

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

Instituto de Ciências Exatas (ICEX)

Programa de Pós-graduação em Estatística

Renato Hoffmam de Oliveira

**PREVISÃO DE VENDAS DE PEÇAS PARA MÁQUINAS PESADAS  
POR MEIO DE SÉRIES TEMPORAIS**

Belo Horizonte  
2022

Renato Hoffmam de Oliveira

**PREVISÃO DE VENDAS DE PEÇAS PARA MÁQUINAS PESADAS  
POR MEIO DE SÉRIES TEMPORAIS**

Monografia de especialização, apresentada ao Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção de título de Especialista em Estatística.

Orientadora: Ela Mercedes M. de Toscano

Coorientador: Luis Alberto T. Medrano

Belo Horizonte

2022

2022, Renato Hoffmam de Oliveira.  
Todos os direitos reservados

Oliveira, Renato Hoffmam de.

O48p      Previsão de vendas de peças para máquinas pesadas  
por meio de séries temporais [manuscrito] / Renato  
Hoffmam de Oliveira. — 2022.  
85.f. il.

Orientadora: Ela Mercedes Medrano de Toscano  
Coorientador: Luís Alberto Toscano Medrano  
Monografia (especialização) - Universidade Federal  
de Minas Gerais, Instituto de Ciências Exatas,  
Departamento de Estatística.  
Referências 84-85.

1. Estatística. 2. Análise de séries temporais.  
3. Análise multivariada. 4. Peças de máquinas – Previsão  
de vendas. I. Toscano, Ela Mercedes Medrano de.  
II. Medrano Toscano, Luís Alberto. III. Universidade  
Federal de Minas Gerais, Instituto de Ciências Exatas,  
Departamento de Estatística. IV. Título.

CDU 519.2 (043)

Ficha catalográfica elaborada pela bibliotecária Irenquer Vismeg Lucas  
Cruz, CRB 6/819 - Universidade Federal de Minas Gerais - ICEx



Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Estatística  
Programa de Pós-Graduação / Especialização  
Av. Pres. Antônio Carlos, 6627 - Pampulha  
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br  
Tel: 3409-5923 – FAX: 3409-5924

## ATA DO 245ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE RENATO HOFFMAM DE OLIVEIRA.

Aos quinze dias do mês de setembro de 2022, às 08:30 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso do aluno **Renato Hoffmam de Oliveira**, intitulado: “*Previsão de vendas de peças para máquinas pesadas por meio de séries temporais*”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, a Presidente da Comissão, Professora Ela Mercedes Medrano – Orientadora, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 15 de setembro de 2022.

Prof.<sup>a</sup> Ela Mercedes Medrano (Orientadora)  
Departamento de Estatística / UFMG

Prof. Luis Alberto Toscano M. (Coorientador)  
Departamento de Matemática / UFRJ

Prof.<sup>a</sup> Sueli Aparecida Mingoti  
Departamento de Estatística / UFMG

Prof. Mario Ernesto Piscocoya Diaz  
IME / UFG



**Universidade Federal de Minas Gerais**  
**Instituto de Ciências Exatas**  
**Departamento de Estatística**  
**Programa de Pós-Graduação / Especialização**  
Av. Pres. Antônio Carlos, 6627 - Pampulha  
31270-901 – Belo Horizonte – MG

E-mail: [pgest@ufmg.br](mailto:pgest@ufmg.br)  
Tel: 3409-5923 – FAX: 3409-5924

## **DECLARAÇÃO DE CUMPRIMENTO DE REQUISITOS PARA CONCLUSÃO DO CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA.**

Declaro para os devidos fins que Renato Hoffmam de Oliveira, número de registro 2017699068, cumpriu todos os requisitos necessários para conclusão do curso de Especialização em Estatística e que me entregou a versão final corrigida. O trabalho foi apresentado no dia 15 de setembro de 2022 com o título “*Previsão de vendas de peças para máquinas pesadas por meio de séries temporais*”.

Belo Horizonte, 24 de novembro de 2022

Roberto da  
Costa  
Quinino:80  
871291720

Assinado de forma  
digital por Roberto  
da Costa  
Quinino:8087129172  
0  
Dados: 2022.11.24  
11:41:48 -03'00'

*Prof. Roberto da Costa Quinino*  
Coordenador da Comissão  
do Curso de Especialização  
em Estatística

**Prof. Roberto da Costa Quinino**  
**Coordenador do curso de**  
**Especialização em Estatística**  
**Departamento de Estatística / UFMG**

## RESUMO

Esta monografia tem como objetivo realizar a previsão, para 12 meses, da quantidade de vendas de peças de reposição para máquinas pesadas, em função de seu próprio histórico de vendas e de potenciais variáveis correlacionadas. Assim, foram estimados modelos univariados, nos quais a variável regressora é a própria variável defasada no tempo, e modelos multivariados, nos quais, além da introdução da variável dependente defasada no tempo, como regressora, foram adicionadas variáveis exógenas. Foram aplicados os modelos ARIMA e SARIMA para os modelos univariados e a técnica de regressão dinâmica para os modelos multivariados, que é a combinação de um modelo de regressão com um modelo adicional para estimar os resíduos da regressão, uma vez que, em séries temporais, os resíduos são autocorrelacionados. Nesta monografia, optou-se por estimar os modelos ARIMAX para os modelos multivariados, que usam de uma função de transferência para introduzir covariáveis, ou variáveis exógenas, ao modelo. A previsão de vendas para peças de reposição é importante para que seja possível equilibrar o estoque em níveis satisfatórios, manter os custos em níveis adequados e atender o cliente em tempo hábil.

Palavras-chave: séries temporais; ARIMA; SARIMA; ARIMAX; regressão dinâmica.

## **ABSTRACT**

This monograph aims to forecast, for 12 months, the amount of sales of spare parts for heavy machinery, based on its own sales history and potential correlated variables. Thus, univariate models were estimated, in which the regression variable is the time-lagged variable itself, and multivariate models, in which exogenous variables were added along with the introduction of the dependent variable lagged in time, as a regression variable. ARIMA and SARIMA models were applied for the univariate models, and dynamic regression technique for the multivariate models, which is the combination of a regression model with an additional model to estimate the regression residuals, since, in time series, the residuals are autocorrelated. In this monograph, it was decided to estimate the ARIMAX models for the multivariate models, which use a transfer function to introduce covariates or exogenous variables to the model. The sales forecast for spare parts is important so that it is possible to balance the stock at satisfactory levels, keep costs at adequate levels and serve the customer in a timely manner.

Keywords: time series; ARIMA; SARIMA; ARIMAX; dynamic regression.

## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>9</b>
1.1. Justificativa.....	9
1.2. Objetivo geral .....	10
1.3. Objetivos específicos.....	10
<b>2. METODOLOGIA.....</b>	<b>11</b>
2.1. Conceitos de série temporal.....	12
2.1.1. Componentes de uma série temporal.....	13
2.1.2. Processo estocástico .....	14
2.1.3. Autocorrelação.....	16
2.1.4. Estacionariedade.....	18
2.1.5. Operadores.....	20
2.2. Modelos de série temporal.....	21
2.2.1. Modelos Autorregressivos (AR).....	21
2.2.2. Modelos de Médias Móveis (MA) .....	22
2.2.3. Modelos ARMA .....	23
2.2.4. Modelos ARIMA .....	23
2.2.5. Modelos SARIMA.....	24
2.2.6. Metodologia Box & Jenkins (1976) .....	24
2.2.7. Modelos ARIMAX.....	27
2.2.8. Metodologia Box & Jenkins (1976) para modelos dinâmicos .....	29
2.3. Critérios de informação .....	31
2.3.1 AIC .....	31
2.3.2 BIC .....	31
2.4. Medidas de acurácia .....	32
2.4.1. ME (Erro Médio).....	33
2.4.2. MSE (Erro Quadrado Médio ou Desvio Quadrado Médio) .....	33
2.4.3. RMSE (Raiz do Erro Quadrado Médio).....	33
2.4.4. MAE (Erro Médio Absoluto) ou MAD (Desvio Padrão Absoluto Médio).....	33
2.4.5. MPE ou EMP (Erro Médio Percentual).....	34
2.4.6. MAPE (Erro Percentual Absoluto Médio) .....	34
2.4.7. Theil's U (Coeficiente de Desigualdade de Theil).....	34
<b>3. ANÁLISE DOS DADOS .....</b>	<b>35</b>
3.1. Série Quantidade de Vendas de Peças.....	35
3.2. Série Quantidade de Serviços de Oficina .....	39
3.3. Série Quantidade de Máquinas Vendidas.....	42

3.4. Variáveis indicadoras .....	44
3.5. Correlação entre as séries .....	46
3.6. Estimação dos parâmetros dos modelos .....	47
3.6.1. Modelo 1 .....	48
3.6.2. Modelo 2.....	51
3.6.3. Modelo 3.....	55
3.6.4. Modelo 4.....	59
3.6.5. Modelo 5.....	63
3.6.6. Modelo 6.....	67
3.6.7. Estatísticas dos ajustes dos modelos.....	71
3.7. Previsões.....	73
3.7.1. Modelo 1 .....	74
3.7.2. Modelo 2.....	75
3.7.3. Modelo 3.....	76
3.7.4. Modelo 4.....	77
3.7.5. Modelo 5.....	78
3.7.6. Modelo 6.....	79
3.7.7. Estatísticas dos erros de previsão .....	80
<b>4. CONSIDERAÇÕES FINAIS.....</b>	<b>82</b>
REFERÊNCIAS .....	84

## 1. INTRODUÇÃO

Máquinas pesadas, ou máquinas de linha amarela, como as retroescavadeiras, escavadeiras e motoniveladoras, são equipamentos que podem ser utilizados em vários segmentos, como construção civil, agricultura e mineração, e em diferentes aplicações, como terraplanagem, construção ou reforma de casas e prédios, limpeza de grandes áreas e carregamento de insumos e resíduos.

As máquinas de linha amarela são bens de capital, ou bens de produção, e são utilizadas por empresas na prestação de serviços para outras empresas ou órgãos governamentais. Como bens de capital, sua comercialização caracteriza um mercado B2B, de empresa para empresa, e sua demanda é fortemente afetada pela economia e políticas governamentais, principalmente por grande parte dessas máquinas ser utilizada na prestação de serviços ao governo.

Tais máquinas precisam de manutenção contínua para manterem sua performance original de fábrica e, assim como os veículos de passeio, têm tabelas de revisões preventivas periódicas, conforme indicação dos fabricantes. Além das revisões preventivas, as máquinas também precisam de manutenção corretiva constantemente, pois podem ser danificadas, seja pelo mau uso, acidentes ou mesmo por falta de manutenção preventiva.

Dessa forma, os donos de máquinas pesadas precisam constantemente dos serviços de oficinas especializadas, sejam estes realizados nas concessionárias autorizadas pelos fabricantes, em oficinas multimarcas ou mesmo por mecânicos terceirizados ou contratados exclusivamente pelos donos das máquinas.

Em qualquer uma dessas possibilidades há a demanda por peças de reposição, que são desgastadas pelo uso contínuo da máquina em suas diversas aplicações.

### 1.1. Justificativa

Esta monografia é baseada em dados reais de uma rede de lojas do setor. Para manter os dados em sigilo, não foram citados o nome da empresa e a marca de seus produtos.

Em 2020, a demanda por peças de reposição teve um aumento considerável, em todas as lojas da rede, sendo necessário manter altos estoques para atender os clientes com agilidade, consequentemente aumentando os custos da empresa. Por outro lado, ter um estoque baixo prejudica o atendimento ao cliente, uma vez que as peças são solicitadas para pronta entrega ou para entrega em curto período de tempo.

Assim, a metodologia de previsão de vendas para peças de reposição se faz relevante, para que seja possível equilibrar o estoque em níveis satisfatórios, manter os custos em níveis adequados e atender o cliente em tempo hábil.

## **1.2. Objetivo geral**

O objetivo do trabalho é desenvolver modelos de séries temporais para previsão da quantidade de vendas de peças de reposição, em função de seu próprio histórico de vendas e de potenciais variáveis correlacionadas.

## **1.3. Objetivos específicos**

- Acompanhar a variabilidade da série que representa a quantidade de vendas de peças;
- Avaliar as interrelações entre a série de interesse e as séries correlacionadas;
- Produzir estimativas para o volume mensal de peças para os próximos 12 meses;
- Apresentar os modelos mais adequados para estimar o volume de vendas de peças;
- Avaliar a capacidade preditiva e as limitações dos modelos ajustados.

## 2. METODOLOGIA

A demanda por máquinas pesadas, bem como por suas peças de reposição, pode ser influenciada por diversas variáveis externas, como programas de governo, incentivos fiscais, taxa de juros, taxa de câmbio, entre outras. Porém, neste estudo, serão consideradas variáveis internas à empresa para prever o volume de vendas de peças para os próximos 12 meses.

Dessa forma, foram consideradas 3 variáveis para a elaboração do estudo, com dados de janeiro de 2016 a dezembro de 2021, retirados do sistema de gestão da empresa, considerando todas as lojas da rede, em diversas regiões do Brasil, conforme descritas a seguir:

- $yp$ : Quantidade de Vendas de Peças (notas fiscais);
- $xo$ : Quantidade de Serviços de Oficina Realizados (ordens de serviço);
- $xm$ : Quantidade de Máquinas Vendidas (unidades vendidas).

Como o principal objetivo do trabalho é a previsão do volume de vendas de peças, foram separados os períodos amostral e de validação das 3 variáveis.

A seguir, encontram-se as séries do período amostral, com dados de janeiro de 2016 a dezembro de 2020.

- $ypA$ : Quantidade de Vendas de Peças (notas fiscais);
- $xoA$ : Quantidade de Serviços de Oficina Realizados (ordens de serviço);
- $xmA$ : Quantidade de Máquinas Vendidas (unidades vendidas).

A seguir, encontram-se as séries do período de validação, que segue de janeiro de 2021 a dezembro de 2021.

- $ypV$ : Quantidade de Vendas de Peças (notas fiscais);
- $xoV$ : Quantidade de Serviços de Oficina Realizados (ordens de serviço);
- $xmV$ : Quantidade de Máquinas Vendidas (unidades vendidas).

A metodologia utilizada nesta monografia pode ser dividida em três etapas.

Primeiramente, foi realizada uma análise descritiva das séries no período amostral, quando cada série foi avaliada individualmente, de forma a entender o comportamento de cada

variável. Também foi analisada a correlação entre as variáveis para identificar as possíveis regressoras a serem incluídas nos modelos. Ainda na análise descritiva, foi identificada a necessidade de definir 3 variáveis indicadoras para diferenciar períodos de variabilidades distintas ao longo da Série  $ypA$ .

Na segunda etapa, foram ajustados dois modelos univariados de séries temporais, seguindo a metodologia de *Box & Jenkins (1976)*, e quatro modelos multivariados, utilizando a técnica de regressão dinâmica.

Na terceira etapa, conforme o objetivo do estudo, foi analisada a variabilidade da Série Quantidade de Vendas de Peças ( $ypA$ ), bem como realizada a sua previsão para 12 meses subsequentes, além de identificados os melhores modelos, tanto para acompanhamento da variabilidade dos dados quanto para a previsão da série.

## 2.1. Conceitos de série temporal

As séries temporais são originadas de dados observados ao longo do tempo, em intervalos regulares, como dias, meses e anos, de forma que as observações em períodos de tempo próximos são dependentes entre si. Por exemplo, em uma dada empresa observa-se a quantidade de novos clientes por dia, em um período de 30 dias.

Uma série univariada é representada como segue:

$$\{Z_t; t = 1, 2, 3, \dots\}, \{Z_t\} \text{ ou } Z_{(t)} \quad (1)$$

em que  $Z$  é a variável em estudo e  $(t)$  o instante de tempo.

Já uma série multivariada pode ser representada como:

$$\{Z_1(t), \dots, X_k(t); t = 1, 2, 3, \dots\} \quad (2)$$

em que  $Z_1$  é a variável em estudo,  $X_k$  a covariável e  $(t)$  o instante de tempo.

Com a análise de séries temporais, espera-se extrair informações para investigar comportamentos passados e prever comportamentos futuros. Para isso, é necessário observar características típicas desse tipo de dado, de forma que modelar tais características é o objetivo principal da análise.

### 2.1.1. Componentes de uma série temporal

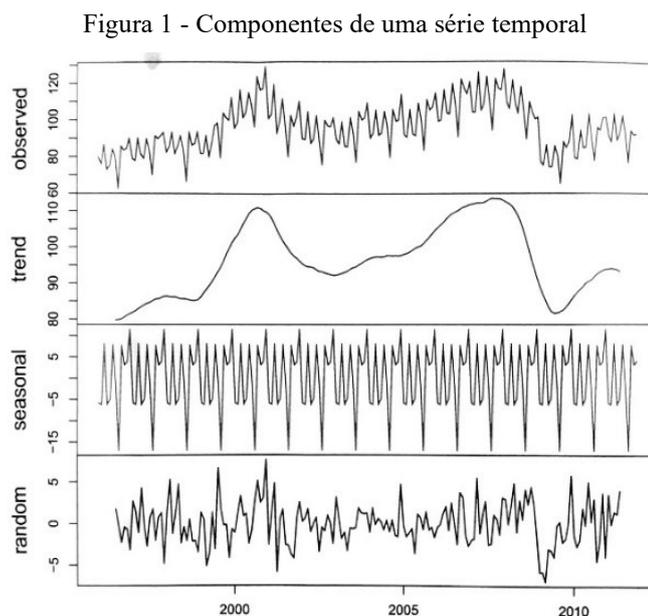
Entre essas principais características estão as componentes de uma série temporal. Pode-se dizer que uma série temporal é o conjunto de três componentes, não observáveis, sendo representada como:

$$Z_t = T_t + S_t + a_t \quad (3)$$

onde:

- A tendência,  $T_t$ , é o comportamento de crescimento ou decrescimento dos valores da variável ao longo do tempo;
- A sazonalidade,  $S_t$ , refere-se a comportamentos que se repetem, em torno da tendência, em intervalos de tempo iguais, como dias, meses e anos, por exemplo;
- A componente de erro,  $a_t$ , é a variabilidade não explicada pelas variáveis do modelo e que tem um comportamento aleatório.

Dessa forma, uma série temporal pode ser decomposta nessas três componentes, conforme Figura 1. A primeira faixa representa a série temporal, enquanto a segunda, terceira e quarta faixas representam, respectivamente, as componentes tendência, sazonalidade e erro.



Fonte: Ferreira *et al.* (2020)

Também é observada a componente ciclo,  $C_t$ , que pode ser entendida como as variações que ocorrem periodicamente, mas não necessariamente em intervalos de tempo iguais. Porém, é comum contemplar essa componente integrada à tendência, caracterizando a componente ciclo-tendência.

O modelo  $Z_t = T_t + S_t + a_t$ , conforme apresentado anteriormente, é chamado de modelo aditivo, mas também é possível encontrar séries temporais que são representadas por modelos multiplicativos, apresentando a estrutura como segue:

$$Z_t = T_t \times S_t \times a_t \quad (4)$$

O modelo multiplicativo é utilizado para representar séries com variância crescente ao longo do tempo. Já o modelo aditivo é utilizado em séries com variância que permanece constante com o tempo.

Conforme será visto adiante, para modelar uma série temporal é preciso que sua variância seja constante ao longo do tempo, caso contrário será necessário transformar os dados para tornar o efeito aditivo.

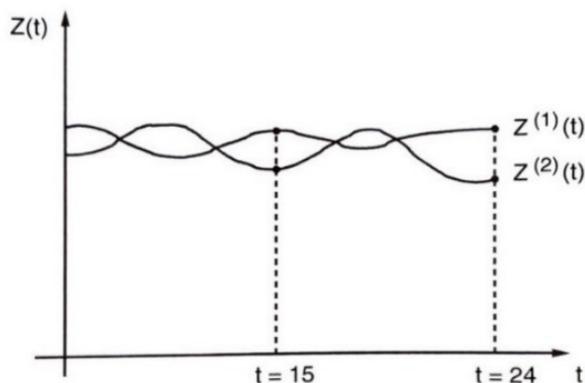
Além das componentes de uma série temporal, também é preciso considerar os pressupostos de autocorrelação e estacionariedade, essenciais para a modelagem dos dados em séries temporais, os quais serão vistos adiante, após a conceituação do processo que dá origem à série temporal.

### 2.1.2. Processo estocástico

“Segundo Morettin e Tolo (2018, p. 25), os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos controlados por leis probabilísticas”.

Pode-se entender um processo estocástico como o processo gerador da série temporal, ou ainda que uma série temporal é uma realização de um processo estocástico, conforme exemplificado pela Figura 2, em que está ilustrada a medição da temperatura do ar, de dado local, durante 24 horas.

Figura 2 - Trajetórias de um processo estocástico



Fonte: Morettin e Toloi (2018)

Segundo Morettin e Toloi (2018, p. 2):

Vamos designar por  $Z_{(t)}$  a temperatura no instante  $t$  (dado em horas, por exemplo). Notamos que para dois dias diferentes obtemos duas curvas que não são, em geral, as mesmas. Essas curvas são chamadas trajetórias do processo físico que está sendo observado e este (o processo estocástico) nada mais é do que o conjunto de todas as trajetórias que poderíamos observar. Cada trajetória é também chamada de uma série temporal ou função amostral. Designando-se por  $Z^{(1)}_{(15)}$  o valor da temperatura no instante  $t=15$ , para a primeira trajetória (primeiro dia de observação), teremos um número real; para o segundo dia, teremos outro número real,  $Z^{(2)}_{(15)}$ .

Para um processo estocástico,  $\{Y_t, t = 1, 2, \dots\}$ , a média e a variância podem ser representadas como segue:

$$\text{Média} \Rightarrow E(Y_t) = \mu \quad (5)$$

$$\text{Variância} \Rightarrow \sigma_t^2 = E(Y_t - \mu)^2 \quad (6)$$

Já a função de autocovariância de um processo estocástico pode ser representada como:

$$\text{Covariância} \Rightarrow \gamma(s, t) = \gamma_k = E[(Y_s - \mu_s)(Y_t - \mu_t)] \quad (7)$$

onde:

$s, t \in Z$  são dois períodos de tempo distintos;

$E(Y_t) = \mu_t$  é a esperança matemática da variável aleatória  $Y_t$ , para todo  $t \in Z$ .

Pode-se dizer que a função de autocovariância mede o grau de dependência linear entre duas observações da série, separadas por um intervalo, que é chamado de lag ou defasagem.

### 2.1.3. Autocorrelação

“Segundo Gujarati e Porter (2011, p. 415), as observações de tais dados (séries temporais) seguem um ordenamento natural, de modo que observações sucessivas costumam apresentar intercorrelações”.

Os dados de séries temporais apresentam uma dependência serial entre si, o que é chamado de autocorrelação, que representa o fato de uma observação estar relacionada com uma outra observação em instante de tempo próximo. Como exemplo, a quantidade de peças de reposição vendida em Julho pode estar relacionada à quantidade vendida em Junho, em Maio, ou mesmo estar relacionada a quantidade vendida em Julho do ano anterior, em virtude da sazonalidade das vendas.

A diferença de tempo entre as observações é chamada de defasagem. No exemplo, se o mês de Julho é considerado como instante ( $t$ ), o mês de Junho é o instante ( $t-1$ ), defasado em um período de tempo, e Maio é o instante ( $t-2$ ), defasado em dois períodos de tempo.

A correlação entre as defasagens é definida pela ordem da autocorrelação. A observação que está correlacionada com a observação anterior, com uma defasagem, é chamada de autocorrelação de primeira ordem, ou seja, a autocorrelação entre Julho e Junho, por exemplo. Já a observação que está correlacionada com observações de dois períodos de tempo anteriores, com duas defasagens, é chamada de autocorrelação de segunda ordem, ou seja, a autocorrelação de Julho com Junho e Maio anterior, por exemplo, e assim por diante.

Um dos objetivos da análise de séries temporais é modelar sua dependência serial e, para essa modelagem, a Função de Autocorrelação (FAC) e a Função de Autocorrelação Parcial (FACP) são muito úteis.

A Função de Autocorrelação (FAC) mede a correlação das observações no tempo ( $t$ ) com as observações nas defasagens anteriores ( $t-k$ ), e pode ser representada como segue:

$$\rho_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)Var(Y_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad (8)$$

onde:

$-1 \leq \rho(k) \leq 1$  é o coeficiente de autocorrelação;

$k = 1, \dots, N$  é a ordem da defasagem ou lag de tempo.

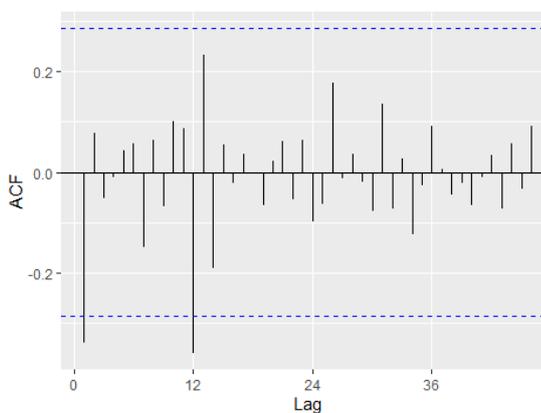
A Função de Autocorrelação permite identificar a ordem de um processo estocástico, por meio da análise de seu Gráfico, o Correlograma, conforme exemplificado na Figura 3.

“Conforme Bueno (2011, p. 44), a função de autocorrelação é o gráfico da autocorrelação contra a defasagem”.

O Correlograma apresenta os coeficientes de autocorrelação  $\rho(k)$  em função das defasagens ( $k$ ), possibilitando entender se a série possui tendência ou sazonalidade e qual a ordem da autocorrelação do processo estocástico que originou a série.

No Correlograma, o eixo horizontal indica a defasagem, o eixo vertical apresenta a autocorrelação e a linha tracejada em azul indica o intervalo de confiança para as respectivas autocorrelações, que uma vez fora do intervalo podem ser classificadas como significativamente diferentes de zero. Aqui, tem-se um teste de hipótese no qual a hipótese nula é a de que a correlação na defasagem ( $k$ ) é igual a zero contra a hipótese alternativa de que é diferente de zero. Assim, uma correlação amostral acima ou abaixo dos limites do intervalo de confiança indicam a rejeição da hipótese nula, de acordo com o nível de significância especificado para realização do teste.

Figura 3 - Correlograma da FAC



Fonte: elaborado pelo autor (2022)

Na Figura 3, o Correlograma da FAC, é possível verificar os lags 1 e 12 como significativamente diferentes de zero, os quais devem ser considerados para a identificação de modelos a serem estimados.

Já a Função de Autocorrelação Parcial (FACP) é a correlação entre instantes diferentes de tempo, retirando a autocorrelação entre os tempos intermediários. Como exemplo, seria a autocorrelação entre Abril e Julho, retirando a autocorrelação com os meses intermediários, Maio e Junho.

Segundo Toscano (2021), tem-se a autocorrelação parcial entre  $Y_t$  e  $Y_{t+k}$ , definida da regressão linear,  $Y_t = \phi_{k1}Y_{t+1} + \phi_{k2}Y_{t+2} + \dots + \phi_{kk-1}Y_{t+k-1} + \phi_{kk}Y_{t+k} + a_t$ , onde a Função de Autocorrelação Parcial (FACP) é:  $\phi_{kk} = \text{Corr}(Y_t, Y_{t+k} / Y_{t+1}, \dots, Y_{t+k-1})$ .

Assim como a Função de Autocorrelação (FAC), a Função de Autocorrelação Parcial (FACP) também pode ser visualizada pelo seu Gráfico, o Correlograma.

Dessa forma, pela análise da FAC e da FACP, e seus respectivos lags significativos, além de outras análises, como a forma do decaimento das autocorrelações, é possível identificar possíveis modelos a serem estimados, conforme será visto mais adiante.

#### 2.1.4. Estacionariedade

Segundo Gujarati e Porter (2011, p. 734):

Um processo estocástico será chamado de estacionário se sua média e variância forem constantes ao longo do tempo e o valor da covariância entre os dois períodos de tempo depender apenas da distância, do intervalo ou da defasagem entre os dois períodos e não o tempo real ao qual a covariância é computada”.

A estacionariedade pode ser entendida como uma espécie de equilíbrio do processo, em que a média e variância não variam com o tempo, ou seja, o processo estocástico é estacionário quando se encontra em um estado de equilíbrio estatístico.

Graficamente, percebe-se uma série temporal estacionária como uma linha que evolui, sem grandes variações, em torno de uma média fixa, com variância constante ao longo do tempo.

Na literatura de séries temporais é visto que o processo estocástico é estritamente estacionário se a distribuição de probabilidade dos dados não muda com o tempo. Porém, como é possível observar apenas uma realização do processo estocástico, ou seja, a série temporal em estudo, é utilizado o conceito de processo fracamente estacionário, ou estacionário de segunda ordem, que diz respeito a um processo com média e variância constante.

Pode-se dizer que um processo estocástico é fracamente estacionário, ou estacionário de segunda ordem, se:

$$E(Y_t) = \mu \text{ é constante para todo } (t)$$

$$\text{Var}(Y_t) = \sigma^2 \text{ é finita e constante para todo } (t)$$

$$\text{Cov}(Y_t, Y_s) = \gamma_k = |t - s| \text{ é função que depende apenas da diferença do tempo}$$

Um importante processo estocástico é o chamado processo puramente aleatório ou de ruído branco, que tem média igual a zero, variância constante e é serialmente não correlacionado, conforme formalmente representado como segue:

$$E(a_t) = 0 \quad (9)$$

$$E[(a_t)^2] = \sigma^2 \quad (10)$$

$$E[(a_t - E(a_t))(a_{t-s} - E(a_t))] = E[(a_t a_{t-s})] = 0 \quad (11)$$

Ou seja, se as variáveis  $(a_t)$  são não correlacionadas, podem ser consideradas um processo de ruído branco, representado por  $RB(0, \sigma^2)$ .

Um processo puramente aleatório especial é o de ruído branco gaussiano, o qual além das características próprias de um ruído branco ( $RB(0, \sigma^2)$ ), como descrito anteriormente, ainda possui uma distribuição normal.

O objetivo da análise de uma série temporal é encontrar o melhor modelo que descreva sua dependência serial e gere um vetor de erros de ruído branco, de preferência um ruído branco gaussiano, mas não necessariamente. Dessa forma, é possível afirmar que foi retirada toda a autocorrelação significativa dos resíduos e assim o modelo está adequado para uso, já que a variabilidade dos dados foi absorvida pelas variáveis do modelo.

O conceito de estacionariedade é essencial para a modelagem de séries temporais, pois é a partir do equilíbrio estatístico que se pode confiar nos testes estatísticos e, com isso, desenvolver inferências, previsões e conclusões a respeito dos dados.

“Conforme Bueno (2011, p. 16), é fundamentalmente a constatação de estacionariedade que permitirá proceder inferências estatísticas sobre os parâmetros estimados com base na realização de um processo estocástico”.

Porém, como a maioria das séries temporais observadas não são estacionárias, com a presença de tendência ou sazonalidade nos dados, é necessário usar técnicas adequadas para tornar a série estacionária antes da análise e modelagem dos dados.

Para definir a técnica ideal a ser utilizada para tornar a série estacionária, é importante detectar qual a causa da estacionariedade e se essa é determinística ou estocástica.

Se a causa da estacionariedade for determinística, basta estimar a tendência ou sazonalidade e removê-las, processo o qual não será detalhado nesta monografia.

Considerando que a estacionariedade seja estocástica, as técnicas são divididas em:

- Aplicação de diferenças na série para eliminar a tendência e sazonalidade;
- Transformações para estabilizar a variância e tornar o modelo aditivo.

“Segundo Morettin e Tolo (2018, p. 5), a transformação mais comum consiste em tomar diferenças sucessivas da série original. A primeira diferença de  $Z_{(t)}$  é definida por  $\Delta Z_{(t)} = Z_{(t)} - Z_{(t-1)}$ . A segunda diferença é  $\Delta^2 Z_{(t)} = Z_{(t)} - 2Z_{(t-1)} + Z_{(t-2)}$ ”.

Assim, a diferenciação é, como o próprio nome diz, a diferença de valores no tempo, sendo que o valor da série diferenciada em um tempo ( $t$ ) é o valor da série original no tempo ( $t$ ) menos o valor no tempo ( $t-k$ ), sendo  $k$  qualquer defasagem de tempo.

Caso a variância aumente ao longo do tempo, será necessário fazer uma transformação logarítmica, que seria transformar os valores em logaritmos dos valores para transformar o efeito multiplicativo em efeito aditivo.

“Para Morettin e Tolo (2018, p. 9), a transformação logarítmica é apropriada se o desvio padrão da série (ou outra medida de dispersão) for proporcional à média”.

### 2.1.5. Operadores

Toscano (2021) apresenta o operador de retardo ( $B$ ), usado por *Box & Jenkins (1976)* na descrição de modelos, como  $B^k Y_t = Y_{t-k}$ , onde  $k$  é a ordem da autocorrelação do processo estocástico.

Para a notação da diferenciação, Toscano (2021) apresenta o operador diferença. Seja a série  $Z_t$ ,  $t = 1, 2, 3, \dots, N$ , tal que após a aplicação de ( $d$ ) diferenças torna-se estacionária, a diferenciação não sazonal é definida como  $Z_t = (I - B)^d Y_t$  e a sazonal como  $Z_t = (I - B^s)^D Y_t$ .

Dessa forma, as diferenças não sazonal e sazonal têm a representação descrita como:

$$Z_t = (I - B)^d (I - B^s)^D Y_t \quad (12)$$

onde:

$s$  é a ordem da sazonalidade;

$d$  é a ordem da diferença não sazonal;

$D$  é a ordem da diferença sazonal.

## 2.2. Modelos de série temporal

“Conforme Morettin e Tolo (2018, p. 40), podemos classificar os modelos para séries temporais em duas classes, segundo o número de parâmetros envolvidos: (a) modelos paramétricos; (b) modelos não paramétricos”.

Nesta monografia, serão estudados os modelos paramétricos, com número finito de parâmetros e cuja análise é feita no domínio do tempo, como os modelos autorregressivos e de médias móveis (ARMA) e os modelos autorregressivos integrados de médias móveis (ARIMA).

Nos modelos não paramétricos, que não são alvo desta monografia, além de um número infinito de parâmetros, a série é descrita no domínio de frequências, como a análise espectral, por exemplo.

Inicialmente são apresentados os modelos univariados, constituídos de apenas uma série temporal, os quais foram estimados pelos modelos paramétricos ARIMA, em que a variável  $Y_t$  é explicada pelos valores passados, ou defasados, da própria variável e dos termos de erro, por meio dos parâmetros autorregressivos e de médias móveis.

Posteriormente, serão apresentados também os modelos multivariados, que englobam mais de uma série temporal, os quais foram estimados pelos modelos ARIMAX, também chamados de modelos de *Box-Tiao (1975)*, que são uma extensão dos modelos paramétricos ARIMA. Nos modelos ARIMAX, além dos parâmetros autorregressivos e de médias móveis, estima-se também uma componente linear referente às covariáveis, também chamadas de variáveis exógenas ou variáveis regressoras.

### 2.2.1. Modelos Autorregressivos (AR)

Nos modelos autorregressivos, os valores presentes de uma série temporal são função de seus valores passados, de forma que o modelo faz uma regressão em valores passados para prever os valores futuros.

Um modelo autorregressivo (AR) de ordem (p) é representado como segue:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t \quad (13)$$

Usando o operador de retardo, temos que:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = a_t \quad (14)$$

$$\Phi(B) Y_t = a_t$$

onde:

$\phi_j$  é o j-ésimo parâmetro autorregressivo;

$\Phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) = 0$  é o polinômio autorregressivo de ordem (p);

$a_t$  é um ruído branco com média  $E(a_t) = 0$  e variância constante  $Var(a_t) = \sigma^2$ ;

o processo é estacionário se  $-1 < \phi < 1$ .

### 2.2.2. Modelos de Médias Móveis (MA)

Os modelos de médias móveis (MA) também são uma função do passado da série, porém estes se baseiam nos erros dos valores passados, e não nos valores observados da série temporal. Assim, o modelo tem semelhança com o modelo autorregressivo, com a diferença que os termos incluídos no modelo são referentes aos erros presentes e passados.

Um modelo de médias móveis (MA) de ordem (q) é representado como:

$$Y_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (15)$$

Usando o operador de retardo, segue que:

$$Y_t - \mu = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t = \Theta(B) a_t \quad (16)$$

onde:

$\mu$  é a média do processo;

$\theta_j$  é o j-ésimo parâmetro da componente de médias móveis;

$\Theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) = 0$  é o polinômio de médias móveis de ordem (q);

$a_t$  é um ruído branco com média  $E(a_t) = 0$  e variância constante  $Var(a_t) = \sigma^2$ ;

o processo é invertível se  $-1 < \theta < 1$ .

### 2.2.3. Modelos ARMA

Segundo Morettin e Tolo (2018, p. 131):

Os modelos autorregressivos são bastante populares em algumas áreas, como em economia, onde é natural pensar o valor de alguma variável no instante  $t$  como função de valores defasados da mesma variável. [...] Por outro lado, representar um processo por um modelo de médias móveis puro parece não ser natural ou intuitivo.

Surgem, então, os modelos ARMA (p,q), que são a junção dos modelos autorregressivos (AR) com os modelos de médias móveis (MA).

Um modelo autorregressivo e de médias móveis de ordem (p,q) é representado conforme adiante:

$$Y_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (17)$$

Usando o operador de retardo, o modelo pode ser representado como:

$$\Phi(B)Y_t = \Theta(B)a_t \quad (18)$$

onde:

$\Phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p) = 0$  é o polinômio autorregressivo de ordem (p);

$\Theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q) = 0$  é o polinômio de médias móveis de ordem (q);

$a_t$  é um processo ruído branco com  $E(a_t) = 0$  e  $Var(a_t) = \sigma^2$ ;

o processo é estacionário se  $-1 < \phi < 1$ ; e invertível se  $-1 < \theta < 1$ .

### 2.2.4. Modelos ARIMA

Um modelo ARIMA (p,d,q) é nada mais que um modelo ARMA que recebeu diferenciações para se tornar estacionário. É chamado de modelo autorregressivo integrado de média móvel ou, simplesmente, ARIMA.

Utilizando o operador de retardo, o modelo ARIMA pode ser representado da seguinte forma:

$$\Phi(B)Z_t = \Theta(B)a_t \quad (19)$$

onde:

$$Z_t = (1 - B)^d Y_t;$$

$$\Phi(B) = (1 - \varphi_1 B - \dots - \varphi_p B^p) = 0;$$

$$\Theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q) = 0;$$

$a_t$  é um processo ruído branco com  $E(a_t) = 0$  e variância constante  $Var(a_t) = \sigma^2$ .

### 2.2.5. Modelos SARIMA

Os modelos SARIMA são aplicados em séries temporais que apresentam sazonalidade em intervalos de tempo semelhantes.

Segundo Toscano (2021), as séries não estacionárias, após a aplicação de “D” diferenças sazonais e/ou “d” diferenças não sazonais, tornam-se estacionárias em média e apresentam componentes autorregressivos de ordem (p) e de médias móveis de ordem (q).

Um modelo SARIMA (p,d,q)(P,D,Q)<sub>s</sub> pode ser representado como:

$$\Phi_p(B^S) \Phi_p(B) W_t = \Theta_q(B) \Theta_Q(B^S) a_t \quad (20)$$

onde:

$$W_t = (1 - B)^d (1 - B^S)^D Y_t;$$

$\Phi_p(B) = 0$  é o polinômio autorregressivo de ordem (p);

$\Theta_q(B) = 0$  é o polinômio de médias móveis de ordem (q);

$\Phi_p(B^S) = 0$  é o polinômio autorregressivo sazonal de ordem (P);

$\Theta_Q(B^S) = 0$  é o polinômio de médias móveis sazonal de ordem (Q);

$a_t$  é um processo ruído branco com  $E(a_t) = 0$  e  $Var(a_t) = \sigma^2$ .

$s$  é o período da sazonalidade (semanal, mensal, anual, etc).

### 2.2.6. Metodologia Box & Jenkins (1976)

Conforme Gujarati e Porter (2011, p. 771):

Observando uma série temporal, [...] como podemos saber se ela segue um processo AR puro (e, se isso acontece, qual o valor de p) ou um processo MA puro (e, se isso acontece, qual o valor de q), um processo ARMA (e, se isso acontece, quais os valores de p e q) ou um processo ARIMA, no caso de precisarmos conhecer os valores de p, d e q? A metodologia de BJ (*Box & Jenkins*) é muito útil para se responder a questão anterior.

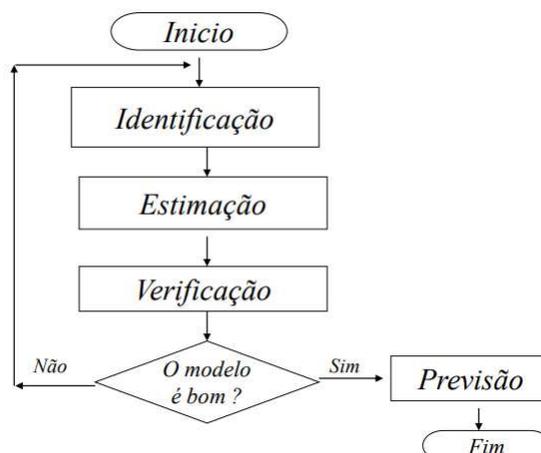
Existem diversas abordagens para modelagem de séries temporais, como Análise de Fourier, Modelos Estruturais e também metodologias mais recentes, como Redes Neurais Artificiais (RNA). Porém, nesta monografia será desenvolvida a metodologia clássica de *Box & Jenkins (1976)*, além de uma variação desta, que incorpora o conceito de função de transferência, conforme será visto adiante.

Segundo Morettin e Toloi (2018, p. 111):

Uma metodologia bastante utilizada na análise de modelos paramétricos é conhecida como abordagem de *Box e Jenkins (1976)*. Tal metodologia consiste em ajustar modelos autorregressivos integrados de médias móveis, ARIMA (p,d,q) a um conjunto de dados. A estratégia para a construção do modelo será baseada em um ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados.

Para modelagem de séries temporais por meio dos modelos ARIMA, usa-se o ciclo iterativo sugerido por *Box & Jenkins (1976)*, conforme Figura 4.

Figura 4 - Fluxograma Metodologia Box & Jenkins (1976)



Fonte: Toscano (2021)

Na etapa de identificação busca-se descobrir a estrutura dos possíveis modelos que melhor se ajustam aos dados, que consiste no uso da Função de Autocorrelação (FAC) e da Função de Autocorrelação Parcial (FACP), para verificar se a série temporal é estacionária e determinar a ordem dos modelos ARIMA.

Conforme visto anteriormente, a FAC e FACP são representadas graficamente pelo Correlograma. De forma geral, caso o Correlograma apresente as autocorrelações decrescentes de forma linear, a série pode ser considerada não estacionária. Nesse caso, será necessário realizar diferenciações na série até que o comportamento do Correlograma apresente um decaimento exponencial das autocorrelações, identificando assim uma série estacionária.

Já para a identificação da ordem do modelo, é necessário avaliar o parâmetro (p), pelos lags significativos na FACP, que indicam a ordem das componentes autorregressivas, e o parâmetro (q), pelos lags significativos na FAC, que indicam a ordem das componentes de média móvel.

Após a identificação de possíveis modelos para o ajuste da série temporal, é necessário estimar seus parâmetros, por meio de técnicas estatísticas apropriadas, como o método dos momentos, mínimos quadrados ordinários e máxima verossimilhança.

Segundo Hyndman e Athanasopoulos (2021), o modelo ARIMA é estimado pela técnica de máxima verossimilhança, encontrando os valores dos parâmetros que maximizam a probabilidade de obter os dados observados.

Na função de verossimilhança, a ideia básica é estimar os parâmetros desconhecidos de forma a encontrar os valores mais prováveis de serem observados.

Segundo Morettin e Toloi (2018, p.185):

Consideremos um modelo ARIMA (p,d,q) e coloquemos seus p+q+1 parâmetros no vetor  $\mathcal{E} = (\phi, \theta, \sigma^2)$ , onde  $\phi = (\phi_1, \dots, \phi_p)$ ,  $\theta = (\theta_1, \dots, \theta_q)$ . [...] Para estimar  $\mathcal{E}$ , um dos métodos empregados será o de máxima verossimilhança: dadas as  $N$  observações  $Z_1, \dots, Z_N$ , consideramos a função de verossimilhança  $L(\mathcal{E} | Z_1, \dots, Z_N)$  encarada como função de  $\mathcal{E}$ . Os estimadores de máxima verossimilhança (EMV) de  $\mathcal{E}$  serão os valores que maximizam  $L$  ou  $l = \log L$ .

Dessa forma, pode-se entender a função de verossimilhança como uma função de probabilidade condicional, que depende dos parâmetros desconhecidos e dos dados observados, sendo os estimadores de máxima verossimilhança definidos como os valores dos parâmetros que são os mais prováveis de serem observados.

“Conforme Bueno (2012, p. 57), em geral, o pesquisador assume distribuição normal ou *t-Student* e procura estimar o vetor de parâmetros  $\Psi = (c, \phi_1, \phi_2, \dots, \phi_p; \theta_0, \theta_1, \theta_2, \dots, \theta_q)$  que maximiza a probabilidade de a amostra pertencer a distribuição especificada”.

Posteriormente à estimação dos parâmetros, deve-se verificar a aleatoriedade dos resíduos, sendo o teste de *Ljung e Box (1978)* bastante utilizado para esse objetivo. O teste de *Ljung e Box (1978)* é um teste de hipótese que tem como hipótese nula que os resíduos são independentes e identicamente distribuídos, não havendo autocorrelação significativa, sendo os resíduos de ruído branco ( $RB(0, \sigma^2)$ ), contra a hipótese alternativa que os resíduos não são independentes e identicamente distribuídos, havendo autocorrelação significativa ao nível de significância do teste.

Após o ajuste do modelo aos dados e verificação dos resíduos, conforme etapas anteriores, procede-se à etapa de previsão por meio do melhor modelo encontrado, conforme análise de medidas de acurácia, que medem a capacidade preditiva do modelo.

As etapas apresentadas anteriormente, e suas principais análises e testes, serão desenvolvidas de forma prática na sessão de análise de dados.

### 2.2.7. Modelos ARIMAX

Os modelos citados até aqui englobam apenas uma variável, de forma que os valores presentes são função dos valores passados da própria variável. Assim, pode-se dizer que, na análise de séries temporais univariadas, o objetivo é construir modelos a partir do passado da série.

Se o modelo englobar mais de uma série temporal, incluindo a variável dependente defasada no tempo e variáveis exógenas, como regressoras, será denominado modelo autorregressivo dinâmico, modelo dinâmico ou ainda regressão dinâmica.

Na regressão dinâmica temos uma variável dependente  $Y_{(t)}$ , que pode ser explicada por variáveis exógenas e também pela própria variável dependente defasada no tempo, sendo representada como uma função:

$$Y_{(t)} = f [X_{(t)}, Y_{(t-1)}, Y_{(t-2)}, \dots, Y_{(t-p)}]; \quad (21)$$

onde:

$Y_t$  representa a variável dependente;

$X_t$  é uma variável ou vetor de variáveis independentes.

Assim como em séries temporais univariadas, a teoria de regressão linear não pode ser diretamente aplicada, em virtude da presença de correlação serial nos dados. Dessa forma, é necessário realizar ajustes ou combinações de modelos para ser possível absorver a estrutura dos dados e dos resíduos dos modelos, sendo possível utilizar uma adaptação dos modelos ARIMA para essa modelagem.

Uma das formas de modelar séries temporais multivariadas é usar os modelos ARIMAX, que incorporam variáveis exógenas ao modelo por meio de uma função de transferência. Os modelos ARIMAX usam da técnica de regressão linear para modelar as séries

temporais, mas englobam os modelos ARIMA para modelar os erros do modelo, que por serem derivados de séries temporais são autocorrelacionados. Por tais características, os modelos ARIMAX são também chamados de modelos de regressão com erros ARIMA, embora essa definição não seja um consenso entre os estudiosos.

Conforme Bolanle e Oluwadare (2017):

Os modelos de séries temporais ARIMAX são uma combinação entre regressão e ARIMA. Para o modelo de série temporal ARIMAX, a variável dependente é explicada não apenas pelos valores passados da variável de previsão, mas também por choques aleatórios e exógenos. O modelo ARIMAX é uma combinação de modelo autorregressivo, modelo de média móvel e modelo de regressão, sendo um caso especial de modelo de função de transferência, onde covariáveis externas, ou variáveis exógenas, podem ser adicionadas ao modelo.

Os modelos ARIMAX também são chamados de modelos *Box-Tiao (1975)*, sendo considerados uma combinação dos modelos ARIMA com variáveis exógenas  $X(r)$ , simbolizados por ARIMAX(p,d,q,r), em que o termo  $(r)$  representa o grau da variável exógena, incorporada no modelo ARIMA por meio de uma função de transferência, que pode receber valores presentes ou passados de uma ou mais variáveis independentes.

Segundo Durka e Silvia (2012):

Supondo duas séries temporais,  $Y_t$  e  $X_t$ , sendo ambas estacionárias e seguindo um modelo ARIMA, o modelo ARIMAX pode ser representado por:  $Y_t = C + v(B)X_t + N_t$ , onde  $Y_t$  é a série da variável dependente;  $X_t$  é a série da variável exógena;  $C$  é o termo constante;  $N_t$  é o vetor de resíduos;  $v(B)X_t$  é a função de transferência que permite que  $X$  influencie  $Y$ . Pode-se escrever a função de transferência como:  $v(B)X_t = (v_0 + v_1B + v_2B^2 + \dots)X_t$ . A série de resíduos,  $N_t$ , pode ser escrita na forma de um modelo ARIMA como  $N_t = \phi_1N_{t-1} + \phi_pN_{t-p} - \theta_1a_{t-1} - \theta_qa_{t-q} + a_t$ .

Dessa forma, o modelo ARIMAX, com defasagem denotada por  $(k)$ , é representado como:

$$Y_t = C + v_0X_t + v_1X_{t-1} + v_2X_{t-2} + \dots + v_KX_{t-K} + \phi_1N_{t-1} + \phi_pN_{t-p} - \theta_1a_{t-1} - \theta_qa_{t-q} + a_t \quad (22)$$

Ou ainda, conforme Durka e Silva (2012), utilizando o operador de retardo para a série  $N_t$ , o modelo ARIMAX pode ser descrito como:

$$Y_t = C + v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots + v_K X_{t-K} + \frac{\theta(B) \Theta(B^S)}{\phi(B) \Phi(B^S)(1-B)^d(1-B^S)^D} a_t \quad (23)$$

### 2.2.8. Metodologia Box & Jenkins (1976) para modelos dinâmicos

Conforme Morettin e Tolo (2018):

Os modelos *Box & Jenkins (1976)* são casos particulares de um modelo de filtro linear, cuja entrada é um ruído branco, representado por  $Z_t = u + a_t + v_1 a_{t-1} + v_2 a_{t-2} + \dots = u + v(B)a_t$ , em que  $v(B) = 1 + v_1 B + v_2 B^2 + \dots$  é denominada função de transferência do filtro e  $u$  é um parâmetro determinando o nível da série.

“Segundo Hyndman (2010), o modelo ARIMAX pode ser considerado um caso de modelo de função de transferência, popularizado por *Box e Jenkins (1976)*”.

Os modelos de função de transferência, como os modelos ARIMAX, descrevem o comportamento de uma série considerando o comportamento passado da própria série, mas incorporando também o comportamento de demais séries temporais correlacionadas com a série de interesse. Por tais características, como visto anteriormente, os modelos de função de transferência podem ser chamados de modelos dinâmicos ou modelos de regressão dinâmica.

Conforme Ferreira *et al.* (2020):

Os modelos de Box & Jenkins (1970) podem incorporar variáveis auxiliares ( $X_{it}$ ) que influenciam a variável resposta ( $Y_t$ ) por meio de uma função de transferência  $f(X_t)$ :  $Y_t = f(X_t) + et$ , onde  $et$  pode ser um ruído branco ou um modelo ARIMA completo. A função pode agrupar valores passados ou presentes de uma ou mais séries temporais, sendo a função de transferência  $f(X_t)$ , para séries quantitativas, representada pela equação  $f(X_t) = \frac{(w_0 + w_1 L + w_2 L^2 + \dots + w_s L^s)}{(1 - \delta_1 L - \delta_2 L^2 - \dots - \delta_r L^r)} X_{t-b}$ , onde é preciso descobrir os valores de  $r$ ,  $s$  e  $b$  e estimar  $w_i$ ,  $i = 0, \dots, s$  e  $\delta_j$ ,  $j = 1, \dots, r$ , para ser possível identificar a estrutura da função de transferência.

Dessa forma, entende-se que a metodologia de *Box & Jenkins (1976)* pode ser utilizada para estimar modelos dinâmicos, como os modelos ARIMAX, por meio de uma função de transferência, cujos os passos da metodologia serão brevemente apresentados a seguir.

Ferreira *et al.* (2020), diz que a metodologia *Box & Jenkins* com função de transferência segue quatro etapas: 1) Calcular a função de correlação cruzada entre  $Y_t$  e  $X_t$ ; 2) Identificar  $r$ ,  $s$  e  $b$ ; 3) Estimar o modelo com função de transferência; 4) Verificar os resíduos.

Para identificar um modelo de função de transferência, é preciso descobrir os valores de  $r$ ,  $s$  e  $b$ , que representam a defasagem da variável exógena que entrará no modelo. Isso é feito

pelo cálculo da função de correlação cruzada (*fcc*) entre a série de interesse e as séries exógenas, com o objetivo de medir a dependência entre as séries do modelo.

A função de correlação cruzada é representada como:

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \quad (25)$$

tal que:

$k \in Z$ ;

$\sigma_x$  e  $\sigma_y$  são os desvios padrões das séries  $X_t$  e  $Y_t$ ;

$\gamma_{xy}(k)$  representa a covariância entre as duas variáveis no lag  $k$ .

Caso as séries não sejam estacionárias, *Box & Jenkins (1976)* sugerem o método de pré-branqueamento para eliminar a estrutura de tendência presente nas séries temporais. Tal método não será detalhado nesta monografia, mas pode ser consultado em *Ferreira et al (2020)*.

Conforme *Bolanle e Oluwadare (2017)*:

A função de correlação cruzada (*ccf*) é uma ferramenta exploratória, que pode ser empregada no processo de identificação do modelo ARIMAX, e geralmente para séries temporais com variável exógena, sendo a função de correlação cruzada uma generalização da função de autocorrelação para o caso multivariado. [...] Para verificar se será necessário determinar o número de defasagens explicativas a serem incluídas no modelo ARIMA, que resultará no modelo ARIMAX, a correlação cruzada deve ser verificada. Após a identificação da forma da função de transferência, se faz necessário estimar os parâmetros do modelo, o que é realizado por meio do método de máxima verossimilhança, que busca maximizar a probabilidade de encontrar valores dos parâmetros mais próximos dos valores observados.

Após a identificação da ordem do modelo por meio da função de correlação cruzada e, posteriormente, a estimação do modelo por meio do método de máxima verossimilhança, chega-se a etapa de verificação dos resíduos, em que é necessário calcular as autocorrelações dos resíduos e a correlação cruzada entre os resíduos e a variável auxiliar, também por meio do Correlograma e do teste de *Ljung & Box (1978)*.

Assim, entende-se que as etapas do fluxograma da Figura 4 também podem ser aplicadas para o ajuste de modelos de função de transferência, como os modelos ARIMAX. Porém, além da identificação e estimação dos coeficientes autorregressivos e de médias móveis, deve-se identificar e estimar também os coeficientes das covariáveis.

Segundo *Durka e Silvia (2012)*, a construção do modelo de função de transferência é um processo iterativo semelhante à construção do modelo ARIMA de *Box-Jenkins* univariado.

### 2.3. Critérios de informação

É provável que exista mais de um modelo que se ajusta adequadamente aos dados da série temporal, sendo necessário uma forma de mensurar qual dos modelos tem o melhor ajuste.

Os critérios de informação podem ser utilizados para escolher o melhor entre os modelos propostos. Esses critérios comparam modelos estimados com base na maximização do logaritmo da função de verossimilhança, levando em consideração a complexidade do modelo no critério de seleção, penalizando o modelo com mais parâmetros.

O melhor modelo será aquele que apresentar menor valor de tais critérios de informação.

#### 2.3.1 AIC

Akaike (1973,1974) introduziu um critério de informação, o AIC, ou Critério de Informação de Akaike, que avalia a qualidade do ajuste do modelo, sendo definido por:

$$AIC(q) = n \text{Ln} \hat{\sigma}_a^2 + 2M \quad (26)$$

onde:

$n$  é o número de observações da série;

$\hat{\sigma}_a^2$  é a variância dos resíduos;

$M$  é o número de parâmetros estimados;

$L$  é a função de máxima verossimilhança.

#### 2.3.2 BIC

Desenvolvido por Schwarz (1978), e semelhante ao AIC, o BIC adiciona uma penalização maior pela inclusão de coeficientes adicionais a serem estimados, definido por:

$$BIC(p,q) = \text{Ln}(\hat{\sigma}^2) + (p+q)[\text{Ln}(n)/n] \quad (27)$$

onde:

$n$  é o número de observações da série;

$\hat{\sigma}_a^2$  é a variância dos resíduos;

$L$  é a função de máxima verossimilhança.

O BIC tem o efeito de aumentar o peso do termo de penalidade de acordo com o número de parâmetros no modelo, e conseqüentemente minimizar o seu valor.

## 2.4. Medidas de acurácia

Um dos principais objetivos do estudo de séries temporais é a previsão dos valores futuros, sendo esta a etapa final da análise. Nessa fase, é necessário estipular critérios para definir o modelo com melhor capacidade preditiva, o que significa o quão perto estão as previsões dos dados observados.

Conforme Morettin e Toloi (2018, p. 6):

Há algo chamado método de previsão de mínimos quadrados, e este, de fato, fornece a base para virtualmente todos os estudos teóricos. Além disso, todos os métodos de previsão são simplesmente diferentes procedimentos computacionais para calcular a mesma quantidade, a saber, a previsão de mínimos quadrados de um valor futuro a partir de combinações lineares de valores passados.

As medidas de acurácia são usadas para avaliar o desempenho de um modelo de previsão e baseiam-se no cálculo de um erro de  $h$  passos à frente no tempo ( $t$ ), dado pela diferença entre o valor no tempo  $y_t(h)$  e a previsão no mesmo instante  $Y_{t+h}$ .

Dessa forma, o erro de previsão pode ser representado como segue:

$$e_t(h) = Y_t(h) - Y_{t+h}, h \geq 1 \quad (28)$$

onde:

$Y_t(h)$  são os valores observados;

$Y_{t+h}$  são os valores previstos;

$t$  é a origem da previsão;

$h$  é o horizonte da previsão.

Na literatura, há diversas dessas medidas, como apresentadas a seguir. O objetivo é que o modelo escolhido minimize seus valores, sendo o melhor modelo aquele com menor valor de tais medidas de acurácia.

### 2.4.1. ME (Erro Médio)

É a média da série de erros de previsão e deve ser muito próxima de zero, sendo representada por:

$$EM = \frac{\sum_{t=1}^n e_t}{n} \quad (29)$$

### 2.4.2. MSE (Erro Quadrado Médio ou Desvio Quadrado Médio)

É a média dos quadrados dos erros, sendo representada por:

$$MSE = \frac{1}{t-h} \sum_{j=1}^{t-h} e_j^2 \quad (30)$$

### 2.4.3. RMSE (Raiz do Erro Quadrado Médio)

É a raiz quadrada da média dos erros quadráticos e é representada por:

$$RMSE = \sqrt{\frac{1}{t-h} \sum_{j=1}^{t-h} e_j^2} \quad (31)$$

### 2.4.4. MAE (Erro Médio Absoluto) ou MAD (Desvio Padrão Absoluto Médio)

É o desvio padrão do ajuste em relação à média nas mesmas unidades dos dados, sendo representado por:

$$MAD = \frac{1}{t-h} \sum_{j=1}^{t-h} |e_j| \quad (32)$$

#### 2.4.5. MPE ou EMP (Erro Médio Percentual)

É o erro percentual médio, obtido relativizando os erros (ME) pelos valores observados, sendo representado por:

$$EMP = \frac{\sum_{t=1}^n EP_t}{n} \quad (33)$$

#### 2.4.6. MAPE (Erro Percentual Absoluto Médio)

É a média do erro percentual em valor absoluto. Expressa a porcentagem média dos erros (em valor absoluto) cometidos na previsão e é representada por:

$$MAPE = \left[ \frac{1}{h} \sum_{j=1}^{T-h} \left| \frac{e_j}{y_j} \right| \right] 100\% \quad (34)$$

#### 2.4.7. Theil's U (Coeficiente de Desigualdade de Theil)

É a comparação entre o modelo ajustado e um passeio aleatório. Quanto menor o índice, melhor ajustado é o modelo. O U de Theil é representado por:

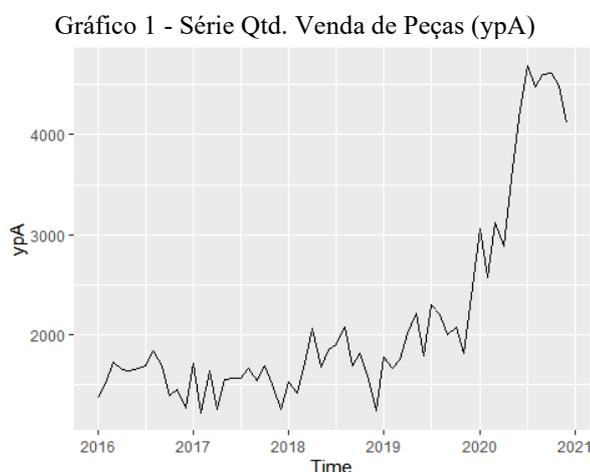
$$TIC = \frac{\sqrt{\frac{1}{T-h} \sum_{j=1}^{T-h} e_j^2}}{\sqrt{\frac{1}{T-h} \sum_{j=1}^{T-h} \hat{y}_j^2 + \frac{1}{T-h} \sum_{j=1}^{T-h} y_j^2}} \quad (35)$$

### 3. ANÁLISE DOS DADOS

Conforme já visto, foram consideradas 3 variáveis para a elaboração do estudo:  $yp$  (Quantidade de Vendas de Peças),  $xo$  (Quantidade de Serviços de Oficina Realizados) e  $xm$  (Quantidade de Máquinas Vendidas). Porém, como o objetivo do trabalho é a previsão do volume de vendas de peças, a análise será realizada sobre os dados do período amostral, cujas as séries foram denominadas como  $ypA$ ,  $xoA$  e  $xmA$ , respectivamente.

#### 3.1. Série Quantidade de Vendas de Peças

Inicialmente, analisando o Gráfico 1, da Série Quantidade de Vendas de Peças, observa-se um comportamento estacionário de 2016 a início de 2018, quando acontece uma tendência crescente, com dois picos, posteriormente apresentando uma tendência decrescente até 2019. A partir de 2019, há uma tendência crescente até meados de 2020, com alguns picos e vales ao longo do tempo, voltando a apresentar uma tendência decrescente até final de 2020.

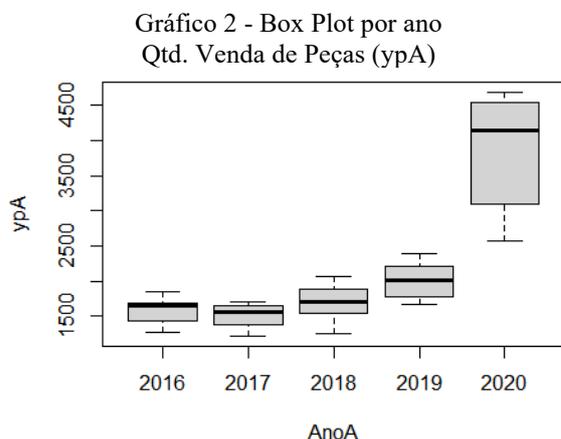


Fonte: elaborado pelo autor (2022)

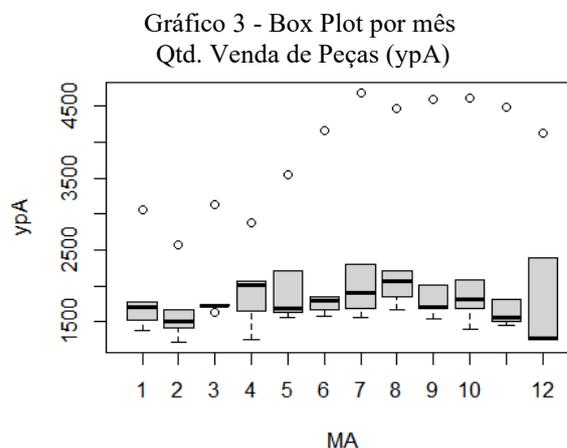
No Gráfico 2, observa-se pequena variabilidade entre os anos de 2016, 2017 e 2018, uma vez que as medianas estão praticamente alinhadas entre si. Há um leve aumento da mediana em 2019 e um grande aumento em 2020, observando uma tendência crescente ao longo da série. Já analisando a variabilidade em cada ano, também pelo Gráfico 2, de 2016 a 2019 há pequena variabilidade nos dados, mas o ano de 2020 destaca-se como o ano com maior variabilidade.

Considerando o Gráfico 3, observa-se que há uma certa sazonalidade nos dados, pois as medianas dos meses estão desalinhadas. Em todos os meses há valores discrepantes e nos meses

de maio, julho e dezembro há maior variabilidade dentro de cada mês, com destaque para dezembro, que, além da alta variabilidade, tem uma distribuição de dados bastante assimétrica, assim como a maioria dos meses, já que se observa as medianas concentradas nos extremos.



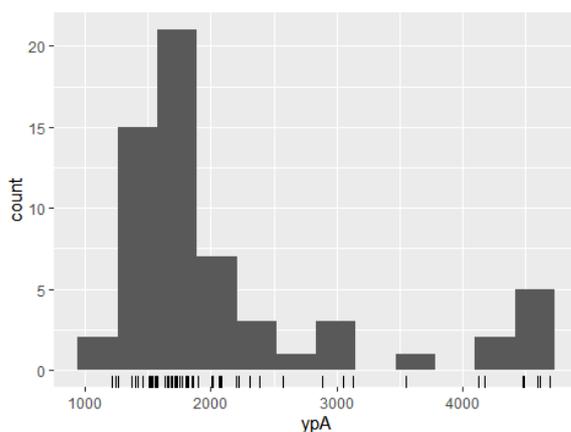
Fonte: elaborado pelo autor (2022)



Fonte: elaborado pelo autor (2022)

O Gráfico 4, Histograma da Série *ypA*, mostra que os dados apresentam grande assimetria positiva com cauda à direita, com média de 2.135 peças e mediana de 1.726 unidades, o que reforça a informação sobre a presença de valores extremos na distribuição. É observado que 25% dos dados concentram-se entre 1.220 e 1.565 peças, 50% estão entre 1.565 e 2.208 e 25% estão entre 2.208 e 4.688.

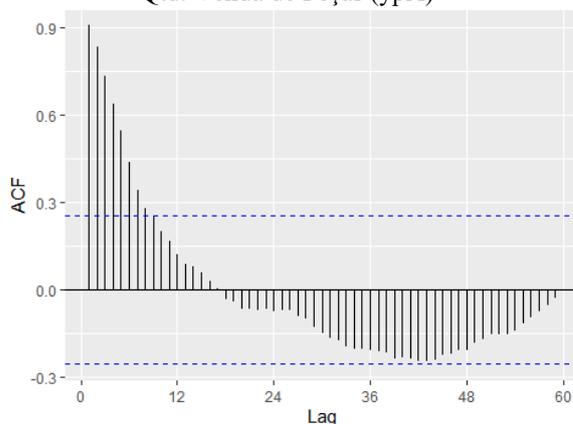
Gráfico 4 - Histograma Qtd. Venda de Peças (ypA)



Min.	1220.000000
1st Qu.	1565.250000
Median	1726.500000
Mean	2134.800000
3rd Qu.	2207.750000
Max.	4688.000000
Var	33214.738983
Sd	966.030403
skewness	1.629141
kurtose	4.403970

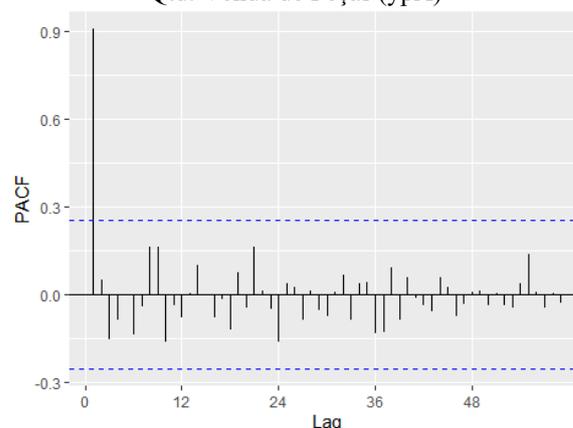
Fonte: elaborado pelo autor (2022)

Gráfico 5 - Correlograma FAC  
Qtd. Venda de Peças (ypA)



Fonte: elaborado pelo autor (2022)

Gráfico 6 - Correlograma FACP  
Qtd. Venda de Peças (ypA)



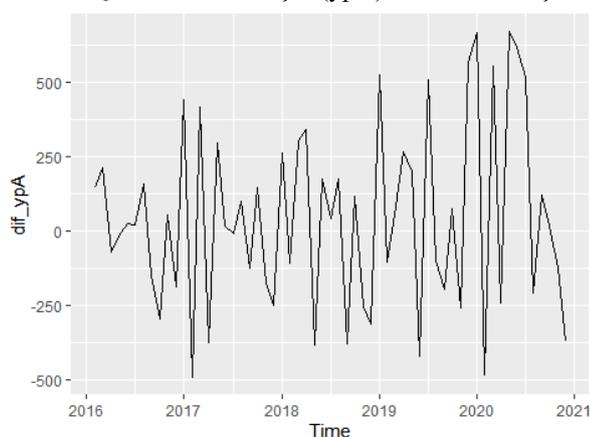
Fonte: elaborado pelo autor (2022)

Pelo Correlograma da FAC, representado no Gráfico 5, há um decaimento senoidal relativamente lento, o que caracteriza uma série não estacionária.

Já pelo Gráfico 6, o Correlograma da FACP, observa-se que há indícios de sazonalidade na Série *ypA*, uma vez que, apesar de não ser um indicio significativo, há um leve aumento nos lags 12, 24 e 36. Informação que é reforçada pelo Gráfico 3, o *Box Plot* por mês, que também demonstra indícios de sazonalidade, já que as medianas não estão alinhadas entre si, apesar do desalinhamento não ser considerável.

Como a Série *ypA* apresenta tendência, conforme Gráfico 1, é necessário aplicar uma diferença não sazonal para torná-la estacionária, cujo resultado pode ser verificado no Gráfico 7, onde é possível perceber a remoção da tendência após a diferenciação.

Gráfico 7 - Série Qtd. Venda de Peças (ypA) com 1 diferença não sazonal

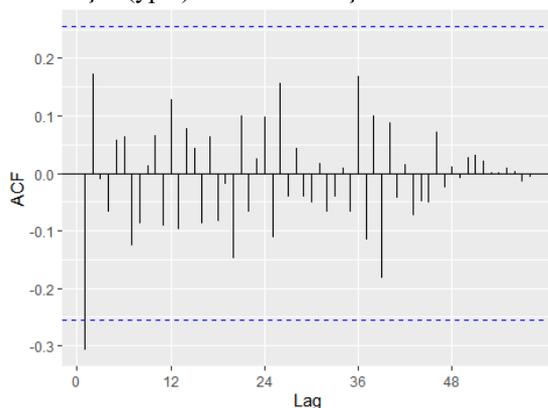


Fonte: elaborado pelo autor (2022)

Após a diferenciação, pelos Gráficos 8 e 9, Correlogramas da FAC e FACP, respectivamente, é possível perceber que, em ambos os Gráficos, é apresentado apenas o primeiro lag significativo.

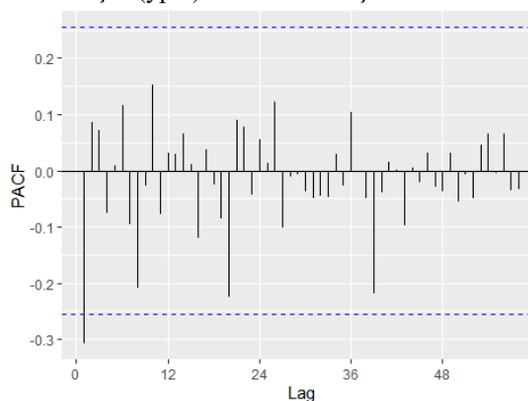
Além disso, como ambos os Gráficos têm características semelhantes, um possível modelo a ser estimado é o ARIMA (1,1,1).

Gráfico 8 - Correlograma FAC - Série Qtd. Venda de Peças (ypA) com 1 diferença não sazonal



Fonte: elaborado pelo autor (2022)

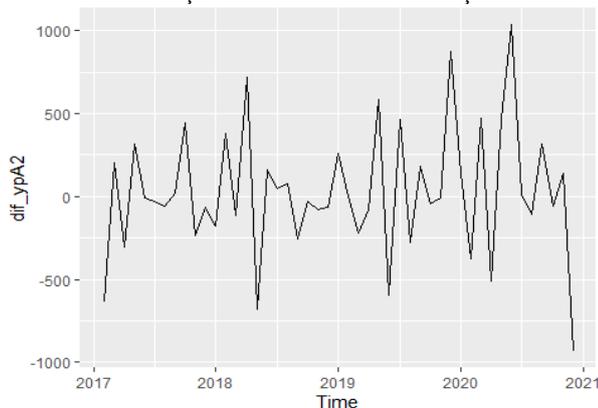
Gráfico 9 - Correlograma FACP - Série Qtd. Venda de Peças (ypA) com 1 diferença não sazonal



Fonte: elaborado pelo autor (2022)

No Gráfico 8, observa-se leves aumentos nos lags 12, 24 e 36, apesar de não serem significativos. Dessa forma, para observar melhor essa sazonalidade, será realizada uma diferenciação sazonal anual na série já diferenciada, cujo resultado pode ser observado no Gráfico 10.

Gráfico 10 - Série Qtd. Venda de Peças (ypA) com 1 diferença não sazonal e 1 diferença sazonal

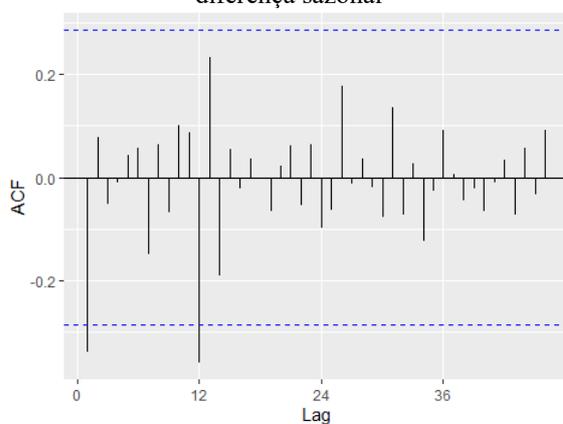


Fonte: elaborado pelo autor (2022)

Após a diferenciação sazonal na série já diferenciada, pelos Gráficos 11 e 12, Correlogramas da FAC e FACP, respectivamente, é possível perceber que, em ambos os Gráficos, é apresentado o primeiro lag não sazonal e o lag 12 sazonal significativos.

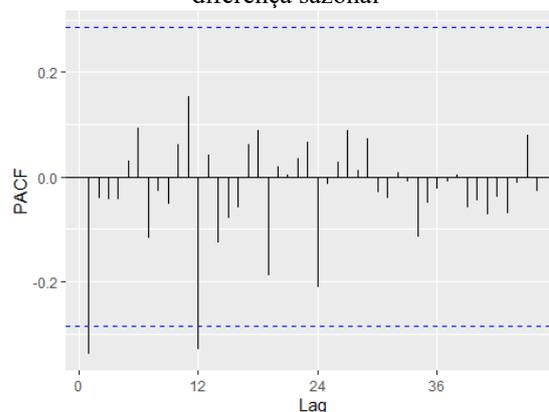
No Gráfico 11, observa-se um leve decaimento nos lags não sazonais e, no Gráfico 12, observa-se um decaimento nos lags sazonais. Assim, um possível modelo a ser ajustado é o SARIMA (1,1,0)(0,1,1)<sub>12</sub>.

Gráfico 11 - Correlograma FAC - Série Qtd. Venda de Peças (ypA) com 1 diferença não sazonal e 1 diferença sazonal



Fonte: elaborado pelo autor (2022)

Gráfico 12 - Correlograma FACP - Série Qtd. Venda de Peças (ypA) com 1 diferença não sazonal e 1 diferença sazonal

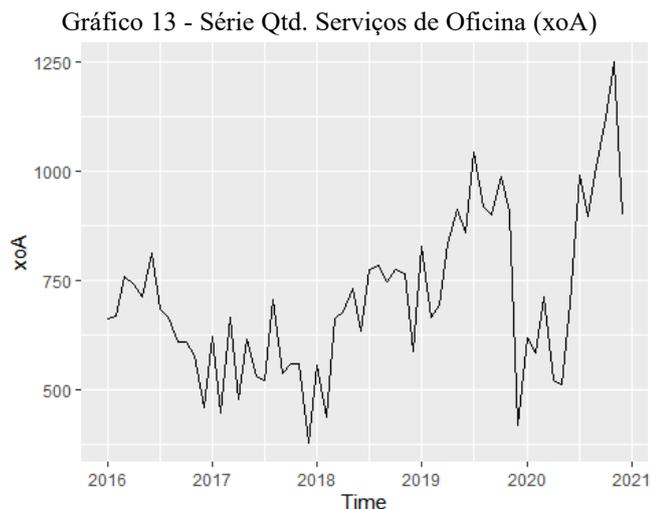


Fonte: elaborado pelo autor (2022)

Considerando o comportamento das vendas na rede de concessionárias, essa sazonalidade realmente faz sentido, pois entre novembro e fevereiro há uma queda nas vendas de peças. Muitos clientes pausam suas atividades, por diversos motivos, como a parada das obras em períodos de chuva e pausas nas atividades no período de entressafra nas usinas de álcool e açúcar, por exemplo, o que causa a redução na demanda por peças de reposição.

### 3.2. Série Quantidade de Serviços de Oficina

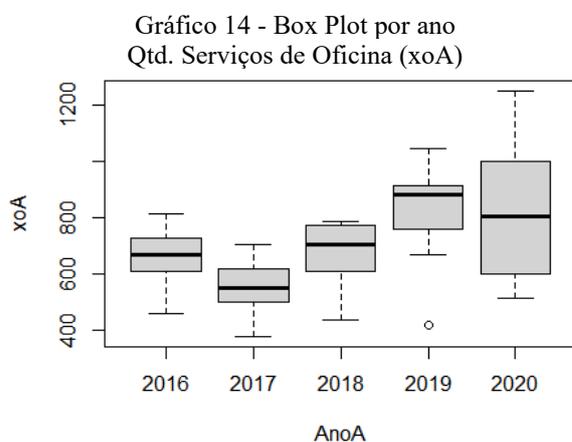
Já analisando a série Quantidade de Serviços de Oficina, no Gráfico 13, observa-se um comportamento não estacionário ao longo de toda a série. É possível visualizar uma tendência decrescente de 2016 a 2017, um período estacionário no ano de 2017 e outro período com tendência crescente de 2018 a final de 2019, quando acontece uma queda significativa com posterior tendência crescente até final de 2020. Observa-se vários picos e vales ao longo de toda a série, o que pode demonstrar certa sazonalidade.



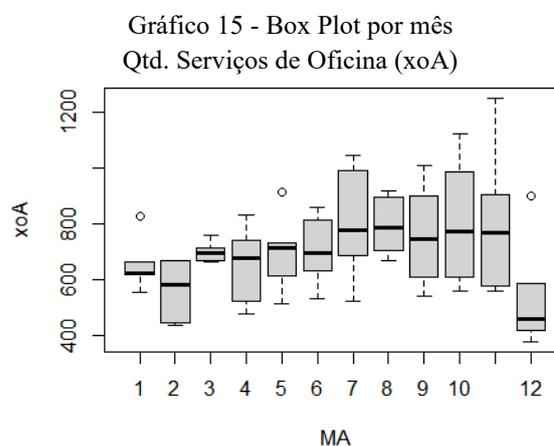
Fonte: elaborado pelo autor (2022)

No Gráfico 14, percebe-se que há grande variabilidade entre os anos, com tendência crescente da mediana, à exceção de 2016 para 2017 e de 2019 para 2020, mostrando a característica não estacionária da série. Considerando a variabilidade em cada ano, ainda pelo Gráfico 14, 2020 apresenta maior variabilidade e 2019 apresenta uma distribuição bastante assimétrica, com valores discrepantes.

Pelo Gráfico 15, pode-se observar que apesar de haver meses com medianas próximas, a série apresenta certa sazonalidade, com destaque para os meses de julho e dezembro, quando há o máximo e mínimo de vendas, respectivamente. Observa-se que na maioria dos meses há grande variabilidade dentro do mês, com destaque para fevereiro, abril, julho, setembro, outubro e novembro, ao contrário dos meses de janeiro e março, que apresentam pequena variabilidade dentro do mês. Observa-se também valores discrepantes nos meses de janeiro, maio e dezembro.

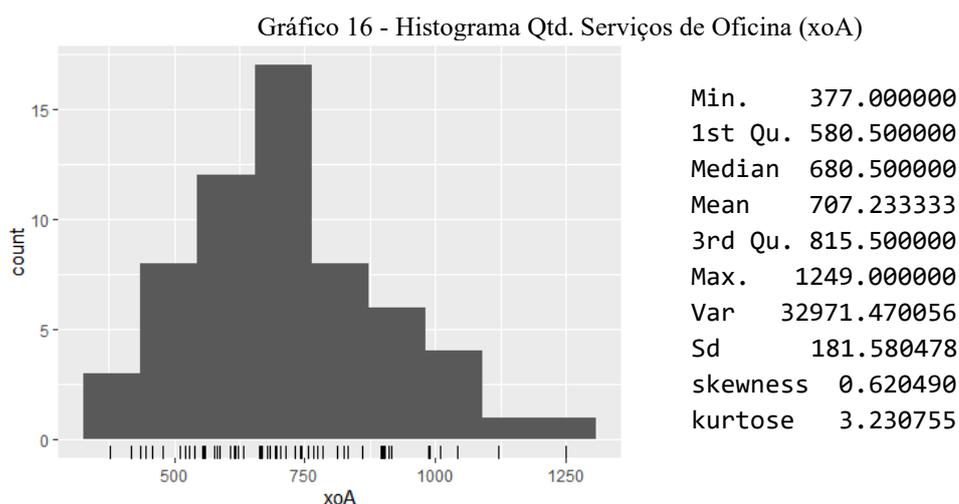


Fonte: elaborado pelo autor (2022)

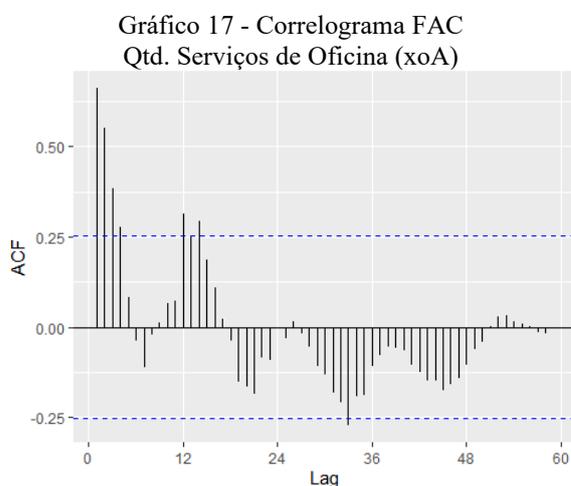


Fonte: elaborado pelo autor (2022)

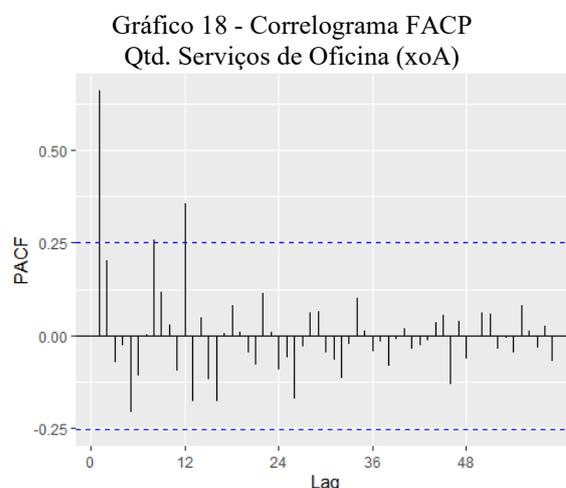
No Gráfico 16, o Histograma e Estatísticas Descritivas mostram que os dados apresentam menor assimetria que a série anterior, mas ainda com uma distribuição assimétrica, com média de 707 e mediana de 680 serviços de oficina. Observa-se que 25% dos dados concentram-se entre 377 e 580 ordens de serviço, 50% estão entre 580 e 815 e 25% estão entre 815 e 1.249.



Fonte: elaborado pelo autor (2022)



Fonte: elaborado pelo autor (2022)

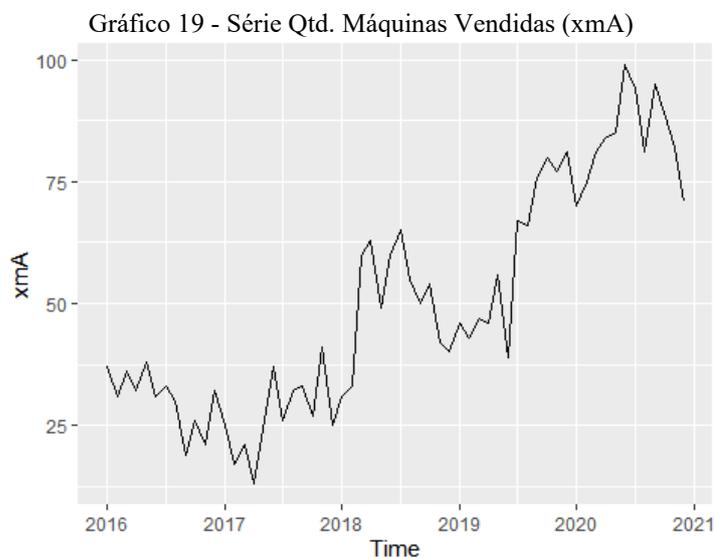


Fonte: elaborado pelo autor (2022)

O Gráfico 17, Correlograma da FAC, mostra que se trata de uma série não estacionária, uma vez que não há um decaimento exponencial, além de haver indícios de sazonalidade, com aumento de autocorrelação no lag 12.

O Gráfico 18, o Correlograma da FACP, também apresenta uma série com sazonalidade, pois há uma correlação significativa no lag 12.

### 3.3. Série Quantidade de Máquinas Vendidas

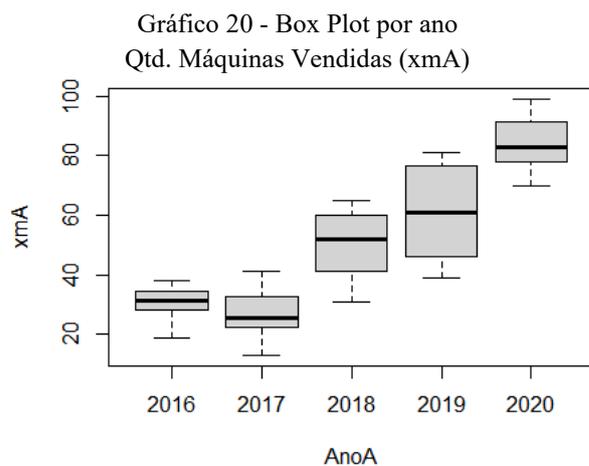


Fonte: elaborado pelo autor (2022)

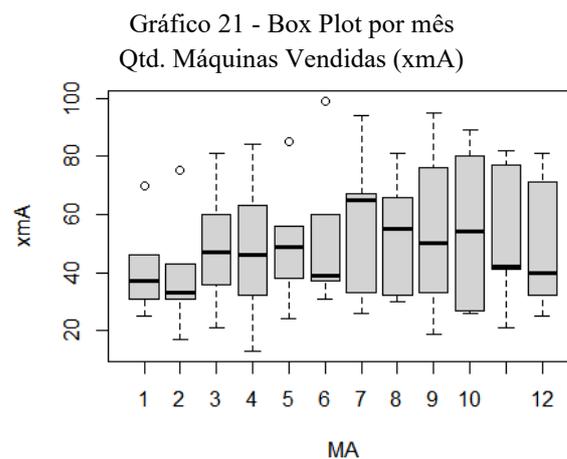
A Série de Máquinas Vendidas, no Gráfico 19, tem um comportamento não estacionário, apresentando uma tendência decrescente de 2016 até meados de 2017, quando começa a apresentar uma tendência crescente até meados de 2018, com quedas intermediárias. Em meados de 2018 ocorre uma queda nas vendas, com uma tendência decrescente até início de 2019, quando ocorre novamente uma tendência crescente até meados de 2020, com quedas intermediárias. Posteriormente a esse período ocorre uma tendência decrescente até o final de 2020.

No Gráfico 20, observa-se que há grande variabilidade entre os anos, sendo possível observar a tendência crescente da mediana, com leve queda apenas de 2016 a 2017, o que demonstra que se trata de uma série não estacionária. O ano de 2019 é o que apresenta maior variabilidade dentro do mês, já o ano de 2016 é o que apresenta menor variabilidade.

Pelo Gráfico 21, observa-se variabilidade entre os meses, indicando a presença de sazonalidade na série, já que as medianas estão desalinhadas. Na maioria dos meses, há grande variabilidade dentro do mês, com destaque para abril, julho, agosto, setembro, outubro, novembro e dezembro. Observa-se grande assimetria na distribuição dos dados dentro dos meses de janeiro, junho, julho, novembro e dezembro, uma vez que a mediana desses meses está concentrada nos extremos.

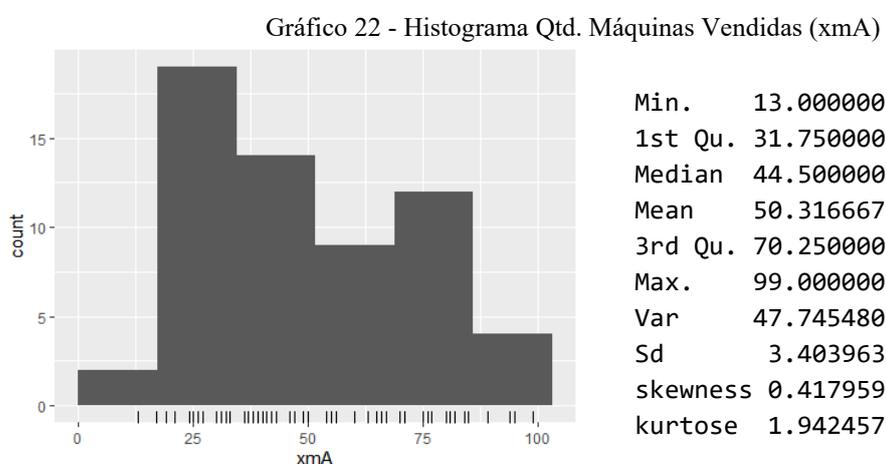


Fonte: elaborado pelo autor (2022)



Fonte: elaborado pelo autor (2022)

No Gráfico 22, o Histograma e Estatísticas Descritivas mostram que os dados apresentam uma distribuição assimétrica, com média de 50 máquinas vendidas e mediana de 44, o que mostra que não há presença de valores extremos na distribuição. Observa-se que 25% dos dados concentram-se entre 13 e 31 máquinas, 50% estão entre 31 e 70 e 25% estão entre 70 e 99 máquinas.

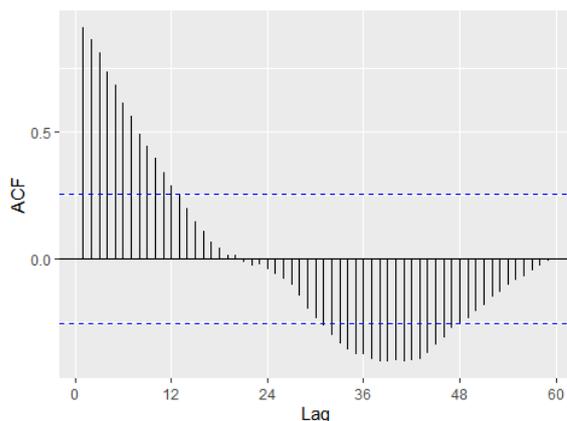


Fonte: elaborado pelo autor (2022)

O Gráfico 23, Correlograma da FAC, mostra que também se trata de uma série não estacionária, uma vez que não há um decaimento exponencial. Não há indícios de sazonalidade, uma vez que não há aumento de autocorrelação nos lags 12, 24 e em diante.

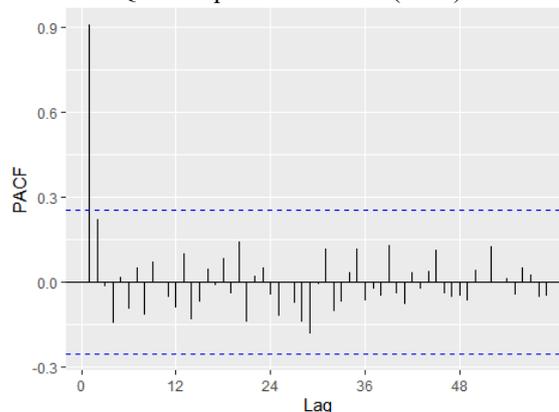
O Gráfico 24, o Correlograma da FACP, também apresenta uma série sem sazonalidade, pois não há correlação significativa nos lags 12, 24 e em diante, com apenas o lag 1 significativo.

Gráfico 23 - Correlograma FAC  
Qtd. Máquinas Vendidas ( $xmA$ )



Fonte: elaborado pelo autor (2022)

Gráfico 24 - Correlograma FACP  
Qtd. Máquinas Vendidas ( $xmA$ )



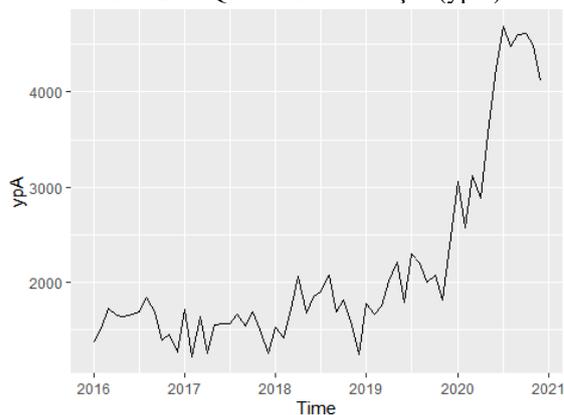
Fonte: elaborado pelo autor (2022)

Para as séries  $xoA$  e  $xmA$  não foi realizada diferenciação para identificação da ordem dos modelos, uma vez que serão estimadas, no software R, pelos modelos de regressão com erros ARIMA, que consideram a mesma ordem, e diferenciação, da série de interesse  $ypA$ .

### 3.4. Variáveis indicadoras

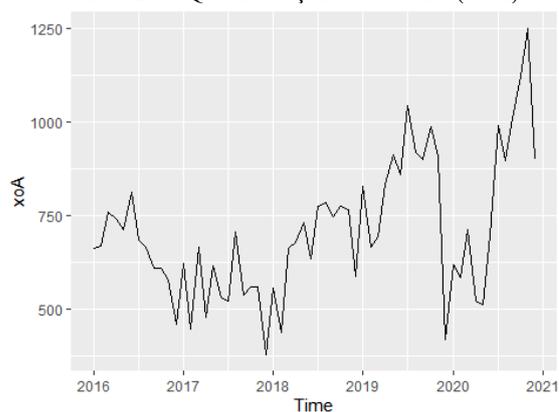
Ao longo da análise descritiva, percebeu-se a necessidade de definir variáveis indicadoras para diferenciar períodos de variabilidades distintas ao longo da Série  $ypA$ .

Gráfico 25 - Qtd. Venda de Peças ( $ypA$ )



Fonte: elaborado pelo autor (2022)

Gráfico 26 - Qtd. Serviços de Oficina ( $xoA$ )



Fonte: elaborado pelo autor (2022)

Conforme pode ser verificado pelo Gráfico 25, da Série Quantidade de Venda de Peças, nas observações até dezembro de 2019 a variabilidade é menor. Já a partir de janeiro de 2020, a tendência é crescente, com maior variabilidade entre as observações.

Assim, a variável  $ind1$ , definida adiante, diferencia dois períodos com variabilidades distintas na Série  $ypA$ , tendo o valor 1 quando  $t < 01/20$  e 0 quando  $t \geq 01/20$ .

$$\bullet \quad ind1 = \begin{cases} 1 & \text{se } t < 01/20 \\ 0 & \text{caso contrário} \end{cases}$$

Essa variável indicadora foi incluída pois em janeiro de 2020 foram incorporados novos produtos no portfólio da linha de peças. Com isso, a venda aumentou consideravelmente, se comparado aos meses anteriores. Dessa forma, foi necessário distinguir os dois períodos: até dezembro de 2019, sem os novos produtos, e a partir de janeiro de 2020, com os novos produtos.

Foram também definidas outras duas variáveis indicadoras. Observando a Série  $ypA$ , no Gráfico 25, percebe-se duas quedas no início de 2020, em fevereiro e abril, sendo observações que não seguem a tendência crescente.

Assim foram usadas duas variáveis indicadoras para esses períodos, sendo a  $ind2$  com valor 1 quando  $t = 02/20$  e 0 se  $t \neq 02/20$ , e a  $ind3$  com valor 1 quando  $t = 04/20$  e 0 quando  $t \neq 04/20$ .

$$\bullet \quad ind2 = \begin{cases} 1 & \text{se } t = 02/20 \\ 0 & \text{caso contrário} \end{cases}$$

$$\bullet \quad ind3 = \begin{cases} 1 & \text{se } t = 04/20 \\ 0 & \text{caso contrário} \end{cases}$$

Essas duas variáveis indicadoras foram incluídas pois, anteriormente à 2020, a venda de peças dependia muito dos serviços de oficina, de forma que uma mudança no volume de serviços realizados impactava diretamente na venda de peças. Em janeiro de 2020, conforme já citado, foram incorporados novos produtos. Porém, um fato importante é que vários deles não são relacionados diretamente a serviços de oficina de máquinas, como pneus e motores para ônibus e caminhão e acessórios para máquinas. Dessa forma, a venda de peças se tornou menos dependente da realização dos serviços de oficina.

Esse fato pode ser observado nos Gráficos 25 e 26, que mostram a Série Quantidade de Venda de Peças com certa correlação com a Série Quantidade de Serviços de Oficina até final de 2019. Posteriormente, a Série de Venda de Peças começa a se deslocar, mas ainda mantendo certa relação, já que na maioria dos serviços de oficina é necessário alguma quantidade de peças.

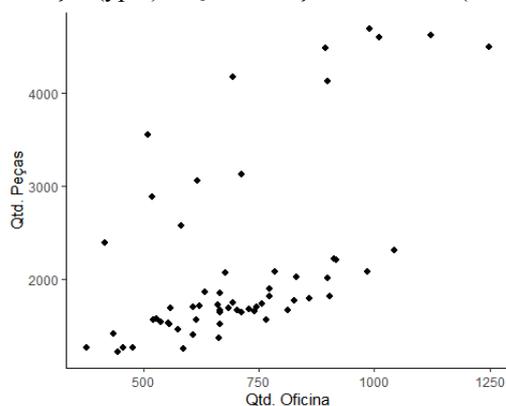
Então, acredita-se que, mesmo com a entrada de novos produtos, as quedas nesses dois períodos têm relação com a queda de oficina, já que houve também quedas na Série de Serviços de Oficina próximas a esses mesmos períodos.

Além disso, em abril, a queda pode ter relação com a pandemia da Covid 19, que iniciou-se em março de 2020 no Brasil. Nesse mercado, acreditava-se que a demanda cairia muito, mas meses depois foi detectado que essa percepção não se tornou realidade no segmento, como pode ser observado pelo crescimento posterior nos Gráficos 25 e 26.

### 3.5. Correlação entre as séries

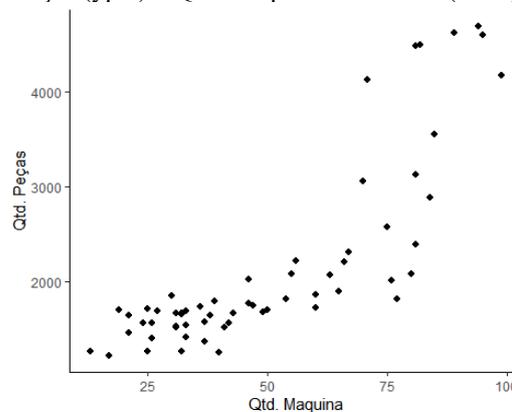
Após as análises individuais das séries, analisou-se também a correlação entre as variáveis. Dessa forma, os Gráficos de Dispersão entre a variável Quantidade de Vendas de Peças ( $ypA$ ) e as variáveis Quantidade de Serviços de Oficina Realizados ( $xoA$ ) e Quantidade de Máquinas Vendidas ( $xmA$ ) podem ser visualizados a seguir.

Gráfico 27 - Correlação entre as séries Qtd. Venda de Peças ( $ypA$ ) x Qtd. Serviços de Oficina ( $xoA$ )



Fonte: elaborado pelo autor (2022)

Gráfico 28 - Correlação entre as séries Qtd. Venda de Peças ( $ypA$ ) x Qtd. Máquinas Vendidas ( $xmA$ )



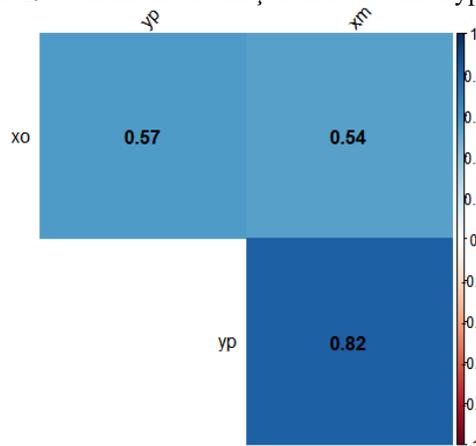
Fonte: elaborado pelo autor (2022)

Pelos Gráficos 27 e 28 parece haver correlação linear entre as variáveis. Porém, no Gráfico 27, que apresenta a correlação entre as variáveis  $ypA$  (Peças) e  $xoA$  (Oficina), há um deslocamento dos dados para valores superiores, parecendo haver dois grupos de dados. Já no Gráfico 28, que indica a correlação entre variáveis  $ypA$  (Peças) e  $xmA$  (Máquinas), a variância aumenta muito a partir de 75 máquinas, mostrando uma correlação menor entre as variáveis.

No Gráfico 29, é possível visualizar os coeficientes de Correlação de Pearson entre as variáveis, com um ponto de atenção para a correlação linear entre as variáveis regressoras, que

pode apresentar o problema da multicolinearidade nos dados. Porém, como serão realizadas diferenciações nas séries, antes da estimação dos modelos, esse risco será minimizado.

Gráfico 29 - Matriz de correlação entre as séries  $yp$ ,  $xm$  e  $xo$



Fonte: elaborado pelo autor (2022)

Quadro 1 - Saída Software R - Correlação de Pearson entre a série de interesse  $ypA$  e as séries  $xmA$  e  $xoA$

Pearson's product-moment correlation data: $ypA$ and $xoA$ $t = 5.2938$ , $df = 58$ , $p\text{-value} = 0.00000192$ alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval: 0.3705452 0.7202971 sample estimates: cor 0.5707662	Pearson's product-moment correlation data: $ypA$ and $xmA$ $t = 10.746$ , $df = 58$ , $p\text{-value} = 0.000000000000002018$ alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval: 0.7087931 0.8862049 sample estimates: cor 0.8158763
---	---

Fonte: elaborado pelo autor (2022)

No Quadro 1, observa-se que a correlação linear entre as variáveis é significativa, uma vez que o  $p\text{-valor}$  é menor que o nível de significância de 0,05, o que leva a rejeitar a hipótese nula de correlação igual a zero.

### 3.6. Estimação dos parâmetros dos modelos

Conforme análise dos Gráficos 5, 6, 8 e 9, como uma tentativa inicial, foi ajustado um modelo ARIMA(1,1,1). Porém, o coeficiente relativo a média móvel não foi significativo, por isso esse modelo foi abandonado e, posteriormente, estimado um modelo sem a componente MA, conforme apresentado a seguir.

### 3.6.1. Modelo 1

M1: ARIMA(1,1,0)

$$(Y_t - Y_{t-1}) - \phi(Y_{t-1} - Y_{t-2}) = a_t \quad (36)$$

$$Y_t = (1 + \phi)Y_{t-1} - \phi Y_{t-2} + a_t$$

$$ypA_t = 0,7211 ypA_{t-1} + 0,2789 ypA_{t-2} + a_t$$

Quadro 2 - Saída Software R - Estimação parâmetros - Modelo 1

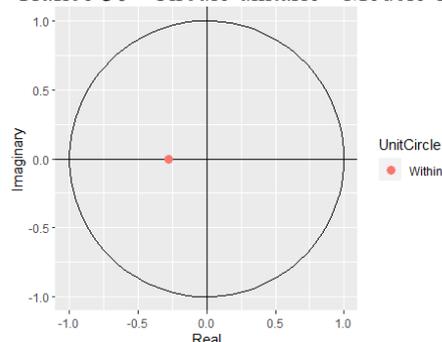
Series: ypA						
ARIMA(1,1,0)						
z test of coefficients:						
	Estimate	Std. Error	z value	Pr(> z )		
ar1	-0.27887	0.12552	-2.2218	0.0263		
AIC	BIC					
843.89	848.04					
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
60.23609	295.7891	224.1306	1.161276	11.02728	0.3485589	-0.01710657
Media	Sd	skewness	kurtose	p_valor		
60.23609	292.0346	0.6068253	3.09881	0.04405277		

Fonte: elaborado pelo autor (2022)

No Quadro 2, observa-se o coeficiente autorregressivo significativo a 5% de significância, já que o *p-valor* é menor que 0,05, o que faz com que a hipótese nula, de que o coeficiente AR é igual a 0, seja rejeitada.

No Quadro 2, também constam as medidas que serão usadas para comparar os modelos deste estudo, após a fase de estimação dos parâmetros.

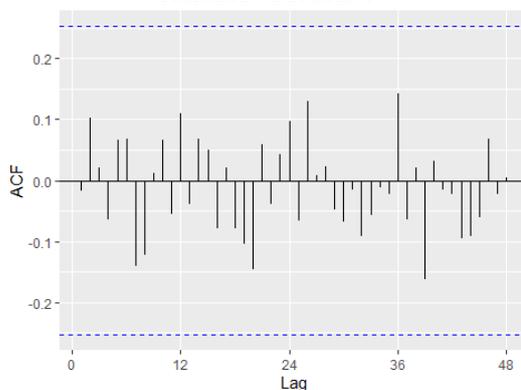
Gráfico 30 - Círculo unitário - Modelo 1



Fonte: elaborado pelo autor (2022)

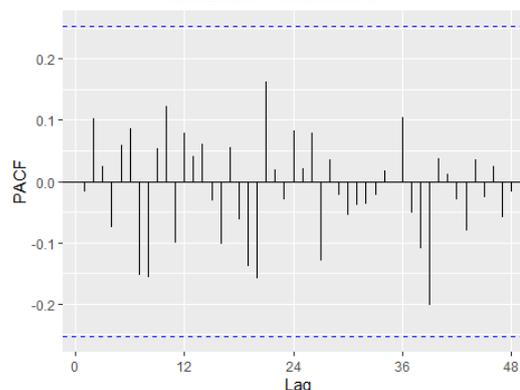
O modelo cumpre a condição de estacionariedade, estando a componente AR entre -1 e 1, conforme o Gráfico 30.

Gráfico 31 - Correlograma FAC  
resíduos - Modelo 1



Fonte: elaborado pelo autor (2022)

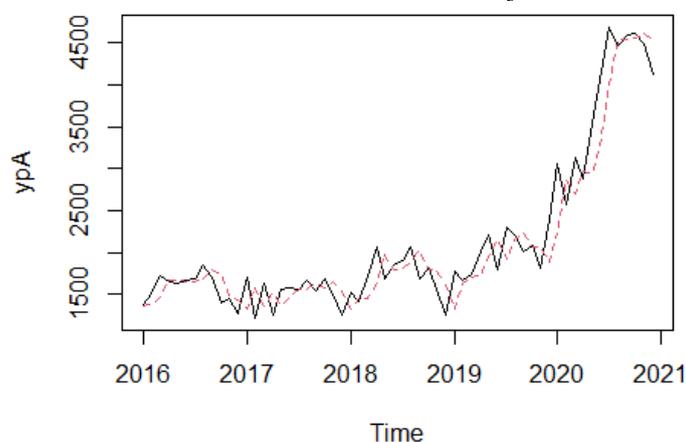
Gráfico 32 - Correlograma FACP  
resíduos - Modelo 1



Fonte: elaborado pelo autor (2022)

Pelos Gráficos 31 e 32, percebe-se que o vetor de resíduos é um ruído branco, pois não existe correlação significativa a 5% de significância.

Gráfico 33 - Valores observados vs ajustados - Modelo 1

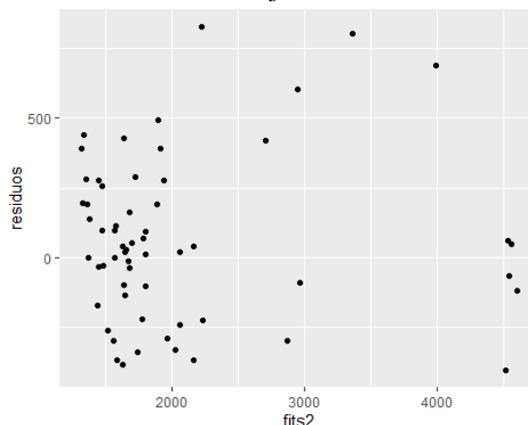


Fonte: elaborado pelo autor (2022)

O modelo acompanha bem a variabilidade dos dados observados, apesar de leve distorção em alguns períodos de tempo, como pode ser visto no Gráfico 33, que apresenta os valores ajustados na linha pontilhada vermelha e os valores observados na linha preta.

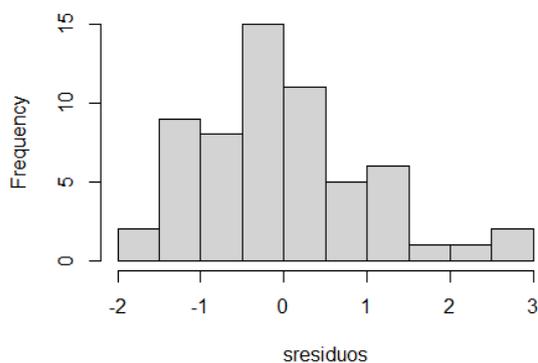
No Gráfico 34, dos resíduos versus valores ajustados, percebe-se que apesar de haver uma concentração dos dados em torno dos menores valores, pode-se dizer que eles estão variando aleatoriamente em torno de 0.

Gráfico 34 - Valores ajustados vs resíduos - Modelo 1

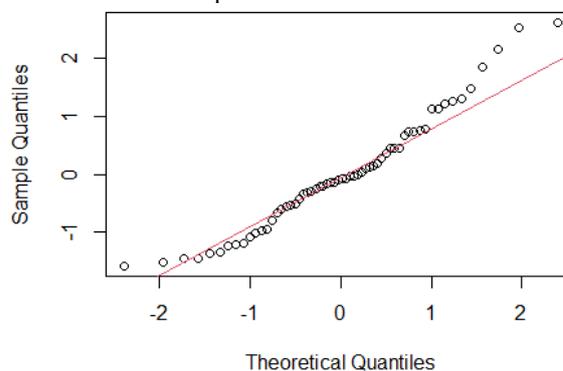


Fonte: elaborado pelo autor (2022)

Pelo Gráfico 35, o Histograma dos resíduos padronizados, percebe-se que não se trata de uma distribuição normal, com valores discrepantes à direita, o que é comprovado pelo Gráfico 36, *QQ Plot*, já que os resíduos não estão seguindo a linha teórica da distribuição normal.

Gráfico 35 - Histograma  
resíduos padronizados - Modelo 1

Fonte: elaborado pelo autor (2022)

Gráfico 36 - *QQ Plot*  
resíduos padronizados - Modelo 1

Fonte: elaborado pelo autor (2022)

No Quadro 3, o teste de *Ljung-Box* confirma o vetor de resíduos como ruído branco, já que pelo *p-valor*, maior que o nível de significância de 0,05, não rejeita-se a hipótese nula de correlações iguais a 0.

Pelo teste de *Shapiro Wilk*, também no Quadro 3, rejeita-se a hipótese nula de normalidade a 5% de significância, já que o *p-valor* é menor que 0,05. Já pelo teste de *Anderson Darling* é possível dizer que trata-se de uma distribuição normal. Porém, observando o Histograma e o *QQ Plot*, será considerado o teste de *Shapiro Wilk*.

Quadro 3 - Saída Software R - Testes autocorrelação e normalidade resíduos - Modelo 1

```

Ljung-Box test
data:  residuos
X-squared = 0.018451, df = 1, p-value = 0.892
X-squared = 0.68917, df = 2, p-value = 0.7085
X-squared = 0.71755, df = 3, p-value = 0.8691
X-squared = 1.0001, df = 4, p-value = 0.9098
X-squared = 1.2911, df = 5, p-value = 0.9358
X-squared = 1.6189, df = 6, p-value = 0.9512

Shapiro-Wilk normality test
data:  sresiduos
W = 0.95941, p-value = 0.04405

Anderson-Darling normality test
data:  sresiduos
A = 0.58231, p-value = 0.1241

```

Fonte: elaborado pelo autor (2022)

### 3.6.2. Modelo 2

Conforme citado na descrição da Série  $ypA$ , observou-se sazonalidade no Gráfico 3, *Box Plot* por mês, e nos Gráficos 8, 9, 11 e 12, Correlogramas da Série  $ypA$  diferenciada, que foi mais evidente após a aplicação das duas diferenciações. Dessa forma, foram estimados modelos SARIMA com o objetivo de acompanhar a sazonalidade da série.

Foi estimado um modelo SARIMA(1,1,0)(0,1,1)<sub>12</sub>, porém os coeficientes não foram significativos em ambas as componentes. Dessa forma, abandonou-se esse modelo e foi estimado um modelo SARIMA com a componente AR sazonal, a seguir.

M2: SARIMA(1,1,0)(1,1,0)<sub>12</sub>

$$(1 - \phi B)(1 - \Phi B^S)(1 - B)(1 - B^S)Y_t = a_t \quad (37)$$

$$Y_t = (1 + \phi)Y_{t-1} - \phi Y_{t-2} + (1 + \Phi)Y_{t-s} - (1 + \phi\Phi)Y_{t-(s+1)} + (\phi\Phi - \phi)Y_{t-(s+2)} - \Phi Y_{t-2s} + (\phi + \phi\Phi)Y_{t-(2s+1)} + \phi\Phi Y_{t-(2s+2)} + a_t$$

$$ypA_t = 0,6951 ypA_{t-1} - 0,3049 ypA_{t-2} + 0,5310 ypA_{t-12} - 1,1430 ypA_{t-13} + 0,4479 ypA_{t-14} + 0,4690 ypA_{t-24} - 0,1619 ypA_{t-25} + 0,1430 ypA_{t-26} + a_t$$

Quadro 4 - Saída Software R - Estimação parâmetros - Modelo 2

```

Series: ypA
ARIMA(1,1,0)(1,1,0)[12]

z test of coefficients:
      Estimate Std. Error  z value  Pr(>|z|)
ar1  -0.30490    0.14503   -2.1022  0.035532
sar1  -0.46899    0.15062   -3.1138  0.001847

AIC      BIC
686.0181 691.5685

ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
44.22521 287.2782 186.9121 0.5907106 8.390284 0.2906781 -0.01976523

Media    Sd      skewness  kurtose  p_valor
44.22521 286.2491 0.864464 4.66383 0.0008308179

```

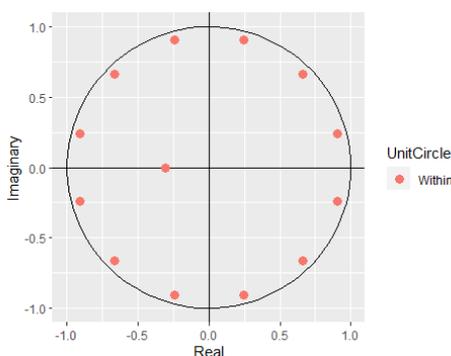
Fonte: elaborado pelo autor (2022)

No Quadro 4, observa-se que os coeficientes autorregressivos sazonais e não sazonais são significativos a 5% de significância, já que o *p-valor* é menor que 0,05, o que faz com que a hipótese nula, de que os coeficientes são iguais a 0, seja rejeitada.

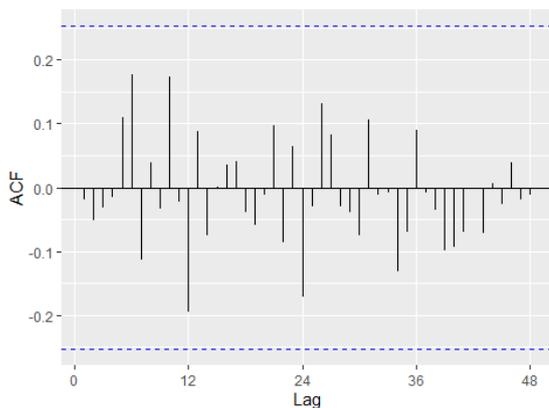
As medidas que serão usadas para comparar os modelos deste estudo, após a fase de estimação dos parâmetros, também se encontram no Quadro 4.

O modelo cumpre a condição de estacionariedade, estando as componentes AR, sazonal e não sazonal, entre -1 e 1, conforme Gráfico 37.

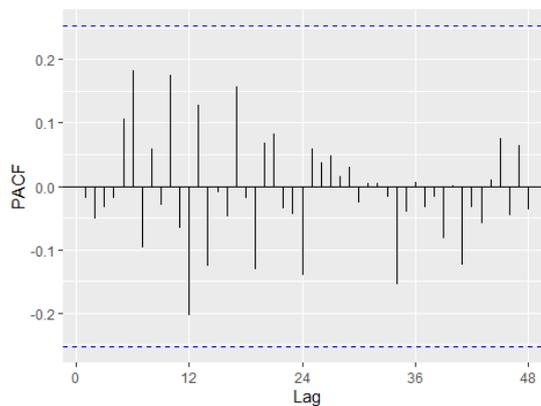
Gráfico 37 - Círculo unitário - Modelo 2



Fonte: elaborado pelo autor (2022)

Gráfico 38 - Correlograma FAC  
resíduos - Modelo 2

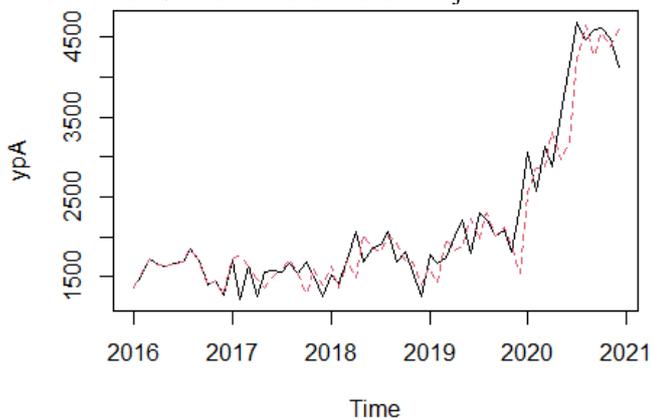
Fonte: elaborado pelo autor (2022)

Gráfico 39 - Correlograma FACP  
resíduos - Modelo 2

Fonte: elaborado pelo autor (2022)

Percebe-se, pelos Gráficos 38 e 39, que o vetor de resíduos é um ruído branco, pois não existe correlação significativa a 5% de significância.

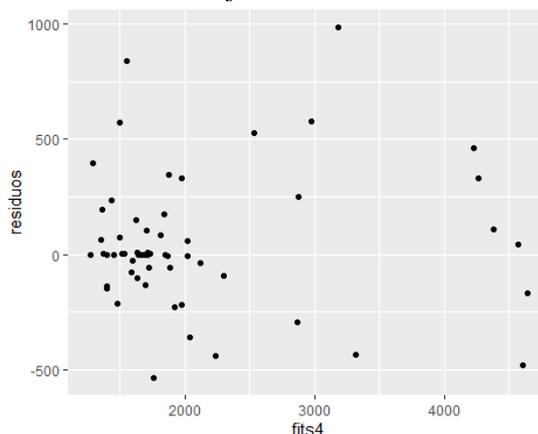
Gráfico 40 - Valores observados vs ajustados Modelo



Fonte: elaborado pelo autor (2022)

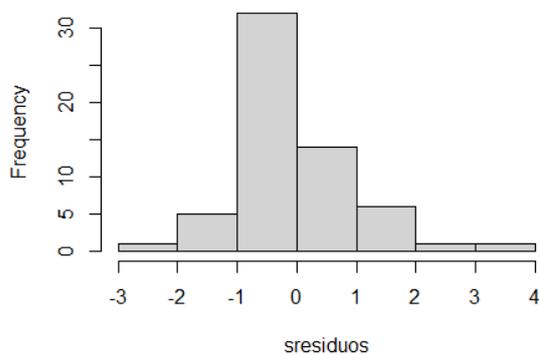
Pelo Gráfico 40, é visto que o modelo parece se ajustar adequadamente aos dados, acompanhando sua variabilidade, apesar de leve distorção em alguns períodos de tempo, já que a linha pontilhada vermelha, que representa os valores ajustados, está acompanhando a linha preta, que representa os valores observados.

Gráfico 41 - Valores ajustados vs resíduos - Modelo 2

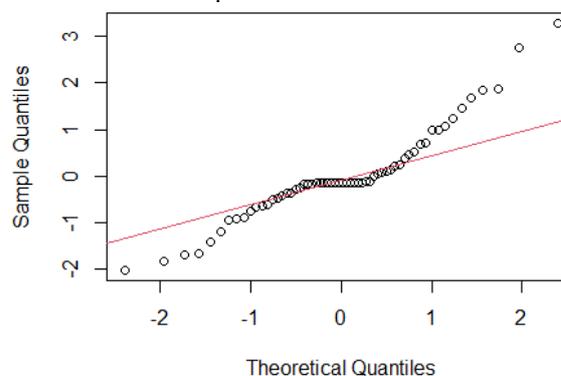


Fonte: elaborado pelo autor (2022)

No Gráfico 41, dos resíduos versus valores ajustados, percebe-se que apesar de haver uma concentração dos dados em torno dos menores valores, pode-se dizer que eles estão variando aleatoriamente em torno de 0.

Gráfico 42 - Histograma  
resíduos padronizados - Modelo 2

Fonte: elaborado pelo autor (2022)

Gráfico 43 - QQ Plot  
resíduos padronizados - Modelo 2

Fonte: elaborado pelo autor (2022)

Pelo Gráfico 42, Histograma dos resíduos padronizados, observa-se uma distribuição assimétrica com valores discrepantes à direita, não tratando-se de uma distribuição normal, o que é comprovado pelo Gráfico 43, *QQ Plot*, já que os resíduos não estão seguindo a linha teórica da distribuição normal.

No Quadro 5, o teste de *Ljung-Box* confirma o vetor de resíduos como ruído branco, já que pelo *p-valor*, maior que o nível de significância de 0,05, não rejeita-se a hipótese nula de correlações iguais a 0.

Pelos testes de *Shapiro Wilk* e *Anderson Darling*, também no Quadro 5, rejeita-se a hipótese nula de normalidade a 5% de significância, já que o *p-valor* é menor que 0,05.

Quadro 5 - Saída Software R - Testes autocorrelação e normalidade resíduos - Modelo 2

```

Ljung-Box test
data:  residuos
X-squared = 0.024632, df = 1, p-value = 0.8753
X-squared = 0.19334, df = 2, p-value = 0.9079
X-squared = 0.2595, df = 3, p-value = 0.9675
X-squared = 0.2735, df = 4, p-value = 0.9915
X-squared = 1.0969, df = 5, p-value = 0.9544
X-squared = 3.2597, df = 6, p-value = 0.7756

Shapiro-Wilk normality test
data:  sresiduos
W = 0.92086, p-value = 0.0008308

Anderson-Darling normality test
data:  sresiduos
A = 1.9761, p-value = 4.271e-05

```

Fonte: elaborado pelo autor (2022)

### 3.6.3. Modelo 3

A partir do modelo 3, a seguir, as estimações foram realizadas com a técnica de regressão dinâmica, por meio dos modelos ARIMAX, em que utiliza-se tanto a variável dependente defasada no tempo, quanto potenciais variáveis exógenas, como regressoras.

Como primeira tentativa, foi estimado um modelo ARIMAX (1,1,0) com a variável *xoA* como regressora, que representa a quantidade de serviços de oficina realizados no período. A variável *xoA* foi incluída pois as máquinas pesadas precisam de serviços de oficina constantemente. Como na maioria desses serviços são usadas peças de reposição, a tendência é que o aumento de ordens de serviço gere o aumento da quantidade de vendas de peças. Porém, ao estimar o modelo, observou-se que a variável *xoA* foi significativa, mas a componente AR não foi significativa. Como foram introduzidos novos produtos, não relacionados à serviços de oficina, a Série Quantidade de Venda de Peças se tornou menos dependente da Série de Serviços de Oficina, o que talvez possa ter gerado essa não significância estatística na estimação do modelo.

Posteriormente, então, foi estimado um modelo com a variável regressora *xmA*, conforme adiante. A variável *xmA* foi incluída pois após o cliente realizar a compra da máquina, há a necessidade de realizar revisões preventivas nesses equipamentos. Como, após a compra da máquina, o cliente tem direito a um ano de garantia, caso realize todas as revisões em alguma concessionária da rede, normalmente a maioria das máquinas compradas são levadas para revisões no período de um ano, o que faz com que a demanda por peças aumente.

M3: ARIMAX(1,1,0) com regressora  $xmA_t$

$$(Y_t - Y_{t-1}) - \phi(Y_{t-1} - Y_{t-2}) + \beta(xmA_t - xmA_{t-1}) = a_t \quad (38)$$

$$ypA_t = 0,6976 ypA_{t-1} + 0,3024 ypA_{t-2} + 13,2310 (xmA_t - xmA_{t-1}) + a_t$$

Quadro 6 - Saída Software R - Estimação parâmetros - Modelo 3

```

Series: ypA
Regression with ARIMA(1,1,0) errors

z test of coefficients:
      Estimate   Std. Error   z value   Pr(>|z|)
ar1  -0.30238    0.12455    -2.4278   0.0151892
xreg  13.23101    3.99019     3.3159   0.0009135

AIC      BIC
835.7661 841.9987

ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
50.87245 271.4456 199.1762 0.7847134 9.750561 0.3097507 0.01146212
Media    Sd          skewness kurtose   p_valor
50.87245 268.8861  1.01815  4.383501 0.006158583

```

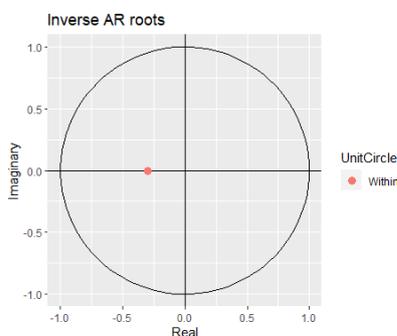
Fonte: elaborado pelo autor (2022)

No Quadro 6, observa-se que tanto a variável regressora  $xmA$  como a componente AR foram significativas a 5% de significância.

As medidas que serão usadas para comparar os modelos deste estudo, após a fase de estimação dos parâmetros, também se encontram no Quadro 6.

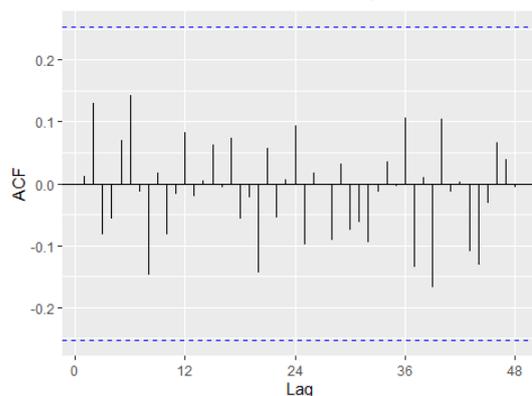
O modelo cumpre a condição de estacionariedade, estando a componente AR entre -1 e 1, conforme Gráfico 44.

Gráfico 44 - Círculo unitário - Modelo 3



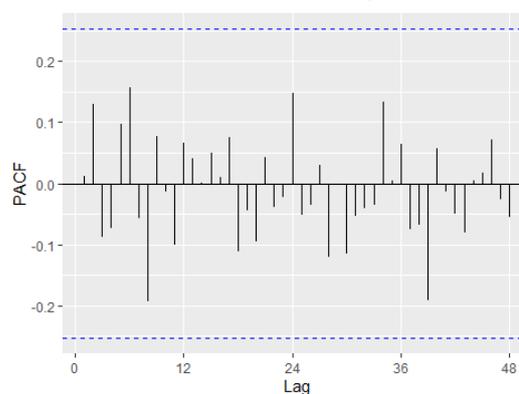
Fonte: elaborado pelo autor (2022)

Gráfico 45 - Correlograma FAC  
resíduos - Modelo 3



Fonte: elaborado pelo autor (2022)

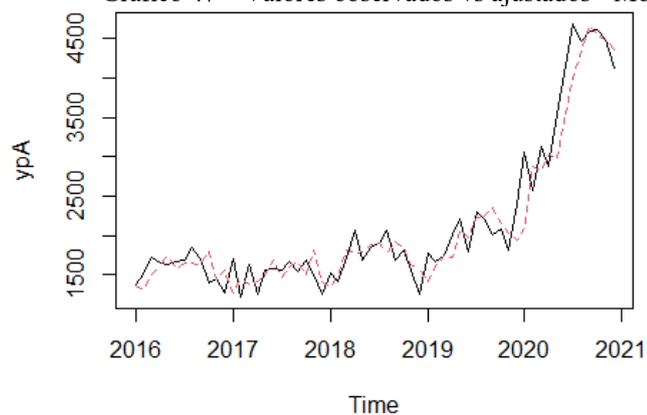
Gráfico 46 - Correlograma FACP  
resíduos - Modelo 3



Fonte: elaborado pelo autor (2022)

Percebe-se, pelos Gráficos 45 e 46, que o vetor de resíduos é um ruído branco, pois não existe correlação significativa a 5% de significância.

Gráfico 47 – Valores observados vs ajustados - Modelo 3

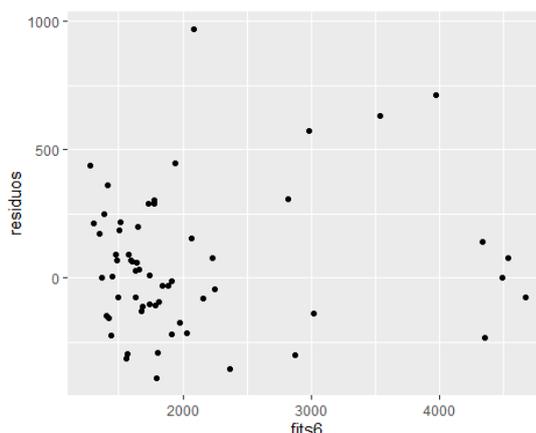


Fonte: elaborado pelo autor (2022)

No Gráfico 47 é possível observar que o modelo se ajusta adequadamente aos dados, acompanhando sua variabilidade, apesar de leve distorção em alguns períodos de tempo.

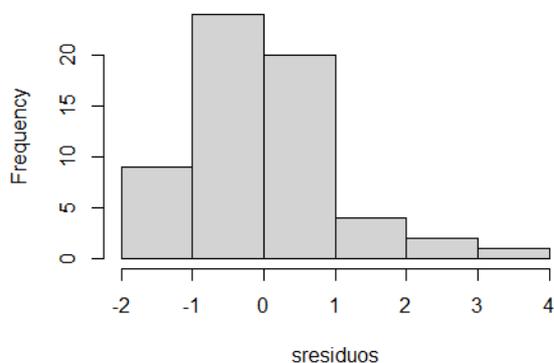
No Gráfico 48, apesar dos dados estarem concentrados nos menores valores, não parece haver um padrão nos dados, estando os erros aleatoriamente distribuídos em torno de 0.

Gráfico 48 - Valores ajustados vs resíduos - Modelo 3

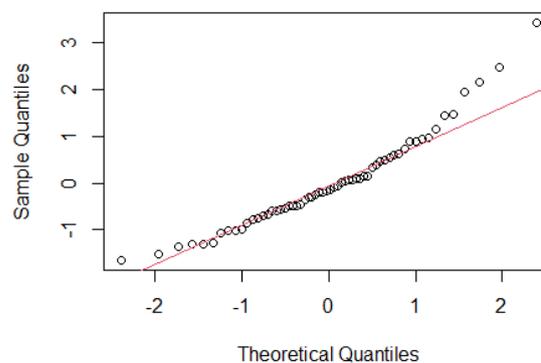


Fonte: elaborado pelo autor (2022)

No Gráfico 49, o Histograma dos resíduos padronizados, percebe-se uma distribuição assimétrica, com cauda e valores discrepantes a direita, não tratando-se de uma distribuição normal, o que é comprovado pelo Gráfico 50, *QQ Plot*, já que os resíduos não estão seguindo a linha teórica da distribuição normal.

Gráfico 49 - Histograma  
resíduos padronizados - Modelo 3

Fonte: elaborado pelo autor (2022)

Gráfico 50 - QQ Plot  
resíduos padronizados - Modelo 3

Fonte: elaborado pelo autor (2022)

No Quadro 7, o teste de *Ljung-Box* confirma o vetor de resíduos como ruído branco, já que pelo *p-valor*, maior que o nível de significância de 0,05, não rejeita-se a hipótese nula de correlações iguais a 0.

Pelos testes de *Shapiro Wilk* e *Anderson Darling*, também no Quadro 7, rejeita-se a hipótese nula de normalidade a 5% de significância, já que o *p-valor* é menor que 0,05.

Quadro 7 - Saída Software R - Testes autocorrelação e normalidade resíduos - Modelo 3

```

Ljung-Box test
data:  residuos
X-squared = 0.0082836, df = 1, p-value = 0.9275
X-squared = 1.0894, df = 2, p-value = 0.58
X-squared = 1.5367, df = 3, p-value = 0.6738
X-squared = 1.7543, df = 4, p-value = 0.7808
X-squared = 2.0926, df = 5, p-value = 0.8362
X-squared = 3.5026, df = 6, p-value = 0.7436

Shapiro-Wilk normality test
data:  sresiduos
W = 0.94125, p-value = 0.006159

Anderson-Darling normality test
data:  sresiduos
A = 0.84879, p-value = 0.02727

```

Fonte: elaborado pelo autor (2022)

### 3.6.4. Modelo 4

M4: ARIMAX (1,1,0) com a regressora  $xmA_t$  e uma indicadora  $ind1$

$$(Y_t - Y_{t-1}) - \phi(Y_{t-1} - Y_{t-2}) + \beta_1(xmA_t - xmA_{t-1}) + \beta_2 ind1 = a_t \quad (39)$$

$$ypA_t = 0,651 ypA_{t-1} + 0,3490 ypA_{t-2} + 14,9505 (xmA_t - xmA_{t-1}) - 817,4163 ind1 + a_t$$

Quadro 8 - Saída Software R - Estimação parâmetros - Modelo 4

```

Series: ypA
Regression with ARIMA(1,1,0) errors

z test of coefficients:
      Estimate  Std. Error  z value  Pr(>|z|)
ar1    -0.34899   0.12248   -2.8494  0.004380
xmA    14.95051   3.67950    4.0632  4.841e-05
ind1   -817.41627 238.07618  -3.4334  0.000596

AIC      BIC
827.1584 835.4685

ME      RMSE    MAE      MPE      MAPE     MASE     ACF1
32.97167 248.037 190.1651 0.1654713 9.432997 0.295737 0.02691308

Media    Sd      skewness  kurtose  p_valor
32.97167 247.9104 0.4513563 3.507597 0.4458184

```

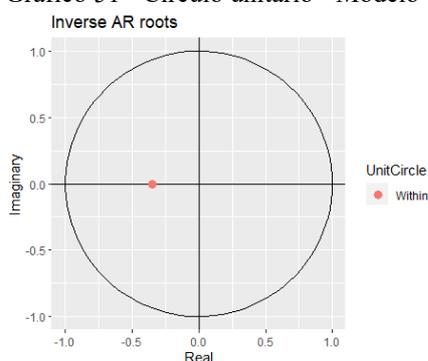
Fonte: elaborado pelo autor (2022)

Neste modelo, além da variável regressora  $xmA$ , que representa a quantidade de máquinas vendidas, foi também incluída a variável indicadora,  $ind1$ , para diferenciar dois períodos com variabilidades distintas na Série  $ypA$ .

No Quadro 8, percebe-se que ambos os coeficientes são significativos a 5% de significância, além de constar as medidas que serão usadas para comparar os modelos deste estudo, após a fase de estimação dos parâmetros.

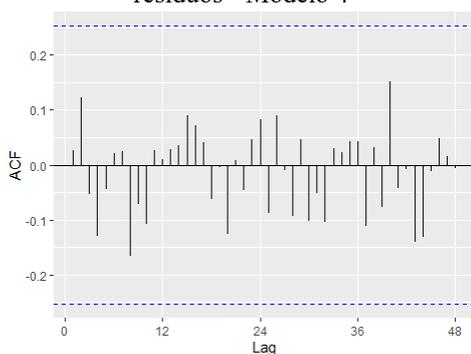
No Gráfico 51, é visto que a componente autorregressiva cumpre a condição de estacionariedade, estando entre -1 e 1.

Gráfico 51 - Círculo unitário - Modelo 4



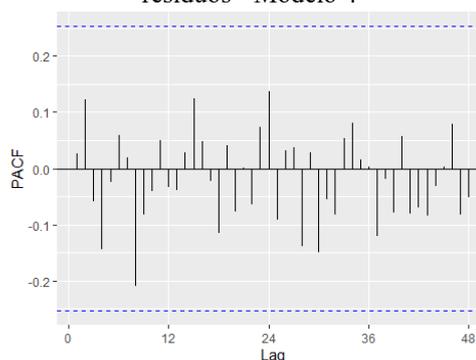
Fonte: elaborado pelo autor (2022)

Gráfico 52 - Correlograma FAC  
resíduos - Modelo 4



Fonte: elaborado pelo autor (2022)

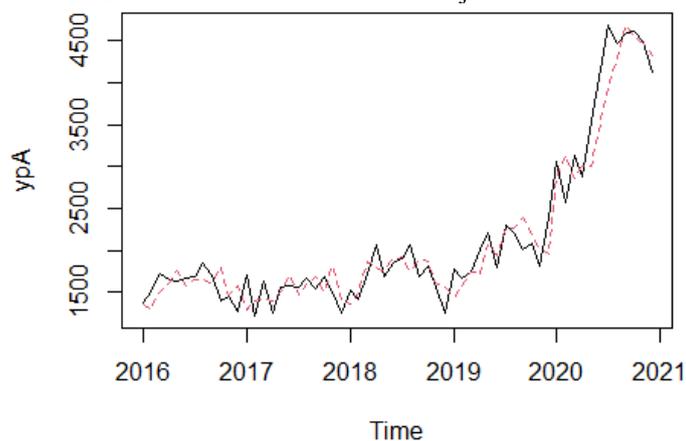
Gráfico 53 - Correlograma FACP  
resíduos - Modelo 4



Fonte: elaborado pelo autor (2022)

Percebe-se, pelos Gráficos 52 e 53, que o vetor de resíduos é um ruído branco, pois não existe correlação significativa a 5% de significância.

Gráfico 54 - Valores observados vs ajustados - Modelo 4

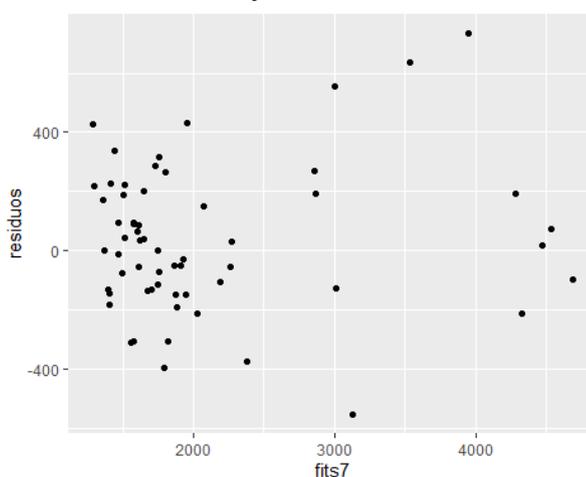


Fonte: elaborado pelo autor (2022)

Pelo Gráfico 54, o modelo parece se ajustar adequadamente aos dados, acompanhando sua variabilidade, apesar de leve distorção em alguns períodos de tempo.

No Gráfico 55, apesar dos dados estarem concentrados nos menores valores, não parece haver um padrão nos dados, estando os erros aleatoriamente distribuídos em torno de 0.

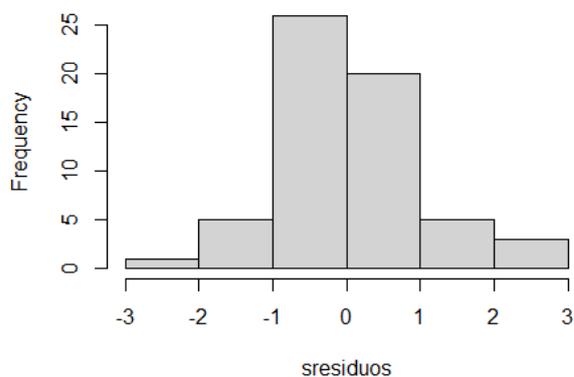
Gráfico 55 – Valores ajustados vs resíduos - Modelo 4



Fonte: elaborado pelo autor (2022)

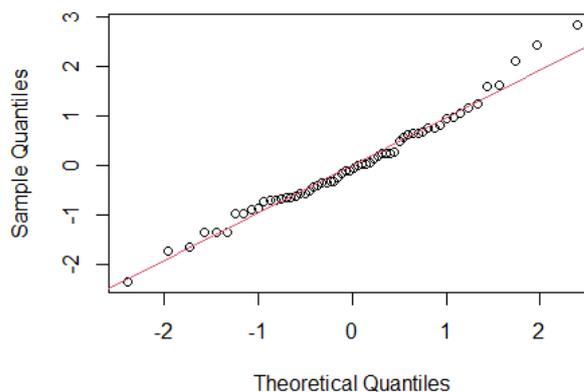
No Gráfico 56, percebe-se uma distribuição bem próxima da simetria, sem valores discrepantes, o que é comprovado pelo Gráfico 57, *QQ Plot*, com praticamente todos os resíduos seguindo a linha teórica da distribuição normal.

Gráfico 56 - Histograma  
resíduos padronizados - Modelo 4



Fonte: elaborado pelo autor (2022)

Gráfico 57 - QQ Plot  
resíduos padronizados - Modelo 4



Fonte: elaborado pelo autor (2022)

No Quadro 9, o teste de *Ljung-Box* confirma o vetor de resíduos como ruído branco, já que pelo *p-valor*, maior que o nível de significância de 0,05, não rejeita-se a hipótese nula de correlações iguais a 0.

Pelos testes de *Shapiro Wilk* e *Anderson Darling*, também no Quadro 9, não rejeita-se a hipótese nula de normalidade a 5% de significância, com o *p-valor* maior que 0,05, de forma que os resíduos podem ser considerados um processo de ruído branco gaussiano.

Quadro 9 - Saída Software R - Testes autocorrelação e normalidade resíduos - Modelo 4

```
Ljung-Box test
data: residuos
X-squared = 0.045669, df = 1, p-value = 0.8308
X-squared = 1.014, df = 2, p-value = 0.6023
X-squared = 1.1944, df = 3, p-value = 0.7544
X-squared = 2.2996, df = 4, p-value = 0.6808
X-squared = 2.4354, df = 5, p-value = 0.7862
X-squared = 2.4682, df = 6, p-value = 0.872

Shapiro-Wilk normality test
data: sresiduos
W = 0.9804, p-value = 0.4458

Anderson-Darling normality test
data: sresiduos
A = 0.402, p-value = 0.3486
```

Fonte: elaborado pelo autor (2022)

### 3.6.5. Modelo 5

Neste modelo, além da inclusão da variável indicadora, ind1, foram incluídas outras duas variáveis indicadoras: ind2 e ind3.

M5: ARIMAX (1,1,0) com a regressora  $xmA_t$  e indicadoras ind1, ind2 e ind3

$$(Y_t - Y_{t-1}) - \phi(Y_{t-1} - Y_{t-2}) + \beta_1(xmA_t - xmA_{t-1}) + \beta_2 ind1 + \beta_3 ind2 + \beta_4 ind3 = a_t \quad (40)$$

$$ypA_t = 0,5992 ypA_{t-1} + 0,4008 ypA_{t-2} + 14,3959 (xmA_t - xmA_{t-1}) - 1038,2473 ind1 - 657,2909 ind2 - 630,9272 ind3 + a_t$$

Quadro 10 - Saída Software R - Estimação parâmetros - Modelo 5

```
Series: ypA
Regression with ARIMA(1,1,0) errors

z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1    -0.40081    0.14781  -2.7116  0.006696
xmA     14.39594    3.21287   4.4807  7.440e-06
ind1  -1038.24731   234.62654  -4.4251  9.639e-06
ind2   -657.29089   222.92982  -2.9484  0.003194
ind3   -630.92723   207.53578  -3.0401  0.002365

AIC      BIC
815.5763 828.0415

ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
29.62992 217.2705 167.922  0.2011118  8.660672  0.2611456  0.02693105

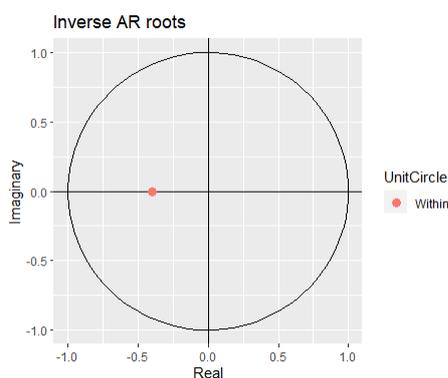
Media    Sd      skewness kurtose  p_valor
29.62992 217.057 0.5900048 3.815384 0.2447036
```

Fonte: elaborado pelo autor (2022)

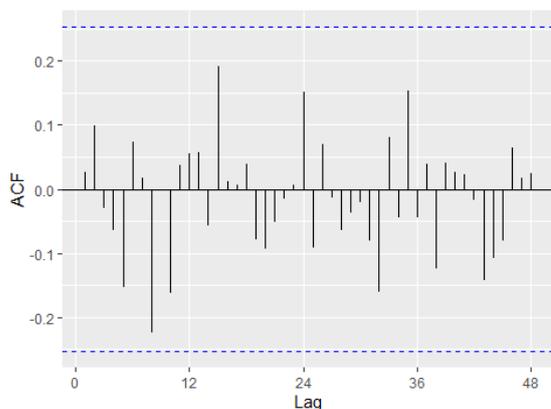
No Quadro 10, percebe-se que ambos os coeficientes são significativos a 5% de significância, além de constar as medidas que serão usadas para comparar os modelos deste estudo, após a fase de estimação dos parâmetros.

Conforme Gráfico 58, a componente autorregressiva cumpre a condição de estacionariedade, estando entre - 1 e 1.

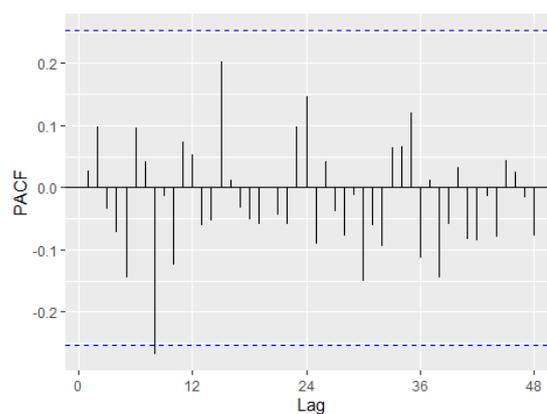
Gráfico 58 - Círculo unitário - Modelo 5



Fonte: elaborado pelo autor (2022)

Gráfico 59 - Correlograma FAC IC 95%  
resíduos - Modelo 5

Fonte: elaborado pelo autor (2022)

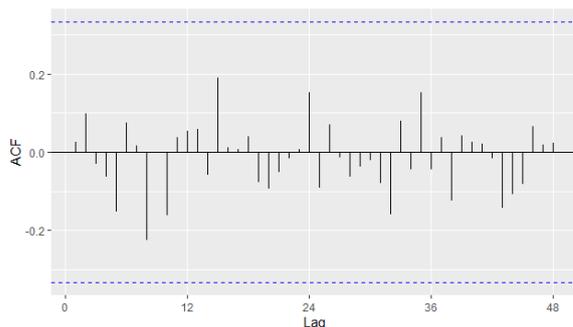
Gráfico 60 - Correlograma FACP IC 95%  
resíduos - Modelo 5

Fonte: elaborado pelo autor (2022)

Percebe-se, pelo Gráfico 60, Correlograma da FACP, que há uma correlação significativa, de forma que não há independência entre os resíduos a um nível de 5% de significância, não sendo os resíduos um processo de ruído branco.

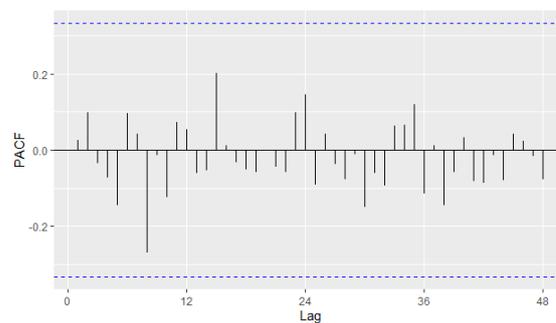
Como apenas um lag é significativo, para não inviabilizar o modelo, optou-se por aumentar o intervalo para um nível de 99% de confiança. Dessa forma, percebe-se, pelos Gráficos 61 e 62, que os resíduos são independentes, podendo ser considerados como ruído branco a 1% de significância.

Gráfico 61 - Correlograma FAC IC 99%  
resíduos - Modelo 5



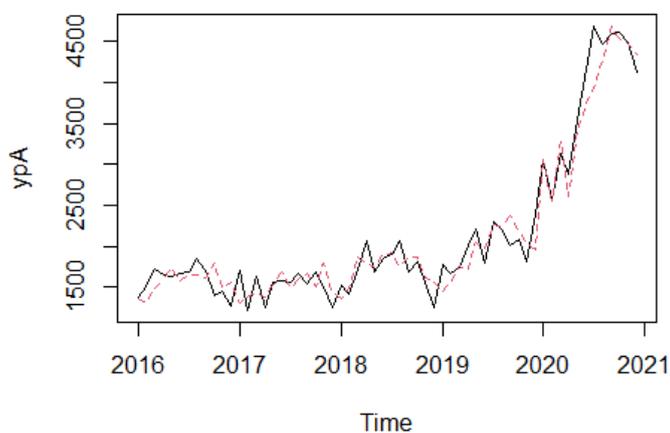
Fonte: elaborado pelo autor (2022)

Gráfico 62 - Correlograma FACP IC 99%  
resíduos - Modelo 5



Fonte: elaborado pelo autor (2022)

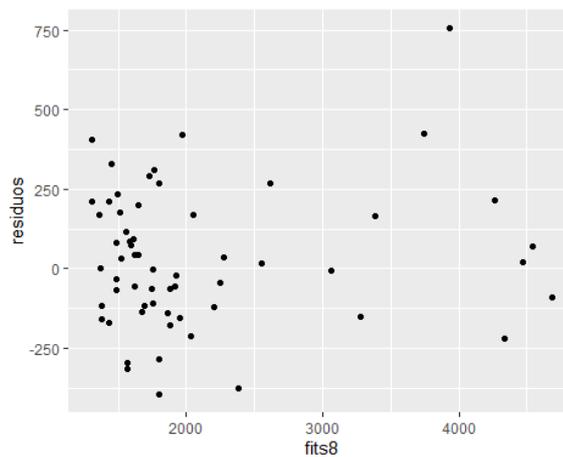
Gráfico 63 – Valores observados vs ajustados - Modelo 5



Fonte: elaborado pelo autor (2022)

No Gráfico 63, é possível observar que o modelo se ajusta bem aos dados, acompanhando sua variabilidade, apesar de leve distorção em alguns períodos de tempo.

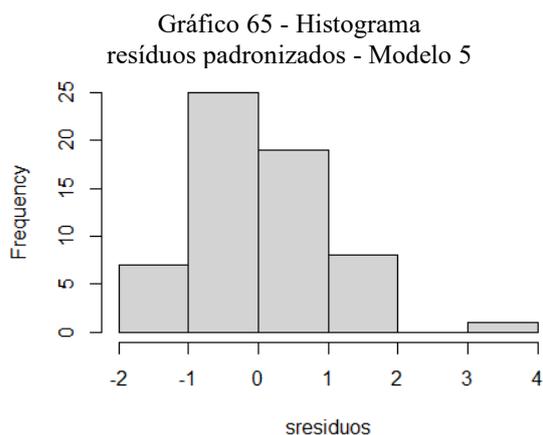
Gráfico 64 – Valores ajustados vs resíduos - Modelo 5



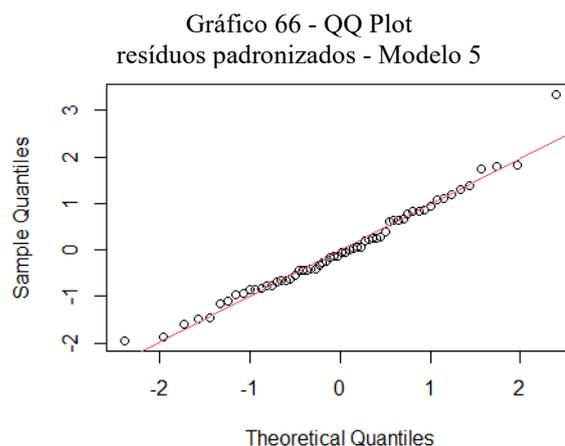
Fonte: elaborado pelo autor (2022)

No Gráfico 64, dos resíduos versus valores ajustados, não parece haver um padrão relevante nos dados, estando os erros aleatoriamente distribuídos em torno de 0.

No Gráfico 65, apesar do Histograma apresentar uma curva com leve assimetria, com valores discrepantes, o Gráfico 66, *QQ Plot*, apresenta uma curva normal, com praticamente todos os resíduos seguindo a linha teórica da distribuição normal.



Fonte: elaborado pelo autor (2022)



Fonte: elaborado pelo autor (2022)

O teste de *Ljung-Box*, no Quadro 11, confirma o vetor de resíduos como ruído branco, já que pelo *p-valor*, maior que o nível de significância de 0,05, não rejeita-se a hipótese nula de correlações iguais a 0.

Quadro 11 - Saída Software R - Testes autocorrelação e normalidade resíduos - Modelo 5

```
Ljung-Box test
data:  residuos
X-squared = 0.04573, df = 1, p-value = 0.8307
X-squared = 0.67036, df = 2, p-value = 0.7152
X-squared = 0.7305, df = 3, p-value = 0.866
X-squared = 1.0054, df = 4, p-value = 0.909
X-squared = 2.5845, df = 5, p-value = 0.7637
X-squared = 2.9622, df = 6, p-value = 0.8136

Shapiro-Wilk normality test
data:  sresiduos
W = 0.97465, p-value = 0.2447

Anderson-Darling normality test
data:  sresiduos
A = 0.32708, p-value = 0.5124
```

Fonte: elaborado pelo autor (2022)

Pelos testes de *Shapiro Wilk* e *Anderson Darling*, também no Quadro 11, não rejeita-se a hipótese nula de normalidade a 5% de significância, já que o *p-valor* é maior que 0,05, sendo o vetor de resíduos considerado um ruído branco gaussiano.

### 3.6.6. Modelo 6

Neste modelo, foram utilizadas todas as variáveis definidas no início do trabalho, incluindo a variável *xoA*, pois mesmo esta não apresentando a componente AR significativa no modelo inicial, ao se trabalhar com as variáveis indicadoras e a regressora *xmA*, pode haver uma melhora na estimação do modelo.

M6: ARIMAX (1,1,0) com regressoras  $xoA_t$ ,  $xmA_t$  e indicadoras *ind1*, *ind2* e *ind3*

$$(Y_t - Y_{t-1}) - \phi(Y_{t-1} - Y_{t-2}) + \beta_1(xoA_t - xoA_{t-1}) + \beta_2(xmA_t - xmA_{t-1}) + \beta_3 ind1 + \beta_4 ind2 + \beta_5 ind3 = a_t \quad (41)$$

$$ypA_t = 0,5626 ypA_{t-1} + 0,4374 ypA_{t-2} + 0,651 (xoA_t - xoA_{t-1}) + 11,614$$

$$(xmA_t - xmA_{t-1}) - 1037,47 ind1 - 692,97 ind2 - 591,09 ind3 + a_t$$

Quadro 12 - Saída Software R - Estimação parâmetros - Modelo 6

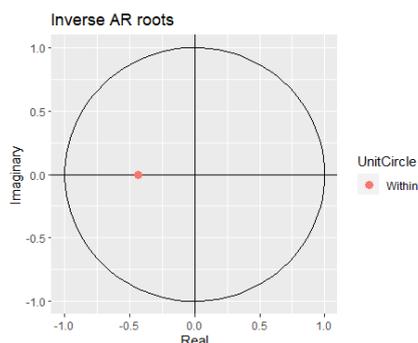
Series: ypA						
Regression with ARIMA(1,1,0) errors						
z test of coefficients:						
	Estimate	Std. Error	z value	Pr(> z )		
ar1	-0.43744	0.21208	-2.0626	0.0391474		
xo	0.65101	0.20922	3.1116	0.0018610		
xm	11.61408	3.09822	3.7486	0.0001778		
ind1	-1037.47053	285.63350	-3.6322	0.0002810		
ind2	-692.96768	293.89983	-2.3578	0.0183818		
ind3	-591.08758	236.33977	-2.5010	0.0123840		
AIC	BIC					
808.5597	823.1024					
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
27.62084	201.2232	149.8164	0.4230955	7.508556	0.2329884	0.04197369
Media	Sd	skewness	kurtose	p_valor		
27.62084	201.0006	0.9631509	4.837931	0.01313397		

Fonte: elaborado pelo autor (2022)

No Quadro 12, percebe-se que todos os coeficientes são significativos a 5% de significância, além de constar as medidas que serão usadas para comparar os modelos deste estudo, após a fase de estimação dos parâmetros.

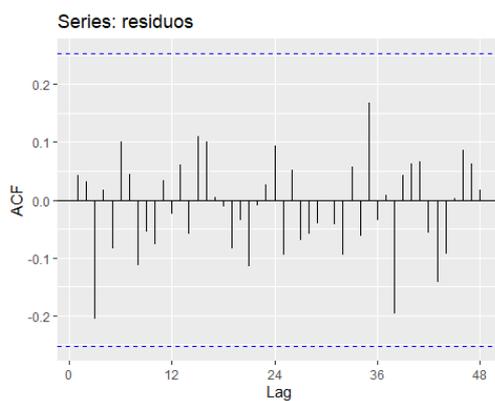
Conforme Gráfico 67, a componente autorregressiva cumpre a condição de estacionariedade, estando entre - 1 e 1.

Gráfico 67 - Círculo unitário - Modelo 6



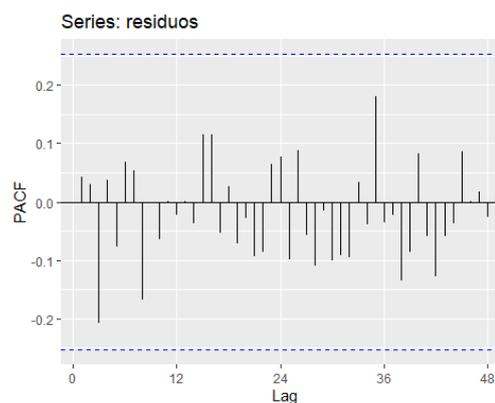
Fonte: elaborado pelo autor (2022)

Gráfico 68 - Correlograma FAC  
resíduos - Modelo 6



Fonte: elaborado pelo autor (2022)

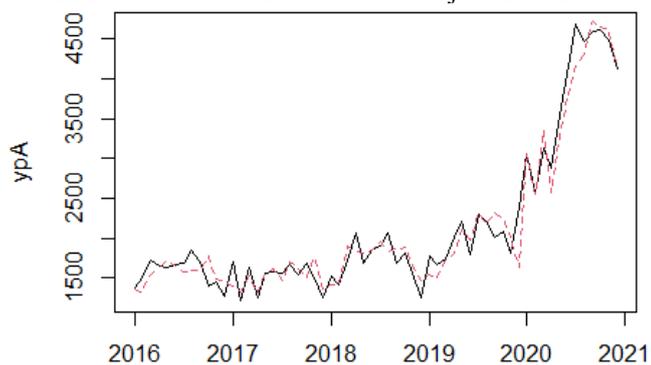
Gráfico 69 - Correlograma FACP  
resíduos - Modelo 6



Fonte: elaborado pelo autor (2022)

Pelos Gráficos 68 e 69, observa-se que o vetor de resíduos é um ruído branco, pois não existe correlação significativa a 5% de significância.

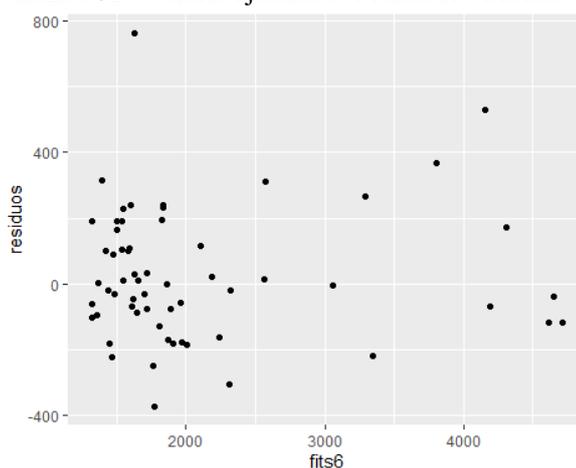
Gráfico 70 – Valores observados vs ajustados Modelo 6



Fonte: elaborado pelo autor (2022)

Pelo Gráfico 70, é possível dizer que o modelo se ajusta adequadamente aos dados, acompanhando sua variabilidade, apesar de leve distorção em alguns períodos de tempo.

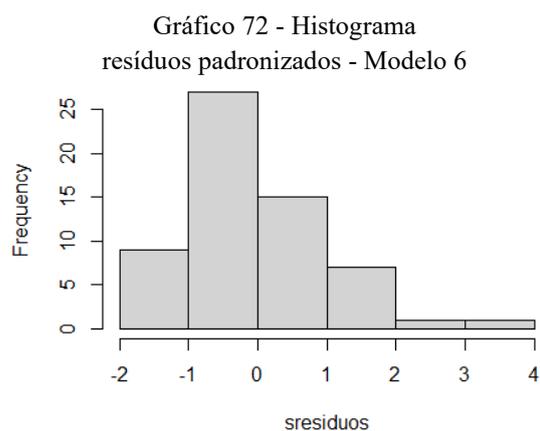
Gráfico 71 - Valores ajustados vs resíduos - Modelo 6



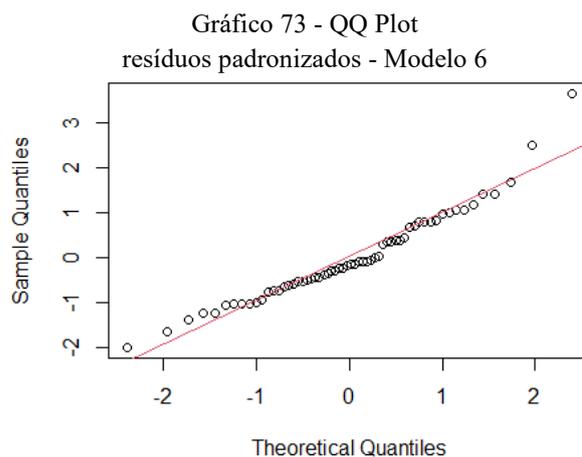
Fonte: elaborado pelo autor (2022)

No Gráfico 71, dos resíduos versus valores ajustados, também não parece haver um padrão relevante nos dados, estando os erros aleatoriamente distribuídos em torno de 0.

Pelo Gráfico 72, o Histograma dos resíduos padronizados, percebe-se uma distribuição assimétrica, com cauda e valores discrepantes à direita, não tratando-se de uma distribuição normal, o que é comprovado pelo Gráfico 73, o *QQ Plot*, já que os resíduos não estão seguindo totalmente a linha teórica da distribuição normal.



Fonte: elaborado pelo autor (2022)



Fonte: elaborado pelo autor (2022)

O teste de *Ljung-Box*, no Quadro 13, confirma o vetor de resíduos como ruído branco, já que pelo *p-valor*, maior que o nível de significância de 0,05, não rejeita-se a hipótese nula de correlações iguais a 0.

Pelo teste de *Shapiro Wilk*, também no Quadro 13, rejeita-se a hipótese nula de normalidade a 5% de significância, já que o *p-valor* é menor que 0,05. No teste de *Anderson Darling*, apesar do *p-valor* ser maior que o nível de significância de 0,05, a diferença é muito pequena, de forma que será considerado o teste de *Shapiro Wilk*.

Quadro 13 - Saída Software R - Testes autocorrelação e normalidade resíduos - Modelo 6

```
Ljung-Box test
X-squared = 0.11108, df = 1, p-value = 0.7389
X-squared = 0.17418, df = 2, p-value = 0.9166
X-squared = 2.9284, df = 3, p-value = 0.4028
X-squared = 2.9491, df = 4, p-value = 0.5664
X-squared = 3.4383, df = 5, p-value = 0.6327
X-squared = 4.1347, df = 6, p-value = 0.6585

Shapiro-Wilk normality test
data: sresiduos
W = 0.94842, p-value = 0.01313

Anderson-Darling normality test
data: sresiduos
A = 0.72612, p-value = 0.05524
```

Fonte: elaborado pelo autor (2022)

Também foram estimados modelos SARIMAX (1,1,0)(1,1,0)<sub>12</sub>. Porém, optou-se por não apresentar esses modelos no trabalho, já que obtiveram valores de parâmetros e testes muito próximos aos dos modelos ARIMAX (1,1,0), que são mais parcimoniosos.

### 3.6.7. Estatísticas dos ajustes dos modelos

A seguir, na Tabela 1, apresenta-se o resumo das estatísticas dos ajustes dos modelos estimados, como forma de comparar e definir o melhor modelo para acompanhar a variabilidade da Série  $ypA$ .

Tabela 1 - Resumo das estatísticas dos ajustes dos modelos

Estatísticas	M1 ARIMA (1,1,0)	M2 SARIMA (1,1,0) (1,1,0) <sub>12</sub>	M3 ARIMAX (1,1,0) com regressora xmA	M4 ARIMAX (1,1,0) com regressoras xmA e Ind1	M5 ARIMAX (1,1,0) com regressoras xmA e Ind1, Ind2 e Ind3	M6 ARIMAX (1,1,0) com regressoras xoA, xmA e Ind1, Ind2 e Ind3
<b>Estatísticas do Modelo</b>						
$\phi_1$	-0,2789	-0,3049	-0,3024	-0,3490	-0,4008	-0,4374
$\Phi_1$		-0,4690				
$\beta_1$			13,2310	14,9505	14,3959	0,6510
$\beta_2$				-817,4162	-1038,2473	11,6141
$\beta_3$					-657,2908	-1037,4705
$\beta_4$					-630,9272	-692,9677
$\beta_5$						-591,0876
<b>RMSE</b>	295,7891	287,2782	271,4456	248,0370	217,2705	<b>201,2232</b>
<b>AIC</b>	843,8856	<b>686,0181</b>	835,7661	827,1584	815,5763	808,5597
<b>BIC</b>	848,0407	<b>691,5685</b>	841,9987	835,4685	828,0415	823,1024
<b>Estatísticas dos Resíduos</b>						
<b>Média</b>	60,2360	44,2252	50,8724	32,9716	29,6299	27,6208
<b>Desvio padrão</b>	292,0346	286,2491	268,8861	247,9104	217,0570	201,0006
<b>Assimetria</b>	0,6068	0,8644	1,0181	0,4513	0,5900	0,9631
<b>Curtose</b>	3,0988	4,6638	4,3835	3,5075	3,8153	4,8379
<b>Teste Shapiro-Wilk - Valor P</b>	0,0440	0,0008	0,0061	<b>0,4458</b>	<b>0,2447</b>	0,0131

Fonte: elaborado pelo autor (2022)

Na Tabela 1, observa-se que o modelo 2 obteve o menor valor dos índices AIC e BIC e o modelo 6 obteve o menor valor da estatística de erro RMSE, ambos com o problema da não normalidade dos resíduos.

O modelo 5 tem os valores da estatística de erro RMSE e índices AIC e BIC pouco maiores que o modelo 6. Tal modelo ainda tem a distribuição normal dos resíduos, conforme pode-se verificar pelo *p-valor*, maior que o nível de significância, de forma que não rejeita-se a hipótese nula de normalidade, obtendo assim um ruído branco gaussiano.

Na Tabela 2, pela equação dos modelos 5 e 6, observa-se que os modelos são parecidos. O modelo 6 combina o passado da série no instante ( $t-1$ ), a informação das variáveis regressoras e três variáveis indicadoras. O modelo 5 combina o passado da série no instante ( $t-1$ ), a variável regressora  $xmA$  e também as três variáveis indicadoras.

Tabela 2 - Equações dos modelos

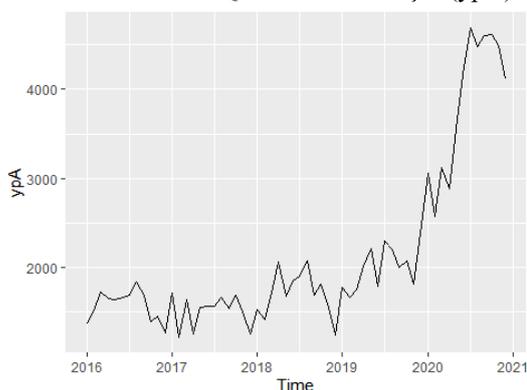
<b>M1</b>	$ypA_t = 0,7211 ypA_{t-1} + 0,2789 ypA_{t-2} + a_t$
<b>M2</b>	$ypA_t = 0,6951 ypA_{t-1} - 0,3049 ypA_{t-2} + 0,5310 ypA_{t-12} - 1,1430 ypA_{t-13} + 0,4479 ypA_{t-14} + 0,4690 ypA_{t-24} - 0,1619 ypA_{t-25} + 0,1430 ypA_{t-26} + a_t$
<b>M3</b>	$ypA_t = 0,6976 ypA_{t-1} + 0,3024 ypA_{t-2} + 13,2310 (xmA_t - xmA_{t-1}) + a_t$
<b>M4</b>	$ypA_t = 0,651 ypA_{t-1} + 0,3490 ypA_{t-2} + 14,9505 (xmA_t - xmA_{t-1}) - 817,4163 ind1 + a_t$
<b>M5</b>	$ypA_t = 0,5992 ypA_{t-1} + 0,4008 ypA_{t-2} + 14,3959 (xmA_t - xmA_{t-1}) - 1038,2473 ind1 - 657,2909 ind2 - 630,9272 ind3 + a_t$
<b>M6</b>	$ypA_t = 0,5626 ypA_{t-1} + 0,4374 ypA_{t-2} + 0,651 (xoA_t - xoA_{t-1}) + 11,614 (xmA_t - xmA_{t-1}) - 1037,47 ind1 - 692,97 ind2 - 591,09 ind3 + a_t$

Fonte: elaborado pelo autor (2022)

A diferença entre os modelos 5 e 6 está na influência das variáveis regressoras  $xmA$  e  $xoA$  na Série  $ypA$ . No modelo 5, o coeficiente da variável  $xmA$  (14,39) é maior que a soma dos coeficientes das variáveis  $xoA$  e  $xmA$  (12,265) no modelo 6, o que mostra que a variável  $xmA$  tem uma grande influência nos dois modelos, porém é mais relevante no modelo 5.

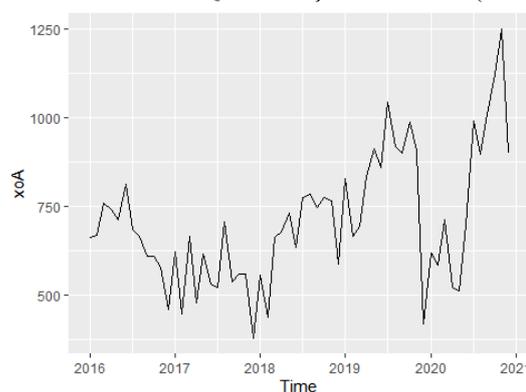
Além disso, mesmo a variável  $xoA$  influenciando a variável  $ypA$  no modelo 6, percebe-se que essa influência não é relevante, se comparado a variável  $xmA$ . Isso pode se dar pelo deslocamento que a Série  $ypA$  sofre a partir do ano de 2020, mostrando que a venda de peças, a partir deste ano, depende menos da venda de oficina, conforme Gráficos 74 e 75, das variáveis  $ypA$  e  $xoA$ , respectivamente.

Gráfico 74 - Série Qtd. Venda de Peças (ypA)



Fonte: elaborado pelo autor (2022)

Gráfico 75 - Série Qtd. Serviços de Oficina (xoA)



Fonte: elaborado pelo autor (2022)

Já observando a equação do modelo 2, na Tabela 2, observa-se que, com este modelo, é possível acompanhar a variabilidade da série apenas observando o passado dos dados. Assim, mesmo o modelo 2 não gerando um ruído branco gaussiano nos resíduos, é o melhor modelo para acompanhar a variabilidade da série, tendo os menores valores das estatísticas AIC e BIC.

### 3.7. Previsões

A seguir, serão apresentadas as previsões de cada modelo estimado. Porém, considerando um intervalo de 95% de confiança, observou-se que apenas o modelo 2, SARIMA  $(1,1,0)(1,1,0)_{12}$ , teve o intervalo de confiança da previsão englobando os valores observados.

Dessa forma, para ter oportunidade de considerar outros modelos, optou-se por realizar as previsões considerando um intervalo de confiança de 99%, os quais serão apresentados a seguir.

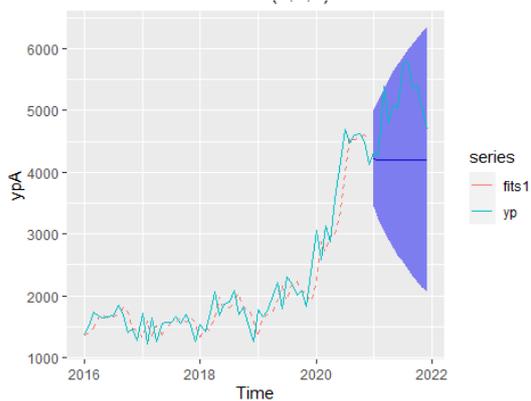
Em uma oportunidade futura, pode-se verificar a possibilidade de aumentar o período de análise, para datas anteriores a 2016, para que se obtenha amostras maiores e com isso ser possível reduzir o nível de confiança.

### 3.7.1. Modelo 1

M1: ARIMA(1,1,0)

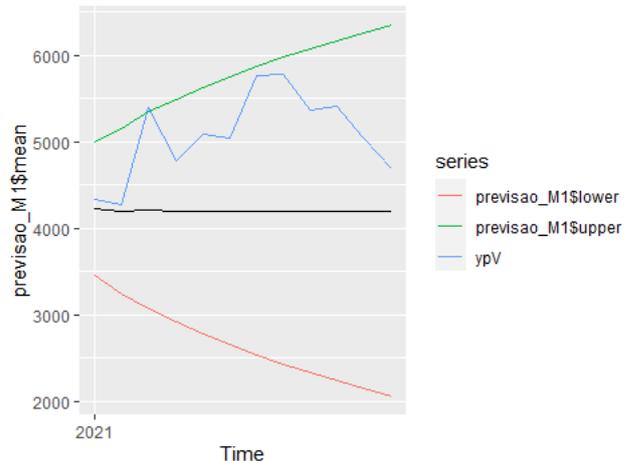
$$ypA_t = 0,7211 ypA_{t-1} + 0,2789 ypA_{t-2} + a_t \quad (42)$$

Gráfico 76 – Valores observados vs valores previstos - Modelo 1



Fonte: elaborado pelo autor (2022)

Gráfico 77 - Valores observados vs valores previstos - período de validação - Modelo 1



Fonte: elaborado pelo autor (2022)

Quadro 14 - Saída Software R - Previsão pontual 12 meses e IC 99% - Modelo 1

	Point Forecast	Lo 99	Hi 99
Jan 2021	4223.902	3448.975	4998.829
Feb 2021	4195.206	3239.80	5150.611
Mar 2021	4203.208	3064.75	5341.659
Apr 2021	4200.977	2913.026	5488.927
May 2021	4201.599	2777.791	5625.407
Jun 2021	4201.425	2654.15	5748.699
Jul 2021	4201.474	2539.750	5863.198
Aug 2021	4201.460	2432.711	5970.209
Sep 2021	4201.464	2331.797	6071.131
Oct 2021	4201.463	2236.056	6166.870
Nov 2021	4201.463	2144.767	6258.160
Dec 2021	4201.463	2057.361	6345.565

Fonte: elaborado pelo autor (2022)

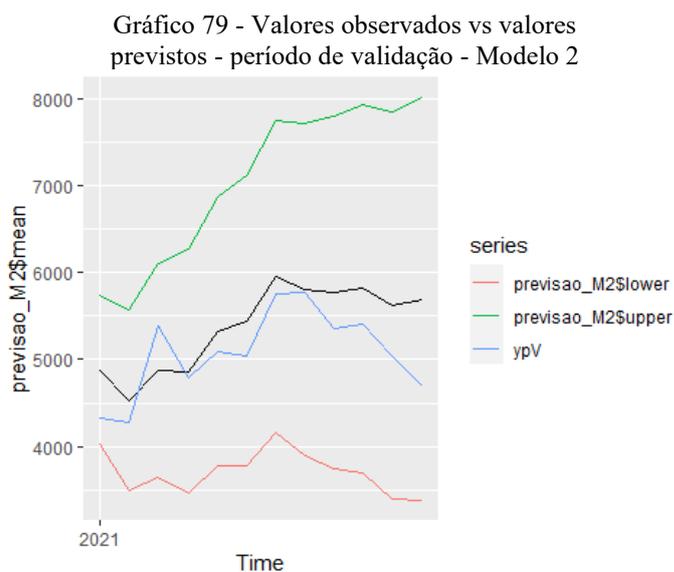
### 3.7.2. Modelo 2

M2: SARIMA(1,1,0)(1,1,0)<sub>12</sub>

$$ypA_t = 0,6951 ypA_{t-1} - 0,3049 ypA_{t-2} + 0,5310 ypA_{t-12} - 1,1430 ypA_{t-13} + 0,4479 ypA_{t-14} + 0,4690 ypA_{t-24} - 0,1619 ypA_{t-25} + 0,1430 ypA_{t-26} + a_t \quad (43)$$



Fonte: elaborado pelo autor (2022)



Fonte: elaborado pelo autor (2022)

Quadro 15 - Saída Software R - Previsão pontual 12 meses e IC 99% - Modelo 2

	Point Forecast	Lo 99	Hi 99
Jan 2021	4882.932	4028.477	5737.386
Feb 2021	4528.103	3487.501	5568.704
Mar 2021	4875.037	3635.571	6114.502
Apr 2021	4866.233	3467.066	6265.399
May 2021	5318.653	3773.111	6864.196
Jun 2021	5447.417	3769.089	7125.745
Jul 2021	5960.855	4159.254	7762.456
Aug 2021	5801.937	3885.045	7718.828
Sep 2021	5773.215	3747.564	7798.867
Oct 2021	5815.944	3687.088	7944.801
Nov 2021	5628.693	3401.407	7855.980
Dec 2021	5698.202	3376.656	8019.748

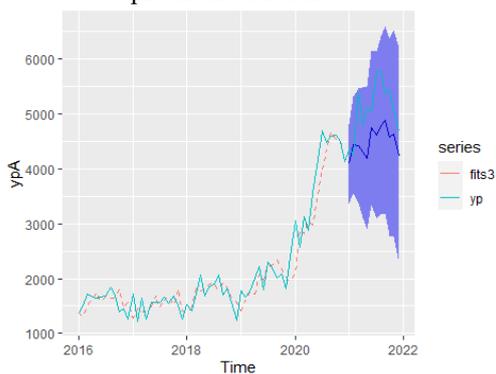
Fonte: elaborado pelo autor (2022)

### 3.7.3. Modelo 3

M3: ARIMAX(1,1,0) com regressora  $xmA_t$

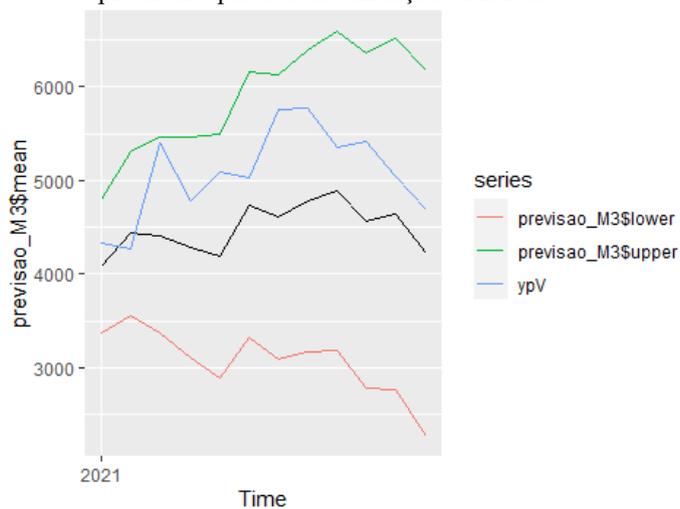
$$ypA_t = 0,6976 ypA_{t-1} + 0,3024 ypA_{t-2} + 13,2310 (xmA_t - xmA_{t-1}) + a_t \quad (44)$$

Gráfico 80 - Valores observados vs valores previstos - Modelo 3



Fonte: elaborado pelo autor (2022)

Gráfico 81 - Valores observados vs valores previstos - período de validação - Modelo 3



Fonte: elaborado pelo autor (2022)

Quadro 16 - Saída Software R - Previsão pontual 12 meses e IC 99% - Modelo 3

	Point Forecast	Lo 99	Hi 99
Jan 2021	4082.721	3365.359	4800.082
Feb 2021	4432.758	3558.083	5307.433
Mar 2021	4412.474	3370.622	5454.325
Apr 2021	4291.526	3115.178	5467.875
May 2021	4186.243	2886.735	5485.752
Jun 2021	4741.775	3330.516	6153.033
Jul 2021	4609.516	3094.531	6124.502
Aug 2021	4781.504	3169.508	6393.500
Sep 2021	4887.357	3183.850	6590.863
Oct 2021	4569.811	2779.470	6360.152
Nov 2021	4635.966	2762.810	6509.123
Dec 2021	4225.805	2273.344	6178.267

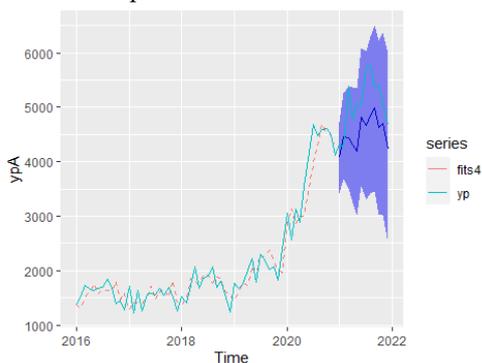
Fonte: elaborado pelo autor (2022)

### 3.7.4. Modelo 4

M4: ARIMAX (1,1,0) com a regressora  $xmA_t$  e uma indicadora  $ind1$

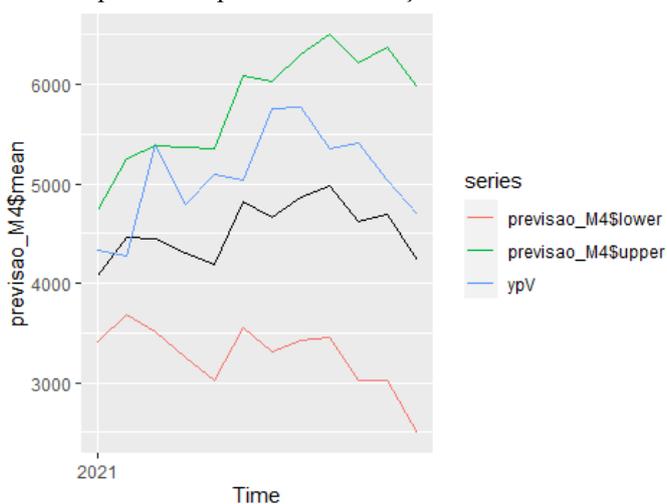
$$ypA_t = 0,651 ypA_{t-1} + 0,3490 ypA_{t-2} + 14,9505 (xmA_t - xmA_{t-1}) - 817,4163 ind1 + a_t \quad (45)$$

Gráfico 82 - Valores observados vs valores previstos - Modelo 4



Fonte: elaborado pelo autor (2022)

Gráfico 83 - Valores observados vs valores previstos - período de validação - Modelo 4



Fonte: elaborado pelo autor (2022)

Quadro 17 - Saída Software R - Previsão pontual 12 meses e IC 99% - Modelo 4

	Point Forecast	Lo 99	Hi 99
Jan 2021	4072.780	3411.455	4734.105
Feb 2021	4466.482	3677.365	5255.599
Mar 2021	4445.275	3505.114	5385.437
Apr 2021	4307.686	3250.729	5364.644
May 2021	4189.141	3022.957	5355.325
Jun 2021	4816.693	3551.999	6081.387
Jul 2021	4667.317	3310.819	6023.815
Aug 2021	4861.629	3419.300	6303.958
Sep 2021	4981.248	3457.870	6504.627
Oct 2021	4622.431	3022.118	6222.744
Nov 2021	4697.185	3023.465	6370.905
Dec 2021	4233.719	2489.680	5977.757

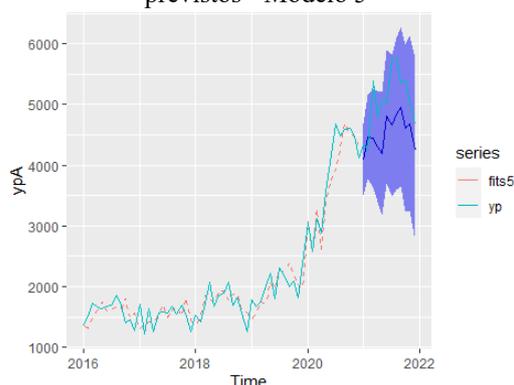
Fonte: elaborado pelo autor (2022)

### 3.7.5. Modelo 5

M5: ARIMAX (1,1,0) com a regressora  $xmA_t$  e indicadoras ind1, ind2 e ind3

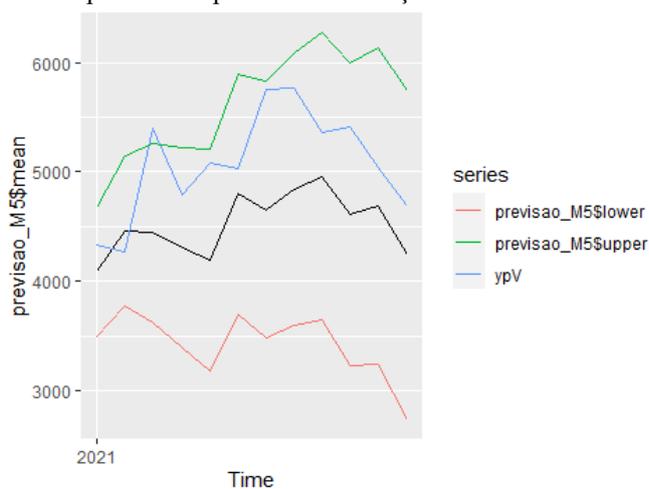
$$ypA_t = 0,5992 ypA_{t-1} + 0,4008 ypA_{t-2} + 14,3959 (xmA_t - xmA_{t-1}) - 1038,2473 ind1 - 657,2909 ind2 - 630,9272 ind3 + a_t \quad (46)$$

Gráfico 84 - Valores observados vs valores previstos - Modelo 5



Fonte: elaborado pelo autor (2022)

Gráfico 85 - Valores observados vs valores previstos - período de validação - Modelo 5



Fonte: elaborado pelo autor (2022)

Quadro 18 - Saída Software R - Previsão pontual 12 meses e IC 99% - Modelo 5

	Point Forecast	Lo 99	Hi 99
Jan 2021	4090.262	3500.337	4680.187
Feb 2021	4459.508	3771.790	5147.225
Mar 2021	4444.280	3623.377	5265.182
Apr 2021	4309.280	3391.568	5226.991
May 2021	4196.291	3184.741	5207.841
Jun 2021	4800.047	3705.006	5895.088
Jul 2021	4656.438	3482.958	5829.918
Aug 2021	4843.445	3596.781	6090.108
Sep 2021	4958.668	3642.760	6274.577
Oct 2021	4613.143	3231.504	5994.782
Nov 2021	4685.132	3240.732	6129.532
Dec 2021	4238.854	2734.316	5743.393

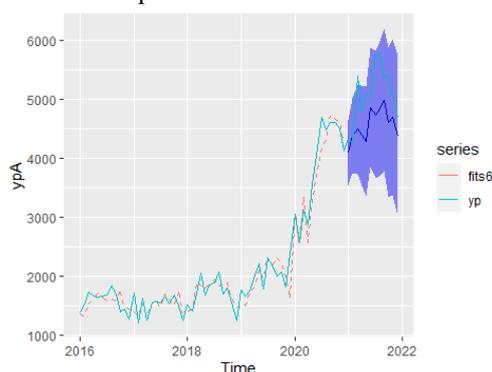
Fonte: elaborado pelo autor (2022)

### 3.7.6. Modelo 6

M6: ARIMAX (1,1,0) com regressoras  $xoA_t$ ,  $xmA_t$  e indicadoras ind1, ind2 e ind3

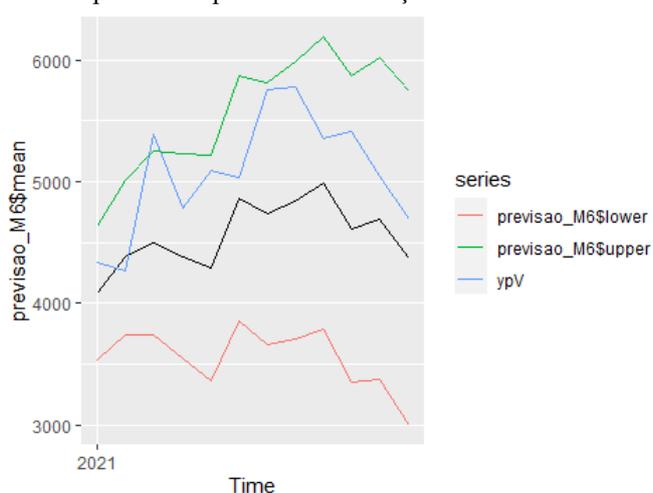
$$ypA_t = 0,5626 ypA_{t-1} + 0,4374 ypA_{t-2} + 0,651 (xoA_t - xoA_{t-1}) + 11,614 (xmA_t - xmA_{t-1}) - 1037,47 ind1 - 692,97 ind2 - 591,09 ind3 + a_t \quad (47)$$

Gráfico 86 - Valores observados vs valores previstos - Modelo 6



Fonte: elaborado pelo autor (2022)

Gráfico 87 - Valores observados vs valores previstos - período de validação - Modelo 6



Fonte: elaborado pelo autor (2022)

Quadro 19 - Saída Software R - Previsão pontual 12 meses e IC 99% - Modelo 6

	Point Forecast	Lo 99	Hi 99
Jan 2021	4084.358	3532.874	4635.842
Feb 2021	4378.872	3746.112	5011.631
Mar 2021	4497.438	3740.305	5254.570
Apr 2021	4386.538	3544.006	5229.070
May 2021	4289.293	3360.958	5217.628
Jun 2021	4862.920	3859.455	5866.385
Jul 2021	4735.755	3661.028	5810.482
Aug 2021	4845.054	3704.074	5986.034
Sep 2021	4986.150	3782.322	6189.978
Oct 2021	4609.756	3346.302	5873.211
Nov 2021	4694.520	3374.087	6014.953
Dec 2021	4371.590	2996.556	5746.625

Fonte: elaborado pelo autor (2022)

### 3.7.7. Estatísticas dos erros de previsão

A seguir, na Tabela 3, apresenta-se o resumo das estatísticas dos erros de previsão dos modelos estimados, como forma de comparar e definir o melhor modelo para acompanhar o passado e realizar a previsão da Série  $ypA$  para 12 meses.

Tabela 3 - Resumo das estatísticas dos erros de previsão

Período Amostral						
	M1 ARIMA (1,1,0)	M2 SARIMA (1,1,0) (1,1,0) <sub>12</sub>	M3 ARIMAX (1,1,0) com regressora $xmA$	M4 ARIMAX (1,1,0) com regressoras $xmA$ e $Ind1$	M5 ARIMAX (1,1,0) com regressoras $xmA$ e $Ind1$ , $Ind2$ e $Ind3$	M6 ARIMAX (1,1,0) com regressoras $xoA$ , $xmA$ , $Ind1$ , $Ind2$ e $Ind3$
<b>ME</b>	60,2360	44,2252	50,8724	32,9716	29,6299	<b>27,6208</b>
<b>RMSE</b>	295,7890	287,2781	271,4456	248,0369	217,2704	<b>201,2232</b>
<b>MAE</b>	224,1308	186,9120	199,1761	190,1650	167,9220	<b>149,8164</b>
<b>MPE</b>	1,1612	0,5907	0,7847	<b>0,1654</b>	0,2011	0,4230
<b>MAPE</b>	11,0272	8,3902	9,7505	9,4329	8,6606	<b>7,5085</b>
<b>ACF1</b>	-0,0171	-0,0197	<b>0,0114</b>	0,0269	0,0269	0,0419
<b>Theil's U</b>	0,9029	0,7965	0,8216	0,7579	0,7056	<b>0,6477</b>
Período de Validação						
<b>ME</b>	875,1579	<b>-305,0184</b>	589,9620	547,9511	553,4710	516,2298
<b>RMSE</b>	995,6919	<b>466,4149</b>	695,6035	658,3011	663,1628	619,9419
<b>MAE</b>	875,1579	<b>391,6789</b>	617,0883	580,6981	585,0556	534,3750
<b>MPE</b>	16,4834	<b>-6,3282</b>	11,1182	10,3224	10,4191	9,7088
<b>MAPE</b>	16,4834	<b>7,9345</b>	11,7535	11,0893	11,1588	10,1337
<b>ACF1</b>	0,3707	0,2980	<b>-0,1474</b>	-0,2294	-0,1990	-0,2299
<b>Theil's U</b>	1,9127	<b>0,8681</b>	1,3696	1,3020	1,3106	1,2189

Fonte: elaborado pelo autor (2022)

Conforme Tabela 3, o melhor modelo para acompanhar o passado da Série Quantidade de Venda de Peças é o modelo 6, pois possui os menores valores na maioria das estatísticas de erro no período amostral, considerando a Série  $ypA$  no instante  $(t-1)$ , as variáveis regressoras  $xoA$  e  $xmA$  e variáveis indicadoras.

Já para previsão, o melhor modelo é o 2, com os menores valores na maioria das estatísticas de erro no período de validação, o que é reforçado pelo Gráfico 79, que mostra a previsão pontual do modelo 2 mais próxima dos dados no período de validação. Com o modelo 2 é possível obter boas previsões combinando apenas o passado da série.

Na Tabela 3, observa-se também que o modelo 3 tem a menor estatística ACF1, tanto no período amostral quanto no período de validação.

Porém, o critério adotado nesta monografia será pela maioria das estatísticas de erro de previsão, conforme Tabela 3. Assim, pode-se resumir que os melhores modelos são o modelo 2 para previsão e o modelo 6 para acompanhar o passado da série.

#### 4. CONSIDERAÇÕES FINAIS

Esta monografia teve como principal objetivo desenvolver modelos de séries temporais para a previsão do volume de vendas de peças para máquinas pesadas, para equilibrar o estoque em níveis satisfatórios considerando a lucratividade da empresa e o adequado atendimento ao cliente.

Além da variável que representa o histórico de vendas de peças ( $yp$ ), foram consideradas duas variáveis exógenas, a Quantidade de Máquinas Vendidas ( $xm$ ) e a Quantidade de Serviços de Oficina Realizados ( $xo$ ), bem como 3 variáveis indicadoras para diferenciar períodos de variabilidades distintas ao longo da Série Quantidade de Vendas de Peças ( $yp$ ).

Dessa forma, foram estimados modelos ARIMA univariados, considerando a variável ( $yp$ ) defasada no tempo, como regressora, e modelos ARIMAX multivariados, adicionando as variáveis exógenas ao modelo, por meio de uma função de transferência.

Após a análise dos dados, avaliando todas as etapas do trabalho, como modelos gerais destaca-se os modelos 2 e 6, além de ser possível considerar o modelo 5 com um bom modelo.

Na etapa de estimação dos parâmetros dos modelos, observou-se que o modelo 2 combina bem o passado da série nos períodos ( $t-1$ ), ( $t-2$ ) e nos períodos sazonais, obtendo o menor valor dos índices AIC e BIC. Assim, mesmo o modelo 2 não gerando um ruído branco gaussiano nos resíduos, foi o modelo mais adequado para acompanhar a variabilidade da série, apenas observando o passado dos dados.

Na fase de previsão, o modelo 6 obteve os menores valores na maioria das medidas de erro de previsão no período amostral, sendo o mais adequado para acompanhar o passado da série de interesse  $yp$ , considerando a série no instante ( $t-1$ ), as variáveis regressoras  $xo$  e  $xm$  e as variáveis indicadoras.

O modelo 2 foi também o melhor modelo para previsão, com os menores valores na maioria das estatísticas de erro de previsão no período de validação, tendo a previsão pontual mais próxima dos dados, em comparação aos demais modelos.

O modelo 5 pode ser considerado um bom modelo para acompanhar a variabilidade da série, em opção ao modelo 2, combinando o passado da série no instante ( $t-1$ ), a variável regressora  $xm$  e as indicadoras, gerando um vetor de resíduos ruído branco gaussiano.

Em resumo, são três modelos que atendem aos objetivos do trabalho: o modelo 2, que depende apenas do passado da série  $yp$  para acompanhar a variabilidade da série e fazer previsões; o modelo 6, que depende da variável  $xo$  e  $xm$  para acompanhar o passado da série; e o modelo 5, que depende da variável  $xm$ , como opção para acompanhar a variabilidade da série.

Finalmente, pode-se concluir que este estudo obteve êxito em suas análises e se mostrou relevante para os objetivos propostos, sendo uma importante ferramenta para avaliar a variabilidade da Série Quantidade de Vendas de Peças e apresentar uma previsão com boa acurácia para os próximos 12 meses.

## REFERÊNCIAS

- ANDRADE, Wany. Estimação de modelos ARIMA/ARIMAX e aplicação em inferência de perdas de propano. Dissertação de Mestrado - Universidade Federal do Rio Grande do Norte, Natal, 2009.
- BANZATO, Fernando. Uma abordagem para previsão da taxa de câmbio por meio de modelos de séries temporais. Dissertação de Mestrado - Universidade do Porto, Porto, 2021.
- BOLANLE, Adekanmbi; OLUWADARE, Akinyemi. ARIMA and ARIMAX stochastic models for fertility in Nigeria. *International Journal of Mathematics and Computer Applications Research (IJMCAR)*, Ogbomoso, Ado-Ekiti, v.7, n.5, p.1-20, 2017.
- BUENO, Rodrigo. *Econometria de séries temporais*. 2ª edição. São Paulo: Editora Cengage, 2011.
- CAMELO, Henrique *et al.* Proposta para previsão de velocidade do vento através de modelagem híbrida elaborada a partir dos modelos ARIMAX e RNA. *Revista Brasileira de Meteorologia*, Natal, v.33, n.1, p.115-129, 2018.
- CRUZ, Rogério. Uso de funções de transferência e da transformada wavelet aplicadas a sinais meteorológicos da floresta nacional de Caxiuanã. Dissertação de Mestrado - Universidade Federal do Pará, Belém, 2008.
- ESPÍNDOLA, André. Inteligência competitiva e modelos de séries temporais para previsão de consumo: o estudo de uma empresa do setor metalúrgico. Dissertação de Mestrado - Universidade de Caxias do Sul, Caxias do sul, 2013.
- FERREIRA, Pedro *et al.* *Análise de séries temporais em R: um curso introdutório*. 1ª edição. São Paulo: Editora Atlas, FGV IBRE, 2020.
- GUJARATI, Damodar; PORTER, Dawn. *Econometria básica*. 5ª edição. Porto Alegre: Editora Bookman, 2011.
- HYNDMAN, Rob; ATHANASOPOULOS, George. *Forecasting: principles and practice*. 3rd edition. Melbourne: OTexts, 2021.
- HYNDMAN, Rob. The ARIMAX model muddle. 2010. Disponível em: <https://robjhyndman.com/hyndsight/arimax>. Acesso em: 10 out. 2022.
- MORETTIN, Pedro; TOLOI, Clélia. *Análise de séries temporais. Modelos lineares univariados*. 3ª edição. São Paulo: Editora Blucher, 2018.
- NIELSEN, Aileen. *Análise prática de séries temporais: Predição com estatística e aprendizado de máquina*. 1ª edição. Rio de Janeiro: Editora Alta Books, 2021.

- PETER, Ďurka; SILVIA, Pastoreková. ARIMA vs. ARIMAX – which approach is better to analyze and forecast macroeconomic time series? Proceedings of 30th International Conference Mathematical Methods in Economics, Karviná, 2012.
- SANTOS, José. Modelação de séries temporais em R. Dissertação de Mestrado - Universidade de Évora, Évora, 2011.
- TOSCANO, Ela. Análise de séries temporais. Belo Horizonte: Apostila do curso de Especialização em Estatística da UFMG, 2021.
- VALER, Leila. Modelo matemático ARIMAX de um propulsor eletromecânico utilizado em naves do tipo multirrotor. Dissertação de Mestrado - Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí, 2016.