

**UNIVERSIDADE ABERTA**

Departamento de Ciências e Tecnologia



MODELO DE PREVISÃO: UMA APLICAÇÃO A UMA INDÚSTRIA DE  
CALÇADO

Nadine Miranda Laranjeira

Mestrado em Bioestatística e Biometria

2025

**UNIVERSIDADE ABERTA**

Departamento de Ciências e Tecnologia



MODELO DE PREVISÃO: UMA APLICAÇÃO A UMA INDÚSTRIA DE  
CALÇADO

Nadine Miranda Laranjeira

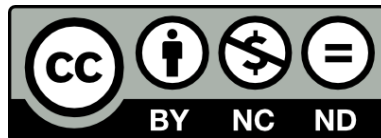
Mestrado em Bioestatística e Biometria

Dissertação de mestrado orientada pela Prof. Doutora Maria do Rosário Olaia  
Duarte Ramos

maio 2025

## DECLARAÇÃO RELATIVA ÀS CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

MODELO DE PREVISÃO: UMA APLICAÇÃO A UMA INDÚSTRIA DE CALÇADO © 2025 by Nadine Miranda Laranjeira is licensed under [Attribution-NonCommercial-NoDerivs 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/)



## Agradecimentos

Agradeço à Instituição Universidade Aberta pela oportunidade e especialmente à orientadora professora Doutora Maria do Rosário Olaia Duarte Ramos pelo seu apoio em todas as fases deste projeto e por todas as sugestões de melhoria que me permitiram alcançar os resultados que aqui apresento.

Agradeço também à empresa ALOFT, que gentilmente cedeu os dados utilizados nesta pesquisa. Sem a sua colaboração não seria possível aprofundar a análise e os resultados que constituem o cerne deste estudo.

Por fim, o meu agradecimento estende-se a todos que, de alguma forma, estiveram envolvidos ao longo dos dois anos de mestrado, oferecendo-me apoio e palavras de incentivo.

A todos, o meu muito obrigado.

## Declaração de Integridade

### Statement of Integrity

Declaro ter atuado com integridade na elaboração da presente dissertação/tese. Confirmando que em todo o trabalho conducente à sua elaboração não recorri à prática de plágio ou a qualquer outra forma de falsificação de resultados.

Mais declaro que tomei conhecimento integral do Regulamento Disciplinar da Universidade Aberta, publicado no Diário da República, 2.ª série, n.º 215, de 6 de novembro de 2013.

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration.

I further declare that I have fully acknowledged Disciplinary Regulations of the Universidade Aberta (regulation published in the official journal Diário da República, 2.ª série, N.º 215, de 6 de novembro de 2013).

Universidade Aberta, 28 de novembro de 2024

Nome completo/Full name: Nadine Miranda Laranjeira

Assinatura/Signature:

Nadine Miranda Laranjeira

---

manuscrita ou digital / handwritten or digital

## Resumo

Com o aumento da faturação, surge a necessidade das empresas possuírem ferramentas de análise de dados que lhes forneçam informação útil, para o apoio do processo de tomada de decisão. Neste âmbito, os modelos de previsão são importantes para gerir o planeamento, controlar o stock e antever a faturação.

O presente trabalho visa a aplicação de modelos de previsão e regressão para séries temporais a uma indústria de calçado, criando para tal um modelo de previsão das vendas de maneira a controlar fatores produtivos. Os modelos explorados têm por base dados históricos das vendas (variável de resposta) bem como as previsões mensais que o cliente partilha no início de cada ano (variável exógena). Considerando que estes dados podem apresentar tendência e efeito sazonal, foram aplicados modelos autorregressivo integrado e médias móveis (ARIMA) e de regressão dinâmica linear.

De modo a garantir os pressupostos para avaliação da qualidade dos modelos, foram aplicadas a transformação Box-Cox e a diferenciação sazonal.

Os resultados demonstraram que a combinação de técnicas aumenta significativamente a precisão preditiva e a robustez dos modelos para séries temporais.

Foram estudados modelos de previsão para 5 artigos de calçado produzidos pela empresa, tendo sido o modelo de regressão dinâmica o que se mostrou mais adequado para previsão em quatro dos artigos da produção e o modelo ARIMA para um dos artigos.

**Palavras-chave:** Modelos ARIMA; Previsão; Regressão Dinâmica; Séries Temporais de Vendas

## Abstract

With the increase in revenue, companies face the need for data analysis tools that provide useful information to support the decision-making process. In this context, forecasting models are essential for managing planning, controlling inventory, and predicting revenue.

This study aims to apply forecasting and regression models for time series to a footwear industry, developing a forecasting model for revenue to manage production factors. The models explored are based on historical sales data (response variable) as well as monthly forecasts shared by the client at the beginning of each year (exogenous variable). Considering that this data may exhibit trends and seasonal effects, autoregressive integrated moving average (ARIMA) models and dynamic linear regression models were applied.

To ensure the assumptions required for evaluating the quality of the models, the Box-Cox transformation and seasonal differencing were applied.

The results demonstrated that combining techniques significantly improves predictive accuracy and the robustness of time series models.

Forecasting models were studied for five footwear products manufactured by the company, with the dynamic regression model proving to be the most suitable for forecasting four of the products, while the ARIMA model was better suited for one of the products.

**Keywords:** ARIMA Models; Dynamic Regression; Forecasting; Sales Time Series

# Índice

Agradecimentos .....	I
Declaração de Integridade .....	II
Resumo.....	III
Abstract.....	IV
Índice de Gráficos .....	VIII
Índice de Tabelas .....	XI
Índice de Figuras.....	XII
Lista de Abreviaturas.....	XV
1. Introdução .....	1
2. Revisão da Literatura.....	3
3. Enquadramento Teórico: Séries Temporais .....	6
3.1. Representação de uma Série Temporal.....	6
3.2. Propriedades de uma Série Temporal.....	8
3.2.1. Estacionaridade .....	8
3.2.1.1. Transformação para a estacionaridade: Transformação Box-Cox	9
3.2.1.2. Diferenciação Sazonal.....	9
3.2.1.3. Teste ADF (Dickey-Fuller Aumentado).....	10
3.2.2. Função de Autocorrelação e Função de Autocorrelação Parcial .....	10
3.3. Modelação de Séries Temporais.....	11
3.3.1. Modelo Autorregressivo (AR).....	11
3.3.2. Modelo de Médias Móveis (MA) .....	12
3.3.3. Modelo Autorregressivo e de Médias Móveis (ARMA).....	12
3.3.4. Modelo Autorregressivo Integrado e de Médias Móveis (ARIMA)....	13

3.3.5.	Modelo Autorregressivo e de Médias Móveis Integrado Sazonal (SARIMA).....	14
3.3.6.	Modelo de Regressão Dinâmica.....	15
3.3.6.1.	Modelo Autorregressivo Integrado e de Médias Móveis com Variáveis Exógenas (ARIMAX).....	15
3.3.6.2.	Modelo Autorregressivo e de Médias Móveis Integrado Sazonal com Variáveis Exógenas (SARIMAX).....	16
3.4.	Medidas de Desempenho.....	17
3.4.1.	Erro Absoluto Médio – MAE.....	17
3.4.2.	Erro Quadrático Médio – MSE.....	17
3.4.3.	Raiz Quadrada do Erro Quadrático Médio – RMSE.....	18
3.4.4.	Erro Absoluto Médio Percentual – MAPE.....	18
3.4.5.	Erro Absoluto Médio Escalado – MASE.....	19
4.	Análise das Séries da Produção e Vendas de Calçado.....	20
4.1.	A Empresa ALOFT.....	20
4.2.	Dados e Métodos.....	21
4.3.	Estudo do artigo 8B001.....	27
4.3.1.	Análise exploratória do artigo 8B001.....	27
4.3.2.	Decomposição e Representação das Séries Temporais.....	29
4.3.3.	Modelos de Previsão.....	32
4.3.4.	Análise dos Resíduos do Modelo de Previsão Selecionado.....	36
4.3.5.	Previsão do Modelo Ajustado.....	38
4.4.	Estudo do artigo 8B007.....	41
4.4.1.	Análise exploratória do artigo 8B007.....	41
4.4.2.	Decomposição e Representação Gráfica das Séries Temporais.....	41
4.4.3.	Modelos de Previsão.....	45
4.4.4.	Análise dos Resíduos do Modelo de Previsão Selecionado.....	49

4.4.5.	Previsão do Modelo Ajustado .....	51
4.5.	Estudo do artigo 8B008 .....	53
4.5.1.	Análise exploratória do artigo 8B008 .....	53
4.5.2.	Decomposição e Representação Gráfica das Séries Temporais .....	53
4.5.3.	Modelos de Previsão .....	56
4.5.4.	Análise dos Resíduos do Modelo de Previsão Selecionado .....	61
4.5.5.	Previsão do Modelo Ajustado .....	62
4.6.	Estudo do artigo 8B010 .....	65
4.6.1.	Análise exploratória do artigo 8B010 .....	65
4.6.2.	Decomposição e Representação Gráfica das Séries Temporais .....	65
4.6.3.	Modelos de Previsão .....	68
4.6.4.	Análise dos Resíduos do Modelo de Previsão Selecionado .....	73
4.6.5.	Previsão do Modelo Ajustado .....	75
4.7.	Estudo do artigo 8B014 .....	78
4.7.1.	Análise exploratória do artigo 8B014 .....	78
4.7.2.	Decomposição e Representação Gráfica das Séries Temporais .....	78
4.7.3.	Modelos de Previsão .....	81
4.7.4.	Análise dos Resíduos do Modelo de Previsão Selecionado .....	86
4.7.5.	Previsão do Modelo Ajustado .....	87
5.	Conclusão .....	90
	Bibliografia .....	92
	Anexos e Apêndices.....	94

## Índice de Gráficos

Gráfico 3.1 - Representação gráfica de exemplos de Decomposição Multiplicativa (gráfico superior) e Decomposição aditiva (gráfico inferior). (Bauer, 2020) .....	8
Gráfico 4.1 - Distribuição e decomposição da série temporal original relativa ao número de pares vendidos do artigo 8B001, num período mensal de 2019 a 2023. ....	28
Gráfico 4.2 – Box-plot das vendas reais mensais do artigo 8B001, num período de 2019 a 2023. ....	28
Gráfico 4.3 – Distribuição anual do número de pares vendidos do artigo 8B001, após alteração dos valores dos meses abril e maio de 2020 pela mediana, num período mensal de 2019 a 2023.....	28
Gráfico 4.4 – Distribuição e decomposição da série temporal relativa ao número de pares vendidos do artigo 8B001, num período mensal de 2019 a 2023. ....	29
Gráfico 4.5 – Distribuição anual do número de pares vendidos do artigo 8B001, após alteração dos valores dos meses abril e maio de 2020 pela mediana, num período mensal de 2019 a 2023.....	30
Gráfico 4.6 – Distribuição mensal de 2019 a 2023 dos pares previstos e vendidos para o artigo 8B001.....	30
Gráfico 4.7 – Correlação cruzada das vendas e previsões do cliente para o artigo 8B001, num período de 2019 a 2023.....	31
Gráfico 4.8 – Box-plot das vendas mensais do artigo 8B001, num período de 2019 a 2023. ....	32
Gráfico 4.9 – Gráficos de autocorrelação e autocorrelação parcial das vendas no artigo 8B001.....	33
Gráfico 4.10 - Box-plot das vendas após a aplicação da diferenciação sazonal para o artigo 8B001.....	34
Gráfico 4.11 – Gráficos de autocorrelação e autocorrelação parcial da série temporal ‘vendas’ ajustada do artigo 8B001. ....	35
Gráfico 4.12 – Análise dos resíduos do modelo de previsão ajustado para o artigo 8B001.....	37
Gráfico 4.13 – Representação gráfica do modelo de previsões para o artigo 8B001. ....	39

Gráfico 4.14 – Distribuição e decomposição da série temporal relativa ao número de pares vendidos do artigo 8B007, num período mensal de 2019 a 2023. ....	42
Gráfico 4.15 – Distribuição anual do número de pares vendidos do artigo 8B007, num período mensal de 2019 a 2023.....	43
Gráfico 4.16 – Distribuição mensal de 2019 a 2023 dos pares previstos e vendidos para o artigo 8B007.....	43
Gráfico 4.17 – Correlação cruzada das vendas e previsões do cliente para o artigo 8B007, num período de 2019 a 2023.....	44
Gráfico 4.18 – Box-plot das vendas mensais do artigo 8B007, num período de 2019 a 2023. ....	45
Gráfico 4.19 – Gráficos de autocorrelação e autocorrelação parcial das vendas no artigo 8B007.....	46
Gráfico 4.20 – Box-plot das vendas após a aplicação da transformação Box-Cox e da diferenciação sazonal para o artigo 8B007. ....	47
Gráfico 4.21 – Gráficos de autocorrelação e autocorrelação parcial da série temporal ‘vendas’ ajustada do artigo 8B007. ....	48
Gráfico 4.22 – Análise dos resíduos do modelo de previsão ajustado para o artigo 8B007.....	49
Gráfico 4.23 – Representação gráfica do modelo de previsões para o artigo 8B007. ....	51
Gráfico 4.24 – Distribuição e decomposição da série temporal relativa ao número de pares vendidos do artigo 8B008, num período mensal de 2019 a 2023. ....	54
Gráfico 4.25 – Distribuição anual do número de pares vendidos do artigo 8B008, num período mensal de 2019 a 2023.....	54
Gráfico 4.26 – Distribuição mensal de 2019 a 2023 dos pares previstos e vendidos para o artigo 8B008.....	55
Gráfico 4.27 – Correlação cruzada das vendas e previsões do cliente para o artigo 8B008, num período de 2019 a 2023.....	56
Gráfico 4.28 – Box-plot das vendas mensais do artigo 8B008, num período de 2019 a 2023. ....	56
Gráfico 4.29 – Gráficos de autocorrelação e autocorrelação parcial das vendas no artigo 8B008.....	57

Gráfico 4.30 – Box-plot das vendas após a aplicação da transformação Box-Cox e da diferenciação sazonal para o artigo 8B008. ....	58
Gráfico 4.31 – Gráficos de autocorrelação e autocorrelação parcial da série temporal ‘vendas’ ajustada do artigo 8B008. ....	59
Gráfico 4.32 – Análise dos resíduos do modelo de previsão ajustado para o artigo 8B008.....	61
Gráfico 4.33 – Representação gráfica do modelo de previsões para o artigo 8B008 sem e com intervalo de confiança.....	63
Gráfico 4.34 – Distribuição e decomposição da série temporal relativa ao número de pares vendidos do artigo 8B010, num período mensal de 2019 a 2023. ....	66
Gráfico 4.35 – Distribuição anual do número de pares vendidos do artigo 8B010, num período mensal de 2019 a 2023.....	66
Gráfico 4.36 – Distribuição mensal de 2019 a 2023 dos pares previstos e vendidos para o artigo 8B010.....	67
Gráfico 4.37 – Correlação cruzada das vendas e previsões do cliente para o artigo 8B010, num período de 2019 a 2023.....	68
Gráfico 4.38 – Box-plot das vendas mensais do artigo 8B010, num período de 2019 a 2023. ....	68
Gráfico 4.39 – Gráficos de autocorrelação e autocorrelação parcial das vendas no artigo 8B010.....	69
Gráfico 4.40 – Box-plot das vendas após a aplicação da transformação Box-Cox e da diferenciação sazonal para o artigo 8B010. ....	71
Gráfico 4.41 – Gráficos de autocorrelação e autocorrelação parcial da série temporal ‘vendas’ ajustada do artigo 8B010. ....	71
Gráfico 4.42 – Análise dos resíduos do modelo de previsão ajustado para o artigo 8B010.....	73
Gráfico 4.43 – Representação gráfica do modelo de previsões para o artigo 8B010 sem e sem intervalo de confiança.....	76
Gráfico 4.44 – Distribuição e decomposição da série temporal relativa ao número de pares vendidos do artigo 8B014, num período mensal de 2019 a 2023. ....	79
Gráfico 4.45 – Distribuição anual do número de pares vendidos do artigo 8B014, num período mensal de 2019 a 2023.....	79

Gráfico 4.46 – Distribuição mensal de 2019 a 2023 dos pares previstos e vendidos para o artigo 8B014.....	80
Gráfico 4.47 – Correlação cruzada das vendas e previsões do cliente para o artigo 8B014, num período de 2019 a 2023.....	81
Gráfico 4.48 – Box-plot das vendas mensais do artigo 8B014, num período de 2019 a 2023.....	81
Gráfico 4.49 – Gráficos de autocorrelação e autocorrelação parcial das vendas no artigo 8B014.....	82
Gráfico 4.50 – Box-plot das vendas após a aplicação da transformação Box-Cox e da diferenciação sazonal para o artigo 8B014.....	83
Gráfico 4.51 – Gráficos de autocorrelação e autocorrelação parcial da série temporal ‘vendas’ ajustada do artigo 8B014. ....	84
Gráfico 4.52 – Análise dos resíduos do modelo de previsão ajustado para o artigo 8B014.....	86
Gráfico 4.53 – Representação gráfica do modelo de previsões para o artigo 8B014.....	88

## Índice de Tabelas

Tabela 4.1 - Comparação dos valores calculados pela função ‘auto.arima’ do R para os quatro modelos aplicados no artigo 8B001. ....	36
Tabela 4.2 - Comparação dos valores calculados pela função ‘auto.arima’ do R para os quatro modelos aplicados no artigo 8B007. ....	49
Tabela 4.3 – Comparação dos valores calculados pela função ‘auto.arima’ do R para os quatro modelos aplicados no artigo 8B008. ....	60
Tabela 4.4 - Comparação dos valores calculados pela função ‘auto.arima’ do R para os dois modelos aplicados no artigo 8B010.....	73
Tabela 4.5 – Comparação dos valores calculados pela função ‘auto.arima’ do R para os quatro modelos aplicados no artigo 8B014. ....	85

## Índice de Figuras

Figura 4.1 - Estatística descritiva da série temporal dos pares vendidos, para o artigo 8B001.....	27
Figura 4.2 – Estatística descritiva da série temporal ajustada dos pares vendidos, para o artigo 8B001.....	29
Figura 4.3 - Correlação de Pearson entre as séries sazonais de previsão do cliente e vendas, para o artigo 8B001. ....	31
Figura 4.4 – Resultado da aplicação do teste de estacionaridade aplicado no artigo 8B001.....	32
Figura 4.5 – Resultado da aplicação do modelo A na série temporal ‘vendas’ do artigo 8B001.....	33
Figura 4.6 – Resultado da aplicação do modelo B na série temporal ‘vendas’ do artigo 8B001 com a inclusão da variável exógena ‘previsões’ do cliente. ....	33
Figura 4.7 – Resultado da aplicação do teste de estacionaridade aplicado à série temporal ‘vendas’ ajustada pela função ‘decompose’ no artigo 8B001. ....	34
Figura 4.8 – Resultado da aplicação do modelo D na série temporal ajustada ‘vendas’ do artigo 8B001.....	35
Figura 4.9 – Resultado da aplicação do modelo C na série temporal ajustada ‘vendas’ do artigo 8B001 com a inclusão da variável exógena ‘previsões’ do cliente. ....	35
Figura 4.10 – Estatística descritiva da série temporal dos pares vendidos, para o artigo 8B007.....	41
Figura 4.11 - Correlação de Pearson entre as séries sazonais de previsão do cliente e vendas, para o artigo 8B007. ....	44
Figura 4.12 – Resultado da aplicação do teste de estacionaridade aplicado no artigo 8B007.....	45
Figura 4.13 – Resultado da aplicação do modelo A na série temporal ‘vendas’ do artigo 8B007.....	46
Figura 4.14 – Resultado da aplicação do modelo B na série temporal ‘vendas’ do artigo 8B007 com a inclusão da variável exógena ‘previsões’ do cliente. ....	46
Figura 4.15 – Resultado da aplicação do teste de estacionaridade aplicado à série temporal ‘vendas’ ajustada pela função ‘decompose’ no artigo 8B007. ....	47

Figura 4.16 – Resultado da aplicação do modelo D na série temporal ajustava ‘vendas’ do artigo 8B007.....	48
Figura 4.17 – Resultado da aplicação do modelo C na série temporal ajustada ‘vendas’ do artigo 8B007 com a inclusão da variável exógena ‘previsões’ do cliente. ....	48
Figura 4.18 – Estatística descritiva da série temporal dos pares vendidos, para o artigo 8B008.....	53
Figura 4.19 - Correlação de Pearson entre as séries sazonais de previsão do cliente e vendas, para o artigo 8B008. ....	55
Figura 4.20 – Resultado da aplicação do teste de estacionaridade aplicado no artigo 8B008.....	57
Figura 4.21 – Resultado da aplicação do modelo A na série temporal ‘vendas’ do artigo 8B008.....	57
Figura 4.22 – Resultado da aplicação do modelo B na série temporal ‘vendas’ do artigo 8B008 com a inclusão da variável exógena ‘previsões’ do cliente. ....	58
Figura 4.23 – Resultado da aplicação do teste de estacionaridade aplicado à série temporal ‘vendas’ ajustada pela função ‘decompose’ no artigo 8B008. ....	59
Figura 4.24 – Resultado da aplicação do modelo D na série temporal ajustava ‘vendas’ do artigo 8B008.....	59
Figura 4.25 – Resultado da aplicação do modelo C na série temporal ajustada ‘vendas’ do artigo 8B008 com a inclusão da variável exógena ‘previsões’ do cliente. ....	60
Figura 4.26 – Estatística descritiva da série temporal dos pares vendidos, para o artigo 8B010.....	65
Figura 4.27 - Correlação de Pearson entre as séries sazonais de previsão do cliente e vendas, para o artigo 8B010. ....	67
Figura 4.28 – Resultado da aplicação do teste de estacionaridade aplicado no artigo 8B010.....	69
Figura 4.29 – Resultado da aplicação do modelo A na série temporal ‘vendas’ do artigo 8B010.....	69
Figura 4.30 – Resultado da aplicação do modelo B na série temporal ‘vendas’ do artigo 8B010 com a inclusão da variável exógena ‘previsões’ do cliente. ....	70

Figura 4.31 – Resultado da aplicação do teste de estacionaridade aplicado à série temporal ‘vendas’ ajustada pela função ‘decompose’ no artigo 8B010. ....	71
Figura 4.32 – Resultado da aplicação do modelo D na série temporal ajustada ‘vendas’ do artigo 8B010. ....	72
Figura 4.33 – Resultado da aplicação do modelo C na série temporal ajustada ‘vendas’ do artigo 8B010 com a inclusão da variável exógena ‘previsões’ do cliente. ....	72
Figura 4.34 – Estatística descritiva da série temporal dos pares vendidos, para o artigo 8B014. ....	78
Figura 4.35 - Correlação de Pearson entre as séries sazonais de previsão do cliente e vendas, para o artigo 8B014. ....	80
Figura 4.36 – Resultado da aplicação do teste de estacionaridade aplicado no artigo 8B014. ....	82
Figura 4.37 – Resultado da aplicação do modelo A na série temporal ‘vendas’ do artigo 8B014. ....	82
Figura 4.38 – Resultado da aplicação do modelo B na série temporal ‘vendas’ do artigo 8B014 com a inclusão da variável exógena ‘previsões’ do cliente. ....	83
Figura 4.39 – Resultado da aplicação do teste de estacionaridade aplicado à série temporal ‘vendas’ ajustada pela função ‘decompose’ no artigo 8B014. ....	84
Figura 4.40 – Resultado da aplicação do modelo D na série temporal ajustada ‘vendas’ do artigo 8B014. ....	84
Figura 4.41 – Resultado da aplicação do modelo C na série temporal ajustada ‘vendas’ do artigo 8B014 com a inclusão da variável exógena ‘previsões’ do cliente. ....	85

## Lista de Abreviaturas

ACF	Função de Autocorrelação
AIC	<i>Akaike's Information Criterion</i>
AICc	<i>Corrected Akaike Information Criterion</i>
AR	Modelo Autorregressivo
ARIMA	Modelo Autorregressivo Integrado e de Médias Móveis
ARIMAX	Modelo Autorregressivo Integrado e de Médias Móveis com Variáveis Exógenas
ARMA	Modelo Autorregressivo e de Médias Móveis
BIC	<i>Bayesian Information Criterion</i>
<i>Lag</i>	Desfasagem/Período
MA	Modelo de Médias Móveis
MAE	Erro Absoluto Médio
MAPE	Erro Absoluto Médio Percentual
MASE	Erro Absoluto Médio Escalado
MSE	Erro Quadrático Médio
PACF	Função de Autocorrelação Parcial
RMSE	Raiz Quadrada do Erro Quadrático Médio
SARIMA	Modelo Autorregressivo e de Médias Móveis Integrado Sazonal
SARIMAX	Modelo Autorregressivo e de Médias Móveis Integrado Sazonal com Variáveis Exógenas

# 1. Introdução

O crescimento económico do setor secundário contribuiu para que as empresas se tornassem mais competitivas. Para alcançar o caminho para o sucesso, há a necessidade de recorrer a previsões, pois quanto melhor o desempenho destas, melhor será realizada a gestão empresarial. Deste modo, a empresa passa a dominar as necessidades do cliente e sabe como atuar sobre as mesmas. Planear é a melhor forma de tentar antever o futuro, com o objetivo de melhorar o desempenho empresarial. Uma análise incorreta poderá levar a consequências muito graves para a empresa.

As previsões desempenham um papel crucial na indústria, permitindo a antecipação de produções, otimização de recursos e redução de custos, o que resulta numa maior eficiência. A previsão das vendas pode ser auxiliada pela utilização de métodos estatísticos, contribuindo assim para reduzir o desperdício de recursos e maximizar as potenciais oportunidades.

A empresa estudada nesta dissertação enfrenta vários desafios relacionados com a gestão de stocks. O seu principal desafio prende-se com a gestão de um planeamento mais perfeccionista, de modo a cumprir quantidades e datas de entrega ao cliente, bem como gerir o stock, isto é, não enfrentar ruturas ou excessos.

Em suma, prever é o ponto de partida para o processo de organização, tomada de decisão e controlo da atividade futura de qualquer organização.

Nesta dissertação propõe-se desenvolver um modelo de previsão aplicado a uma indústria de calçado, considerando cinco dos artigos da sua produção. O objetivo final é gerir a afetação da produção dos modelos de calçado em função da disponibilidade de máquinas de produção, garantindo a data de entrega do produto final ao cliente e a gestão controlada de *stock*.

Os dados analisados neste estudo consistem em cinco séries temporais relativas às cinco principais categorias de calçado vendidas pela empresa ALOFT,

mais concretamente galochas denominadas pelos seguintes códigos: 8B001, 8B007, 8B008, 8B010 e 8B014.

Numa base de dados mensal de 2019 a 2023, serão explorados alguns modelos de previsão, com o apoio do software R e do ambiente RStudio (versão 4.3.2).

Relativamente à estrutura da dissertação, após a presente Introdução, o Capítulo 2 apresenta uma revisão de literatura, que consiste em uma seleção de estudos com objetivos similares. O Capítulo 3 faz o enquadramento teórico e justificação dos métodos de previsão aplicados no Capítulo 4, que apresenta a análise efetuada sobre as séries temporais estudadas, relativas a 5 artigos de calçado. Por fim é apresentada uma síntese das conclusões mais relevantes, algumas limitações encontradas e propostas de investigação futura.

## 2. Revisão da Literatura

Vários estudos recentes sobre previsões de séries temporais têm explorado uma ampla gama de modelos estatísticos, com o objetivo de aprimorar a precisão preditiva. Apresenta-se uma seleção de estudos publicados que se enquadram na indústria do calçado e em áreas do retalho com objetivos similares.

Em De Sousa (2013), é realizado um estudo de previsão anual de vendas por solicitação da empresa Foreva, para cada uma de cinco categorias de calçado comercializado pela empresa. A autora comparou o desempenho de duas metodologias de previsão, os modelos de espaço de estado e os modelos ARIMA, baseando-se nas vendas históricas diárias. Os resultados mostraram que os modelos ARIMA apresentam melhor desempenho em comparação com os modelos de espaço de estado na previsão de séries de vendas do setor do retalho de calçado da empresa.

Patrícia Ramos, Nicolau Santos e Rui Rebelo (Ramos et al., 2014) publicaram um artigo que compara e seleciona o melhor modelo de previsões aplicado ao retalho de calçado feminino, em cinco categorias. Os autores utilizam o critério de informação de Akaike (AIC) para ajustar modelos de previsão: modelo ARIMA (após estabilização da variância das séries por transformações logarítmicas e testes de estacionaridade) e um modelo de alisamento exponencial, ETS (*Error, Trend, Seasonal*). Os resultados mostram que o desempenho da previsão dos modelos ETS e ARIMA avaliados via RMSE, MAE e MAPE é semelhante. O estudo conclui que, na ausência de uma vantagem clara entre ETS e ARIMA em termos de precisão, ambos representam abordagens complementares e eficazes para previsões de vendas no setor do retalho.

Na dissertação desenvolvida por Filipa Veríssimo (Veríssimo, 2016), foi estudado o melhor modelo de séries temporais hierárquicas aplicado ao setor do calçado. Por este ser um produto com um ciclo de vida curto e sazonal, as características de cada artigo variam conforme a tendência da moda em cada estação do ano. Uma vez que há uma grande variedade de tipologias do produto

de calçado (género(M/F), tamanho, cor, material, entre outros), a autora explorou a relação de dependência entre estes níveis e as vendas. Assim, construiu um modelo de previsão que permitiu prever as vendas de uma empresa de calçado alemã, com indústria em Portugal, para cada cliente da empresa, numa dada estação do ano. Explorou a estrutura hierárquica, de modo a criar previsões de vendas para todos os níveis, e comparou com o modelo linear simples ARIMA, sem covariáveis. A conclusão foi que o modelo ARIMA apresentou melhor desempenho na base de dados estudada comparativamente aos modelos hierárquicos.

Em 2018, Susana Lima estudou um conjunto de séries temporais do ramo do retalho, com fortes tendências e padrões sazonais, de modo a ajustar o melhor método de previsão. A autora aplicou os seguintes métodos de previsão que melhor se ajustavam à base de dados: ARMA, modelos de regressão linear múltipla e os métodos de alisamento exponencial (Holt-Winters) e, avaliou a capacidade preditiva das três metodologias através das seguintes medidas: EQM (Erro Quadrático Médio, MSE), o REQM (Raiz do Erro Quadrático Médio, RMSE), o EPAM (Erro Absoluto Médio Percentual, MAPE), o EEAM (Erro Absoluto Médio Escalado, MASE) e a estatística U de Theil. A partir deste estudo, a autora concluiu que para a base de dados em estudo, os modelos de regressão linear múltipla não são os mais adequados, enquanto os modelos SARIMA são os mais precisos. Contudo, apesar de não serem considerados os mais adequados, os modelos de Holt-Winters poderão ser uma alternativa de previsão. (Lima, 2018)

O artigo desenvolvido por Ivan Kaipov e Alexander Nedzvedaborda aborda a importância da previsão de vendas para o retalho do calçado, especialmente com o crescimento das lojas online. O objetivo dos autores foi manter um stock de segurança confortável (sem excesso ou falta de stock), de modo a cumprir com a disponibilidade de produto no momento da compra do cliente. Para tal, foram explorados métodos de previsão baseados em *machine learning* e previsões de séries temporais. Em particular, utilizaram o modelo *Prophet* (desenvolvido pela Facebook/Meta) que se mostrou o mais preciso para prever padrões de vendas tendo em conta a sazonalidade (como por exemplo estações do ano e dias da semana) e eventos específicos (feriados). Foi estudada a hipótese de aplicar o método de redes neuronais, mas o autor concluiu que este método é difícil de

implementar devido à complexidade da arquitetura da rede neuronal e ao facto de o erro de previsão aumentar substancialmente com o alargamento do horizonte da previsão. Concluindo, o modelo *Prophet* oferece um equilíbrio entre precisão e simplicidade de implementação. (Kaipov & Nedzved, 2020)

Um estudo realizado num negócio austríaco de venda de legumes analisou o desempenho da previsão dos modelos SARIMA e LSTM (*Long Short-Term Memory*) para prever as vendas diárias de quatro produtos. Os autores (Falatouria et al., 2022) selecionaram os métodos de previsão com base nas características dos dados e conhecimento de negócios para o produto relacionado. Seguidamente, com o objetivo de melhorar as previsões, usou fatores externos aditivos, como por exemplo as promoções. Considerando que as promoções podem aumentar notavelmente as vendas de vegetais, foi aplicado o método SARIMAX. Foram utilizadas as medidas MAPE e RMSE para avaliar a confiabilidade dos valores de vendas previstos. Assim, os autores puderam concluir que o modelo SARIMAX foi o mais eficaz para produtos com comportamento sazonal, como era o caso do pepino, enquanto o modelo LSTM teve um desempenho superior em produtos com vendas mais estáveis, como era o caso das batatas. O modelo SARIMAX é o mais adequado quando implica a inclusão de variáveis externas.

### 3. Enquadramento Teórico: Séries Temporais

O estudo de séries temporais envolve a análise de dados observados ao longo do tempo, onde a principal característica é a dependência temporal entre as observações. Este tipo de análise tem como objetivo identificar padrões que permitem a construção de modelos preditivos para dados futuros.

Define-se como série temporal a sequência cronológica de um conjunto de observações de uma variável de interesse.

Caracteriza-se a série temporal como sendo contínua ou discreta. Quando o valor da variável é registado continuamente no tempo, a série temporal é contínua. Quando o valor da variável é registado em instantes específicos e separados, geralmente com intervalos de tempo regulares (como dias, meses, trimestres), diz-se que a série temporal é discreta.

Pode-se também definir as séries temporais como sendo univariadas ou multivariadas, se forem constituídas por observações de apenas uma variável, ou várias variáveis, respetivamente. (Lima, 2018)

Uma série temporal é uma realização de um processo estocástico, com determinadas propriedades que serão apresentadas nas secções seguintes.

#### 3.1. Representação de uma Série Temporal

Para representar e analisar uma série temporal, uma metodologia clássica consiste em obter a decomposição da série em componentes com diferentes significados. Assim, as séries temporais podem ser decompostas em quatro componentes: tendência, sazonalidade, efeito cíclico e uma componente de erro aleatório, um processo aleatório estacionário, identificado como ruído. (Cunha, 2022)

A tendência ( $T_t$ ) descreve o sentido do deslocamento (inclinação crescente ou decrescente) ou do comportamento (mudança de nível) da série temporal. A tendência varia em função do tempo  $t$ , ou seja,  $T_t = f(t)$ . Numa série com tendência, a média  $E[Y_t] = T_t$  não é constante, ou seja, a série não é estacionária.

A sazonalidade ( $S_t$ ) corresponde a padrões repetitivos numa série temporal, ou seja, a flutuações periódicas da variável com periodicidade constante.

O efeito cíclico ( $C_t$ ) corresponde a padrões de longo prazo que, ao contrário da sazonalidade, não têm um período fixo.

O termo de ruído ( $\varepsilon_t$ ) corresponde às flutuações aleatórias que não seguem nenhum padrão e estão presentes nos dados ao longo do tempo. É, assim, o resultado de variações aleatórias, erros de medição ou fatores não observados que afetam o comportamento da série.

A decomposição de uma série temporal divide as componentes principais da série, facilitando a análise e a modelação. Os dois métodos mais comuns para decomposição são a aditiva e a multiplicativa.

A decomposição aditiva assume que os efeitos de tendência ( $T_t$ ), sazonalidade ( $S_t$ ) e erro aleatório ( $\varepsilon_t$ ) são aditivos e por isso a série temporal ( $Y_t$ ) é modelada pela soma direta das suas componentes:

$$Y_t = T_t + S_t + \varepsilon_t \quad (1)$$

Esta decomposição é aplicada quando a amplitude das flutuações sazonais não varia com o nível da série. Ou seja, quando as flutuações sazonais são constantes em torno de uma tendência de crescimento linear. No

Gráfico 3.1 (inferior) é possível observar um exemplo.

Já a decomposição multiplicativa assume que os efeitos de tendência ( $T_t$ ), sazonalidade ( $S_t$ ) e de erro ( $\varepsilon_t$ ) são multiplicativos e por isso a série temporal ( $Y_t$ ) é modelada pelo produto das suas componentes:

$$Y_t = T_t \times S_t \times \varepsilon_t \quad (2)$$

Esta decomposição é aplicada quando a amplitude das flutuações sazonais varia com o nível da série. No

Gráfico 3.1 (superior) é possível observar um exemplo.

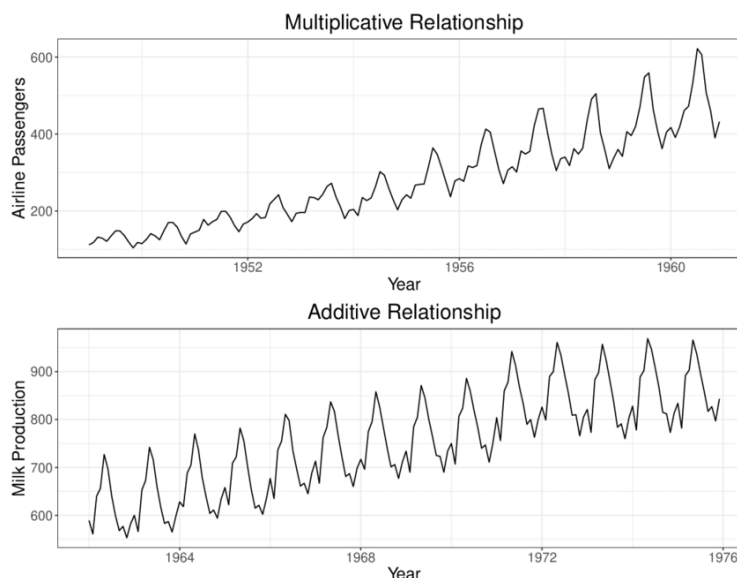


Gráfico 3.1 - Representação gráfica de exemplos de Decomposição Multiplicativa (gráfico superior) e Decomposição aditiva (gráfico inferior). (Bauer, 2020)

### 3.2. Propriedades de uma Série Temporal

As propriedades estudadas relativamente a série temporal incluem a estacionaridade, anteriormente mencionada, que garante que as características estatísticas, como média e variância, se mantêm constantes ao longo do tempo, e a autocorrelação, que mede o grau de correlação entre os valores da série em diferentes intervalos de tempo, sendo ambas fundamentais para a escolha adequada de modelos de previsão.

#### 3.2.1. Estacionaridade

Garantir que uma série temporal é estacionária é um passo importante para aplicar modelos de previsão de séries temporais com eficácia. Assim, um processo estocástico apresenta estacionaridade fraca (ou de segunda ordem), quando a média ( $E[y(t)]$ ) e a variância ( $Var[y(t)]$ ) são constantes ao longo do tempo e a covariância ( $\gamma(k)$ ) é dependente apenas do *lag* (desfasamento)  $k$ .

Um processo estocástico  $y(t)$  é considerado de estacionaridade fraca se:

- $E[y(t)] = \mu$ ;
- $Var[y(t)] = E[y(t) - \mu]^2 = \sigma^2$ ;
- $\gamma(k) = f(|k|) = cov[y(t), y(t + k)] = E\{[y(t) - \mu][y(t + k) - \mu]\}$

Diz-se que a estacionaridade pode ser forte (ou estrita), quando todas as propriedades estatísticas de uma série, em qualquer ponto no tempo, são constantes, ou seja, são invariantes no tempo.

#### 3.2.1.1. Transformação para a estacionaridade: Transformação Box-Cox

Vários métodos de modelação de séries temporais têm como pressuposto a estacionaridade da série, pelo menos a estacionaridade de segunda ordem. Contudo nas séries reais este pressuposto é frequentemente violado. O método de transformação Box-Cox é uma das principais técnicas para tornar uma série temporal estacionária ou quase estacionária em termos da sua variância. Desenvolvida pelos estatísticos George Box e Sir David Cox, esta transformação envolve elevar os dados a uma potência variável, representada como lambda ( $\lambda$ ). A escolha de lambda depende das características dos dados, e a transformação é eficaz para lidar com problemas como heterocedasticidade e não normalidade. (Hyndman & Athanasopoulos; George, 2018)

A transformação de Box-Cox é dada pela seguinte fórmula:

$$w_t = \begin{cases} \log(y_t), & \text{se } \lambda = 0 \\ \frac{y_t^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0 \end{cases}, \text{ onde:} \quad (3)$$

- $w_t$  corresponde à variável transformada;
- $y_t$  corresponde à variável em estudo;
- $\lambda$  corresponde ao parâmetro de transformação que determina a forma da transformação.

#### 3.2.1.2. Diferenciação Sazonal

Nas séries temporais sazonais poderá não haver estacionaridade, uma vez que há flutuações ou padrões que se repetem em intervalos fixos que provocam uma variação na média e variância ao longo do tempo. Uma técnica comum para lidar com a não estacionaridade sazonal é aplicar a diferenciação sazonal, ou seja, calcular a diferença da série em intervalos sazonais sendo assim a sazonalidade removida e tornando-se conseqüentemente uma série mais estacionária. (Hyndman & Athanasopoulos; George, 2018)

A diferenciação sazonal é calculada a partir da seguinte equação:

$$Y'_t = Y_t - Y_{t-1}, \text{ onde:} \quad (4)$$

- $Y'_t$  corresponde ao valor diferenciado;
- $Y_t$  corresponde ao valor da série temporal no tempo.

### 3.2.1.3. Teste ADF (Dickey-Fuller Aumentado)

A partir do teste de Dickey-Fuller Aumentado é possível concluir se a série é estacionária (se apresenta raiz unitária) a partir das seguintes hipóteses:

$H_0$ : A série temporal não é estacionária;

$H_1$ : A série temporal é estacionária.

Baseados na distribuição da estatística de teste de Dickey-Fuller Aumentado, é calculada a significância do teste, valor-p (uso comum do termo inglês p-value). Se esta probabilidade for inferior ao nível de significância fixado (por exemplo  $\alpha = 0,05$ ), então pode-se rejeitar a hipótese nula e concluir que a série temporal é estacionária (Lopes, 2019).

### 3.2.2. Função de Autocorrelação e Função de Autocorrelação Parcial

A função de autocorrelação e autocorrelação parcial são utilizadas como indicador estatístico das características dos dados e determinam o número de defasagens Autorregressivas (AR) e Médias Móveis (MA) para os modelos ARIMA ajudando também a detetar a presença de sazonalidade nos dados.

A função de autocorrelação (ACF-acrônimo do termo em inglês *Autocorrelation Function*), determina a correlação linear que uma série temporal apresenta com ela própria, desfasada em k períodos. Assim, a função ACF é definida por (Vieira, 2022):

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{cov(Y_t, Y_{t+k})}{V(Y_t)}, k = 1, \dots, n - 1 \quad (5)$$

onde  $Y_t$  corresponde ao valor observado no instante t.

Esta função é estimada pela função de correlação empírica, tomando a expressão:

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=k+1}^n (Y_t - \bar{Y})^2}, k = 1, \dots, n - 1 \quad (6)$$

Onde,  $Y_t$  corresponde ao valor da série no instante  $t$ ;  $\bar{Y}$  é a média da série;  $n$  corresponde ao número total de observações; e  $k$  corresponde ao *lag*.

A função de autocorrelação parcial (PACF, do inglês *Partial Autocorrelation Function*), mede a relação entre duas observações na série, eliminando a dependência dos termos intermédios. A função PACF é estimada por (Vieira, 2022):

$$\hat{\rho}_{XY.Z} = \frac{N \sum_{i=1}^N e_{X,i} e_{Y,i} - \sum_{i=1}^N e_{X,i} \sum_{i=1}^N e_{Y,i}}{\sqrt{N \sum_{i=1}^N e_{X,i}^2 - (\sum_{i=1}^N e_{X,i})^2} \sqrt{N \sum_{i=1}^N e_{Y,i}^2 - (\sum_{i=1}^N e_{Y,i})^2}} \quad (7)$$

Onde,  $e_{X,i}$  e  $e_{Y,i}$  correspondem aos resíduos das regressões de X e Y sobre o efeito Z, e N corresponde ao número de observações.

### 3.3. Modelação de Séries Temporais

A modelação de séries temporais desenvolve modelos matemáticos que descrevem as dependências temporais dos dados. Tem o objetivo de capturar padrões presentes na série, como: tendências, sazonalidade, ciclos e ruídos, de modo a prever o futuro. Nesta seção serão explorados os principais métodos utilizados para modelar séries temporais. (Hyndman & Athanasopoulos; George, 2018)

#### 3.3.1. Modelo Autorregressivo (AR)

Diz-se que  $\{Y_t\}$  satisfaz um modelo autorregressivo de ordem  $p$ :  $AR(p)$  ou  $ARIMA(p, 0, 0)$  se,

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t, \text{ onde:}$$

- $Y_t$  é a série temporal estacionária observada no tempo;
- $c$  corresponde à ordenada na origem;
- $\phi_1, \dots, \phi_p$  são os parâmetros autorregressivos não sazonais, sendo  $\phi_p \neq 0$ ;

- $\varepsilon_t$  é o ruído branco com média zero e variância constante. Este:
  - Não possui informação útil;
  - Não está correlacionado;
  - É um bom erro aleatório, como esperado em modelos bem ajustados.

Considerando 'B' como o operador de defasagem (*lag*) e definindo que  $B^p Y_t = Y_{t-p}$ , então a anterior equação é:

$$Y_t = c + \phi_1 B^1 Y_t + \phi_2 B^2 Y_t + \dots + \phi_p B^p Y_t + \varepsilon_t \quad (7)$$

### 3.3.2. Modelo de Médias Móveis (MA)

Diz-se que  $\{Y_t\}$  satisfaz um modelo de médias móveis de ordem  $q$ :  $MA(q)$  ou  $ARIMA(0, 0, q)$  se,

$$Y_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \text{ onde:}$$

- $Y_t$  é a série temporal estacionária;
- $c$  corresponde à ordenada na origem;
- $\theta_1, \dots, \theta_q$  são os parâmetros de médias móveis não sazonais, sendo  $\theta_q \neq 0$ ;
- $\varepsilon_t$  é o ruído branco com média zero e variância constante.

Considerando como 'B' o operador de defasagem (*lag*) e definindo que  $B^q \varepsilon_t = \varepsilon_{t-q}$ , então a anterior equação é:

$$Y_t = c + \theta_1 B^1 \varepsilon_t + \theta_2 B^2 \varepsilon_t + \dots + \theta_q B^q \varepsilon_t + \varepsilon_t \quad (8)$$

### 3.3.3. Modelo Autorregressivo e de Médias Móveis (ARMA)

Diz-se que  $\{Y_t\}$  satisfaz um modelo autorregressivo e de médias móveis de ordem  $p$  e  $q$ :  $ARMA(p, q)$  ou  $ARIMA(p, 0, q)$  se,

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Considerando 'B' o operador de defasagem, então:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = c + \varepsilon_t (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \quad (9)$$

sendo,

- Modelo Autorregressivo de ordem  $p$ :
  - $AR(p) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t$ ;

- Modelo de Médias Móveis de ordem  $q$ :
  - $MA(q) = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$ ;
- $\varepsilon_t$  é o ruído branco com média zero e variância constante;
- $Y_t$  é a série temporal observada no tempo;
- $c$  corresponde à ordenada na origem.

### 3.3.4. Modelo Autorregressivo Integrado e de Médias Móveis (ARIMA)

Um modelo ARIMA não sazonal, denominado por Modelo Autorregressivo e de Médias Móveis Integrado, é classificado como um modelo  $ARIMA(p, d, q)$ , onde:

- $p$  representa a ordem do modelo autorregressivo;
- $d$  representa o grau de diferenciação;
- $q$  representa a ordem do modelo de média móvel.

Os termos  $p, d$  e  $q$  são valores inteiros maiores ou iguais a zero.

Diz-se que  $\{Y_t\}$  satisfaz um modelo autorregressivo e de médias móveis integrado de ordem  $p, d$  e  $q$  se,

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Considerando 'B' o operador de desfasagem, então:

$$\phi(B)(1 - B)^d y'_t = \theta(B)\varepsilon_t \quad (10)$$

sendo,

- Modelo Autorregressivo de ordem  $p$ :
  - $AR(p) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ ;
- Modelo de Médias Móveis de ordem  $q$ :
  - $MA(q) = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$ ;
- Diferenciação de ordem  $d$ :
  - $(1 - B^d)$ ;
- $\varepsilon_t$  é o ruído branco com média zero e variância constante;
- $Y_t$  é a série temporal diferenciada observada no tempo;
- $c$  corresponde à ordenada na origem.

### 3.3.5. Modelo Autorregressivo e de Médias Móveis Integrado Sazonal (SARIMA)

Para além dos modelos ARIMA tradicionais representados por  $ARIMA(p, d, q)$ , existem também os modelos ARIMA sazonais, representados por  $SARIMA(p, d, q)(P, D, Q)[s]$ . Onde:

- $s$  representa o número de observações por ano;
- $p$  representa a ordem do modelo autorregressivo não sazonal;
- $d$  representa o grau de diferenciação não sazonal;
- $q$  representa a ordem do modelo de média móvel não sazonal;
- $P$  representa a ordem do modelo autorregressivo sazonal;
- $D$  representa o grau de diferenciação sazonal;
- $Q$  representa a ordem do modelo de média móvel sazonal.

Diz-se que  $\{Y_t\}$  satisfaz um modelo autorregressivo e de médias móveis integrado sazonal se,

$$\Phi(B^s)\phi(B)(1-B)^d(1-B^s)^D Y_t = \Theta(B^s)\theta(B)\varepsilon_t \quad (11)$$

sendo,

- Modelo Autorregressivo de ordem  $p$ :
  - $AR(p) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ ;
- Modelo Autorregressivo sazonal de ordem  $P$ :
  - $AR(P) = (1 - \Phi_1 B^{1s} - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps})$ ;
- Modelo de Médias Móveis de ordem  $q$ :
  - $MA(q) = (1 + \theta_1 B^1 + \theta_2 B^2 + \dots + \theta_q B^q)$ ;
- Modelo de Médias Móveis sazonais de ordem  $Q$ :
  - $MA(Q) = (1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs})$ ;
- Diferenciação não sazonal  $d$ :
  - $(1 - B^d)$ ;
- Diferenciação sazonal  $D$ :
  - $(1 - B^s)^D$ ;
- $\varepsilon_t$  é o ruído branco com média zero e variância constante;

- $Y_t$  é a série temporal diferenciada observada no tempo;
- $c$  corresponde à ordenada na origem.

### 3.3.6. Modelo de Regressão Dinâmica

#### 3.3.6.1. Modelo Autorregressivo Integrado e de Médias Móveis com Variáveis Exógenas (ARIMAX)

O modelo ARIMA é muito poderoso para prever dados de séries temporais quando estes são univariados. No entanto, quando se quer analisar séries temporais com o uso adicional de covariáveis, pode-se aplicar o modelo ARIMAX, que significa Média Móvel Integrada Autorregressiva com Variáveis Exógenas. Este modelo permite a inclusão de variáveis externas, conhecidas como variáveis exógenas, que poderão ter um efeito de precisão nas séries temporais previstas.

O novo termo adicionado à equação ARIMA consiste no coeficiente  $\beta$  e na variável exógena  $X$ .

Assim, para um modelo ARIMAX a equação matemática é apresentada da seguinte forma:

$$\phi(B)(1 - B)^d Y_t = \theta(B)\varepsilon_t + \beta X_t \quad (13)$$

- Modelo Autorregressivo de ordem  $p$ :
  - $AR(p) = \phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ ;
- Modelo de Médias Móveis de ordem  $q$ :
  - $MA(q) = \theta(B) = (1 + \theta_1 B^1 + \theta_2 B^2 + \dots + \theta_q B^q)$ ;
- Diferenciação não sazonal  $d$ :
  - $(1 - B^d)$ ;
- $\varepsilon_t$  é o ruído branco com média zero e variância constante;
- $Y_t$  é a série temporal diferenciada observada no tempo;
- $\beta$  é o coeficiente da variável exógena;
- $X_t$  é a variável exógena;
- $c$  corresponde à ordenada na origem.

### 3.3.6.2. Modelo Autorregressivo e de Médias Móveis Integrado Sazonal com Variáveis Exógenas (SARIMAX)

À semelhança do descrito anteriormente, o modelo SARIMAX - Modelo Autorregressivo e de Médias Móveis Integrado Sazonal com Variáveis Exógenas – é um modelo ARIMAX mas com componente sazonal.

Assim, para um modelo SARIMAX a equação matemática é apresentada da seguinte forma:

$$\Phi(B^s)\phi(B)(1 - B)^d(1 - B^s)^DY_t = \Theta(B^s)\theta(B)\varepsilon_t + \beta X_t \quad (14)$$

- Modelo Autorregressivo de ordem  $p$ :
  - $AR(p) = \phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ ;
- Modelo Autorregressivo sazonal de ordem  $P$ :
  - $AR(P) = \Phi(B^s) = (1 - \Phi_1 B^{1s} - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps})$ ;
- Modelo de Médias Móveis de ordem  $q$ :
  - $MA(q) = \theta(B) = (1 + \theta_1 B^1 + \theta_2 B^2 + \dots + \theta_q B^q)$ ;
- Modelo de Médias Móveis sazonais de ordem  $Q$ :
  - $MA(Q) = \Theta(B^s) = (1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs})$ ;
- Diferenciação não sazonal  $d$ :
  - $(1 - B^d)$ ;
- Diferenciação sazonal  $D$ :
  - $(1 - B^s)^D$ ;
- $\varepsilon_t$  é o ruído branco com média zero e variância constante;
- $Y_t$  é a série temporal diferenciada observada no tempo;
- $\beta$  é o coeficiente da variável exógena;
- $X_t$  é a variável exógena;
- $c$  corresponde à ordenada na origem.

### 3.4. Medidas de Desempenho

As medidas de precisão em séries temporais são métricas utilizadas para avaliar a precisão de previsões em séries temporais. Estas medidas comparam os valores previstos pelo modelo com os valores reais, de modo a apresentar o quanto o modelo de previsão é capaz de reproduzir a série temporal em estudo. (Carmo & Silva, 2023)

Abaixo estão apresentadas algumas das principais métricas de precisão usadas em séries temporais:

#### 3.4.1. Erro Absoluto Médio – MAE

O erro absoluto médio é uma métrica de erro simples. É calculada a partir da média da diferença absoluta entre os valores previstos e os reais, conforme a seguinte equação:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - p_i|, \text{ onde:} \quad (12)$$

- $n$  representa o número de observações;
- $y$  é o valor observado para cada observação  $i$ ;
- $p$  é o valor previsto pelo modelo para cada observação  $i$ .

Quanto menor o valor de MAE, mais o modelo de previsão se aproxima dos valores reais, ou seja, o modelo é mais preciso.

A principal limitação desta métrica é não permitir a análise comparativa de modelos com unidades de medida diferentes, e tende a penalizar previsões com erros maiores.

Em modelos de séries temporais, esta métrica é utilizada quando se deseja que não seja influenciada por *outliers*.

#### 3.4.2. Erro Quadrático Médio – MSE

O erro quadrático médio é calculado como a média dos quadrados das diferenças entre os valores reais observados e os valores previstos para um modelo, conforme a seguinte equação:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2, \text{ onde:} \quad (13)$$

Quanto menor o valor de MSE, mais o modelo de previsão se aproxima dos valores reais, ou seja, o modelo é mais preciso.

Esta medição é importante ser implementada em modelos de séries temporais quando os grandes erros são particularmente indesejáveis.

### 3.4.3. Raiz Quadrada do Erro Quadrático Médio – RMSE

A raiz quadrada do erro quadrático médio é uma métrica utilizada especialmente em séries temporais e modelos de regressão. Ele é calculado a partir da raiz quadrada do Erro Quadrático Médio (MSE) e fornece uma medida da magnitude média dos erros de previsão, mantendo as unidades de medida dos dados, conforme a seguinte equação:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2}, \text{ onde:} \quad (14)$$

Quanto menor o RMSE, melhor é o modelo de previsão, ou seja, um valor de RMSE=0 indica uma previsão perfeita.

O RMSE é uma métrica importante para avaliar o desempenho de modelos de previsão em séries temporais tendo a capacidade de penalizar grandes erros.

### 3.4.4. Erro Absoluto Médio Percentual – MAPE

O erro absoluto médio percentual é uma métrica utilizada especialmente para realizar comparações entre modelos cujas variáveis apresentem unidades de medidas diferentes, conforme a seguinte equação:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - p_i}{y_i} \right| \times 100\% \quad (15)$$

Quanto menor o MAPE, melhor é o modelo de previsão, ou seja, um valor de MAPE=0% indica uma previsão perfeita.

O MAPE é uma métrica valiosa quando aplicada em modelos de séries temporais, quando se compara séries de diferentes tamanhos.

### 3.4.5. Erro Absoluto Médio Escalado – MASE

O erro absoluto médio escalado (MASE -*Mean Absolute Scaled Error*) é uma métrica proposta por Rob J. Hyndman e Anne B. Koehler em 2006, com o objetivo de comparar várias séries temporais, contrariamente às métricas MAE e MAPE.

$$MASE = \frac{\frac{1}{n} \sum_{i=1}^n |y_i - p_i|}{\frac{1}{n-m} \sum_{i=m+1}^n |y_i - y_{i-m}|}, \text{ onde:} \quad (16)$$

- $m$  representa o número de períodos usados no modelo de referência.

Quando:

- $MASE < 1$ , então o modelo de previsão apresenta erros inferiores ao modelo de referência. Este valor indica que o modelo é altamente eficiente em comparação com o modelo de referência;
- $MASE = 1$ , então o modelo de previsão tem um desempenho equivalente ao modelo de referência. Isto indica que o modelo não está a acrescentar valor adicional em comparação com a simples previsão baseada na última observação;
- $MASE > 1$ , então o modelo de previsão apresenta, em média, erros maiores do que o modelo de referência. Isso sugere que o modelo pode não estar a capturar bem os padrões da série temporal.

O MASE é uma métrica eficaz para avaliar o desempenho de modelos de previsão de séries temporais, em contextos onde se deseja comparar o desempenho de diferentes modelos ou séries temporais de diferentes tamanhos.

## 4. Análise das Séries da Produção e Vendas de Calçado

No âmbito desta dissertação, propõe-se explorar alguns métodos de análise de séries temporais e escolher o melhor modelo de previsão a adotar para cada artigo vendido pela empresa ALOFT.

O presente capítulo inicia com a apresentação da empresa ALOFT, para enquadrar a origem dos dados que serão analisados nas secções seguintes. Em seguida serão apresentados os resultados do estudo de cada uma das séries temporais relativas aos artigos de calçado disponíveis para análise e as respetivas conclusões. O estudo de cada série de vendas inicia com a análise exploratória da série, seguido do ajustamento e validação do melhor modelo de previsão. Por fim, é apresentada a equação e a representação gráfica do modelo selecionado.

O software R é utilizado em todas as análises em ambiente RStudio, que é um ambiente de desenvolvimento integrado de uso de linguagem de programação R. Esta linguagem foi desenvolvida em 1996 por Ross Ihaka e Robert Gentleman, da Universidade de Auckland, na Nova Zelândia. O R permite o armazenamento e manipulação de dados, a aplicação de cálculos matemáticos para tratamento dos dados, bem como exibições gráficas. (Hyndman & Athanasopoulos; George, 2018; W.N.Venables & R core Team, 2023)

### 4.1. A Empresa ALOFT

ALOFT é uma empresa do setor do calçado técnico e de performance, fundada em 2013 e sediada em Canidelo, Vila do Conde.

Com um foco elevado na temática do I&D (Investigação e Desenvolvimento), procura encontrar soluções técnicas para responder às necessidades dos seus clientes.

Dispõe de um setor vocacionado para a produção de solas de borracha, com desenvolvimento das suas próprias formulações através de um departamento de I&D e de um laboratório devidamente equipado, assim como parcerias com laboratórios externos. Adicionalmente, os setores de injeção convencional e injeção em máquinas rotativas com sistemas automatizados permite-lhe desenvolver e concretizar uma série de soluções técnicas.

Capaz de entregar uma solução desde a sua concepção até ao produto final, a ALOFT dispõe de equipas com conhecimento para desenvolvimento do design das peças, definição dos materiais adequados à funcionalidade e/ou técnica pretendida, ajuste e implementação dos processos produtivos a novos materiais ou projetos e até mesmo acabamento das peças com processos de limpeza, pintura ou costura.

A sua política inclui uma forte preocupação ambiental, com implementação de práticas desde os setores da produção até ao embalamento ou, até mesmo à definição do design das peças. A importância desta política para a administração, aliada à implantação de um departamento de sustentabilidade permitiu à empresa tornar a frase “*Green is the new good*” como uma das suas imagens de marca.

#### 4.2. Dados e Métodos

Para a realização deste trabalho, a empresa ALOFT disponibilizou o histórico de vendas mensais, de janeiro de 2019 até dezembro de 2023, referentes a cinco artigos de galochas da sua produção: 8B001, 8B007, 8B008, 8B010 e 8B014.

A análise exploratória dos dados é o ponto de partida de qualquer estudo. Assim, iniciou-se com uma análise estatística descritiva dos artigos em estudo, para os dados mensais registados de 2019 a 2023.

Conforme referido inicialmente, tem-se como objetivo modelar e prever as vendas em função do tempo, a variável dependente é o número de pares de calçado vendido ao longo de determinado período. Num segundo estudo, pretende-se incorporar a informação da série das previsões de vendas enviada pelos clientes, e analisar se esta informação melhora as previsões.

##### 1. Conversão dos dados em séries temporais para tratamento informático:

Procedeu-se ao seguinte pré processamento dos dados realizando a conversão dos dados em séries temporais para tratamento informático dos dados. Para este efeito, foi aplicada a função 'ts' do R em todas as variáveis de modo a converter os objetos em séries temporais. Seguidamente, determinou-se um conjunto de estatísticas da série relativa às vendas a partir da função 'summary'.

## 2. Decomposição e representação gráfica das séries temporais:

Aplicou-se a decomposição aditiva para cada artigo da série temporal das vendas, utilizando a função 'decompose' do R e o respectivo gráfico que contem a série observada (*observed*) e as componentes: linha de tendência (*trend*), sazonalidade (*seasonal*) e ruído (*random*).

Apresentou-se graficamente a distribuição ao longo do tempo das curvas de distribuição das previsões do cliente e das vendas sobrepostas no mesmo período, assim como o gráfico box-plot de modo a facilitar a leitura dos meses com mais procura por parte do cliente.

Uma vez que o cliente disponibiliza uma previsão mensal das “necessidades/encomendas” no início de cada ano, que é tomada como variável exógena ou preditor, apresentou-se graficamente as linhas de tendência sobrepostas das vendas e previsões, de modo a poder avaliar o comportamento dos desvios anuais. Assim, mediu-se a correlação entre duas variáveis, a partir da função 'cor.test' do R que retorna o coeficiente de correlação e a probabilidade de significância (p-value) da correlação. Esta função tem por base a correlação de Pearson que define se as variáveis são significativamente correlacionadas a partir das seguintes hipóteses:

H<sub>0</sub>: Não há correlação entre as variáveis;

H<sub>1</sub>: Há correlação entre as variáveis.

Baseados na distribuição da estatística de teste para o coeficiente de Pearson, é calculada a significância do teste. Se p-value do teste for inferior ao nível de significância fixado ( $\alpha = 0,05$ ), então pode-se rejeitar a hipótese nula (H<sub>0</sub>) e concluir que existe correlação estatisticamente significativa entre as variáveis.

O coeficiente de correlação ( $r$ ) é avaliado a partir da seguinte escala, (Martins, 2014):

- $r = 0$ , então não há correlação entre as variáveis;
- $0 < r \leq 0,3$ , então a correlação entre as variáveis é fraca;
- $0,3 < r \leq 0,7$ , então a correlação entre as variáveis é moderada;
- $0,7 < r \leq 1$ , então a correlação entre as variáveis é forte;

No estudo prático é apresentado graficamente a correlação cruzada entre estas duas variáveis, a partir da função do R 'ccf(Previsão, Vendas, type="correlation")'.

### 3. Teste de estacionariedade

Como descrito na secção anterior, uma série temporal é considerada estacionária se não tiver tendência e se apresentar variância e média constante ao longo do tempo. A partir do teste de Dickey-Fuller aumentado (função 'adf.test' do R) é possível concluir se a série é estacionária (se apresenta raiz unitária) a partir das seguintes hipóteses:

H<sub>0</sub>: A série temporal não é estacionária;

H<sub>1</sub>: A série temporal é estacionária.

Baseados na distribuição da estatística de teste de Dickey-Fuller aumentado, é calculada a significância do teste. Se a probabilidade de significância (p-value) do teste for inferior ao nível de significância fixado ( $\alpha = 0,05$ ), então pode-se rejeitar a hipótese nula (H<sub>0</sub>) e concluir que a série temporal é estacionária. (Lopes, 2019)

No caso das séries temporais em análise serem não estacionárias, nomeadamente quando se observa que a variância não é constante ao longo do tempo (heterocedasticidade), pode aplicar-se a transformação Box-Cox, a partir da função do R 'BoxCox()'.

No caso de se estudar séries temporais sazonais poderá não haver estacionariedade, uma vez que há flutuações ou padrões que se repetem em intervalos fixos que provocam uma variação na média e variância ao longo do tempo. Uma técnica comum para lidar com a não estacionariedade sazonal é aplicar a diferenciação sazonal, a partir da função do R 'seasadj(decompose())'. O objetivo é tornar a série temporal mais estacionária. Esta função remove o padrão sazonal estimado pela função do R 'decompose' da série temporal.

### 4. Apresentação gráfica da Função de Autocorrelação e Função de Autocorrelação Parcial

De acordo com o referido anteriormente, calculou-se a função de autocorrelação 'Acf' e autocorrelação parcial 'Pacf' no R na série temporal de vendas. A sua

representação gráfica inclui com linhas azuis os intervalos de confiança para a autocorrelação. (Alpuim, 1998)

## 5. Modelos de Previsão

De modo a ajustar o modelo ARIMA, foi aplicada a função 'auto.arima' do R, proveniente do pacote 'forecast' que auxilia a identificação de um modelo para cada série, incluindo a existência de componente sazonal. Esta função procura o modelo que melhor se ajusta aos dados, tendo como critério de seleção a combinação de ordens que geram o menor valor de AIC (*Akaike's Information Criterion*), AICc (*Corrected AIC*) e BIC (*Bayesian Information Criterion*) e também, o menor valor de erro MASE. Esta métrica foi a selecionada neste estudo, uma vez que é uma das mais eficazes para comparar o desempenho entre séries temporais. Para o caso onde não foi verificado um padrão de sazonalidade, o modelo será descrito na forma ARIMA(p,d,q). Caso seja sazonal, o modelo será descrito na forma SARIMA(p,d,q)(P,D,Q)[s].

Adicionalmente, são utilizadas variáveis exógenas (fatores externos) que poderão influenciar a série temporal que está sendo analisada ou prevista. Estas variáveis são usadas para melhorar a precisão das previsões. Para o caso onde não foi verificado um padrão de sazonalidade, mas utilizado estes fatores externos, o modelo será descrito na forma ARIMAX(p,d,q). Caso seja sazonal, o modelo será descrito na forma SARIMAX(p,d,q)(P,D,Q)[s]. (Lopes, 2019)

## 6. Análise dos resíduos do modelo

A avaliação dos resíduos em séries temporais é fundamental para compreensão e modelação dos dados, permitindo a verificação de padrões e tendências. Através desta análise, é possível avaliar a adequação do modelo ajustado e identificar possíveis violações de pressupostos.

Os resíduos são o resultado do diferencial entre os valores observados e os valores previstos pelo modelo. Estes deverão cumprir os seguintes pressupostos:

- Independência - Os resíduos deverão ser independentes entre si, ou seja, não deverão apresentar autocorrelação significativa (teste Ljung-Box);

- Normalidade - Os resíduos devem-se aproximar de uma distribuição normal (teste Shapiro-Wilk);
- Homoscedasticidade - A variância dos resíduos deve ser constante ao longo do tempo (teste Breusch-Pagan);
- Ausência de Tendência ou Sazonalidade - Os resíduos devem ser isentos de tendências ou padrões sazonais.

O teste de Ljung-Box é aplicado para verificar independência e a presença de autocorrelação nos resíduos aplicado no modelo e baseia-se nas seguintes hipóteses:

H<sub>0</sub>: Os resíduos são distribuídos independentemente – resíduos não correlacionados;

H<sub>1</sub>: Os resíduos não são distribuídos de forma independente – resíduos autocorrelacionados.

Baseados na distribuição da estatística de teste de Ljung-Box, é calculada a significância do teste. Se a probabilidade de significância (p-value) do teste for inferior ao nível de significância fixado ( $\alpha = 0,05$ ), então pode-se rejeitar a hipótese nula e concluir que a série temporal é distribuída de forma independente e não correlacionada. Quanto mais aleatórios forem os erros, melhor será o modelo ajustado.

O teste de Shapiro-Wilk é aplicado para avaliar se a distribuição dos resíduos se aproxima de uma distribuição normal e baseia-se nas seguintes hipóteses:

H<sub>0</sub>: Os resíduos são normalmente distribuídos;

H<sub>1</sub>: Os resíduos não são normalmente distribuídos.

Baseados na distribuição da estatística de Shapiro-Wilk, é calculada a significância do teste. Se a probabilidade de significância (p-value) do teste for inferior ao nível de significância fixado ( $\alpha = 0,05$ ), então pode-se rejeitar a hipótese nula (H<sub>0</sub>) e concluir que os resíduos não seguem uma distribuição normal. (Lopes, 2019)

O teste de Breusch-Pagan é aplicado para verificar a variância dos resíduos e baseia-se nas seguintes hipóteses:

H<sub>0</sub>: Os resíduos são distribuídos com variância constante - Homoscedasticidade;

H<sub>1</sub>: Os resíduos não são distribuídos com variância constante - Heterocedasticidade.

Baseados na distribuição da estatística de Breusch-Pagan, é calculada a significância do teste. Se a probabilidade de significância (p-value) do teste for inferior ao nível de significância fixado ( $\alpha = 0,05$ ), então pode-se rejeitar a hipótese nula (H<sub>0</sub>) e concluir que a heterocedasticidade está presente no modelo de regressão. (Baum, 2014)

## 7. Previsão do Modelo

A previsão de um modelo ajustado graficamente permite visualizar a relação entre as variáveis preditoras e a variável resposta após o ajuste do modelo. A função 'autoplot(forecast)' do R permite esta visualização no período definido.

A partir da equação matemática de previsão é ainda possível descrever a relação entre variáveis, com o objetivo de prever o valor de uma variável de interesse em função das variáveis preditoras. A equação utiliza um modelo matemático que foi ajustado a partir de dados observados.

### 4.3. Estudo do artigo 8B001

O código de artigo 8B001 corresponde a um produto de calçado destinado ao uso em piscinas.

#### 4.3.1. Análise exploratória do artigo 8B001

A partir dos resultados da estatística descritiva, Figura 4.1, para o artigo 8B001, pode-se inferir que no período de janeiro de 2019 a dezembro de 2023, o número médio de pares vendidos por mês foi de 10091, o desvio padrão foi de 11127 pares e que estava compreendido entre o número mínimo de 0 e um máximo de 48741 pares. Estes valores tão distantes são justificados, pela variação das vendas ao longo dos meses do ano. Trata-se de um artigo para uso em piscinas onde as vendas são mais acentuadas em setembro na época de “regresso às aulas.

```
> summary(Vendas)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0    1894    7042   10091   13871   48741
> sd(Vendas)
[1] 11127.58
```

Figura 4.1 - Estatística descritiva da série temporal dos pares vendidos, para o artigo 8B001.

Tal como se pode observar no Gráfico 4.1, Gráfico 4.2 e Gráfico 4.3 há um alerta visual de falta de vendas nos meses de abril e maio do ano de 2020. Considerando que a evolução das vendas no mês de abril desde 2019 até 2023 foi respetivamente de 16803, 0, 10209, 7482 e 8636 e no mês de maio foi de 25225, 922, 17722, 20281 e 37186 para o mesmo espaço temporal, procedeu-se à substituição destes valores pela mediana. Assim, em abril o valor passará de 0 para 10209 pares e em maio de 922 para 17722 pares. Estes valores são explicados pelo facto de neste ano a população ter sofrido restrições no seu quotidiano causadas pela pandemia Covid-19 e por consequência um encerramento das piscinas. Uma vez que este artigo é para uso em piscinas, as vendas foram condicionadas.

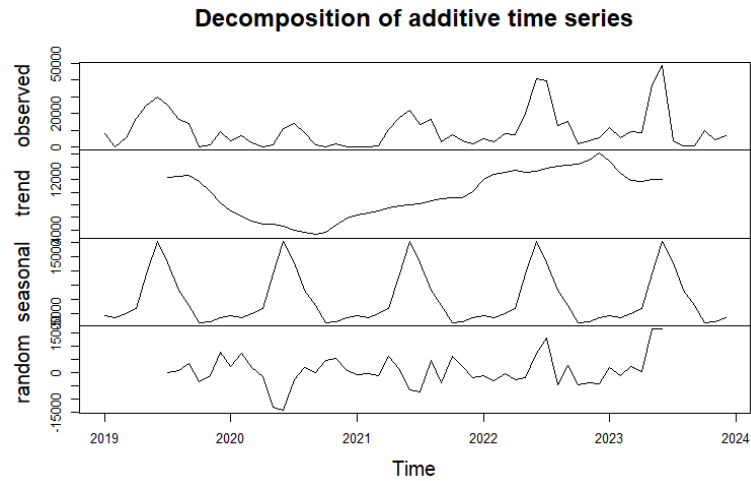


Gráfico 4.1 - Distribuição e decomposição da série temporal original relativa ao número de pares vendidos do artigo 8B001, num período mensal de 2019 a 2023.

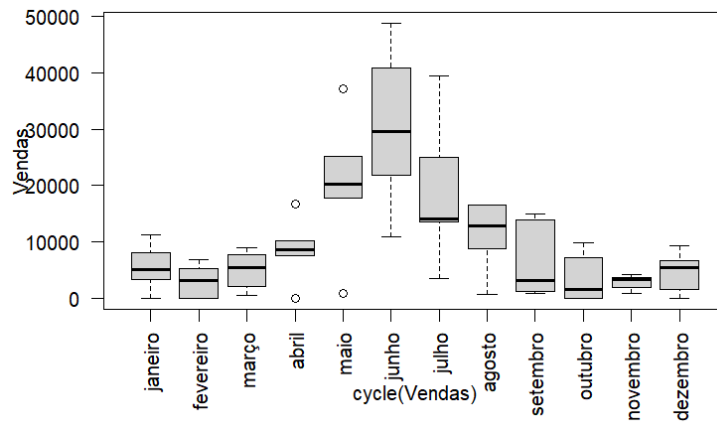


Gráfico 4.2 – Box-plot das vendas reais mensais do artigo 8B001, num período de 2019 a 2023.

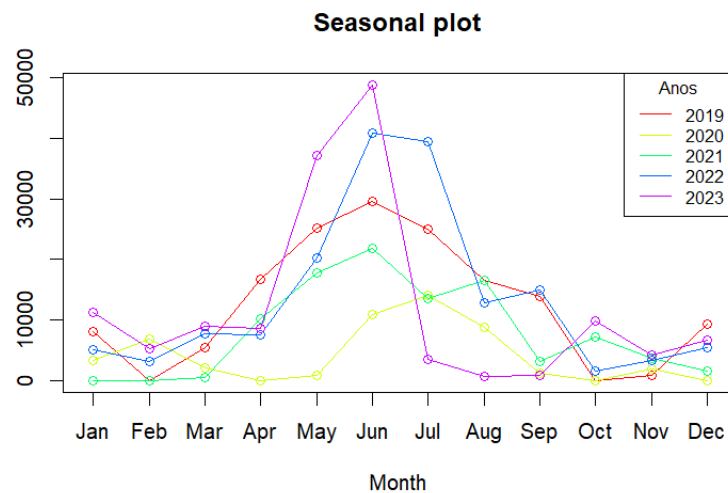


Gráfico 4.3 – Distribuição anual do número de pares vendidos do artigo 8B001, após alteração dos valores dos meses abril e maio de 2020 pela mediana, num período mensal de 2019 a 2023.

Assim, a partir dos resultados da estatística descritiva, após a alteração dos dados mencionadas acima para o artigo 8B001, Figura 4.2, pode-se observar que no período de janeiro de 2019 a dezembro de 2023, o número médio de pares vendidos por mês foi de 10541, o desvio padrão é de 11020 pares e que estava compreendido entre o número mínimo de 0 e um máximo de 48741 pares.

```
> summary(Vendas)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0    2968    7629   10541   14281   48741
> sd(Vendas)
[1] 11020.58
```

Figura 4.2 – Estatística descritiva da série temporal ajustada dos pares vendidos, para o artigo 8B001.

#### 4.3.2. Decomposição e Representação das Séries Temporais

O Gráfico 4.4 apresenta a série temporal do número de pares vendidos ao longo do período considerado (*observed*), a decomposição aditiva da série nas componentes (*trend*), a sazonalidade (*seasonal*) e o ruído (*random*). Pode-se constatar que no ano de 2020 houve uma redução de vendas face ao previsto, explicada pelo facto de neste ano a população ter sofrido restrições no seu quotidiano causadas pela pandemia Covid-19. Uma vez que este artigo é especializado para uso em piscina e estas estiveram encerradas durante este período, a procura deste artigo reduziu cerca de 67%. Contudo, a partir de 2020 verificou-se um aumento das vendas. Em 2023 o cliente optou por minimizar o stock extra em armazém, explicando-se, deste modo, a redução do número de pares fabricados deste artigo. Verifica-se, também, que há um efeito sazonal com repetição anual.

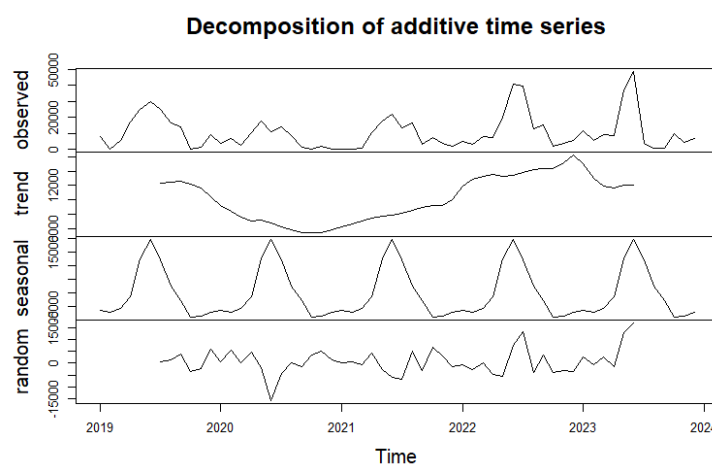


Gráfico 4.4 – Distribuição e decomposição da série temporal relativa ao número de pares vendidos do artigo 8B001, num período mensal de 2019 a 2023.

No Gráfico 4.5 apresenta a sobreposição das curvas de distribuição dos pares vendidos por ano. Confirmando a afirmação anterior, os meses com maiores vendas foram de maio a agosto para todos os anos avaliados. É possível observar que o ano de 2020 desviou-se da tendência anual e as vendas mensais mantiveram-se constantes ao longo do ano.

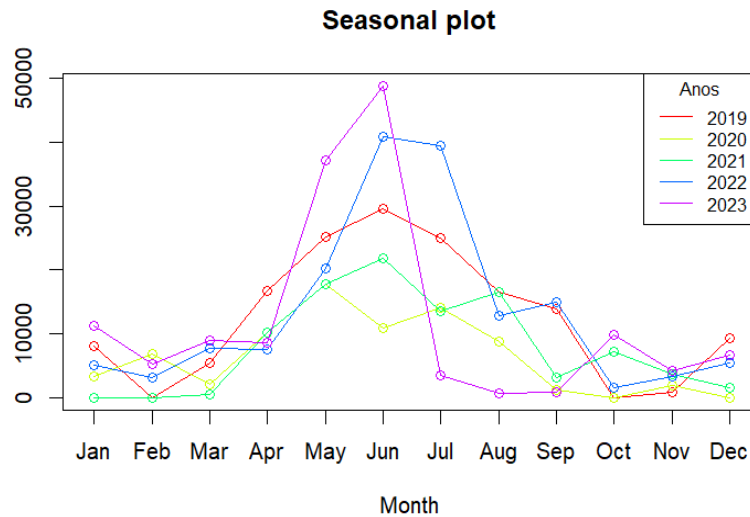


Gráfico 4.5 – Distribuição anual do número de pares vendidos do artigo 8B001, após alteração dos valores dos meses abril e maio de 2020 pela mediana, num período mensal de 2019 a 2023.

No Gráfico 4.6 apresenta a distribuição ao longo do tempo das curvas de distribuição das previsões do cliente e das vendas sobrepostas no mesmo período. É possível observar que houve um ligeiro desvio das previsões do cliente nos anos de 2019, 2020 e 2021, face às vendas.

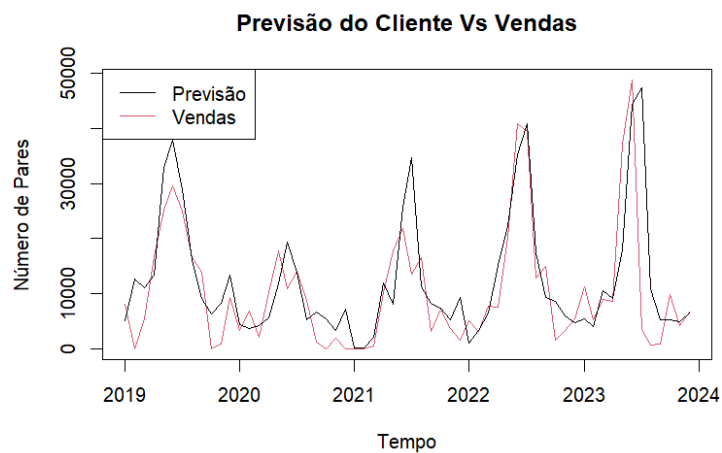


Gráfico 4.6 – Distribuição mensal de 2019 a 2023 dos pares previstos e vendidos para o artigo 8B001.

A partir da função ‘cor.test(Previsão, Vendas)’, Figura 4.3, calcula-se o valor de correlação das duas séries temporais em estudo, com o valor de 0,74. Pode-se inferir que estamos na presença de uma correlação entre variáveis forte. Como o valor de prova é inferior ao nível de significância fixado,  $\text{valor-p}=2,006^{-11} < \alpha=0,05$ , rejeita-se a hipótese nula ( $H_0$ ) e conclui-se que existe correlação significativa entre as variáveis.

```
> cor.test(Previsão, Vendas)

Pearson's product-moment correlation

data: Previsão and Vendas
t = 8.2897, df = 58, p-value = 2.006e-11
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5934547 0.8343212
sample estimates:
 cor
0.736408
```

Figura 4.3 - Correlação de Pearson entre as séries sazonais de previsão do cliente e vendas, para o artigo 8B001.

Recorrendo à função do *software* R ‘`ccf(Previsão, Vendas, type="correlation")`’, apresenta-se graficamente a correlação cruzada entre as variáveis vendas e previsão do cliente (Gráfico 4.7). Pode-se concluir que no *lag* 0 a correlação é a mais elevada e significativa, sugerindo uma forte correlação entre as duas variáveis, o que era esperado. Os *lag* significativos de correlação positiva (ex: -1; -11; -12; 1; 11; 12) sugerem que as vendas passadas dos meses novembro, dezembro e janeiro têm um impacto de 12 meses nas previsões futuras. Os *lag* significativos de correlação negativa (ex: -5; -6; -7; 4; 5; 6) sugerem que as vendas passadas respetivas a estes meses (abril, maio, junho e julho) têm um impacto negativo nas previsões futuras desse mesmo período.

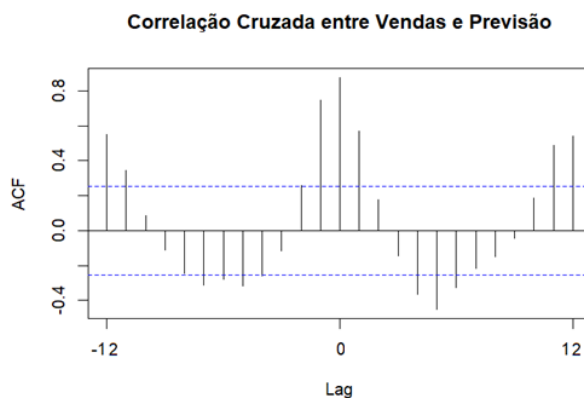


Gráfico 4.7 – Correlação cruzada das vendas e previsões do cliente para o artigo 8B001, num período de 2019 a 2023.

Observa-se, no Gráfico 4.8, que o mês de junho apresenta distribuição de valores e a mediana mais alta, o que confirma o dado conhecido, de que a procura é superior do cliente neste período do ano, uma vez que se trata de um artigo para uso em piscina.

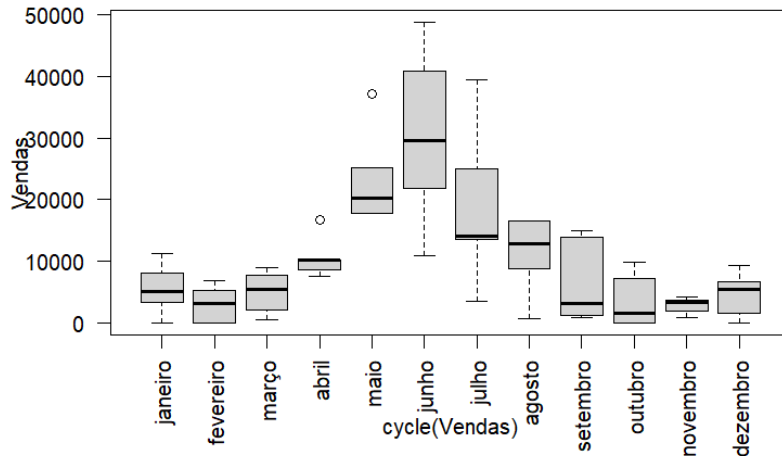


Gráfico 4.8 – Box-plot das vendas mensais do artigo 8B001, num período de 2019 a 2023.

#### 4.3.3. Modelos de Previsão

O resultado da aplicação do teste de Dickey-Fuller, a partir da função ‘adf.test’, conduz à conclusão de que a série das vendas é estacionária, uma vez que se obteve o valor- $p=0.01 < \alpha=0,05$ , conforme resultados apresentados na Figura 4.4.

```
> adf.test(Vendas)

Augmented Dickey-Fuller Test

data: Vendas
Dickey-Fuller = -4.0859, Lag order = 3, p-value = 0.01194
alternative hypothesis: stationary
```

Figura 4.4 – Resultado da aplicação do teste de estacionaridade aplicado no artigo 8B001.

A função de autocorrelação, representada no Gráfico 4.9, apresenta uma sinusoidal, o que indica haver indícios de sazonalidade. Os *lag* 1, 5, 6, 11 e 12 são significativos e repetem-se de acordo com a sazonalidade evidenciada.

Na PACF observa-se que as correlações mais fortes e significativas, por se encontrarem fora do intervalo de confiança são nos *lag* 1, 2 e 11.

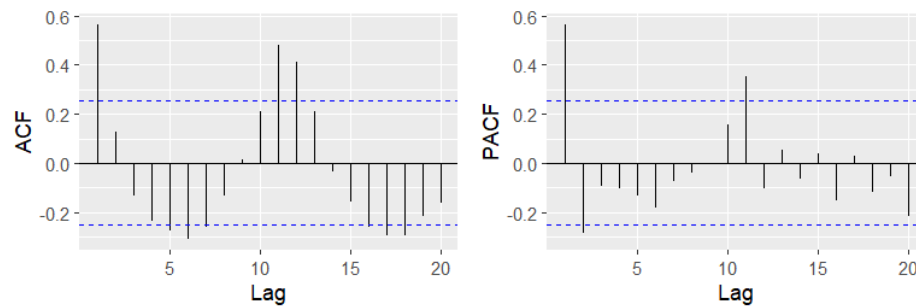


Gráfico 4.9 – Gráficos de autocorrelação e autocorrelação parcial das vendas no artigo 8B001.

A Figura 4.5 apresenta os resultados obtidos após a aplicação da função 'auto.arima' na série temporal 'vendas' (modelo A). Pode-se observar que o resultado é apresentado no formato SARIMA(0,0,1)(0,1,1)[12], o que significa que se tratam de séries com componente sazonal.

```

> a=auto.arima(Vendas)
> summary(a)
Series: Vendas
ARIMA(0,0,1)(0,1,1)[12]

Coefficients:
      ma1      sma1
      0.3012  -0.6662
s.e.  0.1342  0.2700

sigma^2 = 67560193: log likelihood = -503.24
AIC=1012.47  AICc=1013.02  BIC=1018.08

Training set error measures:
              ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set -5.267393  7196.954  4818.071  NaN  Inf  0.6979003  0.02185653

```

Figura 4.5 – Resultado da aplicação do modelo A na série temporal 'vendas' do artigo 8B001.

Seguidamente, uma vez que há uma correlação forte entre as vendas e as previsões do cliente e esta é estatisticamente significativa, aplica-se a função 'auto.arima' (Modelo B), Figura 4.6, na série temporal com a inclusão da variável exógena 'previsões' do cliente. Pode-se observar que o resultado é apresentado no formato ARIMAX(0,0,0), o que significa que há ruído branco.

```

> b=auto.arima(Vendas, xreg=Previsão)
> summary(b)
Series: Vendas
Regression with ARIMA(0,0,0) errors

Coefficients:
      xreg
      0.7837
s.e.  0.0572

sigma^2 = 56738877: log likelihood = -620.25
AIC=1244.5  AICc=1244.71  BIC=1248.69

Training set error measures:
              ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set  710.7724  7469.487  4904.309  -Inf  Inf  0.7103919  0.02787266

```

Figura 4.6 – Resultado da aplicação do modelo B na série temporal 'vendas' do artigo 8B001 com a inclusão da variável exógena 'previsões' do cliente.

Uma vez que a série temporal 'vendas' é estacionária aplicou-se um ajuste sazonal às séries temporais 'vendas' e 'previsão', uma vez que é possível verificar sazonalidade no Gráfico 4.6, ou seja, foi removido o padrão sazonal estimado pela função 'decompose'. Assim, as componentes restantes da série podem ser analisadas mais claramente e, por conseguinte, melhorar a precisão das previsões, tornando assim a tendência da série mais estacionária. Conforme o Gráfico 4.10 é possível verificar que houve um ajuste da sazonalidade da série temporal.

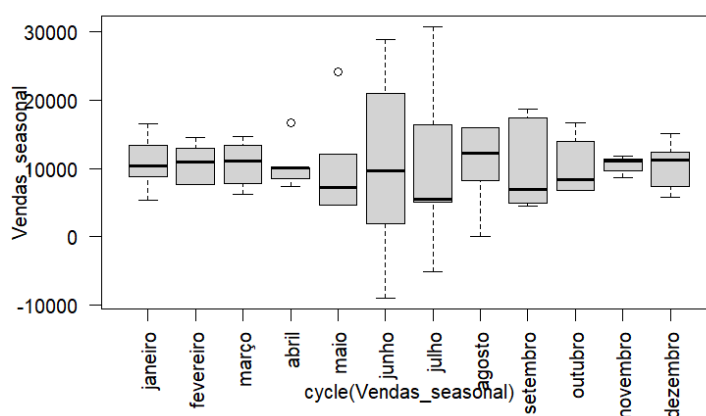


Gráfico 4.10 - Box-plot das vendas após a aplicação da diferenciação sazonal para o artigo 8B001.

Procedeu-se à aplicação do teste Dickey-Fuller, a partir da função 'adf.test' à série ajustada e concluiu-se que a série não é estacionária, uma vez que ao nível de significância de 5% não se rejeita a hipótese nula, o valor-p=0.1509 >  $\alpha=0,05$ , Figura 4.7.

```
> adf.test(Vendas_seasonal)

Augmented Dickey-Fuller Test

data: Vendas_seasonal
Dickey-Fuller = -3.0486, Lag order = 3, p-value = 0.1509
alternative hypothesis: stationary
```

Figura 4.7 – Resultado da aplicação do teste de estacionaridade aplicado à série temporal 'vendas' ajustada pela função 'decompose' no artigo 8B001.

Procedeu-se à análise da função de autocorrelação, representada no Gráfico 4.11, onde se observa que o lag 1 é significativo. No gráfico da função de autocorrelação parcial, observa-se que os lag 1 e 12 são significativos, o que indica que apesar da diferenciação sazonal, há indícios de sazonalidade na série, mas de intensidade fraca.

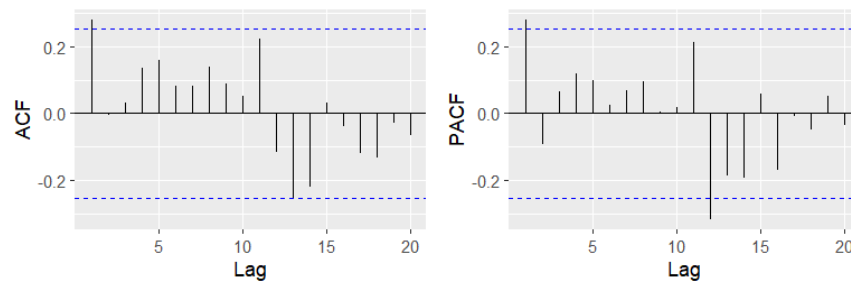


Gráfico 4.11 – Gráficos de autocorrelação e autocorrelação parcial da série temporal 'vendas' ajustada do artigo 8B001.

A Figura 4.8 apresenta os resultados obtidos após a aplicação da função 'auto.arima' na série temporal ajustada 'vendas' (modelo D). Pode-se observar que o resultado é apresentado no formato ARIMA(0,0,1).

```
> d=auto.arima(Vendas_seasonal)
> summary(d)
Series: Vendas_seasonal
ARIMA(0,0,1) with non-zero mean

Coefficients:
      ma1      mean
      0.3088 10570.432
s.e.    0.1224 1068.238

sigma^2 = 41664799: log likelihood = -610.52
AIC=1227.05  AICc=1227.48  BIC=1233.33

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -14.78027 6346.335 4657.914 -122.1311 167.1012 0.6747014 0.00217689
```

Figura 4.8 – Resultado da aplicação do modelo D na série temporal ajustada 'vendas' do artigo 8B001.

Seguidamente, uma vez que há uma correlação forte e estatisticamente significativa entre as vendas e as previsões do cliente, aplica-se a função 'auto.arima' (Modelo C), Figura 4.9, na série temporal ajustada com a inclusão da variável exógena 'previsões' do cliente. Pode-se observar que o resultado é apresentado no formato SARIMAX(0,0,2)(1,0,0)[12].

```
> c= auto.arima(Vendas_seasonal, xreg=Previsão_seasonal)
> summary(c)
Series: Vendas_seasonal
Regression with ARIMA(0,0,2)(1,0,0)[12] errors

Coefficients:
      ma1      ma2      sar1      xreg
      0.0100 -0.4882 -0.4848 0.8434
s.e.    0.1286 0.1450 0.1484 0.0246

sigma^2 = 33221643: log likelihood = -604.52
AIC=1219.04  AICc=1220.15  BIC=1229.51

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -31.52565 5568.381 4342.713 -137.267 183.6965 0.6290445 -0.01287622
```

Figura 4.9 – Resultado da aplicação do modelo C na série temporal ajustada 'vendas' do artigo 8B001 com a inclusão da variável exógena 'previsões' do cliente.

Comparando os quatro modelos aplicados na série temporal em estudo, pode-se concluir que o modelo A é o melhor modelo ajustado. Este apresenta o menor valor de AIC, AICc e BIC, conforme é possível observar na Tabela 4.1. No entanto, após a verificação dos pressupostos dos resíduos, conclui-se que este modelo não cumpre o pressuposto da normalidade residual. Assim, o modelo selecionado é o C.

Tabela 4.1 - Comparação dos valores calculados pela função 'auto.arima' do R para os quatro modelos aplicados no artigo 8B001.

Modelos:	MASE	AIC	AICc	BIC
A: Série temporal 'vendas' SARIMA(0,0,1)(0,1,1)[12]	0,70	1012,47	1013,02	1018,08
B: Série temporal 'vendas' com a inclusão da variável exógena 'previsões'. ARIMAX(0,0,0)	0,71	1244,50	1244,71	1248,69
C: Série temporal 'vendas' ajustada com a inclusão da variável exógena 'previsões' SARIMAX(0,0,2)(1,0,0)[12]	0,63	1219,04	1220,15	1229,51
D: Série temporal 'vendas' ajustada ARIMA(0,0,1)	0,67	1227,05	1227,48	1233,33

#### 4.3.4. Análise dos Resíduos do Modelo de Previsão Selecionado

Os resíduos deste modelo foram calculados a partir da função do R 'checkresiduals', conforme Gráfico 4.12. Observa-se que apresentam valores elevados, apesar de aparentemente centrados em zero. A aplicação prévia da transformação Box-Cox para estabilizar a variância e eliminar possíveis *outliers* podia ser explorada, contudo, esta transformação não é sempre recomendada. O que se pode afirmar é que o modelo não consegue captar suficientemente o comportamento da série como almejado, mas isto poderá estar relacionado com a variabilidade e dimensão da amostra.

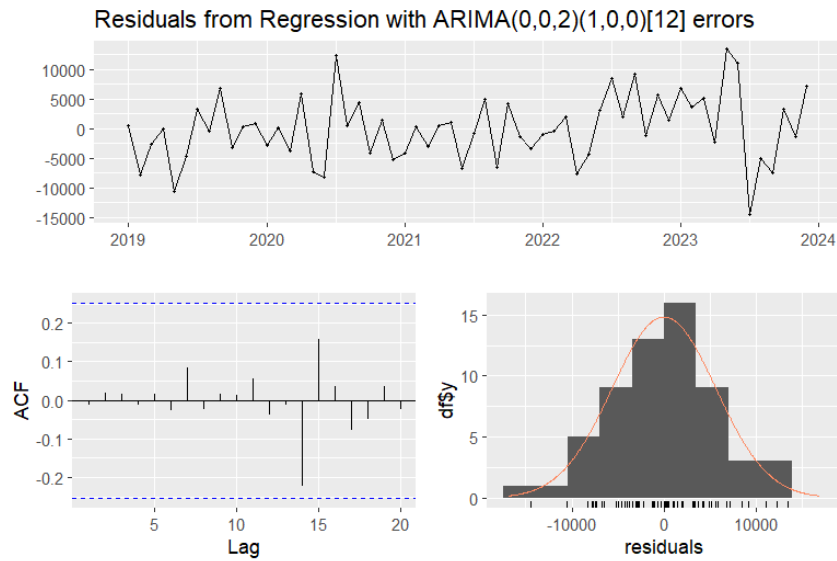


Gráfico 4.12 – Análise dos resíduos do modelo de previsão ajustado para o artigo 8B001.

Estes resíduos deverão cumprir os seguintes pressupostos:

- ✓ Independência – A partir do gráfico de ACF é possível observar que não há autocorrelação significativa fora dos limites de confiança de 95%, logo, assume-se que são não correlacionados. A partir do teste Ljung-Box, ao nível de significância de 5% ( $\alpha = 0,05$ ) confirma-se esta análise, não se rejeita a hipótese nula, uma vez que o valor- $p=0,9993 > \alpha$ .

```
> checkresiduals(c)

      Ljung-Box test

data:  Residuals from Regression with ARIMA(0,0,2)(1,0,0)[12] errors
Q* = 1.0534, df = 9, p-value = 0.9993

Model df: 3.   Total lags used: 12
```

- ✓ Normalidade - A distribuição dos resíduos aproxima-se da normalidade, o que indica que o modelo é o adequado, confirmado pelo teste Shapiro-Wilk. Pode-se concluir que os resíduos seguem uma distribuição normal, uma vez que não se rejeita a hipótese nula, o valor- $p=0,9772 > \alpha=5\%$ .

```
> residual <- residuals(c)
> shapiro_test <- shapiro.test(residual)
> print(shapiro_test)

      Shapiro-wilk normality test

data:  residual
W = 0.99272, p-value = 0.9772
```

✓ Homoscedasticidade – Analisando a distribuição dos resíduos ao longo do tempo, observa-se que a variância dos resíduos é relativamente constante, mas existem alguns picos esporádicos (eventualmente *outliers*). Apesar destes picos, a variância parece aproximadamente estável, sem um aumento ou diminuição gradual ao longo do tempo. Este comportamento sugere homoscedasticidade, no entanto, os *outliers* podem indicar que certos períodos têm uma variância ligeiramente maior, o que pode motivar uma investigação mais profunda com testes estatísticos. Assim, aplicou-se o teste Breusch-Pagan e confirma-se a homoscedasticidade, com  $p=0,4317 > \alpha=5\%$ .

```
> bptest(residual ~ Vendas_seasonal)
      studentized Breusch-Pagan test
data:  residual ~ Vendas_seasonal
BP = 0.61829, df = 1, p-value = 0.4317
```

✓ Ausência de Tendência ou Sazonalidade - Os resíduos apresentam-se como ruído branco, o que significa que estão espalhados aleatoriamente em torno de zero, sem quaisquer padrões ou tendências aparentes.

Em suma, os resíduos do modelo ajustado cumprem com todos os pressupostos, o que indica que se trata de um ajuste adequado.

#### 4.3.5. Previsão do Modelo Ajustado

Após se ter definido o modelo que melhor se ajusta aos dados, apresentou-se graficamente a partir da função do R 'autoplot(forecast)' (Gráfico 4.13), num período de 12 meses.

Dada a elevada variabilidade da série temporal, caracterizada por um elevado nível de ruído e uma base de dados relativamente curta (com apenas cinco anos), as previsões apresentam inevitavelmente um maior grau de incerteza. Consequentemente, os intervalos de confiança tendem a ser mais amplos, refletindo a limitação na capacidade do modelo em capturar padrões robustos e consistentes ao longo do tempo.

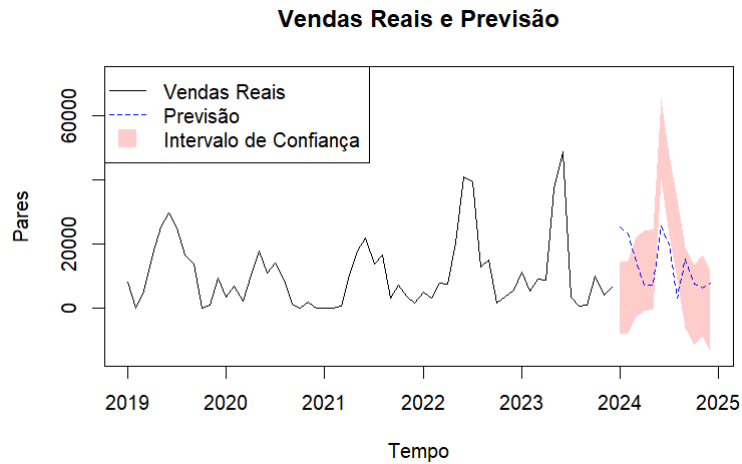


Gráfico 4.13 – Representação gráfica do modelo de previsões para o artigo 8B001.

Assim, foi possível definir a equação matemática correspondente ao modelo ajustado:

Equação Geral do modelo SARIMAX:

$$\Phi(B^s)\phi(B)(1 - B)^d(1 - B^s)^D Y_t = c + \Theta(B^s)\theta(B)\varepsilon_t + \beta X_t$$

Equação do modelo selecionado SARIMAX(0,0,2)(1,0,0)[12]:

- $p = 0: AR(0) = (1 - \phi_0 B) = 1;$
- $P = 1: AR(1) = (1 - \Phi_1 B^{12});$
- $q = 2: MA(2) = (1 + \theta_1 B^1 + \theta_2 B^2);$
- $Q = 0: MA(0) = (1 + \Theta_1 B^{12}) = 1;$
- $d = 0: (1 - B^0) = 1;$
- $D = 0: (1 - B^{12})^0 = 1;$
- $s = 12.$

$$\begin{aligned} \Phi(B^s)Y_t &= c + \theta(B)\varepsilon_t + \beta X_t \\ (1 - \Phi_1 B^{12})Y_t &= c + (1 + \theta_1 B^1 + \theta_2 B^2)\varepsilon_t + \beta X_t \Leftrightarrow \\ Y_t - \Phi_1 Y_{t-12} &= c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \beta X_t \end{aligned}$$

Se:

- $\Phi_1 = -0,48$
- $\theta_1 = 0,01$
- $\theta_2 = -0,49$
- $\beta_1 = 0,84$
- $c = \text{N\~{a}o Aplicado}$

Ent\~{a}o:

$$Y_t = -0,48Y_{t-12} + \varepsilon_t + 0,01\varepsilon_{t-1} - 0,49\varepsilon_{t-2} + 0,84X_t$$

#### 4.4. Estudo do artigo 8B007

O código de artigo 8B007 corresponde a um produto de calçado destinado ao uso em caça desportiva.

##### 4.4.1. Análise exploratória do artigo 8B007

A partir dos resultados da estatística descritiva, Figura 4.10, para o artigo 8B007, pode-se inferir que no período de janeiro de 2019 a dezembro de 2023, o número médio de pares vendidos por mês foi de 2773, o desvio padrão foi de 2412 pares e que estava compreendido entre o número mínimo de 0 e um máximo de 11141 pares. Estes valores tão distantes são justificados, pela variação das vendas ao longo dos meses do ano. Trata-se de um artigo de caça desportiva onde as vendas são mais acentuadas nas estações do ano mais frias.

```
> summary(Vendas)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0     932     2321    2773    3995   11141
> sd(Vendas)
[1] 2411.853
```

Figura 4.10 – Estatística descritiva da série temporal dos pares vendidos, para o artigo 8B007.

##### 4.4.2. Decomposição e Representação Gráfica das Séries Temporais

O Gráfico 4.14 apresenta a série temporal do número de pares vendidos ao longo do período considerado (*observed*), a decomposição aditiva da série nas componentes (*trend*), a sazonalidade (*seasonal*) e o ruído (*random*). Pode-se constatar que as vendas tiveram uma tendência positiva desde 2019 até 2021 e que neste último ano houve uma redução. Isto deveu-se ao facto de o cliente ter investido no desenvolvimento de artigos semelhantes noutras indústrias. Por consequência, este artigo tem vindo, desde essa altura, a seguir um processo lento de descontinuidade. Também, é possível observar um efeito repetitivo da curva de distribuição de vendas ao longo dos meses neste período, o que aparenta ser um artigo de calçado sazonal.

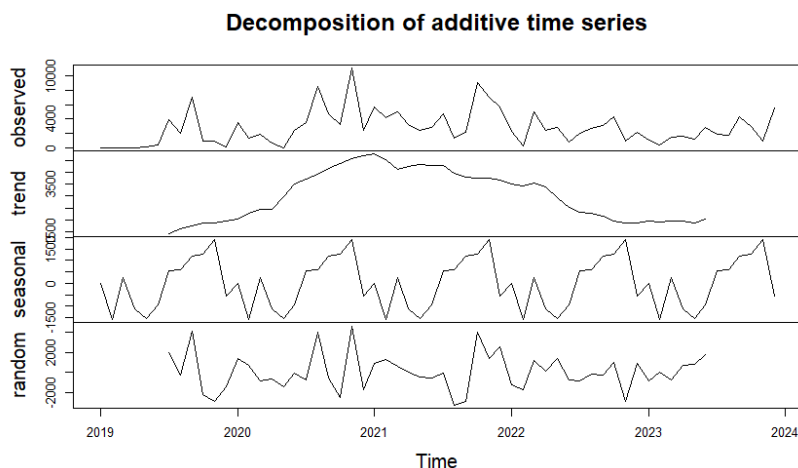


Gráfico 4.14 – Distribuição e decomposição da série temporal relativa ao número de pares vendidos do artigo 8B007, num período mensal de 2019 a 2023.

O Gráfico 4.15 apresenta a sobreposição das curvas de distribuição dos pares vendidos por ano. Confirmando a afirmação anterior, os meses de setembro a dezembro são os que apresentam maior quantidade vendida. Contudo, é possível observar que o ano de 2020 desviou-se da tendência anual. As vendas tendem a aumentar no período de agosto a outubro, reduzem no mês de novembro e voltam a aumentar no mês de dezembro. Já no ano de 2020 houve uma antecipação deste pico para o mês de agosto, seguindo de uma redução nos meses posteriores. No mês de novembro do ano 2020, contrariando os restantes anos, houve uma antecipação de vendas, seguindo de uma redução em dezembro. Estes desvios de vendas mensais são explicados pelo facto de neste ano a população ter sofrido restrições no seu quotidiano causadas pela pandemia Covid-19. Uma vez que as lojas estiveram praticamente encerradas no primeiro semestre do ano, houve a necessidade de o cliente antecipar o reabastecimento de stock.

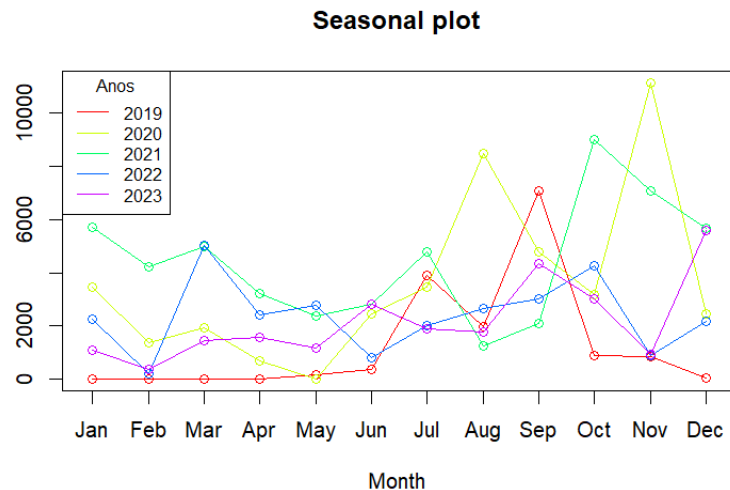


Gráfico 4.15 – Distribuição anual do número de pares vendidos do artigo 8B007, num período mensal de 2019 a 2023.

No Gráfico 4.16 apresenta a distribuição ao longo do tempo das curvas de distribuição das previsões do cliente e das vendas sobrepostas no mesmo período. É possível observar que houve um grande desvio das previsões do cliente, face às vendas, conforme é possível constatar na Figura 4.11 onde o valor da correlação entre variáveis é fraco.

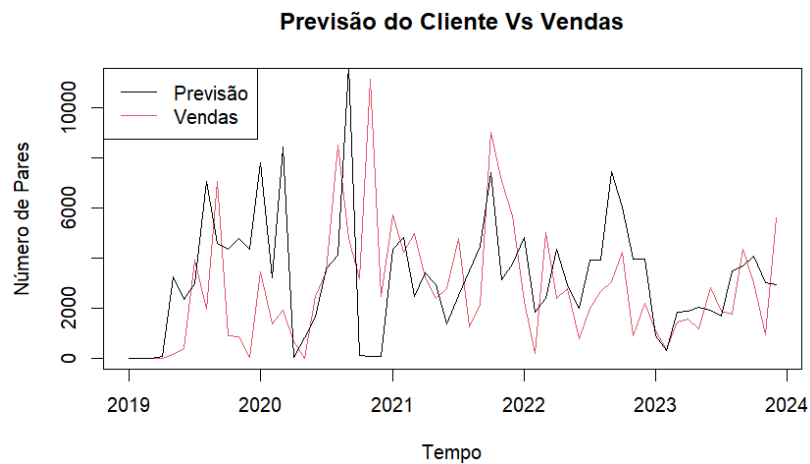


Gráfico 4.16 – Distribuição mensal de 2019 a 2023 dos pares previstos e vendidos para o artigo 8B007.

A Figura 4.11 apresenta o cálculo da correlação das duas séries temporais em estudo, com o valor de 0,28, que corresponde a uma correlação fraca entre as variáveis, mas significativa cujo valor- $p=0,03 < \alpha=0,05$ .

```

> cor.test(Previsão, Vendas)

Pearson's product-moment correlation

data: Previsão and Vendas
t = 2.2097, df = 58, p-value = 0.03109
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.02661776 0.49738926
sample estimates:
cor
0.278659

```

Figura 4.11 - Correlação de Pearson entre as séries sazonais de previsão do cliente e vendas, para o artigo 8B007.

O Gráfico 4.17 apresenta graficamente a correlação cruzada entre as variáveis “vendas” e “previsão” do cliente. Pode-se concluir que no *lag* 0 a correlação é a mais elevada e significativa, sugerindo uma forte correlação entre as duas variáveis, o que era esperado. Contudo, o *lag* -12 é estatisticamente significativo, o que significa que as vendas de há 12 meses têm impacto positivo nas previsões futuras.

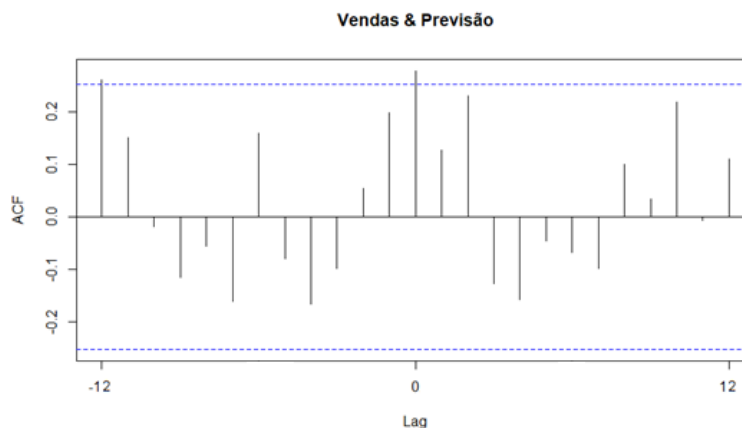


Gráfico 4.17 – Correlação cruzada das vendas e previsões do cliente para o artigo 8B007, num período de 2019 a 2023.

Observa-se, no Gráfico 4.18, que o mês de setembro apresenta a mediana mais alta e o mês de novembro apresenta a distribuição de valores mais alta, o que confirma a procura do cliente neste período do ano, uma vez que se trata de um artigo para uso em caça desportiva. O mês de fevereiro apresenta o valor da mediana mais baixo.

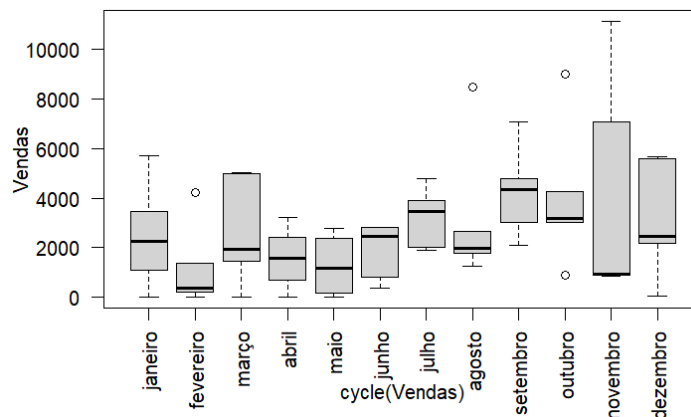


Gráfico 4.18 – Box-plot das vendas mensais do artigo 8B007, num período de 2019 a 2023.

#### 4.4.3. Modelos de Previsão

A partir do teste de Dickey-Fuller conclui-se que a série temporal das vendas não é estacionária, ou seja, não se rejeita a hipótese nula, dado que se obteve o valor- $p=0,1684 > \alpha=0,05$ , conforme resultados apresentados na Figura 4.12.

```
> adf.test(Vendas)

Augmented Dickey-Fuller Test

data: Vendas
Dickey-Fuller = -3.0054, Lag order = 3, p-value = 0.1684
alternative hypothesis: stationary
```

Figura 4.12 – Resultado da aplicação do teste de estacionaridade aplicado no artigo 8B007.

A função de autocorrelação, representada no Gráfico 4.19, apresenta decaimento lento até ao *lag* 5, sendo o *lag* 1 e 2 significativos. No *lag* 17 e 19 observa-se uma inversão para autocorrelações negativas significativas. Este padrão pode indicar a presença de componentes de médias móveis (MA).

Na função de autocorrelação parcial observa-se que as correlações mais fortes e significativas, por se encontrarem fora do intervalo de confiança, são nos *lag* 1 e 17. Após o *lag* 2 há uma tendência dos valores para zero, o que pode indicar a presença de um modelo autorregressivo (AR).

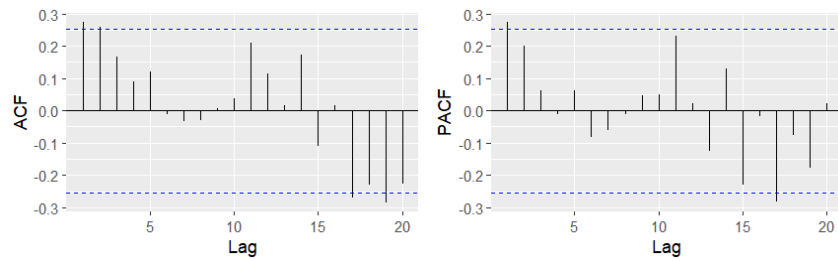


Gráfico 4.19 – Gráficos de autocorrelação e autocorrelação parcial das vendas no artigo 8B007.

A Figura 4.13 apresenta os resultados obtidos após a aplicação da função 'auto.arima' na série temporal 'vendas' (modelo A). Pode-se observar que o resultado é apresentado no formato ARIMA(1,0,1).

```
> a=auto.arima(Vendas)
> summary(a)
Series: Vendas
ARIMA(1,0,1) with non-zero mean

Coefficients:
      ar1      ma1      mean
  0.7717  -0.5339  2713.4163
s.e.  0.1923   0.2458   576.9279

sigma^2 = 5308918; log likelihood = -548.23
AIC=1104.47  AICc=1105.19  BIC=1112.84

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 59.04713 2245.768 1678.137 -Inf  Inf  0.6797862 -0.03369292
```

Figura 4.13 – Resultado da aplicação do modelo A na série temporal 'vendas' do artigo 8B007.

Seguidamente, apesar de haver uma correlação fraca entre as vendas e as previsões do cliente, esta é estatisticamente significativa e por isso aplica-se a função 'auto.arima' (Modelo B), Figura 4.14, na série temporal com a inclusão da variável exógena 'previsões' do cliente. Pode-se observar que o resultado é apresentado no formato ARIMAX(1,0,1).

```
> b=auto.arima(Vendas, xreg=Previsão)
> summary(b)
Series: Vendas
Regression with ARIMA(1,0,1) errors

Coefficients:
      ar1      ma1  intercept      xreg
  0.8266  -0.6346  1910.9871  0.2488
s.e.  0.1503   0.1978   692.5166  0.1244

sigma^2 = 5071591; log likelihood = -546.33
AIC=1102.66  AICc=1103.77  BIC=1113.13

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 54.11371 2175.657 1584.643 -Inf  Inf  0.6419131 -0.04069033
```

Figura 4.14 – Resultado da aplicação do modelo B na série temporal 'vendas' do artigo 8B007 com a inclusão da variável exógena 'previsões' do cliente.

Uma vez que a série temporal 'vendas' não é estacionária e apresenta uma grande variância, conforme é possível observar no mês de novembro do Gráfico 4.18, foi realizada uma transformação Box-Cox na série de modo a remover possíveis *outliers* que possam distorcer as análises e previsões, de modo a estabilizar a variância. De seguida aplicou-se um ajuste sazonal às séries temporais 'vendas' e 'previsão' do cliente, conforme é possível verificar sazonalidade no Gráfico 4.14, ou seja, foi removido o padrão sazonal estimado pela função 'decompose'. Assim, as componentes restantes da série podem ser analisadas mais claramente e, por conseguinte, melhorar a precisão das previsões, tornando assim a tendência da série mais estacionária. No Gráfico 4.20 é possível verificar que houve um ajuste da variância da série temporal, contudo permaneceram *outliers*.

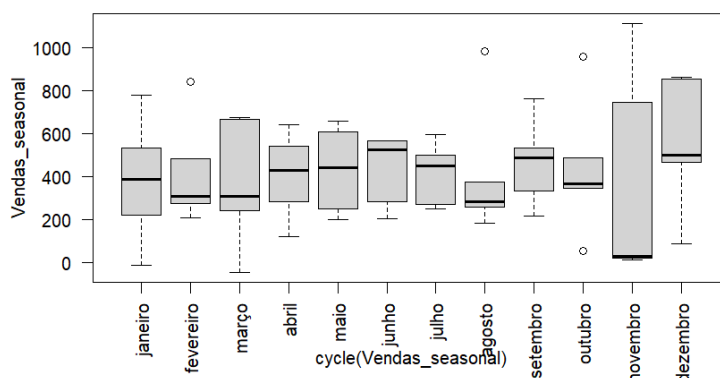


Gráfico 4.20 – Box-plot das vendas após a aplicação da transformação Box-Cox e da diferenciação sazonal para o artigo 8B007.

Procedeu-se à aplicação do teste de Dickey-Fuller à série ajustada e concluiu-se que a série permaneceu não estacionária, uma vez que se rejeita a hipótese nula, valor- $p=0,056 > \alpha=0,05$ , conforme apresentado na Figura 4.15.

```
> adf.test(Vendas_seasonal)

Augmented Dickey-Fuller Test

data: Vendas_seasonal
Dickey-Fuller = -2.9133, Lag order = 3, p-value = 0.2056
alternative hypothesis: stationary
```

Figura 4.15 – Resultado da aplicação do teste de estacionaridade aplicado à série temporal 'vendas' ajustada pela função 'decompose' no artigo 8B007.

Procedeu-se à análise da função de autocorrelação, representada no Gráfico 4.21, onde se observa um decaimento gradual das autocorrelações ao longo das

desfasagens. Este decaimento indica que o padrão de correlação foi suavizado após a aplicação da transformação Box-Cox e da diferenciação sazonal.

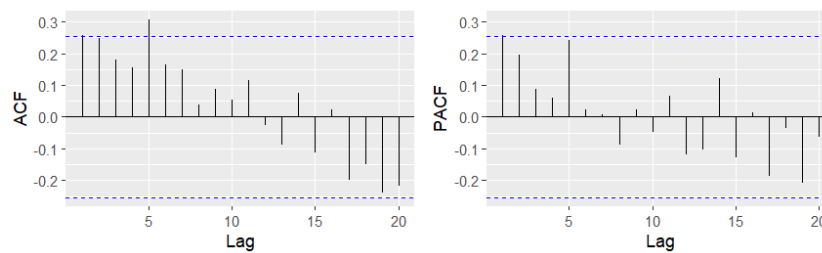


Gráfico 4.21 – Gráficos de autocorrelação e autocorrelação parcial da série temporal ‘vendas’ ajustada do artigo 8B007.

A Figura 4.16 apresenta os resultados obtidos após a aplicação da função ‘auto.arima’ na série temporal ajustada ‘vendas’ (modelo D). Pode-se observar que o resultado é apresentado no formato ARIMA(1,0,1).

```
> d=auto.arima(Vendas_seasonal)
> summary(d)
Series: Vendas_seasonal
ARIMA(1,0,1) with non-zero mean

Coefficients:
      ar1      ma1      mean
  0.9068 -0.7191 400.0785
s.e.  0.0894  0.1298  89.6809

sigma^2 = 62705: log likelihood = -415.16
AIC=838.31 AICc=839.04 BIC=846.69

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 15.17362 244.0696 189.9732 -42.00641 200.0427 0.6247354 -0.04584031
```

Figura 4.16 – Resultado da aplicação do modelo D na série temporal ajustada ‘vendas’ do artigo 8B007.

Seguidamente, uma vez que há uma correlação fraca mas estatisticamente significativa entre as vendas e as previsões do cliente, aplica-se a função ‘auto.arima’ (Modelo C), Figura 4.17/ Figura 4.25, na série temporal ajustada com a inclusão da variável exógena ‘previsões’ do cliente. Pode-se observar que o resultado é apresentado no formato ARIMAX(1,0,1).

```
> c= auto.arima(Vendas_seasonal, xreg=Previsão_seasonal)
> summary(c)
Series: Vendas_seasonal
Regression with ARIMA(1,0,1) errors

Coefficients:
      ar1      ma1  intercept      xreg
  0.8757 -0.7269  434.7297  0.0126
s.e.  0.1109  0.1426   66.2263  0.0069

sigma^2 = 60783: log likelihood = -413.61
AIC=837.22 AICc=838.33 BIC=847.69

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 4.648758 238.1825 190.9023 -96.20158 174.7688 0.6277909 -0.02717908
```

Figura 4.17 – Resultado da aplicação do modelo C na série temporal ajustada ‘vendas’ do artigo 8B007 com a inclusão da variável exógena ‘previsões’ do cliente.

Comparando os quatro modelos aplicados na série temporal em estudo, pode-se concluir que o modelo C é o melhor modelo ajustado. Este apresenta o menor valor de AIC, AICc e BIC, contudo não apresenta menor valor de erro MASE, conforme é possível observar na Tabela 4.2.

Tabela 4.2 - Comparação dos valores calculados pela função 'auto.arima' do R para os quatro modelos aplicados no artigo 8B007.

Modelos:	MASE	AIC	AICc	BIC
A: Série temporal 'vendas' ARIMA(1,0,1)	0,68	1104,47	1105,19	1112,84
B: Série temporal 'vendas' com a inclusão da variável exógena 'previsões'. ARIMAX(1,0,1)	0,64	1102,66	1103,77	1113,13
C: Série temporal 'vendas' ajustada com a inclusão da variável exógena 'previsões' ARIMAX(1,0,1)	0,63	837,22	838,33	847,69
D: Série temporal 'vendas' ajustada ARIMA(1,0,1)	0,62	838,31	839,04	846,69

#### 4.4.4. Análise dos Resíduos do Modelo de Previsão Selecionado

Os resíduos deste modelo foram calculados a partir da função do R 'checkresiduals', conforme Gráfico 4.22.



Gráfico 4.22 – Análise dos resíduos do modelo de previsão ajustado para o artigo 8B007.

Estes resíduos deverão cumprir os seguintes pressupostos:

✓ Independência – A partir do gráfico de ACF é possível observar que não há autocorrelação significativa fora dos limites de confiança de 95%, logo, assume-se que são não correlacionados. A partir do teste Ljung-Box, ao nível de significância de 5% ( $\alpha = 0,05$ ) confirma-se esta análise, não se rejeita a hipótese nula, uma vez que o valor- $p=0,7471 > \alpha$ .

```
> checkresiduals(c)

Ljung-Box test

data: Residuals from Regression with ARIMA(1,0,1) errors
Q* = 6.7682, df = 10, p-value = 0.7471

Model df: 2. Total lags used: 12
```

✓ Normalidade - A distribuição dos resíduos aproxima-se da normalidade, o que indica que o modelo é o adequado, confirmado pelo teste Shapiro-Wilk. Pode-se concluir que os resíduos seguem uma distribuição normal, uma vez que não se rejeita a hipótese nula, o valor- $p=0,07691 > \alpha=5\%$ .

```
> residual <- residuals(c)
> shapiro_test <- shapiro.test(residual)
> print(shapiro_test)

Shapiro-Wilk normality test

data: residual
W = 0.96435, p-value = 0.07691
```

✓ Homoscedasticidade – Analisando a distribuição dos resíduos ao longo do tempo, observa-se que a variância dos resíduos é relativamente constante, mas existem alguns picos esporádicos (eventualmente *outliers*). Apesar destes picos, a variância parece aproximadamente estável, sem um aumento ou diminuição gradual ao longo do tempo. Este comportamento sugere homoscedasticidade, no entanto, os *outliers* podem indicar que certos períodos têm uma variância ligeiramente maior, o que pode motivar uma investigação mais profunda com testes estatísticos. Assim, aplicou-se o teste Breusch-Pagan e confirma-se a homoscedasticidade, com valor- $p=0,2264 > \alpha=5\%$ .

```
> bptest(residual ~ Vendas_seasonal)

studentized Breusch-Pagan test

data: residual ~ Vendas_seasonal
BP = 1.4633, df = 1, p-value = 0.2264
```

✓ Ausência de Tendência ou Sazonalidade - Os resíduos apresentam-se como ruído branco, o que significa que estão espalhados aleatoriamente em torno de zero, sem quaisquer padrões ou tendências aparentes.

Em suma, os resíduos do modelo ajustado cumprem com todos os pressupostos, o que indica que se trata de um ajuste adequado.

#### 4.4.5. Previsão do Modelo Ajustado

Após se ter definido o modelo que melhor se ajusta aos dados, apresentou-se graficamente a partir da função do R 'autoplot(forecast)' (Gráfico 4.23), num período de 12 meses.

Dada a elevada variabilidade da série temporal, caracterizada por um elevado nível de ruído e uma base de dados relativamente curta (com apenas cinco anos), as previsões apresentam inevitavelmente um maior grau de incerteza. Consequentemente, os intervalos de confiança tendem a ser mais amplos, refletindo a limitação na capacidade do modelo em capturar padrões robustos e consistentes ao longo do tempo.

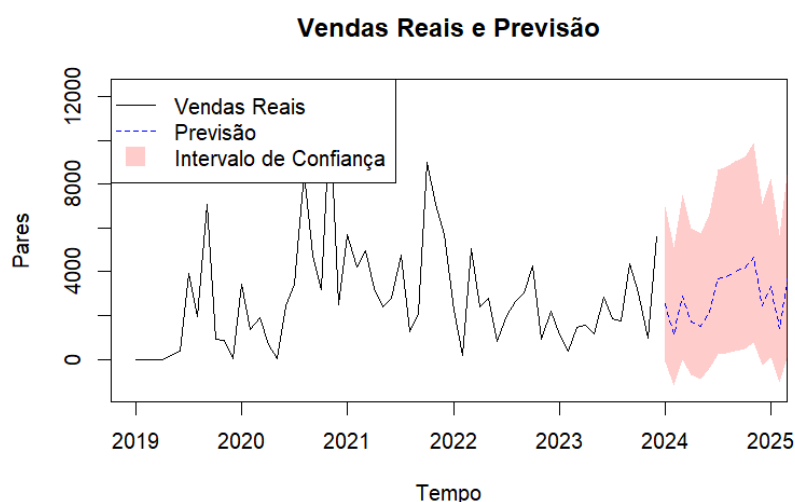


Gráfico 4.23 – Representação gráfica do modelo de previsões para o artigo 8B007.

Assim, foi possível definir a equação matemática correspondente ao modelo ajustado:

Equação Geral do modelo ARIMAX:

$$\phi(B)(1 - B)^d Y_t = c + \theta(B)\varepsilon_t + \beta X_t$$

Equação do modelo selecionado ARIMAX(1,0,1):

- $p = 1: AR(1) = (1 - \phi_1 B)$ ;
- $d = 0: (1 - B^0) = 1$ ;
- $q = 1: MA(1) = (1 + \theta_1 B^1)$ ;

$$(1 - \phi_1 B)Y_t = c + (1 + \theta_1 B^1)\varepsilon_t + \beta X_t \Leftrightarrow$$

$$Y_t = \phi_1 Y_{t-1} + c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \beta X_t$$

Se:

- $\phi_1 = 0,876$
- $\theta_1 = -0,727$
- $c = 434,730$
- $\beta_1 = 0,013$

Então:  $Y_t = 0,876Y_{t-1} + 434,73 + \varepsilon_t - 0,727\varepsilon_{t-1} + 0,013X_t$

## 4.5. Estudo do artigo 8B008

O código de artigo 8B008 corresponde a uma galocha destinada ao uso em jardinagem.

### 4.5.1. Análise exploratória do artigo 8B008

A partir dos resultados da estatística descritiva, Figura 4.18 , para o artigo 8B008, pode-se inferir que no período de fevereiro de 2019 (mês em que foi implementado o artigo em loja) a outubro de 2023 (modelo sem vendas a partir deste momento por motivos de qualidade, bloqueado pelo cliente), o número médio de pares vendidos por mês foi de 14538, o desvio padrão foi de 10222 pares e que estava compreendido entre o número mínimo de 380 e um máximo de 51074 pares. Estes valores tão distantes são justificados, pela variação das vendas ao longo dos meses do ano. Trata-se de um artigo de chuva onde as vendas são mais acentuadas nas estações do ano mais chuvosas.

```
> summary(Vendas)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   380   7361   12226   14538   18508   51074
> sd(Vendas)
[1] 10222.15
```

Figura 4.18 – Estatística descritiva da série temporal dos pares vendidos, para o artigo 8B008.

### 4.5.2. Decomposição e Representação Gráfica das Séries Temporais

O Gráfico 4.24 apresenta a série temporal do número de pares vendidos ao longo do período considerado (*observed*), a decomposição aditiva da série nas componentes (*trend*), a sazonalidade (*seasonal*) e o ruído (*random*). Pode-se constatar que se verificou uma redução de vendas no ano de 2020 e um crescimento de vendas desde 2021 até ao ano de 2023. A redução de vendas a partir de junho de 2023 deveu-se ao facto de haver anomalias produtivas que causaram não conformidades no artigo. Por conseguinte foi suspensa a venda deste artigo até resolução do problema.

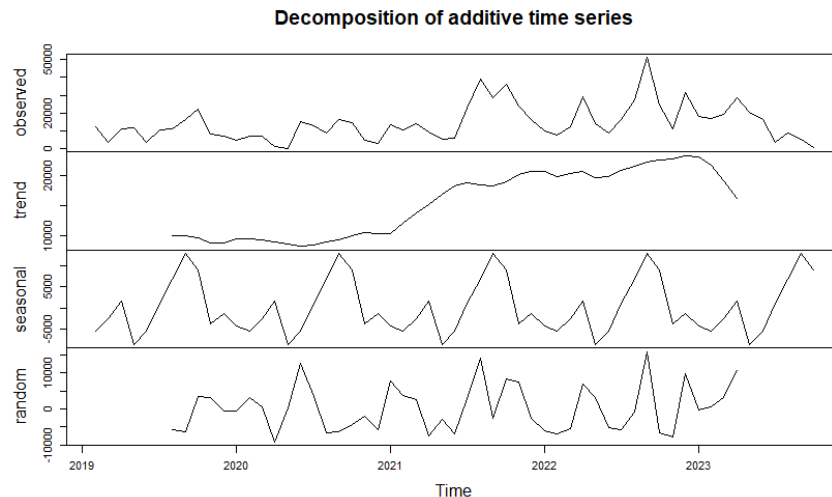


Gráfico 4.24 – Distribuição e decomposição da série temporal relativa ao número de pares vendidos do artigo 8B008, num período mensal de 2019 a 2023.

No Gráfico 4.25 apresenta a sobreposição das curvas de distribuição dos pares vendidos por ano. Verifica-se que as maiores vendas são realizadas no último semestre do ano, justificando a curva de sazonalidade apresentada no Gráfico 4.24.

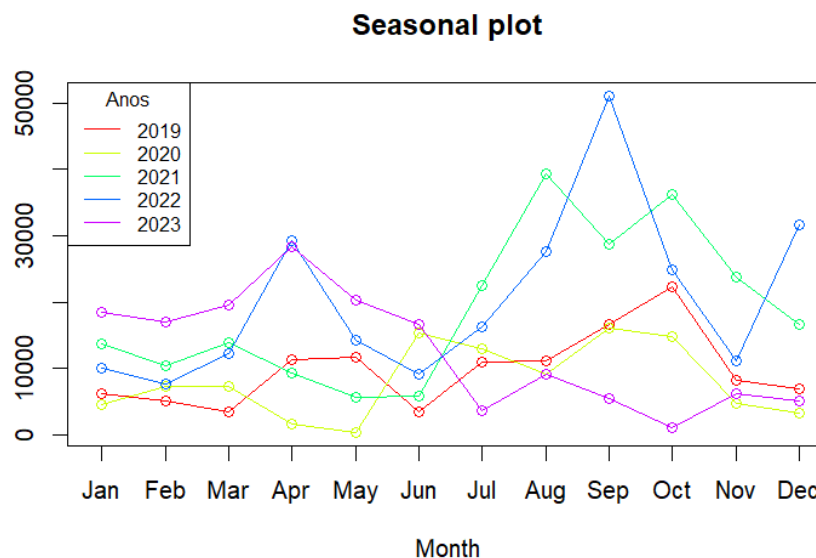


Gráfico 4.25 – Distribuição anual do número de pares vendidos do artigo 8B008, num período mensal de 2019 a 2023.

No Gráfico 4.26 apresenta a distribuição ao longo do tempo das curvas de distribuição das previsões do cliente e das vendas sobrepostas no mesmo período. Pode-se verificar que houve um ligeiro desvio negativo das vendas face às

previsões do cliente, principalmente nos anos de 2019 e 2020. No ano de 2022 o desvio foi positivo, houve mais vendas face ao previsto. A correlação moderada é justificada na Figura 4.19.

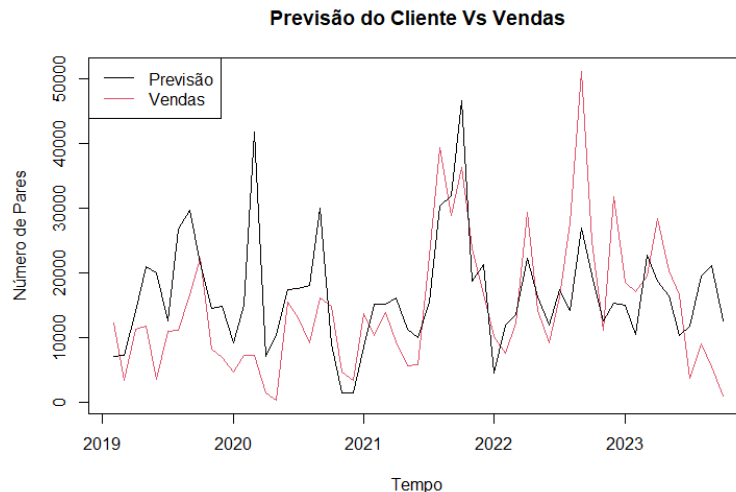


Gráfico 4.26 – Distribuição mensal de 2019 a 2023 dos pares previstos e vendidos para o artigo 8B008.

A Figura 4.19 apresenta o cálculo da correlação das duas séries temporais em estudo, com o valor de 0,53, que corresponde a uma correlação moderada entre as variáveis, mas significativa, cujo valor- $p=0,0000253 < \alpha=0,05$ .

```
> cor.test(Previsão, vendas)

Pearson's product-moment correlation

data: Previsão and vendas
t = 4.633, df = 55, p-value = 2.253e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3123811 0.6945103
sample estimates:
 cor
0.5298231
```

Figura 4.19 - Correlação de Pearson entre as séries sazonais de previsão do cliente e vendas, para o artigo 8B008.

O Gráfico 4.27 apresenta graficamente a correlação cruzada entre as variáveis vendas e previsão do cliente. Pode-se concluir que no *lag* 0 a correlação é a mais elevada e significativa, sugerindo uma relação moderada entre as duas variáveis, o que era esperado. O *lag* significativo -1 sugere que as vendas passadas de um mês têm impacto nas previsões do mês seguinte.

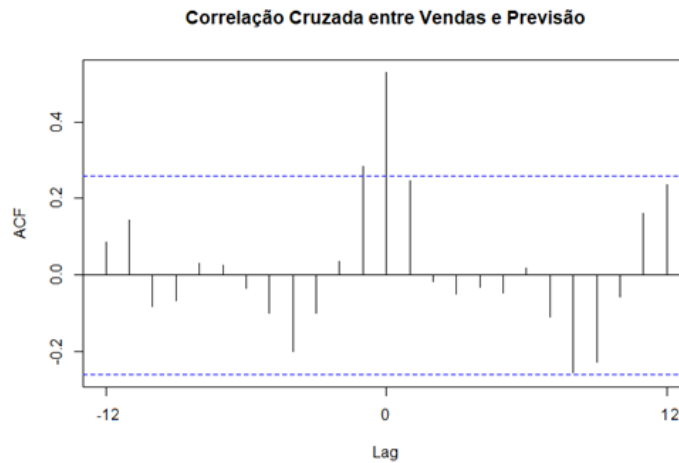


Gráfico 4.27 – Correlação cruzada das vendas e previsões do cliente para o artigo 8B008, num período de 2019 a 2023.

Observa-se, no Gráfico 4.28, que o mês de outubro apresenta a distribuição de valores e a mediana mais alta, seguindo o mês de setembro. Pode-se, assim, inferir que se confirma o dado conhecido, há um claro destaque de procura do cliente neste período do ano. Os meses de junho e novembro apresentam o valor da mediana mais baixo.

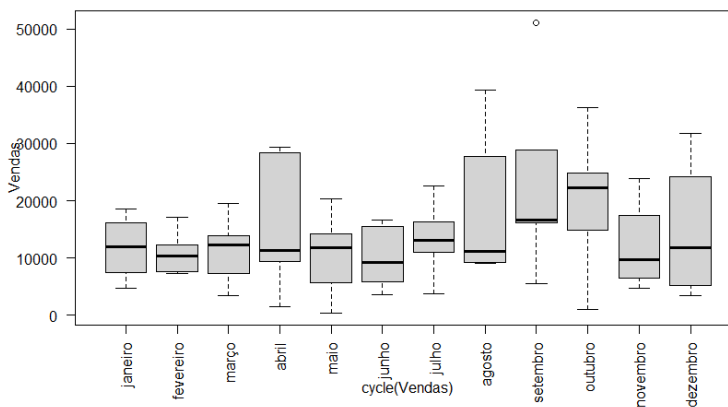


Gráfico 4.28 – Box-plot das vendas mensais do artigo 8B008, num período de 2019 a 2023.

#### 4.5.3. Modelos de Previsão

A partir do teste de Dickey-Fuller conclui-se que a série temporal das vendas não é estacionária, ou seja, não se rejeita a hipótese nula, dado que se obteve o valor- $p=0,2787 > \alpha=0,05$ , conforme resultados apresentados na Figura 4.20.

```

> adf.test(vendas)

Augmented Dickey-Fuller Test

data: vendas
Dickey-Fuller = -2.7334, Lag order = 3, p-value = 0.2787
alternative hypothesis: stationary

```

Figura 4.20 – Resultado da aplicação do teste de estacionaridade aplicado no artigo 8B008.

A função de autocorrelação representada pelo Gráfico 4.29, apresenta o *lag* 1 significativo. No entanto, não há um padrão claro de repetição regular noutros *lag*, o que sugere que não há uma forte indicação de sazonalidade. A função de autocorrelação parcial apresenta o *lag* 1 significativo e depois cai abruptamente nos *lag* seguintes, o que pode indicar a presença de uma estrutura autorregressiva (AR).

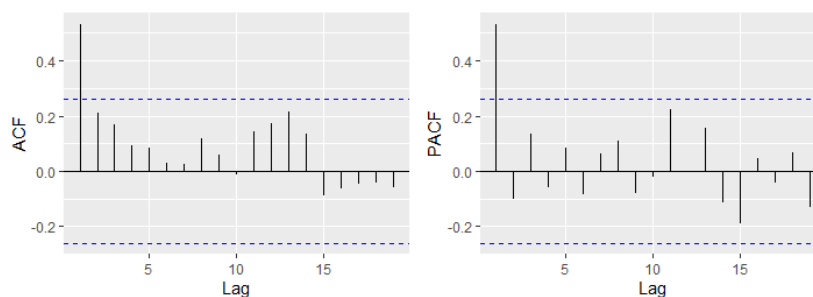


Gráfico 4.29 – Gráficos de autocorrelação e autocorrelação parcial das vendas no artigo 8B008.

A Figura 4.21 apresenta os resultados obtidos após a aplicação da função ‘auto.arima’ na série temporal ‘vendas’ (modelo A). Pode-se observar que o resultado é apresentado no formato ARIMA(0,1,0).

```

> a=auto.arima(vendas)
> summary(a)
Series: vendas
ARIMA(0,1,0)

sigma^2 = 94307353: log likelihood = -593.6
AIC=1189.2 AICc=1189.27 BIC=1191.22

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -195.7503 9625.634 7482.25 -42.92544 78.51991 0.7201118 -0.1561696

```

Figura 4.21 – Resultado da aplicação do modelo A na série temporal ‘vendas’ do artigo 8B008.

Seguidamente, apesar de haver uma correlação moderada entre as vendas e as previsões do cliente, esta é estatisticamente significativa e por isso aplica-se a função ‘auto.arima’ (Modelo B), Figura 4.22, na série temporal com a inclusão da variável exógena ‘previsões’ do cliente. Pode-se observar que o resultado é apresentado no formato ARIMAX(0,1,1).

```

> b=auto.arima(Vendas, xreg=Previsão)
> summary(b)
Series: Vendas
Regression with ARIMA(0,1,1) errors

Coefficients:
      ma1      xreg
      -0.6159  0.5761
s.e.    0.1438  0.1166

sigma^2 = 62780710: log likelihood = -581.43
AIC=1168.85  AICc=1169.31  BIC=1174.93

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -401.2377 7712.098 5790.269 -36.18202 68.01904 0.5572711
              ACF1
Training set 0.09790861

```

Figura 4.22 – Resultado da aplicação do modelo B na série temporal ‘vendas’ do artigo 8B008 com a inclusão da variável exógena ‘previsões’ do cliente.

Uma vez a série temporal ‘vendas’ não é estacionária e aparenta ter padrão sazonal, conforme Gráfico 4.24 mas contrariado pela análise do Gráfico 4.29, foi realizado uma transformação Box-Cox na série de modo a remover possíveis *outliers* que possam distorcer as análises e previsões, de modo a estabilizar a variância. De seguida aplicou-se um ajuste sazonal às séries temporais ‘vendas’ e ‘previsão’, ou seja, foi removido o padrão sazonal estimado pela função ‘decompose’. Assim, as componentes restantes da série podem ser analisadas mais claramente e, por conseguinte, melhorar a precisão das previsões, tornando assim a tendência da série mais estacionária. Conforme o Gráfico 4.30 é possível verificar que houve um ajuste da variância da série temporal, contudo permaneceram *outliers*.

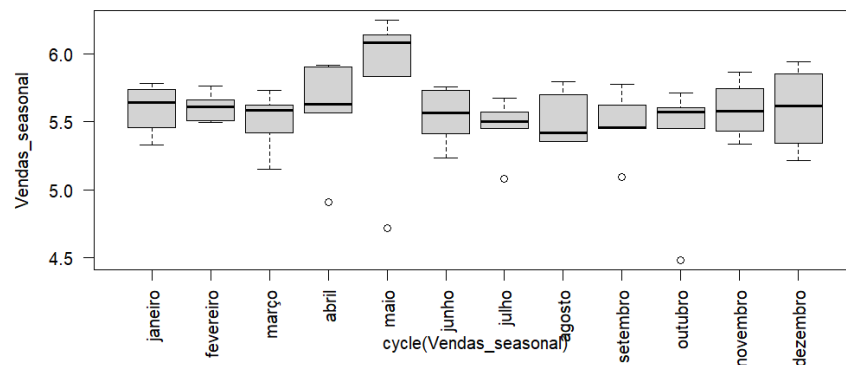


Gráfico 4.30 – Box-plot das vendas após a aplicação da transformação Box-Cox e da diferenciação sazonal para o artigo 8B008.

Procedeu-se à aplicação do teste Dickey-Fuller, à série ajustada e concluiu-se que a série permaneceu não é estacionária, uma vez que ao nível de significância de 5% ( $\alpha=0,05$ ) não se rejeita a hipótese nula, o valor- $p=0,898 > \alpha$ , Figura 4.23.

```
> adf.test(Vendas_seasonal)

Augmented Dickey-Fuller Test

data: Vendas_seasonal
Dickey-Fuller = -1.1985, Lag order = 3, p-value = 0.898
alternative hypothesis: stationary
```

Figura 4.23 – Resultado da aplicação do teste de estacionaridade aplicado à série temporal ‘vendas’ ajustada pela função ‘decompose’ no artigo 8B008.

Procedeu-se à análise da função de autocorrelação, representada no Gráfico 4.31, após a aplicação da transformação Box-Cox e do ajustamento sazonal, há uma autocorrelação significativa no *lag* 1 que decai rapidamente para zero. A ausência de *lag* significativos com indício a padrões sazonais é visível, sugere que a componente sazonal foi removida com sucesso. O decaimento é mais rápido, comparativamente ao Gráfico 4.29, sugerindo que a série está mais estacionária.

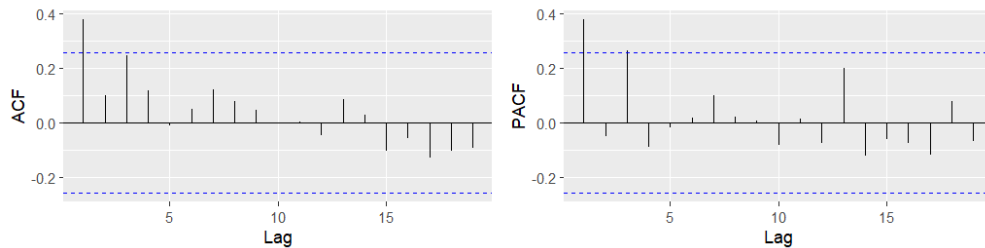


Gráfico 4.31 – Gráficos de autocorrelação e autocorrelação parcial da série temporal ‘vendas’ ajustada do artigo 8B008.

A Figura 4.24 apresenta os resultados obtidos após a aplicação da função ‘auto.arima’ na série temporal ajustada ‘vendas’ (modelo D). Pode-se observar que o resultado é apresentado no formato SARIMA(0,0,1)(0,0,1)[12].

```
> d=auto.arima(Vendas_seasonal)
> summary(d)
Series: Vendas_seasonal
ARIMA(0,0,1)(0,0,1)[12] with non-zero mean

Coefficients:
      ma1      sma1      mean
      0.6211  -0.3915  5.5514
s.e.  0.1037   0.2046  0.0396

sigma^2 = 0.07238: log likelihood = -5.74
AIC=19.49  AICc=20.26  BIC=27.66

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -0.007518696 0.2618564 0.2052596 -0.3722992 3.762837 0.6549765
              ACF1
Training set 0.02015403
```

Figura 4.24 – Resultado da aplicação do modelo D na série temporal ajustada ‘vendas’ do artigo 8B008.

Seguidamente, uma vez que há uma correlação moderada e estatisticamente significativa entre as vendas e as previsões do cliente, aplica-se a função 'auto.arima' (Modelo C), Figura 4.25, na série temporal ajustada com a inclusão da variável exógena 'previsões' do cliente. Pode-se observar que o resultado é apresentado no formato SARIMAX(0,0,1)(0,0,1)[12].

```
> c= auto.arima(Vendas_seasonal, xreg=Previsão_seasonal)
> summary(c)
Series: Vendas_seasonal
Regression with ARIMA(0,0,1)(0,0,1)[12] errors

Coefficients:
      ma1      sma1  intercept      xreg
      0.5736  -0.3506  -951.3167  956.8899
s.e.    0.1116   0.2039   499.1159  499.1275

sigma^2 = 0.07003; log likelihood = -4.02
AIC=18.03  AICc=19.21  BIC=28.25

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -0.006859245  0.2551805  0.195816  -0.3537387  3.596601  0.6248422
              ACF1
Training set  0.01692674
```

Figura 4.25 – Resultado da aplicação do modelo C na série temporal ajustada 'vendas' do artigo 8B008 com a inclusão da variável exógena 'previsões' do cliente.

Comparando os quatro modelos aplicados na série temporal em estudo, pode-se concluir que o modelo C é o melhor modelo ajustado. Este apresenta o menor valor de AIC, AICc e BIC, no entanto não apresenta o menor valor de MASE, conforme é possível observar na Tabela 4.3.

Tabela 4.3 – Comparação dos valores calculados pela função 'auto.arima' do R para os quatro modelos aplicados no artigo 8B008.

Modelos:	MASE	AIC	AICc	BIC
A: Série temporal 'vendas' ARIMA(0,1,0)	0,72	1189,2	1189,27	1191,22
B: Série temporal 'vendas' com a inclusão da variável exógena 'previsões'. ARIMAX(0,1,1)	0,56	1168,85	1169,31	1174,93
C: Série temporal 'vendas' ajustada com a inclusão da variável exógena 'previsões' SARIMAX(0,0,1)(0,0,1)[12]	0,62	18,03	19,21	28,25
D: Série temporal 'vendas' ajustada SARIMA(0,0,1)(0,0,1)[12]	0,65	19,49	20,26	27,66

#### 4.5.4. Análise dos Resíduos do Modelo de Previsão Selecionado

Os resíduos deste modelo foram calculados a partir da função do R 'checkresiduals', conforme Gráfico 4.32.

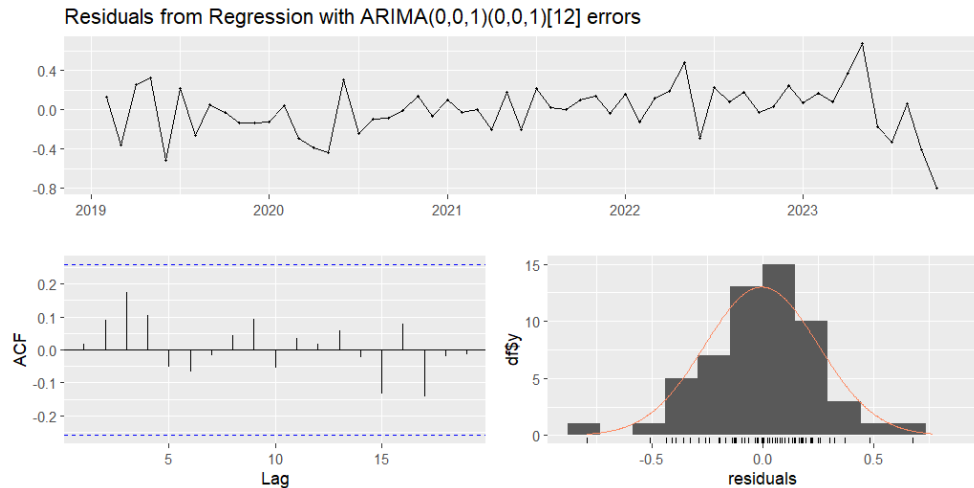


Gráfico 4.32 – Análise dos resíduos do modelo de previsão ajustado para o artigo 8B008.

Estes resíduos deverão cumprir os seguintes pressupostos:

✓ Independência – A partir do gráfico de ACF é possível observar que não há autocorrelação significativa fora dos limites de confiança de 95%, logo, assume-se que são não correlacionados. A partir do teste Ljung-Box, ao nível de significância de 5% ( $\alpha = 0,05$ ) confirma-se esta análise, não se rejeita a hipótese nula, uma vez que o valor- $p=0,8677 > \alpha$ .

```
> checkresiduals(c)
      Ljung-Box test
data:  Residuals from Regression with ARIMA(0,0,1)(0,0,1)[12] errors
Q* = 4.5993, df = 9, p-value = 0.8677
Model df: 2.   Total lags used: 11
```

✓ Normalidade - A distribuição dos resíduos aproxima-se da normalidade, o que indica que o modelo é o adequado, confirmado pelo teste Shapiro-Wilk. Pode-se concluir que os resíduos seguem uma distribuição normal, uma vez que não se rejeita a hipótese nula, o valor- $p=0,5385 > \alpha=5\%$ .

```
> residual <- residuals(c)
> shapiro_test<- shapiro.test(residual)
> print(shapiro_test)
      Shapiro-wilk normality test
data:  residual
W = 0.98168, p-value = 0.5385
```

✓ Homoscedasticidade – Analisando a distribuição dos resíduos ao longo do tempo, observa-se que a variância dos resíduos é relativamente constante, mas existem alguns picos esporádicos (eventualmente *outliers*). Apesar destes picos, a variância parece aproximadamente estável, sem um aumento ou diminuição gradual ao longo do tempo. Este comportamento sugere homoscedasticidade, no entanto, os *outliers* podem indicar que certos períodos têm uma variância ligeiramente maior, o que pode motivar uma investigação mais profunda com testes estatísticos. Assim, aplicou-se o teste Breusch-Pagan e confirma-se a homoscedasticidade, com  $p=0,6515 > \alpha=5\%$ .

```
> bptest(residual ~ Vendas_seasonal)
studentized Breusch-Pagan test
data:  residual ~ Vendas_seasonal
BP = 0.204, df = 1, p-value = 0.6515
```

✓ Ausência de Tendência ou Sazonalidade - Os resíduos apresentam-se como ruído branco, o que significa que estão espalhados aleatoriamente em torno de zero, sem quaisquer padrões ou tendências aparentes, o que indica um ajuste adequado.

Em suma, os resíduos do modelo ajustado cumprem com todos os pressupostos, o que indica que se trata de um ajuste adequado.

#### 4.5.5. Previsão do Modelo Ajustado

Após se ter definido o modelo que melhor se ajusta aos dados, apresentou-se graficamente a partir da função do R 'autoplot(forecast)' (Gráfico 4.33 **Erro! A origem da referência não foi encontrada.**), num período de 12 meses.

Dada a elevada variabilidade da série temporal, caracterizada por um elevado nível de ruído e uma base de dados relativamente curta (com apenas cinco anos), as previsões apresentam inevitavelmente um maior grau de incerteza. Consequentemente, os intervalos de confiança tendem a ser mais amplos, refletindo a limitação na capacidade do modelo em capturar padrões robustos e consistentes ao longo do tempo.

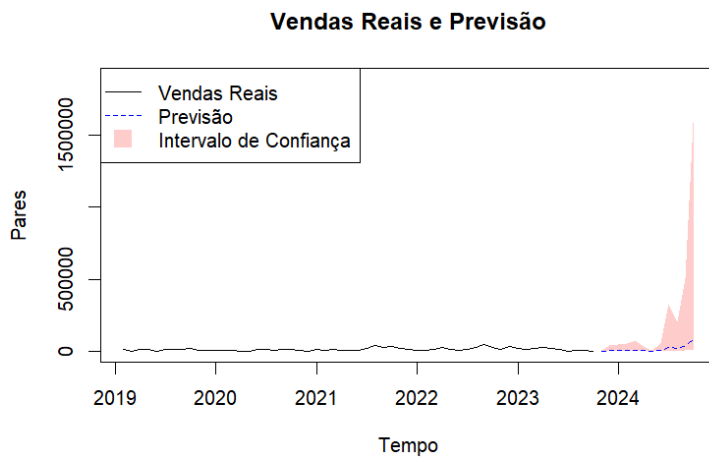
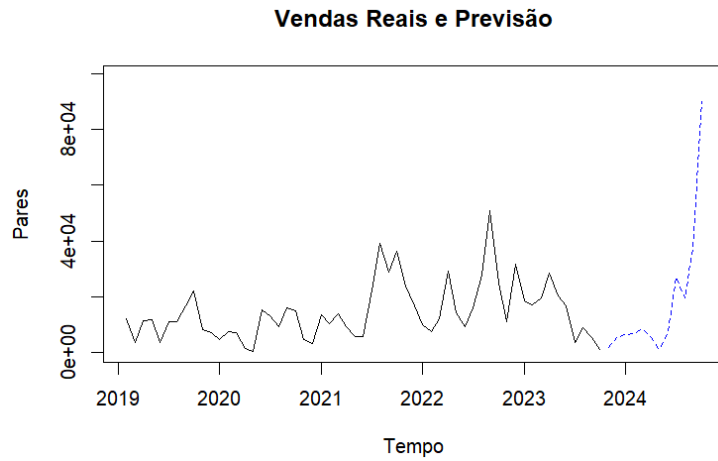


Gráfico 4.33 – Representação gráfica do modelo de previsões para o artigo 8B008 sem e com intervalo de confiança.

Também, foi possível definir a equação matemática correspondente ao modelo ajustado:

Equação Geral do modelo SARIMAX:

$$\Phi(B^s)\phi(B)(1 - B)^d(1 - B^s)^D Y_t = c + \Theta(B^s)\theta(B)\varepsilon_t + \beta X_t$$

Equação do modelo selecionado SARIMAX(0,0,1)(0,0,1)[12]:

- $p = 0: AR(0) = (1 - \phi_0 B) = 1;$
- $d = 0: (1 - B^0) = 1;$
- $q = 1: MA(1) = (1 + \theta_1 B^1);$
- $P = 0: AR(0) = (1 - \Phi_0 B^{0*12}) = 1;$
- $D = 0: (1 - B^{12})^0 = 1.$

○  $Q = 1: MA(1) = (1 + \Theta_1 B^{1*12});$

$$Y_t = c + \Theta(B^S)\theta(B)\varepsilon_t + \beta X_t$$

$$Y_t = c + (1 + \Theta_1 B^{12})(1 + \theta_1 B^1)\varepsilon_t + \beta X_t \Leftrightarrow$$

$$Y_t = c + \varepsilon_t + \Theta_1 \varepsilon_{t-12} + \theta_1 \varepsilon_{t-1} + \Theta_1 \theta_1 \varepsilon_{t-13} + \beta X_t$$

Se:

- $\theta_1 = 0,574$
- $\Theta_1 = -0,351$
- $c = -951,317$
- $\beta_1 = 956,890$

Então:

$$Y_t = -951,317 + \varepsilon_t - 0,351\varepsilon_{t-12} + 0,574\varepsilon_{t-1} - 0,351 * 0,574\varepsilon_{t-13} + 956,89X_t$$

## 4.6. Estudo do artigo 8B010

O código de artigo 8B010 corresponde a uma galocha destinada ao uso em neve.

### 4.6.1. Análise exploratória do artigo 8B010

A partir dos resultados da estatística descritiva, Figura 4.26, para o artigo 8B010, pode-se inferir que no período de janeiro de 2019 a dezembro de 2023, o número médio de pares vendidos por mês foi de 12756, o desvio padrão é de 18722 pares e que estava compreendido entre o número mínimo de 0 e um máximo de 88657 pares. Estes valores tão distantes são justificados, pela variação das vendas ao longo dos meses do ano. Trata-se de um artigo de neve onde as vendas são mais acentuadas nas estações do ano mais frias.

```
> summary(Vendas)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0         0    3602   12756   20443   88657
> sd(Vendas)
[1] 18722.81
```

Figura 4.26 – Estatística descritiva da série temporal dos pares vendidos, para o artigo 8B010.

### 4.6.2. Decomposição e Representação Gráfica das Séries Temporais

O Gráfico 4.34 apresenta a série temporal do número de pares vendidos ao longo do período considerado (*observed*), a decomposição aditiva da série nas componentes (*trend*), a sazonalidade (*seasonal*) e o ruído (*random*). Pode-se constatar que no ano de 2020 houve uma redução de vendas face ao previsto, explicada pelo facto de neste ano a população ter sofrido restrições no seu quotidiano causadas pela pandemia Covid-19. Nos anos seguintes, houve um aumento de vendas até ao final do ano de 2022. No ano de 2023, apesar de haver uma previsão de vendas similares aos restantes anos, o cliente teve a necessidade de adiar algumas quantidades para o início do ano de 2024, uma vez que o frio da Europa ainda não se fazia sentir.

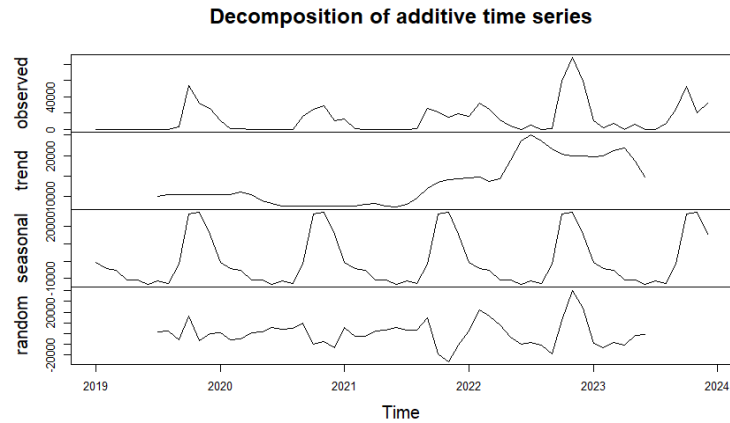


Gráfico 4.34 – Distribuição e decomposição da série temporal relativa ao número de pares vendidos do artigo 8B010, num período mensal de 2019 a 2023.

O Gráfico 4.35 apresenta a sobreposição das curvas de distribuição dos pares vendidos por ano. Pode-se concluir, juntamente com o Gráfico 4.34 que as vendas ocorrem em maior quantidade entre os meses de setembro a março e por isso há uma clara distribuição sazonal deste artigo de calçado.

Seasonal plot

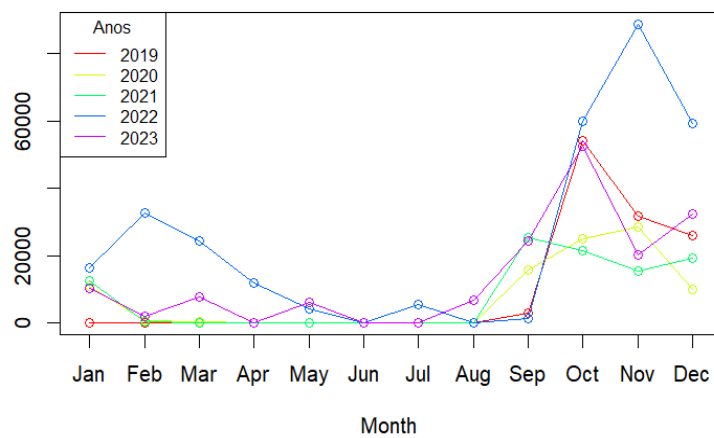


Gráfico 4.35 – Distribuição anual do número de pares vendidos do artigo 8B010, num período mensal de 2019 a 2023.

O Gráfico 4.36 apresenta a distribuição ao longo do tempo das curvas de distribuição das previsões do cliente e das vendas sobrepostas no mesmo período. No ano de 2020, 2021 e 2023 há um desvio claro das vendas face ao previsto, conforme justificado anteriormente.

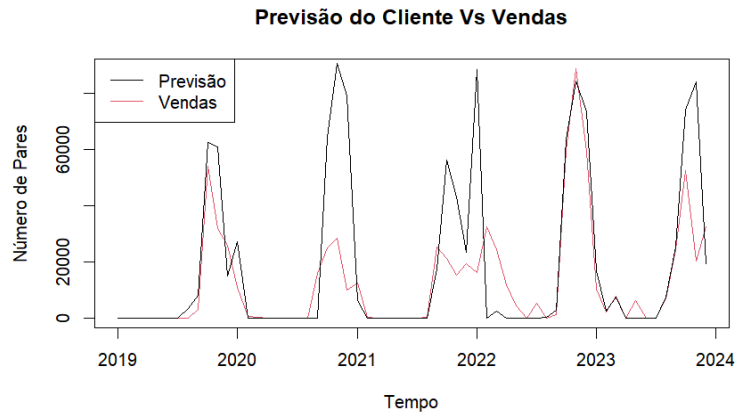


Gráfico 4.36 – Distribuição mensal de 2019 a 2023 dos pares previstos e vendidos para o artigo 8B010.

A Figura 4.27 apresenta o cálculo da correlação das duas séries temporais em estudo, com o valor de 0,74, que corresponde a uma correlação forte entre as variáveis e significativa, valor- $p=2,178^{-11} < \alpha=0,05$ .

```
> cor.test(Previsão, Vendas)

Pearson's product-moment correlation

data: Previsão and Vendas
t = 8.2683, df = 58, p-value = 2.178e-11
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5922183 0.8337409
sample estimates:
cor
0.7355342
```

Figura 4.27 - Correlação de Pearson entre as séries sazonais de previsão do cliente e vendas, para o artigo 8B010.

O Gráfico 4.37 apresenta graficamente a correlação cruzada entre as variáveis “vendas” e “previsão” do cliente. Pode-se concluir que no *lag* 0 a correlação é a mais elevada e significativa, sugerindo uma forte correlação entre as duas variáveis, o que era esperado. Os *lag* significativos, mas de correlação negativa: *lag*  $\pm 5$ ,  $\pm 6$ ,  $\pm 7$  e  $\pm 8$  (correspondente aos meses de maio a agosto, respetivamente) sugerem que as vendas passadas têm impacto negativo nas previsões dos mesmos meses. Já os *lag*  $\pm 1$  (janeiro), 10 (outubro),  $\pm 11$  (novembro) e  $\pm 12$  (dezembro) sugerem que as vendas passadas destes meses têm impacto positivo nas previsões futuras.

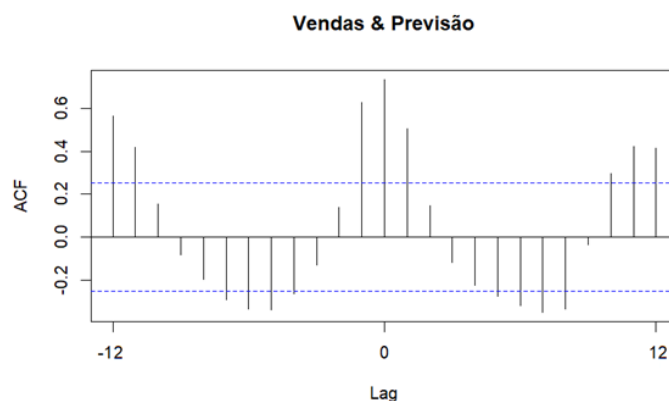


Gráfico 4.37 – Correlação cruzada das vendas e previsões do cliente para o artigo 8B010, num período de 2019 a 2023.

Observa-se, no Gráfico 4.38, que o mês de outubro apresenta a distribuição de valores e a mediana mais alta, seguindo-se os meses de novembro e dezembro. Pode-se, assim, concluir que há um claro destaque da procura do cliente neste período do ano, na medida em que se trata de um artigo para uso no inverno. As condições climatéricas que se fazem sentir na Europa confirmam, no restante período do ano, a falta de procura, por parte do consumidor, deste artigo. Esta afirmação é também confirmada pela curva de sazonalidade apresentada no Gráfico 4.34.

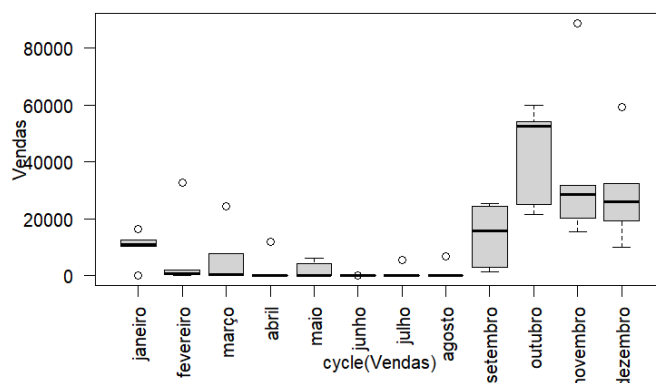


Gráfico 4.38 – Box-plot das vendas mensais do artigo 8B010, num período de 2019 a 2023.

#### 4.6.3. Modelos de Previsão

A partir do teste de Dickey-Fuller conclui-se que a série temporal das vendas é estacionária, ou seja, rejeita-se a hipótese nula, dado que se obteve o valor- $p=0.0139 < \alpha=0,05$ , conforme Figura 4.28.

```

> adf.test(Vendas)

Augmented Dickey-Fuller Test

data: Vendas
Dickey-Fuller = -4.0401, Lag order = 3, p-value = 0.01394
alternative hypothesis: stationary

```

Figura 4.28 – Resultado da aplicação do teste de estacionaridade aplicado no artigo 8B010.

A função de autocorrelação, representada no Gráfico 4.39, apresenta uma estrutura sinusoidal, o que indica haver indícios de sazonalidade com repetição de 12 meses, por este motivo poderá haver componentes de médias móveis (MA). Na função de autocorrelação parcial observa-se que as correlações mais fortes e significativas, por se encontrarem fora do intervalo de confiança são nos *lag* 1 e 2. Este gráfico mostra um pico significativo no *lag* 1 e depois decai abruptamente, o que é um sinal típico de uma componente autorregressivo (AR).

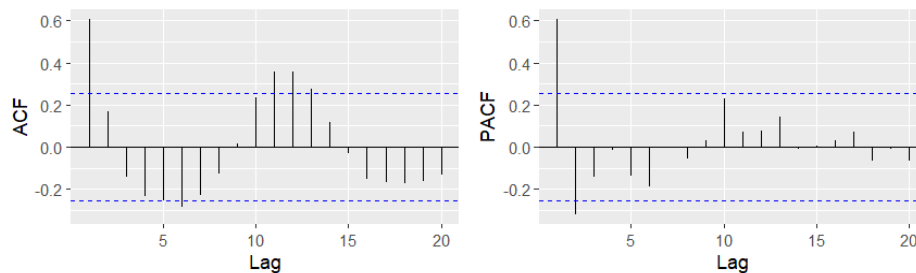


Gráfico 4.39 – Gráficos de autocorrelação e autocorrelação parcial das vendas no artigo 8B010.

A Figura 4.29 apresenta os resultados obtidos após a aplicação da função ‘auto.arima’ na série temporal ‘vendas’ (modelo A). Pode-se observar que o resultado é apresentado no formato SARIMA(2,0,0)(1,1,0)[12], o que significa que se trata de uma série com componente sazonal.

```

> a=auto.arima(Vendas)
> summary(a)
Series: Vendas
ARIMA(2,0,0)(1,1,0)[12]

Coefficients:
      ar1      ar2      sar1
      0.5263 -0.2458 -0.5390
s.e.  0.1402  0.1485  0.1335

sigma^2 = 2.27e+08: log likelihood = -530.55
AIC=1069.11  AICc=1070.04  BIC=1076.59

Training set error measures:
              ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set 1949.473 13048.55 7330.191 NaN  Inf  0.6128015 -0.04600149

```

Figura 4.29 – Resultado da aplicação do modelo A na série temporal ‘vendas’ do artigo 8B010.

Seguidamente, uma vez que há uma correlação forte entre as vendas e as previsões do cliente e esta é estatisticamente significativa, aplica-se a função 'auto.arima' (Modelo B), Figura 4.30, na série temporal com a inclusão da variável exógena 'previsões' do cliente. Pode-se observar que o resultado é apresentado no formato SARIMAX(0,0,0)(0,0,1)[12].

```
> b=auto.arima(Vendas, xreg=Previsão)
> summary(b)
Series: Vendas
Regression with ARIMA(0,0,0)(0,0,1)[12] errors

Coefficients:
      sma1  intercept      xreg
      -0.2831  3593.307   0.4951
s.e.      0.1666  1499.774   0.0460

sigma^2 = 156244255:  log likelihood = -650.11
AIC=1308.21  AICC=1308.94  BIC=1316.59

Training set error measures:
              ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set -155.2862 12183.27 8843.185 -Inf  Inf  0.7392872 0.1567721
```

Figura 4.30 – Resultado da aplicação do modelo B na série temporal 'vendas' do artigo 8B010 com a inclusão da variável exógena 'previsões' do cliente.

Uma vez que a série temporal 'vendas' é estacionária e apresenta uma grande variância, conforme é possível observar no mês de novembro do Gráfico 4.38, foi realizada uma transformação Box-Cox na série de modo a remover possíveis *outliers* que possam distorcer as análises e previsões, de modo a estabilizar a variância. De seguida aplicou-se um ajuste sazonal às séries temporais 'vendas' e 'previsão', que é possível verificar no Gráfico 4.34 sazonalidade, ou seja, foi removido o padrão sazonal estimado pela função 'decompose'. Assim, as componentes restantes da série podem ser analisadas mais claramente e, por conseguinte, melhorar a precisão das previsões, tornando assim a tendência da série mais estacionária. No Gráfico 4.40 é possível verificar que houve um ajuste da variância da série temporal, contudo permaneceram *outliers*.

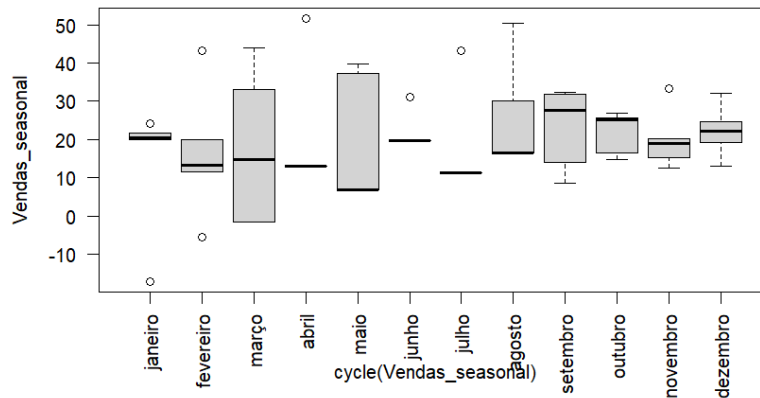


Gráfico 4.40 – Box-plot das vendas após a aplicação da transformação Box-Cox e da diferenciação sazonal para o artigo 8B010.

Procedeu-se à aplicação do teste de Dickey-Fuller à série ajustada e concluiu-se que a série não é estacionária, uma vez não se rejeita a hipótese nula, o valor- $p=0.075 > \alpha=5\%$ , conforme Figura 4.31.

```
> adf.test(Vendas_seasonal)

Augmented Dickey-Fuller Test

data: Vendas_seasonal
Dickey-Fuller = -3.3331, Lag order = 3, p-value = 0.07496
alternative hypothesis: stationary
```

Figura 4.31 – Resultado da aplicação do teste de estacionaridade aplicado à série temporal 'vendas' ajustada pela função 'decompose' no artigo 8B010.

Procedeu-se à análise da função de autocorrelação, representada no Gráfico 4.41, e verifica-se que as *lag* 1, 2 e 3 são significativas o que pode sugerir sazonalidade residual. Enquanto no gráfico da função de autocorrelação parcial, apenas apresenta valores significativos no *lag* 1. O facto de o *lag* cair rapidamente após o *lag* 1 sugere que os dados podem agora seguir um processo autorregressivo (AR).

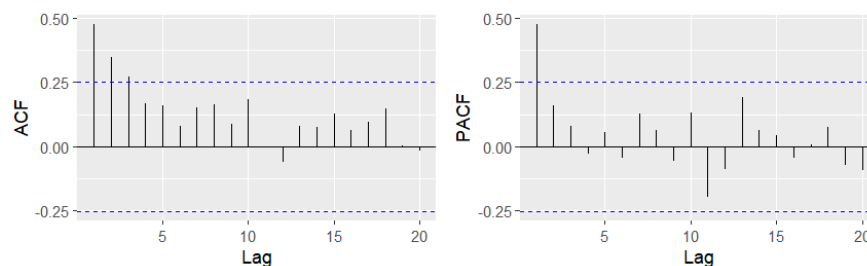


Gráfico 4.41 – Gráficos de autocorrelação e autocorrelação parcial da série temporal 'vendas' ajustada do artigo 8B010.

A Figura 4.32 apresenta os resultados obtidos após a aplicação da função 'auto.arima' na série temporal ajustada 'vendas' (modelo D). Pode-se observar que o resultado é apresentado no formato SARIMA(0,1,1)(0,0,1)[12].

```
> d=auto.arima(Vendas_seasonal)
> summary(d)
Series: Vendas_seasonal
ARIMA(0,1,1)(0,0,1)[12] with drift

Coefficients:
      ma1      sma1      drift
      -0.6306  -0.6377  0.4831
s.e.    0.1328   0.2062  0.2725

sigma^2 = 92.88: log likelihood = -219.21
AIC=446.42  AICc=447.16  BIC=454.73

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.7547034  9.310443  7.391148  4.095951  68.0216  0.5551102  -0.01465915
```

Figura 4.32 – Resultado da aplicação do modelo D na série temporal ajustada 'vendas' do artigo 8B010.

Seguidamente, uma vez que há uma correlação forte e estatisticamente significativa entre as vendas e as previsões do cliente, aplica-se a função 'auto.arima' (Modelo C), Figura 4.33, na série temporal ajustada com a inclusão da variável exógena 'previsões' do cliente. Pode-se observar que o resultado é apresentado no formato ARIMAX(0,1,0).

```
> c= auto.arima(Vendas_seasonal, xreg=Previsão_seasonal)
> summary(c)
Series: Vendas_seasonal
Regression with ARIMA(0,1,0) errors

Coefficients:
      xreg
      0e+00
s.e.    2e-04

sigma^2 = 155.8: log likelihood = -232.14
AIC=468.28  AICc=468.49  BIC=472.43

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.6980422  12.27064  9.851507  -4.591946  71.38775  0.7398948  -0.384642
```

Figura 4.33 – Resultado da aplicação do modelo C na série temporal ajustada 'vendas' do artigo 8B010 com a inclusão da variável exógena 'previsões' do cliente.

Comparando os quatro modelos aplicados na série temporal em estudo, pode-se concluir que o modelo D é o melhor modelo ajustado. Este apresenta o menor valor de AIC, AICc e BIC e também, o menor valor de MASE, conforme é possível observar na Tabela 4.4.

Tabela 4.4 - Comparação dos valores calculados pela função 'auto.arima' do R para os dois modelos aplicados no artigo 8B010.

Modelos:	MASE	AIC	AICc	BIC
A: Série temporal 'vendas' SARIMA(2,0,0)(1,1,0)[12]	0,61	1069,11	1070,04	1076,59
B: Série temporal 'vendas' com a inclusão da variável exógena 'previsões'. SARIMAX(0,0,0)(0,0,1)[12]	0,74	1308,21	1308,94	1316,59
C: Série temporal 'vendas' ajustada com a inclusão da variável exógena 'previsões' ARIMAX(0,1,0)	0,74	468,28	468,49	472,43
D: Série temporal 'vendas' ajustada SARIMA(0,1,1)(0,0,1)[12]	0,56	446,42	447,16	454,73

#### 4.6.4. Análise dos Resíduos do Modelo de Previsão Selecionado

Os resíduos deste modelo foram calculados a partir da função do R 'checkresiduals', conforme Gráfico 4.42.

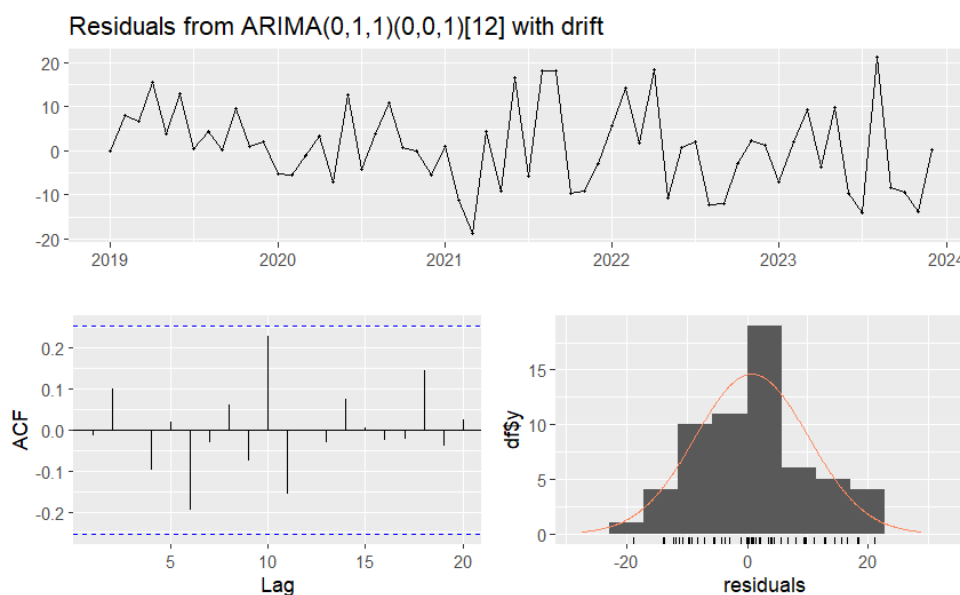


Gráfico 4.42 – Análise dos resíduos do modelo de previsão ajustado para o artigo 8B010.

Estes resíduos deverão cumprir os seguintes pressupostos:

- ✓ Independência – A partir do gráfico de ACF é possível observar que não há autocorrelação significativa fora dos limites de confiança de 95%, logo, assume-se que são não correlacionados. A partir do teste Ljung-Box, ao nível de significância de 5% ( $\alpha = 0,05$ ) confirma-se esta análise, não se rejeita a hipótese nula, uma vez que o valor- $p=0,3996 > \alpha$ .

```
> checkresiduals(d)

Ljung-Box test

data: Residuals from ARIMA(0,1,1)(0,0,1)[12] with drift
Q* = 10.478, df = 10, p-value = 0.3996

Model df: 2. Total lags used: 12
```

- ✓ Normalidade - A distribuição dos resíduos aproxima-se da normalidade, o que indica que o modelo é o adequado, confirmado pelo teste Shapiro-Wilk. Pode-se concluir que os resíduos seguem uma distribuição normal, uma vez que não se rejeita a hipótese nula, o valor- $p=0,2811 > \alpha=5\%$ .

```
> residual <- residuals(d)
> shapiro_test <- shapiro.test(residual)
> print(shapiro_test)

Shapiro-Wilk normality test

data: residual
W = 0.97593, p-value = 0.2811
```

- ✓ Homoscedasticidade – Analisando a distribuição dos resíduos ao longo do tempo, observa-se que a variância dos resíduos é relativamente constante, mas existem alguns picos esporádicos (eventualmente *outliers*). Apesar destes picos, a variância parece aproximadamente estável, sem um aumento ou diminuição gradual ao longo do tempo. Este comportamento sugere homoscedasticidade, no entanto, os *outliers* podem indicar que certos períodos têm uma variância ligeiramente maior, o que pode motivar uma investigação mais profunda com testes estatísticos. Assim, aplicou-se o teste Breusch-Pagan e confirma-se a homoscedasticidade, com valor- $p=0,5386 > \alpha=5\%$ .

```
> bptest(residual ~ Vendas_seasonal)

studentized Breusch-Pagan test

data: residual ~ Vendas_seasonal
BP = 0.3781, df = 1, p-value = 0.5386
```

✓ Ausência de Tendência ou Sazonalidade - Os resíduos apresentam-se como ruído branco, o que significa que estão espalhados aleatoriamente em torno de zero, sem quaisquer padrões ou tendências aparentes, o que indica um ajuste adequado.

Em suma, os resíduos do modelo ajustado cumprem com todos os pressupostos, o que indica que se trata de um ajuste adequado.

#### 4.6.5. Previsão do Modelo Ajustado

Após se ter definido o modelo que melhor se ajusta aos dados, apresentou-se graficamente a partir da função do R 'autoplot(forecast)' (Gráfico 4.43), a variação das previsões num período de 12 meses.

Dada a elevada variabilidade da série temporal, caracterizada por um elevado nível de ruído e uma base de dados relativamente curta (com apenas cinco anos), as previsões apresentam inevitavelmente um maior grau de incerteza. Consequentemente, os intervalos de confiança tendem a ser mais amplos, refletindo a limitação na capacidade do modelo em capturar padrões robustos e consistentes ao longo do tempo.

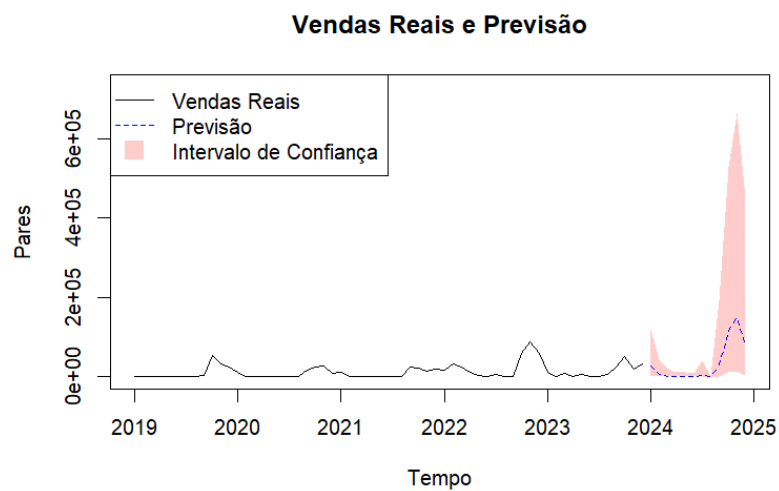
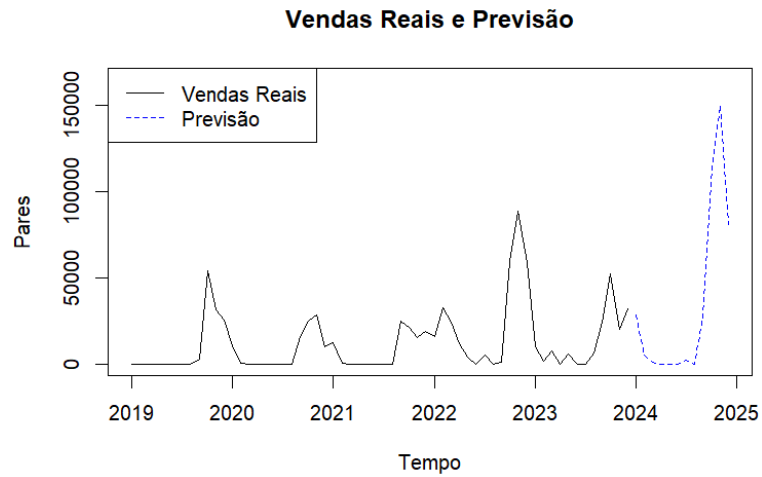


Gráfico 4.43 – Representação gráfica do modelo de previsões para o artigo 8B010 sem e sem intervalo de confiança.

Assim, foi possível definir a equação matemática correspondente ao modelo ajustado:

Equação Geral do modelo SARIMA:

$$\Phi(B^s)\phi(B)(1 - B)^d(1 - B^s)^D Y_t = c + \Theta(B^s)\theta(B)\varepsilon_t$$

Equação do modelo selecionado SARIMA(0,1,1)(0,0,1)[12]

- $p = 0: AR(0) = (1 - \phi_0 B^0) = 1;$
- $d = 1: (1 - B^1);$
- $q = 1: MA(1) = (1 + \theta_1 B^1);$
- $P = 0: AR(0) = (1 - \Phi_0 B^{12}) = 1;$

- $D = 0: (1 - B^{12})^0 = 1;$
- $Q = 1: MA(1) = (1 + \Theta_1 B^{12}).$

$$(1 - B)^d Y_t = c + \Theta(B^s)\theta(B)\varepsilon_t$$

$$(1 - B^1)Y_t = c + (1 + \Theta_1 B^{12})(1 + \theta_1 B^1)\varepsilon_t \Leftrightarrow$$

$$Y_t - Y_{t-1} = c + \varepsilon_t + \Theta_1 \varepsilon_{t-12} + \theta_1 \varepsilon_{t-1} + \Theta_1 \theta_1 \varepsilon_{t-13}$$

Se:

- $\Theta_1 = -0,631$
- $\theta_1 = -0,638$
- $c = \text{N\~{a}o Aplicado}$

Ent\~{a}o:

$$Y_t = Y_{t-1} + \varepsilon_t - 0,631\varepsilon_{t-12} - 0,638\varepsilon_{t-1} + 0,631 * 0,638\varepsilon_{t-13}$$

## 4.7. Estudo do artigo 8B014

O código de artigo 8B014 corresponde a uma galocha destinada ao uso em equitação.

### 4.7.1. Análise exploratória do artigo 8B014

A partir dos resultados da estatística descritiva, Figura 4.34, para o artigo 8B014, pode-se inferir que no período de janeiro de 2019 a dezembro de 2023, o número médio de pares vendidos por mês foi de 14442, o desvio padrão foi de 8211 pares e que estava compreendido entre o número mínimo de 398 e um máximo de 37062 pares. Estes valores tão distantes são justificados, pela variação das vendas ao longo dos meses do ano. Trata-se de um artigo de chuva onde as vendas do cliente são mais acentuadas nas estações do ano mais chuvosas.

```
> summary(Vendas)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   398   9506   13776   14442   18312   37062
> sd(Vendas)
[1] 8211.195
```

Figura 4.34 – Estatística descritiva da série temporal dos pares vendidos, para o artigo 8B014.

### 4.7.2. Decomposição e Representação Gráfica das Séries Temporais

O Gráfico 4.44 apresenta a série temporal do número de pares vendidos ao longo do período considerado (*observed*), e a decomposição aditiva da série nas componentes (*trend*), a sazonalidade (*seasonal*) e o ruído (*random*). Pode-se constatar que visualmente o artigo aparenta ser sazonal a cada 12 meses, com crescimento do número de pares vendidos ao longo dos anos. Contudo, no ano de 2022 houve uma redução de vendas devido à necessidade de o cliente controlar o excesso de stock em armazém. De realçar que, após este equilíbrio, verificou-se uma venda controlada no ano de 2023.

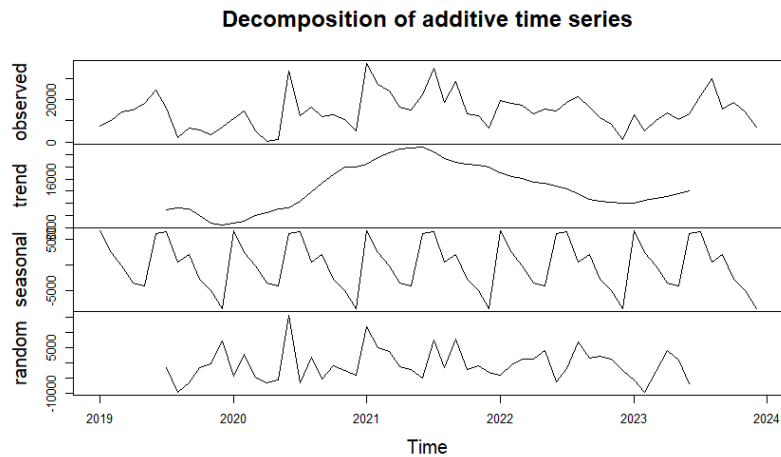


Gráfico 4.44 – Distribuição e decomposição da série temporal relativa ao número de pares vendidos do artigo 8B014, num período mensal de 2019 a 2023.

O Gráfico 4.45 apresenta a sobreposição das curvas de distribuição dos pares vendidos por ano. Destaca-se que é um artigo de venda anual, mas com maior incidência nos meses de junho a setembro (respetivo a vendas do fornecedor-ALOFT).

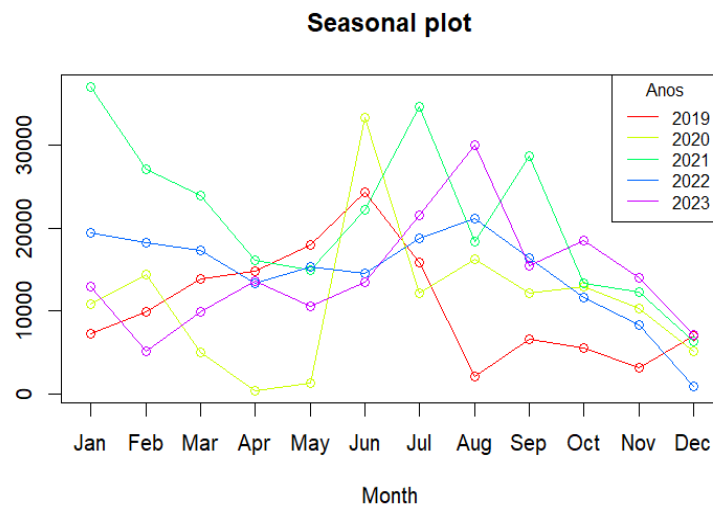


Gráfico 4.45 – Distribuição anual do número de pares vendidos do artigo 8B014, num período mensal de 2019 a 2023.

No Gráfico 4.46 apresenta a distribuição ao longo do tempo das curvas de distribuição das previsões do cliente e das vendas sobrepostas no mesmo período. É notório um desvio das vendas face ao previsto pelo cliente, justificada pela correlação moderada apresentada na Figura 4.35.

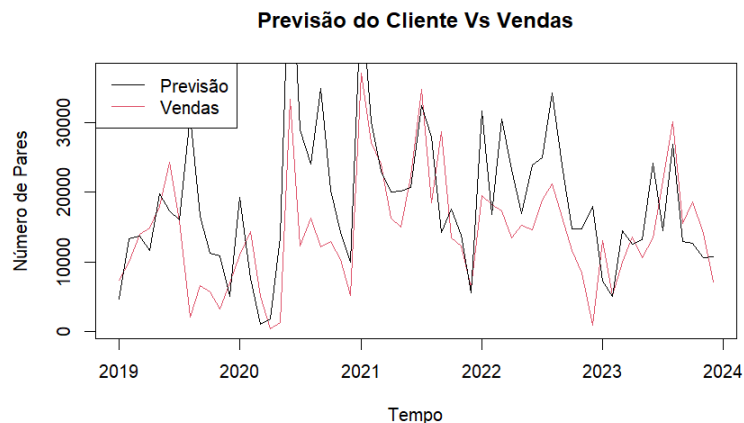


Gráfico 4.46 – Distribuição mensal de 2019 a 2023 dos pares previstos e vendidos para o artigo 8B014.

A Figura 4.35 apresenta o cálculo da correlação das duas séries temporais em estudo, com o valor de 0,65, que corresponde a uma correlação moderada entre variáveis, mas significativa valor- $p=1,5^{-8} < \alpha=0,05$ .

```
> cor.test(Previsão, Vendas)

Pearson's product-moment correlation

data: Previsão and Vendas
t = 6.5769, df = 58, p-value = 1.5e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4792045 0.7783450
sample estimates:
 cor
0.6535982
```

Figura 4.35 - Correlação de Pearson entre as séries sazonais de previsão do cliente e vendas, para o artigo 8B014.

O Gráfico 4.47 apresenta graficamente a correlação cruzada entre as variáveis “vendas” e “previsão” do cliente. Pode-se concluir que no *lag* 0 a correlação é a mais elevada significativa, sugerindo uma moderada correlação entre as duas variáveis, o que era esperado. Os *lag* significativos com valores positivos: 1 (janeiro) e 12 (dezembro) sugerem que as vendas passadas terão um impacto positivo nas previsões de 1 e 12 meses.

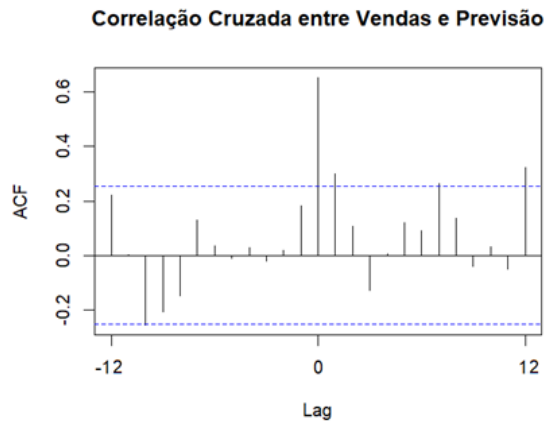


Gráfico 4.47 – Correlação cruzada das vendas e previsões do cliente para o artigo 8B014, num período de 2019 a 2023.

Observa-se, no Gráfico 4.48, que o mês de junho apresenta a distribuição de valores e a mediana mais alta, seguindo-se os meses de julho, agosto e setembro com medianas aproximadas. Pode-se, assim, afirmar que há um claro destaque da procura do cliente neste período do ano. Esta tendência deve-se ao facto de a venda ser direcionada para o chamado “regresso às aulas”. O mês de dezembro apresenta o valor da mediana mais baixo.

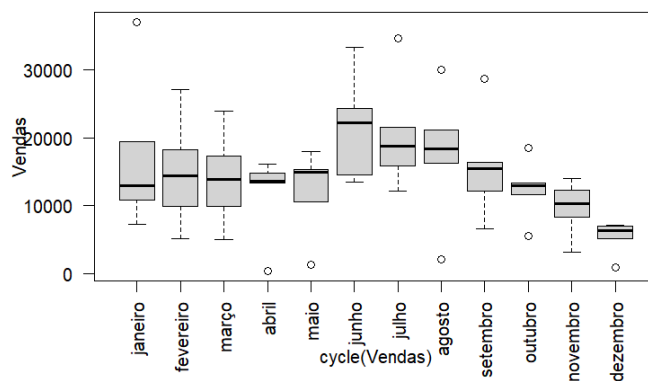


Gráfico 4.48 – Box-plot das vendas mensais do artigo 8B014, num período de 2019 a 2023.

#### 4.7.3. Modelos de Previsão

A partir do teste de Dickey-Fuller conclui-se que a série temporal das vendas não é estacionária, ou seja, não se rejeita a hipótese nula, dado que se obteve o valor- $p=0.1328 > \alpha$ , conforme resultados apresentados na Figura 4.36.

```

> adf.test(Real) #teste de estacionaridade
Augmented Dickey-Fuller Test
data: Real
Dickey-Fuller = -3.0935, Lag order = 3, p-value = 0.1328
alternative hypothesis: stationary

```

Figura 4.36 – Resultado da aplicação do teste de estacionaridade aplicado no artigo 8B014.

A função de autocorrelação, representada no Gráfico 4.49, denota o efeito sazonal, com repetição em 12 meses, justificado pelo *lag* 12 significativo da função de autocorrelação parcial. Observa-se ainda que o *lag* 14 é significativo, o que pode indicar a existência de evidências de ciclos adicionais.

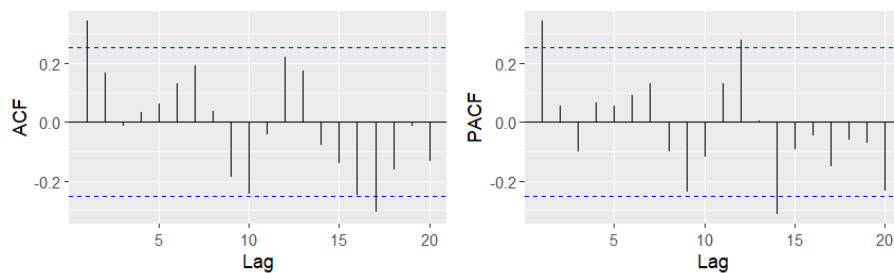


Gráfico 4.49 – Gráficos de autocorrelação e autocorrelação parcial das vendas no artigo 8B014.

A Figura 4.37 apresenta os resultados obtidos após a aplicação da função ‘auto.arima’ na série temporal ‘vendas’ (modelo A). Pode-se observar que o resultado é apresentado no formato SARIMA(2,0,0)(1,0,0)[12], o que significa que se trata de séries com componente sazonal.

```

> a=auto.arima(Vendas)
> summary(a)
Series: Vendas
ARIMA(2,0,0)(1,0,0)[12] with non-zero mean

Coefficients:
      ar1      ar2      sar1      mean
      0.2633  0.2347  0.3568 13913.974
s.e.    0.1266  0.1396  0.1350  2542.116

sigma^2 = 54739149: log likelihood = -618.55
AIC=1247.09  AICC=1248.21  BIC=1257.57

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 231.0189  7147.718  4911.468 -99.01032  119.5521  0.6035753  0.01318372

```

Figura 4.37 – Resultado da aplicação do modelo A na série temporal ‘vendas’ do artigo 8B014.

Seguidamente, uma vez que há uma correlação moderada entre as vendas e as previsões do cliente e esta é estatisticamente significativa, aplica-se a função ‘auto.arima’ (Modelo B), Figura 4.38, na série temporal com a inclusão da variável exógena ‘previsões’ do cliente. Pode-se observar que o resultado é apresentado no formato ARIMAX(1,0,0).

```

> b=auto.arima(Vendas, xreg=Previsão)
> summary(b)
Series: Vendas
Regression with ARIMA(1,0,0) errors

Coefficients:
      ar1  intercept    xreg
      0.2670  5411.999  0.4838
s.e.    0.1249  1767.601  0.0775

sigma^2 = 37111469:  log likelihood = -606.52
AIC=1221.04  AICc=1221.76  BIC=1229.41

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 1.907801  5937.667  4410.278  -79.27737  100.0184  0.5419836  -0.03392478

```

Figura 4.38 – Resultado da aplicação do modelo B na série temporal ‘vendas’ do artigo 8B014 com a inclusão da variável exógena ‘previsões’ do cliente.

Uma vez que a série temporal ‘vendas’ não é estacionária e apresenta uma grande variância e ruído em todos os meses, conforme é possível observar no mês de novembro do Gráfico 4.48, foi realizada uma transformação Box-Cox na série de modo a remover possíveis *outliers* que possam distorcer as análises e previsões, de modo a estabilizar a variância. De seguida, aplicou-se um ajuste sazonal às séries temporais ‘vendas’ e ‘previsão’, como é possível verificar no Gráfico 4.44 sazonalidade, ou seja, foi removido o padrão sazonal estimado pela função ‘decompose’. Assim, as componentes restantes da série podem ser analisadas mais claramente e, por conseguinte, melhorar a precisão das previsões, tornando assim a tendência da série mais estacionária. No Gráfico 4.50 é possível verificar que houve um ajuste da variância da série temporal, contudo permaneceram *outliers*.

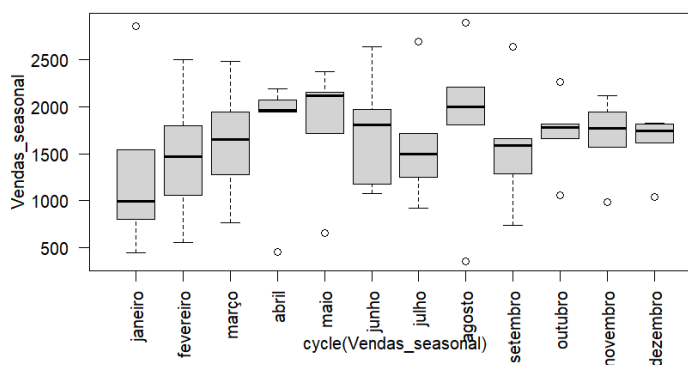


Gráfico 4.50 – Box-plot das vendas após a aplicação da transformação Box-Cox e da diferenciação sazonal para o artigo 8B014.

Procedeu-se à aplicação do teste Dickey-Fuller à série ajustada e concluiu-se que a série permaneceu não estacionária, uma vez que ao nível de significância de 5% ( $\alpha=0,05$ ) não se rejeita a hipótese nula, o valor- $p=0.2703 > \alpha$ , Figura 4.39.

```
> adf.test(Vendas_seasonal)

Augmented Dickey-Fuller Test

data: Vendas_seasonal
Dickey-Fuller = -2.7532, Lag order = 3, p-value = 0.2703
alternative hypothesis: stationary
```

Figura 4.39 – Resultado da aplicação do teste de estacionaridade aplicado à série temporal ‘vendas’ ajustada pela função ‘decompose’ no artigo 8B014.

Procedeu-se à análise da função de autocorrelação, representada no Gráfico 4.51, sendo possível denotar que o *lag* 1 é significativo e que há um decaimento gradual dos coeficientes de autocorrelação. Após a transformação Box-Cox, seguido da diferenciação sazonal, houve uma menor persistência de correlação nos *lag* subsequentes, o que sugere que as tendências sazonais podem ter sido reduzidas ou removidas.

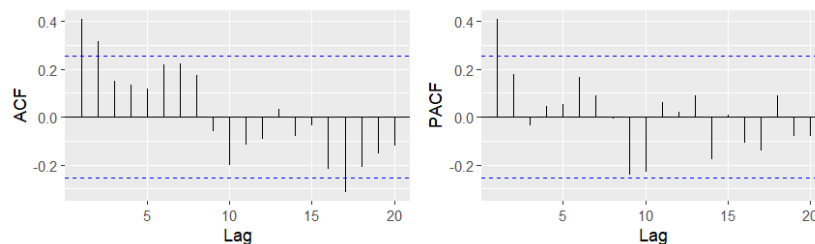


Gráfico 4.51 – Gráficos de autocorrelação e autocorrelação parcial da série temporal ‘vendas’ ajustada do artigo 8B014.

A Figura 4.40 apresenta os resultados obtidos após a aplicação da função ‘auto.arima’ na série temporal ajustada ‘vendas’ (modelo D). Pode-se observar que o resultado é apresentado no formato SARIMA(1,0,0)(0,0,1)[12].

```
> d=auto.arima(Vendas_seasonal)
> summary(d)
Series: Vendas_seasonal
ARIMA(1,0,0)(0,0,1)[12] with non-zero mean

Coefficients:
      ar1      sma1      mean
      0.4032  -0.7566  1671.1338
s.e.    0.1277   0.4061   57.1457

sigma^2 = 269625; log likelihood = -463.71
AIC=935.42  AICC=936.15  BIC=943.8

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -37.96755  506.1064  373.9493 -18.79365  34.29879  0.5097734 -0.05039011
```

Figura 4.40 – Resultado da aplicação do modelo D na série temporal ajustada ‘vendas’ do artigo 8B014.

Seguidamente, uma vez que há uma correlação moderada entre as vendas e as previsões do cliente, mas estatisticamente significativa, aplica-se a função 'auto.arima' (Modelo C), Figura 4.41, na série temporal ajustada com a inclusão da variável exógena 'previsões' do cliente. Pode-se observar que o resultado é apresentado no formato SARIMAX(1,0,0)(0,0,1)[12].

```
> c = auto.arima(Vendas_seasonal, xreg=Previsão_seasonal)
> summary(c)
Series: Vendas_seasonal
Regression with ARIMA(1,0,0)(0,0,1)[12] errors

Coefficients:
      ar1      sma1  intercept      xreg
  0.3718 -0.5224  958.2353  0.0368
s.e.  0.1226  0.2517  166.8262  0.0082

sigma^2 = 226903: log likelihood = -455.02
AIC=920.04  AICc=921.15  BIC=930.51

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -19.0179 460.1913 351.408 -15.68067 31.63081 0.4790447 -0.06289081
```

Figura 4.41 – Resultado da aplicação do modelo C na série temporal ajustada 'vendas' do artigo 8B014 com a inclusão da variável exógena 'previsões' do cliente.

Comparando os quatro modelos aplicados na série temporal em estudo, pode-se concluir que o modelo C é o melhor modelo ajustado. Este apresenta o menor valor de AIC, AICc e BIC e também, o menor valor de MASE, conforme é possível observar na Tabela 4.5.

Tabela 4.5 – Comparação dos valores calculados pela função 'auto.arima' do R para os quatro modelos aplicados no artigo 8B014.

Modelos:	MASE	AIC	AICc	BIC
A: Série temporal 'vendas' SARIMA(2,0,0)(1,0,0)[12]	0,60	1247,09	1248,21	1257,57
B: Série temporal 'vendas' com a inclusão da variável exógena 'previsões'. ARIMAX(1,0,0)	0,54	1221,04	1221,76	1229,41
C: Série temporal 'vendas' ajustada com a inclusão da variável exógena 'previsões' SARIMAX(1,0,0)(0,0,1)[12]	0,48	920,04	921,15	930,51
D: Série temporal 'vendas' ajustada SARIMA(1,0,0)(0,0,1)[12]	0,51	935,42	936,15	943,80

#### 4.7.4. Análise dos Resíduos do Modelo de Previsão Selecionado

Os resíduos deste modelo foram calculados a partir da função do R 'checkresiduals', conforme Gráfico 4.52.

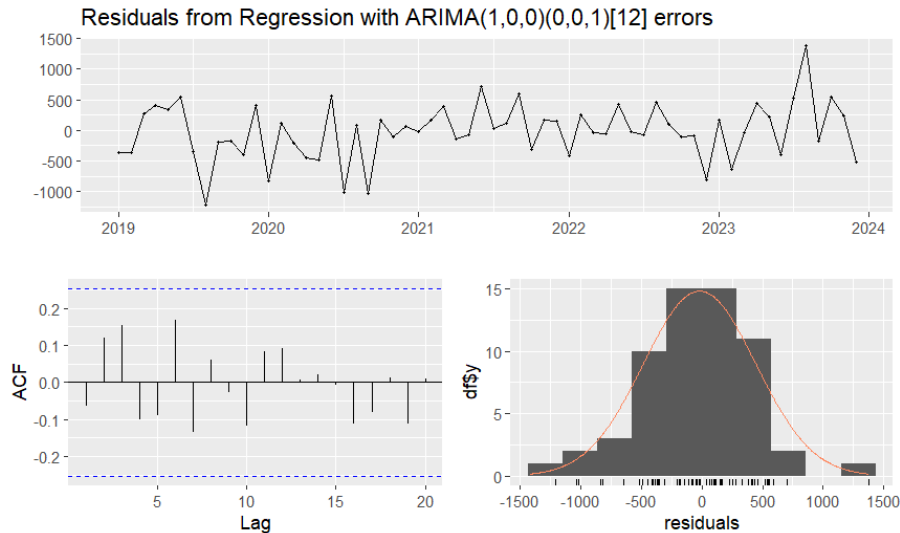


Gráfico 4.52 – Análise dos resíduos do modelo de previsão ajustado para o artigo 8B014.

Estes resíduos deverão cumprir os seguintes pressupostos:

✓ Independência – A partir do gráfico de ACF é possível observar que não há autocorrelação significativa fora dos limites de confiança de 95%, logo, assume-se que são não correlacionados. A partir do teste Ljung-Box, ao nível de significância de 5% ( $\alpha = 0,05$ ) confirma-se esta análise, não se rejeita a hipótese nula, uma vez que o valor- $p=0,4703 > \alpha$ .

```
> checkresiduals(c)

Ljung-Box test

data: Residuals from Regression with ARIMA(1,0,0)(0,0,1)[12] errors
Q* = 9.666, df = 10, p-value = 0.4703

Model df: 2. Total lags used: 12
```

✓ Normalidade - A distribuição dos resíduos aproxima-se da normalidade, o que indica que o modelo é o adequado, confirmado pelo teste shapiro-Wilk. Pode-se concluir que os resíduos seguem uma distribuição normal, uma vez que não se rejeita a hipótese nula, o valor- $p=0,2754 > \alpha$ .

```
> residual <- residuals(c)
> shapiro_test<- shapiro.test(residual)
> print(shapiro_test)

Shapiro-wilk normality test

data: residual
W = 0.97574, p-value = 0.2754
```

✓ Homoscedasticidade – Analisando a distribuição dos resíduos ao longo do tempo, observa-se que a variância dos resíduos é relativamente constante, mas existem alguns picos esporádicos (eventualmente *outliers*). Apesar destes picos, a variância parece aproximadamente estável, sem um aumento ou diminuição gradual ao longo do tempo. Este comportamento sugere homoscedasticidade, no entanto, os *outliers* podem indicar que certos períodos têm uma variância ligeiramente maior, o que pode motivar uma investigação mais profunda com testes estatísticos. Assim, aplicou-se o teste Breusch-Pagan e confirma-se a homoscedasticidade, com  $p=0,6028 > \alpha=5\%$ .

```
> bptest(residual ~ Vendas_seasonal)
      studentized Breusch-Pagan test
data:  residual ~ Vendas_seasonal
BP = 0.27081, df = 1, p-value = 0.6028
```

✓ Ausência de Tendência ou Sazonalidade - Os resíduos apresentam-se como ruído branco, o que significa que estão espalhados aleatoriamente em torno de zero, sem quaisquer padrões ou tendências aparentes, o que indica um ajuste adequado.

Em suma, os resíduos do modelo ajustado cumprem com todos os pressupostos, o que indica que se trata de um ajuste adequado.

#### 4.7.5. Previsão do Modelo Ajustado

Após se ter definido o modelo que melhor se ajusta aos dados, apresentou-se graficamente a partir da função do R 'autoplot(forecast)' (Gráfico 4.53), a variação das previsões num período de 12 meses.

Dada a elevada variabilidade da série temporal, caracterizada por um elevado nível de ruído e uma base de dados relativamente curta (com apenas cinco anos), as previsões apresentam inevitavelmente um maior grau de incerteza. Consequentemente, os intervalos de confiança tendem a ser mais amplos, refletindo a limitação na capacidade do modelo em capturar padrões robustos e consistentes ao longo do tempo.

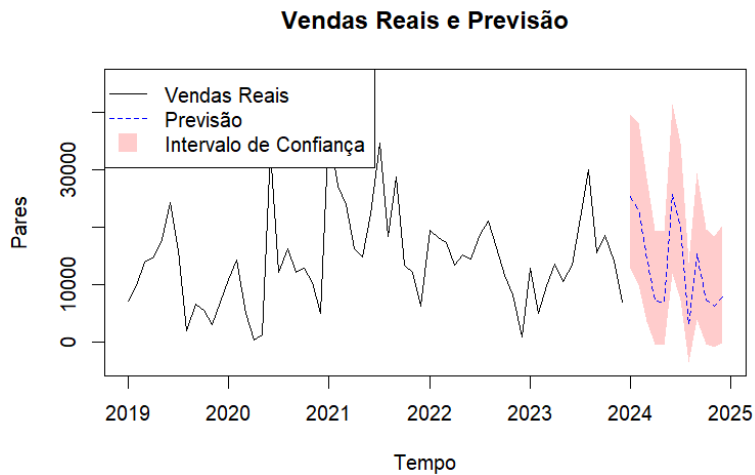


Gráfico 4.53 – Representação gráfica do modelo de previsões para o artigo 8B014.

Assim, foi possível definir a equação matemática correspondente ao modelo ajustado:

Equação Geral do modelo SARIMAX:

$$\Phi(B^s)\phi(B)(1 - B)^d(1 - B^s)^D Y_t = c + \Theta(B^s)\theta(B)\varepsilon_t + \beta X_t$$

Equação do modelo selecionado SARIMAX(1,0,0)(0,0,1)[12]:

- $p = 1: AR(1) = (1 - \phi_1 B)$ ;
- $d = 0: (1 - B^0) = 1$ ;
- $q = 0: MA(0) = (1 + \theta_0 B^0) = 1$ ;
- $P = 0: AR(0) = (1 - \Phi_0 B^{0*12}) = 1$ ;
- $D = 0: (1 - B^{12})^0 = 1$ .
- $Q = 1: MA(1) = (1 + \Theta_1 B^{1*12})$ ;

$$\phi(B)Y_t = c + \Theta(B^s)\varepsilon_t + \beta X_t$$

$$(1 - \phi_1 B)Y_t = c + (1 + \Theta_1 B^{12})\varepsilon_t + \beta X_t \Leftrightarrow$$

$$Y_t = \phi_1 Y_{t-1} + c + \varepsilon_t + \Theta_1 \varepsilon_{t-12} + \beta X_t$$

Se:

- $\phi_1 = 0,372$
- $\theta_1 = -0,522$
- $c = 958,235$
- $\beta_1 = 0,037$

Então:

$$Y_t = 0,372Y_{t-1} + 958,235 + \varepsilon_t - 0,522\varepsilon_{t-12} + 0,037X_t$$

## 5. Conclusão

O presente trabalho teve como principal objetivo modelar e prever as vendas de calçado para o ano 2024, com base no número de pares vendidos mensalmente no período de 2019 a 2023, para cinco produtos diferentes, da produção da empresa ALOFT. Num segundo momento, procurou-se incorporar no modelo a informação das previsões mensais fornecidas pelos clientes. Para atingir estes objetivos, foram aplicados modelos ARIMA e modelos de regressão dinâmica, que se mostraram adequados para captar as dinâmicas temporais da série de dados. De modo a garantir que a série temporal atenda aos pressupostos de validade dos modelos estatísticos, foram aplicados aos dados a transformação Box-Cox e a diferenciação sazonal. Os resultados demonstraram que a combinação destas técnicas aumenta significativamente a precisão preditiva e a robustez dos modelos para séries temporais. Para os artigos 8B001, 8B007, 8B008 e 8B014 o modelo de regressão dinâmica foi o mais adequado enquanto que para o artigo 8B010 foi o modelo autorregressivo ARIMA.

O estudo apresentou algumas limitações tais como o facto de a base de dados ser composta por apenas cinco anos, restringindo a robustez dos modelos. Adicionalmente, as vendas do ano de 2020 mostraram-se atípicas devido ao impacto que a pandemia provocou na sociedade, alterando assim significativamente o padrão de vendas. Como consequência, os intervalos de confiança gerados pelas previsões foram amplos, o que aumenta o risco de que as previsões sejam influenciadas por flutuações não controladas, reduzindo a confiabilidade dos resultados no suporte à tomada de decisão.

Como trabalhos futuros, sugere-se a incorporação de variáveis adicionais à análise que possam complementar a robustez modelar, como por exemplo o destino de cada venda e ainda a aplicação de novos métodos de previsão e comparação com os modelos já utilizados, tais como o método Prophet e o método ETS. Estes métodos têm a vantagem de detetar sazonalidade e tendências e são robustos na situação de dados em falta (missing data) e na presença de outliers.

Para finalizar, será importante comparar as previsões ajustadas pelo modelo com as vendas reais no ano 2024, a fim de validar a precisão dos modelos propostos.

## Bibliografia

- Alpuim, T. (1998). *Séries Temporais*. Associação dos Estudantes da Faculdade de Ciências da Universidade de Lisboa, 2ª Ed.
- Bauer, A. (2020). *Automated Hybrid Time Series Forecasting: Design, Benchmarking, and Use Cases*. Dissertation, Julius-Maximilians-Universität Würzburg.
- Baum, C. (2014). *Heteroskedasticity*. Boston College, Chestnut Hill MA, USA.
- Carmo, C., & Silva, J. (2023). *Machine Learning and Provision of Data Storage Services: Metrics for Analysis and Validation of Predictive Algorithms*. Artigo, GETEC, v.12, n.38, p.123-144/2023.
- Cunha, T. (2022). *Métodos de Previsão: aplicação prática no volume de negócios de uma empresa de consultoria*. Dissertação de Mestrado em Economia e Administração de Empresas, Faculdade de Economia da Universidade do Porto.
- De Sousa, A. I. (2013). *Modelação e Previsão de Vendas no Setor do Retalho de Calçado*. Dissertação de mestrado em logística, Associação de Politécnicos do Norte (APNOR).
- Falaturia, T., Darbaniana, F., Brandtnera, P., & Udokwua, C. (2022). *Predictive Analytics for Demand Forecasting - A Comparison of SARIMA and LSTM in Retail SCM*. *Procedia Computer Science*, 200, 993-1003. <https://doi.org/10.1016/j.procs.2022.01.298>
- Hyndman, R., & Athanasopoulos;George. (2018). *Forecasting: Principles and Practice* (Texts, Ed.; 2nd ed.).
- Kaipov, I., & Nedzved, A. (2020). *Sales Forecasting of Goods in Shoe Retail*. *Central European Researchers Journal*, Vol.6 Issue 1.
- Lima, S. (2018). *Métodos de previsão de séries temporais — uma aplicação a dados do segmento do Retalho*. Dissertação de Mestrado em Estatísticas, Universidade do Minho.

- Lopes, R. (2019). *Utilização de Modelos Estatísticos e Machine Learning Para a Previsão de Vendas no Sector do Retalho - Um Estudo Comparativo*. Dissertação de Mestrado em Ciências Empresariais, Lisbon School of Economics & Management.
- Martins, M. E. (2014). Coeficiente de correlação amostral. *Revista de Ciência Elementar*. Revista de Ciência Elementar.
- Ramos, P., Santos, N., & Rebelo, R. (2014). *Performance of State Space and ARIMA Models for Consumer Retail Sales Forecasting*. Artigo Volume 34, Pages 151-163.
- Veríssimo, F. (2016). *Modelo de Séries Temporais Hierárquicas de Previsão de Vendas Aplicado à Indústria do Calçado*. Dissertação de Mestrado em Modelação, Análise de Dados e Sistemas de Apoio à Decisão, Faculdade de Economia do Porto.
- Vieira, M. (2022). *Previsão dos Indicadores do COVID-19*. Dissertação de Mestrado em Métodos Quantitativos para Decisões Económicas e Empresariais, Lisbon School of Economics & Management.
- W.N.Venables, D. M. S., & R core Team. (2023). *An Introduction to R*. <https://cran.r-project.org/doc/manuals/R-intro.pdf>

## Anexos e Apêndices

### Anexo I

```
library(TSstudio)
library(forecast)
library(tseries)
library(seastests)
library(dynlm)
library(lmtest)
# Leitura dos dados
Dados <- read.csv("Dados_8b010.csv", stringsAsFactors=TRUE)
# Converter o objeto em série temporal
Vendas<- ts(as.vector(Dados$Quantidade), start=c(2019,1),
end=c(2023,12),frequency = 12)
Previsão <-ts(as.vector(Dados$Previsao), start=c(2019,1),
end=c(2023,12),frequency = 12)
# Análise descritiva
Summary(Vendas)
sd(Vendas)
#Gráfico_decomposição
decomp_Vendas <-decompose(Vendas)
plot(decomp_Vendas)
#Gráfico_Boxplot variável "Vendas"
label=c("janeiro","fevereiro","março","abril","maio","junho",
"julho","agosto","setembro","outubro","novembro","dezembro")
boxplot(Vendas~cycle(Vendas), names = label, las=2)
#Gráfico de distribuição anual
seasonplot(Vendas, 12, col=rainbow(12), year.labels=TRUE, main="Sea-
sonal plot")
#Teste de correlação de pearson entre variáveis
cor.test(Previsão, Vendas)
#correlação cruzada entre duas variáveis
ccf(Previsão, Vendas, type="correlation")
#Gráfico_Distribuição Vendas vs Previsão
plot.ts(Vendas, main="Previsão do Cliente Vs Vendas", ylab="Número
de Pares", xlab="Tempo", col=2, pch=21)
lines(Previsão, lwd=1, col=1)
legend("topleft",c("Previsão","Vendas"),col=1:2,lty=1)
#Teste de estacionaridade
adf.test(Vendas)
#Gráfico de Autocorrelação e Autocorrelação Parcial
ggsdisplay(Vendas)
```

```

#Modelo A
a=auto.arima(Vendas)
summary(a)
#Modelo B - Regressão Dinâmica, com regressor externo(xreg=Previsão)
b=auto.arima(Vendas, xreg=Previsão)
summary(b)
#Transformação box-cox para as variáveis "Vendas" e "Previsão"
lambdaV <- BoxCox.lambda(Vendas)
Vendas_BC=BoxCox(Vendas, lambdaV)
lambdaP <- BoxCox.lambda(Previsão)
Previsao_BC=BoxCox(Previsão, lambdaP)#Diferenciação Sazonal
#Diferenciação Sazonal após transformação box-cox
Vendas_seasonal<-seasadj(decompose(Vendas_BC,"additive"))
Previsão_seasonal<-seasadj(decompose(Previsao_BC,"additive"))
#Gráfico_Boxplot após transformação box-cox
boxplot(Vendas_seasonal~cycle(Vendas_seasonal),names = label, las=2)
#Teste de estacionaridade na variável diferenciada
adf.test(Vendas_seasonal)
#Gráfico de ACF e PACF na variável diferenciada
ggttsdisplay(Vendas_seasonal)
#Modelo C - Regressão Dinâmica de variáveis diferenciadas, com re-
gressor externo(xreg= Previsão_seasonal)
c= auto.arima(Vendas_seasonal, xreg=Previsão_seasonal)
summary(c)
#Modelo D de variável diferenciada
d=auto.arima(Vendas_seasonal)
summary(d)
#Análise de Resíduos do método selecionado (d)
checkresiduals(d)
#teste da normalidade dos resíduos
residual <- residuals(d)
shapiro_test<- shapiro.test(residual)
print(shapiro_test)
#teste da homoscedasticidade dos resíduos
Bptest(residual ~ Vendas_seasonal # se modelo C ou D
bptest(residual ~ Vendas) # se modelo A ou B
#Previsões
forecasted_seasonal <- forecast(d, h=12)
# Reintroduzir a sazonalidade
sazonalidade <- tail(decompose(Vendas_BC, type = "additive")$seaso-
nal, frequency(Vendas))
forecasted_with_seasonal <- forecasted_seasonal$mean + rep(sazonali-
dade, length.out=length(forecasted_seasonal$mean))
forecasted_with_seasonal_upper <- forecasted_seasonal$upper[,2] +
rep(sazonalidade, length.out=length(forecasted_seasonal$upper[,2]))
forecasted_with_seasonal_lower <- forecasted_seasonal$lower[,2] +
rep(sazonalidade, length.out=length(forecasted_seasonal$lower[,2]))

```

```

# Inverter a transformação Box-Cox
forecasted_original <- InvBoxCox(forecasted_with_seasonal, lambdaV)
forecasted_original_upper <- InvBoxCox(forecasted_with_seasonal_upper, lambdaV)
forecasted_original_lower <- InvBoxCox(forecasted_with_seasonal_lower, lambdaV)
# Gráfico de Previsões
plot(Vendas, main = "Vendas Reais e Previsão", ylab = "Pares",
      xlab = "Tempo", xlim = c(time(Vendas)[1], max(time(Vendas))+ 1),
      ylim = range(c(Vendas, forecasted_original_upper,
                    forecasted_original_lower)) * 1.1)
polygon(c(time(forecasted_original), rev(time(forecasted_original))),
        c(forecasted_original_upper, rev(forecasted_original_lower)),
        col = rgb(1, 0, 0, 0.2), border = NA)
lines(forecasted_original, col = "blue", lty = 2)
legend("topleft", legend = c("Vendas Reais", "Previsão", "Intervalo
de Confiança"), col = c("black", "blue", rgb(1, 0, 0, 0.2)), lty =
c(1, 2, NA), pch = c(NA, NA, 15), pt.cex = 2)

```