EPP



Fonte: Autoria própria.

Para finalizarmos este capítulo de ajuste de modelos, ajustaremos um modelo para o log do número de EPPs atendidos. Utilizaremos o log para estabilizar a variância, aplicando a mesma metodologia citada anteriormente, com o acréscimo de uma variável regressora de atendimentos a EPP. Portanto, analisaremos a Figura 4.18, que apresenta os gráficos da ACF e da PACF.

Figura 4.18: Gráficos do log da diferença da diferença sazonal da série de EPP atendidas, da autocorrelação e da autocorrelação parcial



Fonte: Autoria própria.

Após a análise dos gráficos da ACF e da PACF, sugerimos o modelo com os valores de  $(p, d, q)(P, D, Q)$  sendo  $(3, 1, 1)(1, 1, 1)$ . Para selecionar o modelo, compararemos o *REQM* dos modelos com os valores observados de EPPs atendidas na parte de teste (ano de 2023). Temos, então, os seguintes modelos:

Tabela 4.7: Modelos propostos para as EPPs atendidas.

Modelo	<i>REQM</i>
SARIMA(0,0,2)(0,1,1)	131
SARIMAE(0,0,2)(0,1,1)	139
SARIMAE(2,1,2)(1,1,1)	140
SARIMAE(3,1,1)(1,1,1)	140

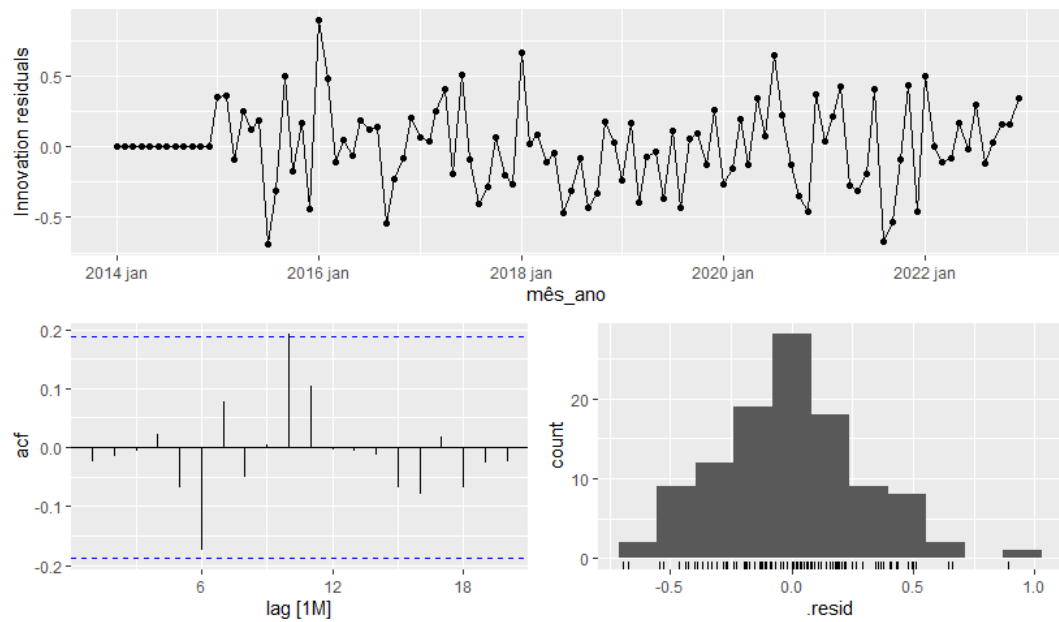
O primeiro modelo a ser escolhido será o SARIMA(0,0,2)(0,1,1). Observamos que, para a série de EPPs, o modelo que apresentou o melhor desempenho conforme nosso critério de escolha foi o modelo sem a variável regressora.

Após a análise diagnóstica do modelo escolhido, a qual pode ser visualizada na Figura 4.19, realizamos o teste de Ljung-Box para o lag 24, que obteve um p-valor de 0,853, e o teste de Kolmogorov-Smirnov para a distribuição normal, com um p-valor de 0,7233. Assim, ao nível de significância de 5%, concluímos que os resíduos são não correlacionados e possuem uma distribuição normal. Portanto, o modelo para as EPPs



atendidas será um SARIMA(0,0,2)(0,1,1).

Figura 4.19: Análise diagnóstica do modelo EPP atendidas



Fonte: Autoria própria.

## 4.3 Previsão

Após a análise das séries temporais e os ajustes dos modelos, realizaremos a previsão das empresas atendidas e dos atendimentos as empresas para o ano de 2024.

Para realizar as previsões necessárias, utilizaremos a função *forecast* do pacote *forecast* do software *R*. Esta função retorna os valores previstos para as variáveis de interesse. Primeiro, escolhemos o modelo adequado; em seguida, definimos a quantidade de valores futuros pela letra *h*. Por fim, a saída da função mostrará as previsões pontuais e os intervalos de confiança.

Ressaltamos que, nos modelos que incluem a variável regressora, os valores previstos para o número de MEIs, MEs e EPPs atendidos dependerão da quantidade prevista de atendimentos a MEIs, MEs e EPPs, respectivamente.

### 4.3.1 Previsão para o MEI

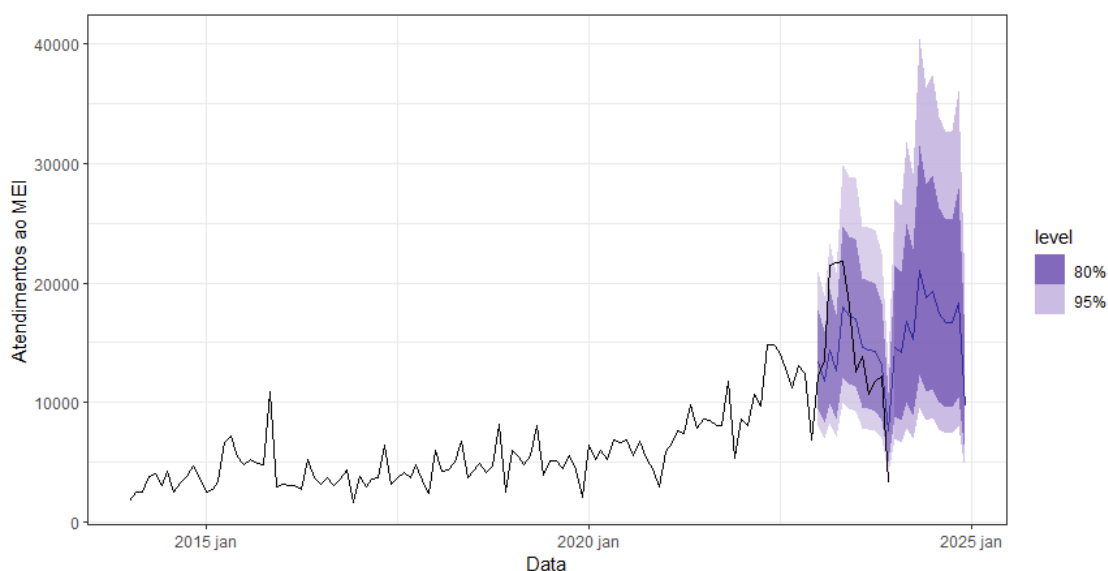
O modelo que selecionamos para fazer a previsão para a quantidade de MEI atendidos foi o SARIMAE(2,1,1)(0,1,3) sendo sua variável regressora uma SARIMA(1,0,1)(2,1,0) para quantidade de atendimentos ao MEI, por isso temos a seguinte tabela de previsão para o ano de 2024:

Tabela 4.8: Previsão de MEIs atendidos e Atendimentos aos MEIs

Mês	MEIs atendidos	Atendimentos aos MEIs
Janeiro	8116	14602
Fevereiro	6433	14136
Março	6816	16811
Abril	5699	15234
Maio	7075	21051
Junho	6567	18770
Julho	5777	19246
Agosto	5200	17401
Setembro	5182	16699
Outubro	4989	16715
Novembro	6254	18349
Dezembro	2610	9698
<b>Total</b>	<b>70717</b>	<b>198711</b>

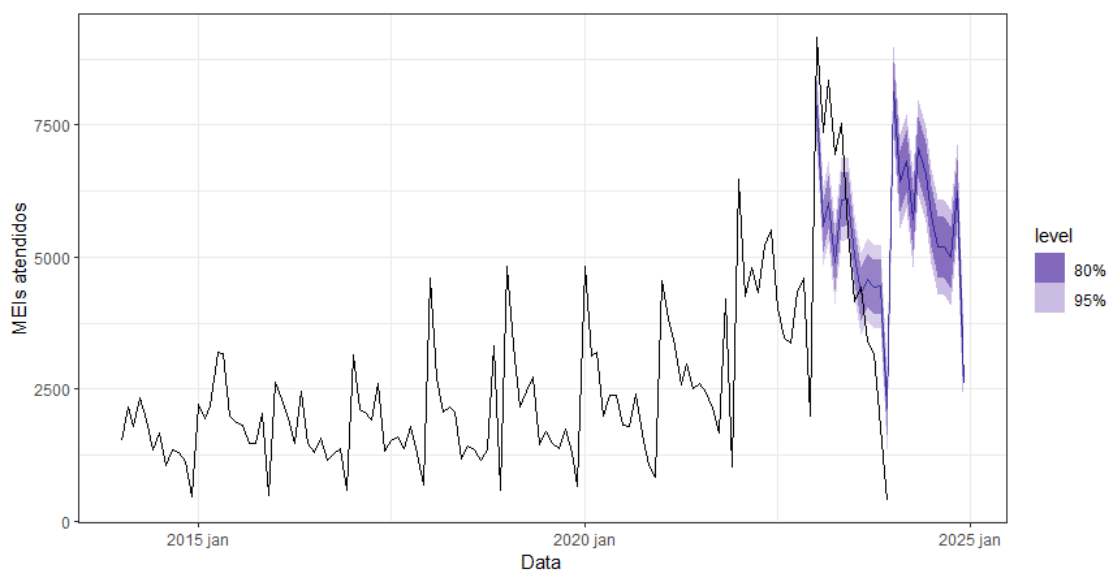
Com base nos modelos escolhidos, a previsão para a quantidade de MEIs atendidos é de 70.717, com 198.711 atendimentos ao MEI. Podemos visualizar o comportamento das séries nas Figuras 4.20 e 4.21. Nessas figuras, os valores em azul representam as previsões do modelo para os anos de 2023 e 2024. Além disso, é possível observar o comportamento da série em ambos os anos e comparar com os valores observados em 2023.

Figura 4.20: Gráfico da previsão de de MEIs atendidos no ano de 2024



Fonte: Autoria própria.

Figura 4.21: Gráfico da previsão de de MEIs atendidos no ano de 2024



Fonte: Autoria própria.

### 4.3.2 Previsão para a ME

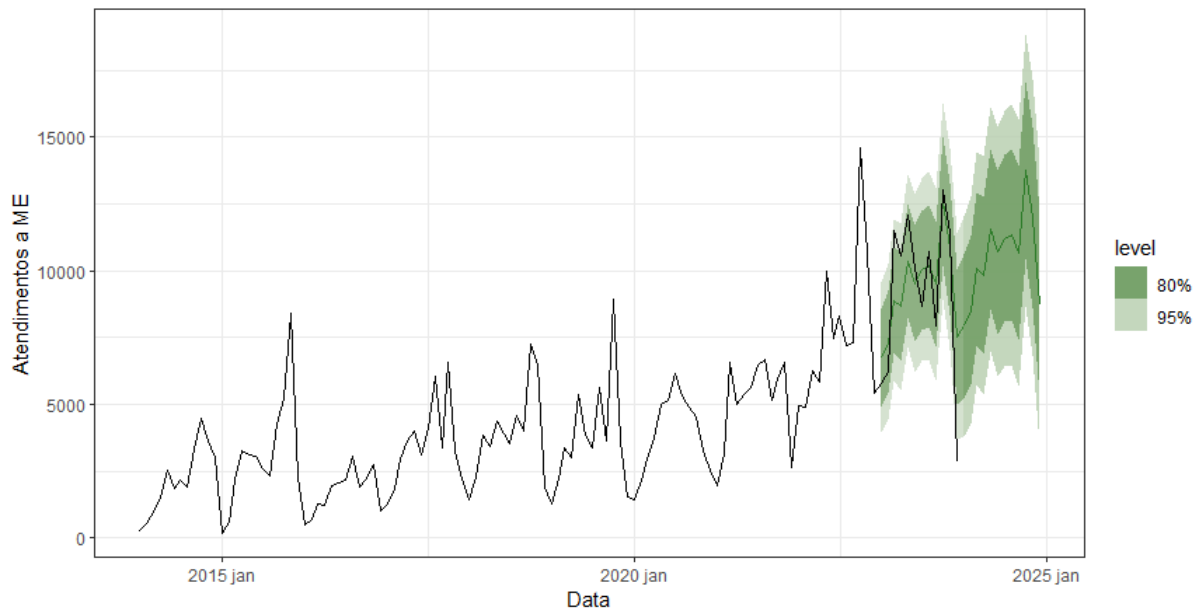
Com base no modelo selecionado, realizaremos a previsão para o ano de 2024 para o número de MEs atendidas, utilizando o modelo SARIMAE(3,1,1)(0,1,1), cuja variável regressora é um SARIMA(0,1,1)(0,1,1) para a quantidade de atendimentos às MEs. Portanto, temos a seguinte tabela de previsão para o ano de 2024:

Tabela 4.9: Previsão de MEs atendidos e Atendimentos as MEs

Mês	MEs atendidas	Atendimentos as MEs
Janeiro	2620	7916
Fevereiro	2350	8504
Março	2989	10069
Abril	2630	9826
Maio	2771	11545
Junho	2657	10700
Julho	2767	11194
Agosto	2591	11350
Setembro	2365	10656
Outubro	2698	13762
Novembro	2515	11989
Dezembro	1874	8708
<b>Total</b>	<b>30827</b>	<b>126220</b>

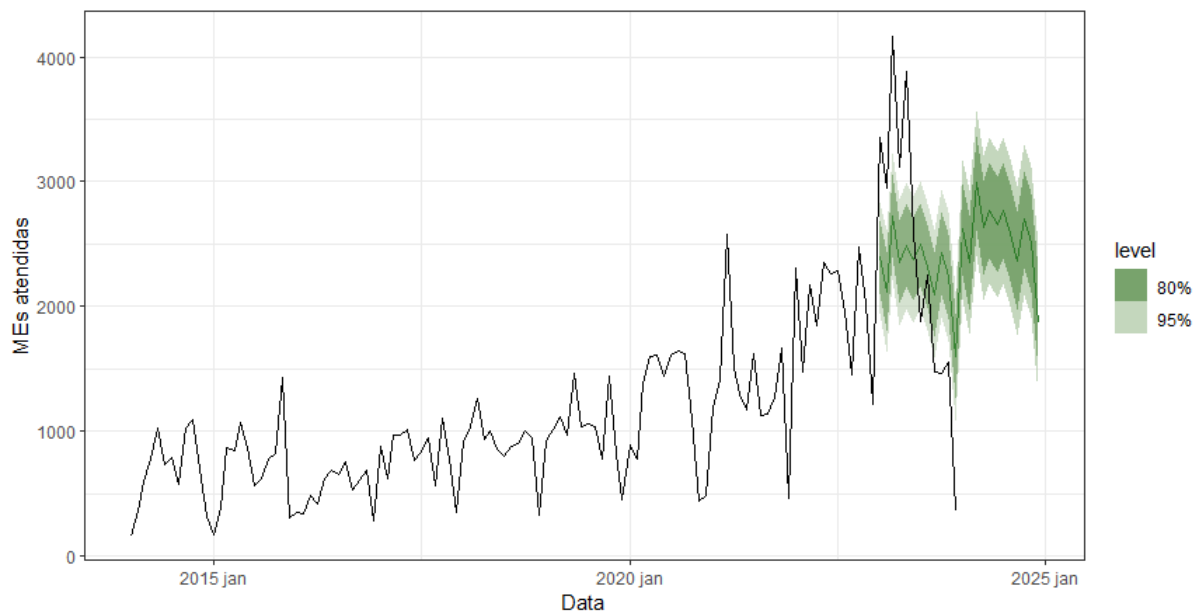
Tendo em vista os modelos escolhidos, a previsão para a quantidade de MEs atendidos no ano de 2024 é de 30.827, com 126.220 atendimentos às MEs. Podemos visualizar o comportamento das séries nas imagens 4.22 e 4.23, onde os valores em verde representam as previsões do modelo para os anos de 2023 e 2024. Além disso, é possível observar o comportamento da série em ambos os anos e comparar com os valores observados em 2023.

Figura 4.22: Gráfico da previsão de de MEIs atendidos no ano de 2024



Fonte: Autoria própria.

Figura 4.23: Gráfico da previsão de de MEIs atendidos no ano de 2024



Fonte: Autoria própria.

### 4.3.3 Previsão para a EPP

Como observamos no ajuste do modelo para as EPPs, notamos um comportamento distinto em relação aos demais (MEI e ME). O melhor modelo para o número de EPPs atendidas é o SARIMA(0,0,2)(0,1,1). Também realizamos a previsão para o número de

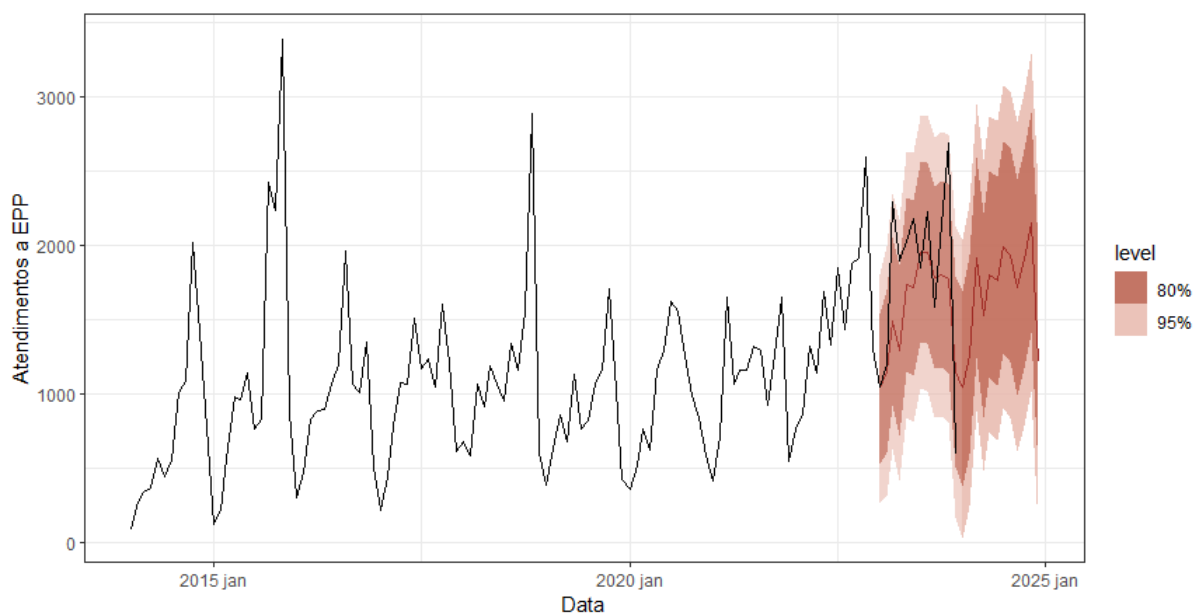
atendimentos a EPPs, que retornou um modelo SARIMA(1,1,1)(2,1,1). Portanto, realizaremos a previsão tanto para as EPPs atendidas quanto para os atendimentos a EPPs. A seguir, apresentamos a tabela de previsão para o ano de 2024:

Tabela 4.10: Previsão de EPPs atendidas e Atendimentos as EPPs

Mês	EPPs atendidas	Atendimentos as EPPs
Janeiro	345	1035
Fevereiro	393	1262
Março	557	1916
Abril	443	1527
Maio	510	1801
Junho	450	1760
Julho	422	1985
Agosto	442	1935
Setembro	392	1718
Outubro	399	1901
Novembro	341	2151
Dezembro	164	1220
<b>Total</b>	<b>4858</b>	<b>20211</b>

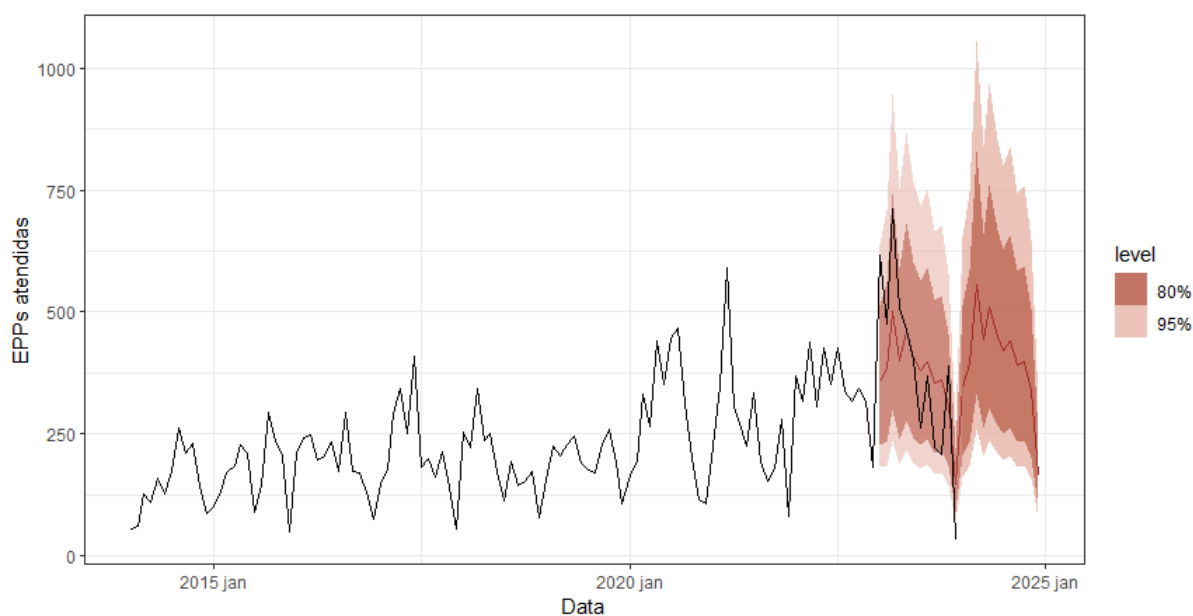
Com base nos modelos escolhidos, a previsão para a quantidade de EPPs atendidas no ano de 2024 é de 30.827, com 126.220 atendimentos às EPPs. Podemos visualizar o comportamento das séries nas Figuras 4.24 e 4.25, onde os valores em verde representam as previsões do modelo para os anos de 2023 e 2024. Nessas figuras, é possível observar o comportamento da série em ambos os anos e comparar com os valores observados em 2023.

Figura 4.24: Gráfico da previsão de de EPPs atendidas no ano de 2024



Fonte: Autoria própria.

Figura 4.25: Gráfico da previsão de de EPPs atendidos no ano de 2024



Fonte: Autoria própria.

#### 4.3.4 Previsão total

Com base no que foi visualizado nos subcapítulos anteriores e nos modelos escolhidos para o MEI, ME e EPP, podemos afirmar que o Sebrae/RN atenderá um total de 106.404 micro e pequenos negócios no RN, com 345.144 atendimentos. A previsão men-

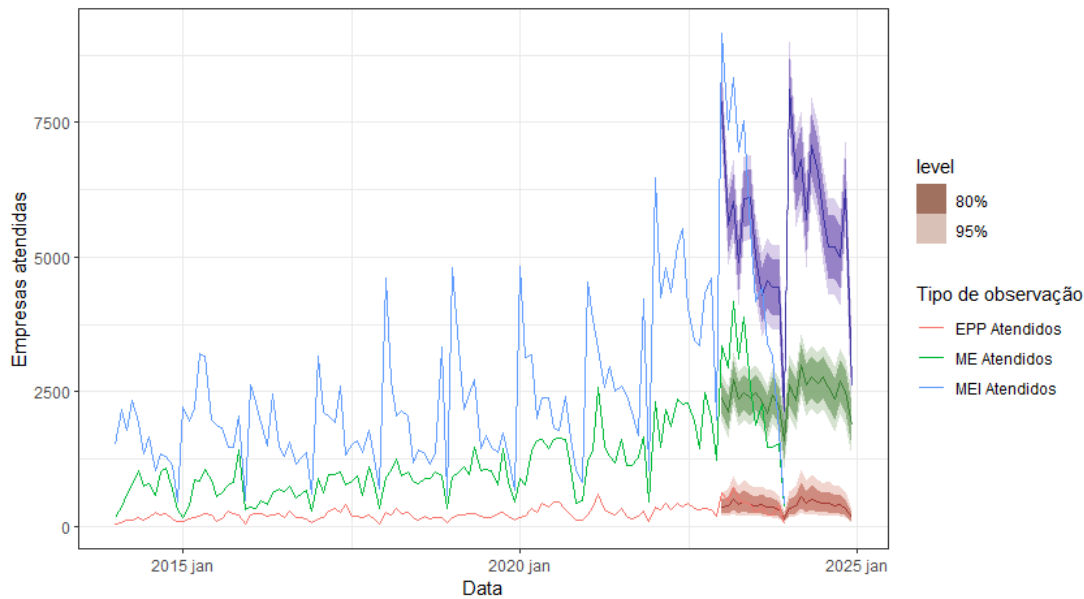
sal desses valores pode ser visualizada na Tabela 4.11. Além disso, o comportamento de cada série é mostrado nas Figuras 4.26 e 4.27 para os anos de 2023 e 2024, permitindo a comparação dos valores previstos em 2023 com os valores observados no mesmo período.

Tabela 4.11: Empresas atendidas e Atendimentos as empresas

Mês	Empresas Atendidas				Atendimentos a empresas			
	MEI	ME	EPP	Total	MEI	ME	EPP	Total
Janeiro	8116	2620	345	11081	14602	7916	1035	23553
Fevereiro	6433	2350	393	9176	14136	8504	1262	23902
Março	6816	2989	557	10362	16811	10069	1916	28796
Abril	5699	2630	443	8772	15234	9826	1527	26587
Maio	7075	2771	510	10357	21051	11545	1801	34397
Junho	6567	2657	450	9674	18770	10700	1760	31231
Julho	5777	2767	422	8966	19246	11194	1985	32425
Agosto	5200	2591	442	8233	17401	11350	1935	30686
Setembro	5182	2365	392	7939	16699	10656	1718	29073
Outubro	4989	2698	399	8086	16715	13762	1901	32379
Novembro	6254	2515	341	9111	18349	11989	2151	32489
Dezembro	2610	1874	164	4647	9698	8708	1220	19626
Total	70717	30827	4858	106402	198712	126220	20211	345143

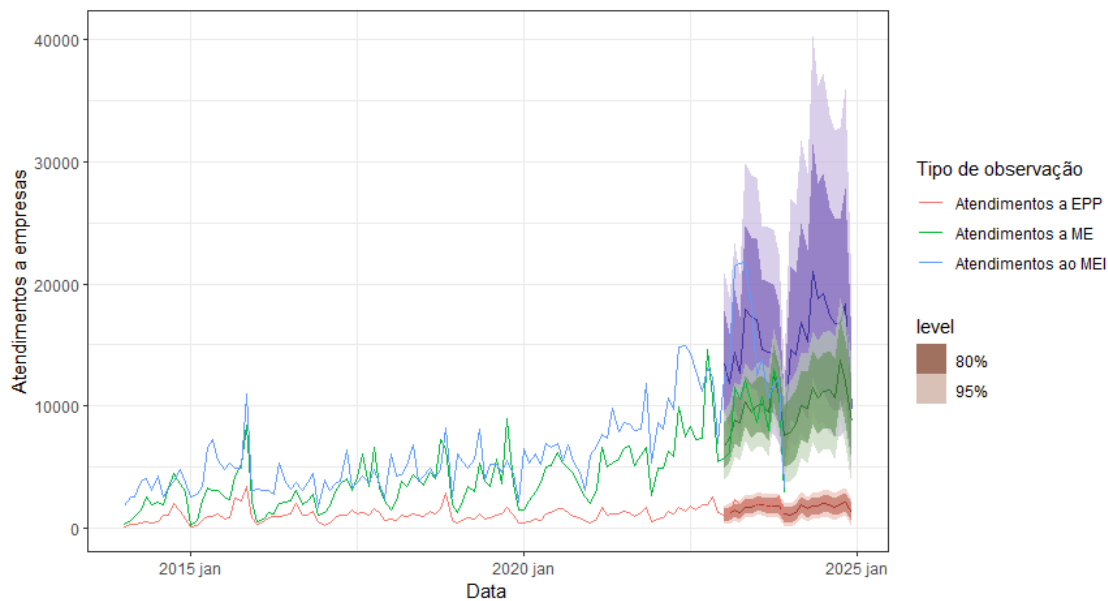


Figura 4.26: Gráfico da previsão de empresas atendidas no ano de 2024



Fonte: Autoria própria.

Figura 4.27: Gráfico da previsão de empresas atendidas no ano de 2024



Fonte: Autoria própria.

## Capítulo 5

### Considerações Finais

Nesta monografia, o objetivo principal foi explorar a abordagem de análise de séries temporais para entender o comportamento do número de micro e pequenos negócios atendidos pelo Sebrae nos anos de 2014 a 2023. Este estudo auxiliará na compreensão do papel do Sebrae no desenvolvimento e no apoio ao empreendedorismo no Brasil. Foram também propostos objetivos específicos, que incluíram a escolha do melhor modelo para cada tipo de micro e pequena empresa, a aplicação da metodologia Box-Jenkins para previsão da série temporal e, por fim, a realização de previsões para os próximos 12 meses.

Ao avaliar os resultados descritos no Capítulo 4, pode-se afirmar que todos os objetivos delineados no Capítulo 1 foram alcançados de maneira satisfatória. As séries temporais do MEI, ME e EPP analisadas, após os ajustes definidos, revelaram um comportamento estacionário ao longo do período estudado, mantendo uma média e variância constantes. Para o número de MEIs atendidos, identificamos um modelo SARIMAE(2,1,1)(0,1,3), com a variável regressora, atendimentos ao MEI, sendo um SARIMA(1,0,1)(2,1,0) como o mais apropriado. Para o número de MEs atendidas, identificamos um modelo SARIMAE(3,1,1)(0,1,1), com a variável regressora, atendimentos a ME, sendo um SARIMA(0,1,1)(0,1,1) como o mais apropriado. Por fim, para o número de EPPs atendidas, identificamos um modelo SARIMA(0,0,2)(0,1,1), com a variável regressora, atendimentos a EPP, sendo um SARIMA(1,1,1)(2,1,1) como o mais apropriado. Esses modelos foram selecionados com base no critério citado no Capítulo 3, apresentando um *REQM* de 1.598 para MEIs atendidos, 890 para MEs atendidas e 131 para EPPs atendidas. Os modelos terão como base os valores previstos da variável regressora para realizar a previsão dos

valores futuros. Esses resultados indicam que os modelos escolhidos foram eficazes na representação dos padrões observados nos dados dos clientes atendidos, fornecendo uma base sólida para análises e previsões futuras nesse contexto. Portanto, conclui-se que este estudo alcançou seus objetivos ao modelar o número de empresas que serão atendidas mensalmente, contribuindo com informações valiosas para a compreensão e gestão desses resultados.

Durante a elaboração deste trabalho, observamos a complexidade envolvida na previsão do número de micro e pequenas empresas atendidas, considerando a influência de diversos fatores. Com base nisso, sugerimos que futuros estudos possam complementar este trabalho com a inclusão de uma nova variável regressora mais adequada, como o número de micro e pequenas empresas existentes no Rio Grande do Norte, podendo ser detalhada conforme os moldes deste estudo.

# Referências Bibliográficas

- [1] DE PINHO, A. L. S. & SIROKY, A. N. **Notas de Aula da Disciplina EST0070 - Séries Temporais. 2023.** , Disponível para alunos matriculados na disciplina;
- [2] Harris, R., & Sollis, R., **Applied time series modelling and forecasting.** John Wiley & Sons, 2003;
- [3] Hyndman, R.J., & Athanasopoulos, G.; **Forecasting: principles and practice, 3rd edition**, OTexts: Melbourne, Australia. OTexts.com/fpp3, 2021;
- [4] Box, George E. P.; **Time series analysis : forecasting and control. – Fifth edition** / George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung, 2016
- [5] G. M. Ljung, G. E. P. Box **On a measure of lack of fit in time series models.** Biometrika, Volume 65, Issue 2, August 1978
- [6] Andrew V. Metcalfe, Paul S.P. Cowpertwait; **Introductory Time Series with R**, Springer New York, NY, 2008
- [7] Jonathan D. Cryer, Kung-Sik Chan **Time Series Analysis With Applications in R**, Second Edition, Springer, 2008;
- [8] R Core Team (2022). **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [9] Robert H. Shumway, David S. Stoffer **Time Series Analysis and Its Applications With R Examples**, Springer International Publishing AG 2017, 4th edition
- [10] TENÓRIO, Ricardo Bruno de Araújo. **Modelagem da cobertura de gelo marinho nos mares Antárticos de Weddell, Bellingshausen e Amundsen**

**com uso de redes neurais artificiais.** 2021. 73f. Tese (Doutorado em Ciências Climáticas) - Centro de Ciências Exatas e da Terra, Universidade Federal do Rio Grande do Norte, Natal, 2021.