



Universidade Federal de Uberlândia
Faculdade de Matemática

Bacharelado em Estatística

**ANÁLISE DE SÉRIES TEMPORAIS
PARA PREVISÃO DO FATURAMENTO
MENSAL DE UMA EMPRESA**

Tiago Rosa Marques Soares

Uberlândia-MG

2019

Tiago Rosa Marques Soares

**ANÁLISE DE SÉRIES TEMPORAIS
PARA PREVISÃO DO FATURAMENTO
MENSAL DE UMA EMPRESA**

Trabalho de conclusão de curso apresentado à Coordenação do Curso de Bacharelado em Estatística como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientadora: Prof^a. Dra. Maria Imaculada de Sousa Silva

Uberlândia-MG

2019



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, _____ de _____ de 20_____

BANCA EXAMINADORA

Prof^a. Dra. Maria Imaculada de Sousa Silva

Prof. Dr. Pedro Franklin Cardoso Silva

Prof. Dr. Ednaldo Carvalho Guimarães

**Uberlândia-MG
2019**

AGRADECIMENTOS

Agradeço, primeiro a Deus pelas conquistas, pelos aprendizados, pela força, pela saúde e por sempre me dar esperanças para nunca desistir dos meus sonhos, que sempre estarão sendo sustentados por Ele.

À minha mãe Maria Soares e ao meu pai Gesse James, que apesar de todas as dificuldades, sempre me ofereceram com o pouco que tinham, minha educação e incentivo para conquistar meus objetivos. Agradeço a essas duas pessoas que mais estiveram presente em minha vida, e que me mostraram o real valor de buscar o que queremos ser.

Meus agradecimentos sinceros aos meus amigos de faculdade, que me apoiaram nesta jornada incrível, onde conheci pessoas que tenho profunda gratidão, carinho e respeito, serão levadas e lembradas assim como grandes amigas por toda minha vida.

À minha amada, Ana Clara, que compartilhou comigo este momento e me ajudou bastante dando dicas e apoio moral para o desenvolvimento e conclusão deste trabalho.

À esta universidade, ao curso de Estatística e seu corpo docente, direção e administração que me propiciaram todo conhecimento, assegurado pela confiança no mérito e ética aqui presentes.

À minha orientadora neste trabalho árduo, Maria Imaculada, pelo suporte no pouco tempo que lhe coube, pelas suas correções, incentivos, momentos de alegria não só durante o desenvolvimento do trabalho, mas também durante minha jornada como graduando.

A todos que estiveram presentes em momentos de minha graduação, o meu muito obrigado.

RESUMO

Dentro do mercado atual, pode-se dizer que a empresa que conhece seus números e que conte com um bom planejamento estratégico, torna-se uma empresa que esteja preparada para mudanças no mercado, tendo portanto mais chances de obter sucesso. O objetivo deste estudo foi obter a previsão para o faturamento mensal de uma empresa atacadista, identificar características inerentes à séries em estudo, como tendência, sazonalidade e ruído branco por meio de testes específicos, observar a presença e os efeitos de possíveis intervenções e ajustar modelos de séries temporais que representaram os dados adequadamente. A metodologia utilizada neste estudo foi a identificação e estimação de um modelo conforme definido por Box e Jenkins mais especificamente, um modelo sazonal representado por $SARIMA(p, d, q)(P, D, Q)$. Sendo possível que vários modelos sejam ajustados para o mesmo conjunto de dados, foi possível utilizar de testes de hipóteses e técnicas de comparação de modelos, como o erro percentual médio absoluto. Após a definição do modelo adequado para a série temporal, as previsões se mostraram realistas quanto aos valores observados da série.

Palavras-chave: Séries Temporais, Box e Jenkins, Previsão, Sarima .

ABSTRACT

Within the current market, it can be said that the company that knows its numbers and that has a good strategic planning, becomes a company that is prepared for changes in the market, and therefore more likely to success. The objective of this study is to obtain the forecast for the monthly invoicing of a wholesale company, to identify characteristics inherent to the series under study, such as trend, seasonality and white noise by means of specific tests, observe the presence and effects of possible interventions, and adjust time series models that represent the data appropriately. The methodology used in this study is the identification and estimation of a model as defined by Box and Jenkins, more specifically, a seasonal model represented by $SARIMA(p, d, q)(P, D, Q)$. If it is possible that several models are adjusted for the same data set, we intend to use hypothesis tests and model comparison techniques, such as the absolute mean percentage error. After defining the appropriate model for the time series, the forecasts were realistic about the observed values of the series.

Keywords: Time Series, Box and Jenkins, Forecast, Sarima.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	III
1 Introdução	1
2 Fundamentação Teórica	3
2.1 Modelos de Séries Temporais	4
2.2 Teste para Tendência e Sazonalidade	5
2.3 Tipos de Modelos	8
2.4 Identificação do Modelo (ARIMA)	9
2.5 Análise de Intervenção	10
2.6 Critérios de seleção de modelos	11
2.7 Testes de Adequação do Modelo	13
2.8 Modelo Sazonal	14
3 Metodologia	17
4 Resultados	19
5 Conclusões	49
Referências Bibliográficas	51

LISTA DE FIGURAS

4.1	Faturamento Mensal dos anos de 2005 à 2018.	19
4.2	FAC da série original de faturamento.	20
4.3	Gráfico da série original com uma diferença.	21
4.4	FAC e FACP da série com uma diferença.	21
4.5	Média pela Amplitude de grupos de observações da série.	23
4.6	Gráfico de linhas dos Resíduos Modelo $(2,1,0)(0,1,0)$	25
4.7	Histograma dos Resíduos do Modelo $(2,1,0)(0,1,0)$	25
4.8	Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(0,1,0)$	26
4.9	Gráfico de linhas dos Resíduos Modelo $(2,1,0)(0,1,0)$ c/ interv. 50.	27
4.10	Histograma dos Resíduos do Modelo $(2,1,0)(0,1,0)$ c/ interv. 50.	28
4.11	Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(0,1,0)$ c/ interv. 50.	28
4.12	Gráfico de linhas dos Resíduos Modelo $(2,1,0)(0,1,0)$ c/ interv. 70.	29
4.13	Histograma dos Resíduos do Modelo $(2,1,0)(0,1,0)$ c/ interv. 70.	30
4.14	Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(0,1,0)$ c/ interv. 70.	30
4.15	Gráfico de linhas dos Resíduos do Modelo $(2,1,0)(1,1,0)$	31
4.16	Histograma dos Resíduos do Modelo $(2,1,0)(1,1,0)$	32
4.17	Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(1,1,0)$	32
4.18	Gráfico de linhas dos Resíduos do Modelo $(2,1,0)(1,1,0)$ c/ interv. 50.	34
4.19	Histograma dos Resíduos do Modelo $(2,1,0)(1,1,0)$ c/ interv. 50.	34
4.20	Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(1,1,0)$ c/ interv. 50.	35
4.21	Gráfico de linhas dos Resíduos do Modelo $(2,1,0)(1,1,0)$ c/ interv. 70.	36
4.22	Histograma dos Resíduos do Modelo $(2,1,0)(1,1,0)$ c/ interv. 70.	37
4.23	Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(1,1,0)$ c/ interv. 70.	37
4.24	Gráfico de linhas dos Resíduos Modelo $(1,1,1)(0,0,2)$	38
4.25	Histograma dos Resíduos do Modelo $(1,1,1)(0,0,2)$	39
4.26	Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(0,1,0)$	39
4.27	Gráfico de linhas dos Resíduos Modelo $(1,1,1)(0,0,2)$ c/ interv. 50	41
4.28	Histograma dos Resíduos do Modelo $(1,1,1)(0,0,2)$ c/ interv. 50.	41
4.29	Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(0,1,0)$ c/ interv. 50.	42
4.30	Gráfico de linhas dos Resíduos Modelo $(1,1,1)(0,0,2)$ c/ interv. 70	43
4.31	Histograma dos Resíduos do Modelo $(1,1,1)(0,0,2)$ c/ interv. 70.	44
4.32	Gráfico da Autocorrelação Residual do Modelo $(1,1,1)(0,0,2)$ c/ interv. 70.	44
4.33	Comparação dos três modelos significativos com a série original.	46

LISTA DE TABELAS

4.1	Teste de Cox-Stuart para tendência	20
4.2	Teste de Cox-Stuart para tendência na Série com Diferença	20
4.3	Teste de Fisher para Sazonalidade	22
4.4	Estimativas dos parâmetros do Modelo $(2,1,0)(0,1,0)$ e do teste t.	24
4.5	Testes de Diagnóstico Modelo $(2,1,0)(0,1,0)$	24
4.6	Estimativas dos parâmetros do Modelo $(2,1,0)(0,1,0)$ c/ interv.50 e do teste t.	27
4.7	Testes de diagnóstico modelo $(2,1,0)(0,1,0)$ c/ interv.50	27
4.8	Estimativas dos parâmetros do modelo $(2,1,0)(0,1,0)$ c/ interv.70 e do teste t.	29
4.9	Testes de diagnóstico modelo $(2,1,0)(0,1,0)$ c/ interv.70.	29
4.10	Estimativas dos parâmetros do modelo $(2,1,0)(1,1,0)$ e do teste t.	31
4.11	Testes de diagnóstico modelo $(2,1,0)(1,1,0)$	31
4.12	Estimativas dos parâmetros do modelo $(2,1,0)(1,1,0)$ c/ interv.50 e do teste t.	33
4.13	Testes de diagnóstico modelo $(2,1,0)(1,1,0)$ c/ interv.50	33
4.14	Estimativas dos parâmetros do modelo $(2,1,0)(1,1,0)$ c/ interv.70 e do teste t.	36
4.15	Testes diagnóstico modelo $(2,1,0)(1,1,0)$ c/ interv.70	36
4.16	Estimativas dos parâmetros do modelo $(1,1,1)(0,0,2)$ e do teste t.	38
4.17	Testes de diagnóstico modelo $(1,1,1)(0,0,2)$	38
4.18	Estimativas dos parâmetros do modelo $(1,1,1)(0,0,2)$ c/ interv. 50 e do teste t.	40
4.19	Testes diagnóstico modelo $(1,1,1)(0,0,2)$ c/ interv. 50.	40
4.20	Estimativas dos parâmetros do modelo $(1,1,1)(0,0,2)$ c/ interv. 70 e do teste t.	42
4.21	Testes de diagnóstico modelo $(1,1,1)(0,0,2)$ c/ interv. 70	43
4.22	Critério de Seleção do Modelo	45
4.23	Faturamento Observado e Previstos.	47
4.24	Previsão Mensal	47

1. INTRODUÇÃO

O desejo de prever o futuro sempre encantou a humanidade. Saber o que vai acontecer antes mesmo de os primeiros sinais se manifestarem pode propiciar melhor aproveitamento dos efeitos positivos de eventos futuros, ou uma preparação antecipada de efeitos negativos. Talvez até mais importante que antecipar os resultados seja reconhecer quais fatores podem interferir no fenômeno em questão [5].

A intensa luta atualmente entre empresas, estimula-as a buscarem métodos para produzirem de maneira mais eficiente e eficaz. Aplicar corretamente metodologias que possam diminuir gastos e aumentar a porcentagem de acertos, faz com que técnicas de previsões sejam atrativas aos olhos empresariais [9].

Quem deseja melhorar seu negócio pode alcançar esse objetivo por meio da criação de uma projeção de vendas e de faturamento para sua organização. Ter uma ferramenta capaz de guiar o empreendedor em relação a quais investimentos e estratégias seguir para dar conhecimento de possíveis aumentos e quedas no faturamento é consideravelmente importante. Por isso para uma boa e adequada gestão não só da área comercial, mas de toda a companhia, para chegar a essas estimativas, consideram-se fatores como a constância nas vendas, a estabilidade da empresa, o cenário atual do mercado e a sazonalidade.

Dentro do mercado atual, pode-se dizer que a empresa que conhece seus números, que esteja preparada para mudanças no mercado, e que conte com um bom planejamento estratégico, tem mais chances de obter sucesso, para que seja possível produzir a quantidade adequada no momento oportuno, se adaptando às variações do mercado. Atualmente, o varejo tem utilizado diversos modelos de previsões por meio de séries temporais. Analisando o comportamento de uma série temporal, é possível ter uma visão do futuro, fazer cenários, elaborar estratégias e tomar decisões.

Como um dos objetivos é realizar a previsão de valores futuros da série, esta metodologia tem várias aplicações. Um exemplo é o estudo dos resultados mensais da taxa de desemprego da Região Metropolitana de Porto Alegre, sendo que essas informações iriam auxiliar os órgãos governamentais em tomadas de decisões na região, referente ao índice de desemprego [6].

De acordo com [16] a maioria dos métodos de previsão de séries temporais se baseia na suposição de que as observações passadas contêm todas as informações sobre o padrão de comportamento da série temporal e esse padrão é recorrente no tempo. A modelagem de séries temporais é uma estratégia comumente utilizada como forma de previsão de valores futuros em séries de vendas, produção, exportação e importação, faturamento entre outras. No estudo em

[8], os autores utilizaram modelos de séries temporais e redes neurais artificiais para previsão de demanda de um determinado produto. No exemplo em questão, os autores encontraram que o modelo de redes neurais artificiais apresentou melhores previsões. De acordo com (Zhang, 2003), o processo gerador da série temporal sob estudo pode ser representado tanto por uma estrutura linear quanto por uma não linear, sendo, portanto, necessário que se utilize mais de um modelo para previsões, na tentativa de descrever melhor o processo gerador da série, e obter previsões mais acuradas [16] [8] [17].

Entre os vários estudos que utilizam os modelos de séries temporais, tem-se aquele descrito por Spanhol (2004) [15], que utilizou o índice de faturamento real do comércio varejista de alimentos da Região Metropolitana de São Paulo, na tentativa de verificar a eficácia e consistência do modelo de previsões de Holt-Winters como método de previsão estatístico [15].

Pinho (2015) desenvolveu um estudo realizando modelos de previsão de vendas para séries de Vestuário e Alimentos. A partir de dados extraídos do site do IBGE, foram realizadas análises estatísticas através de modelos econométricos, encontrando um modelo satisfatório para a previsão de vendas das séries em questão [12].

O histórico de uma variável como o faturamento pode ser utilizado na identificação de períodos de crescimento ou decréscimo, sazonalidade e ainda para "prever" observações futuras. Na verdade, os modelos estatísticos para séries temporais utilizam o passado histórico, ou seja, observações a priori para projetar observações futuras.

Este trabalho é motivado por um conjunto de dados composto por observações do faturamento mensal de uma empresa atacadista da cidade de Uberlândia MG, no período de Janeiro de 2005 à Dezembro de 2018. O objetivo do estudo é obter a previsão para o faturamento mensal da empresa atacadista, identificar características inerentes à séries em estudo, como tendência, sazonalidade e ruído branco por meio de testes específicos, observar a presença e os efeitos de possíveis intervenções e ajustar modelos de séries temporais que representem os dados adequadamente por meio do método Box e Jenkins [7]. De modo aplicado deseja-se prever o faturamento com base em informações passadas e assim tomar decisões que ajudem a empresa a alcançar seus objetivos.

2. FUNDAMENTAÇÃO TEÓRICA

Para estudar séries temporais, é importante entender sobre como processos estocásticos afetam toda a metodologia de séries, uma vez visto que um processo estocástico é uma coleção de variáveis aleatórias que, em geral, são utilizadas para estudar a evolução de fenômenos que são observados ao longo do tempo. Utilizam-se processos estocásticos, para os quais dada uma condição inicial, ainda se tem diversas trajetórias possíveis para a evolução do sistema.

Um processo estocástico é uma família $Z = Z(t), t \in T$, tal que, para cada $t \in T$, $Z(t)$ é uma variável aleatória.

Em outras palavras, um processo estocástico é uma família de variáveis aleatórias, dentro do mesmo espaço de probabilidade, sendo que o conjunto T é usualmente definido dentro do conjunto dos inteiros \mathcal{Z} ou dos reais \mathcal{R} . De acordo com Morettin (2006), se o conjunto T for finito ou enumerável, como $T = 1, 2, \dots, N$ ou $T = \mathcal{Z}$, o processo pode ser dito com parâmetro discreto. Se T tiver o intervalo dentro do conjunto \mathcal{R} obtemos um processo com parâmetro contínuo. No primeiro caso, $Z(t)$ poderia por exemplo, representar o número de demissões durante o primeiro ano de uma determinada empresa. No segundo caso, $Z(t)$ representa uma medida que varia continuamente, como temperatura, bolsa de valores, eletrocardiograma, etc [10].

Um processo Z é estacionário se todas as características do comportamento do processo não são alterados no tempo, ou seja, ele se desenvolve no tempo sendo que a definição de uma origem não é considerável (suas propriedades não dependem do momento em que foi observada).

Resumindo, as características de $Z(t + \tau)$, para todo τ , são as mesmas de $Z(t)$, existindo duas formas de estacionariedade: fraca(ou ampla, ou de segunda ordem) e estrita(ou forte).

Um processo estocástico $Z = Z(t), t \in T$ pode ser classificado em "estritamente estacionário" se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo. Logo,

$$F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n), \quad (2.1)$$

para quaisquer t_1, \dots, t_n, τ de T .

Todas as distribuições unidimensionais são invariantes sob translações do tempo, logo a média $\mu(t)$ e a variância $V(t)$ são constantes, isto é,

$$\mu(t) = \mu, V(t) = \sigma^2, \quad (2.2)$$

para todo $t \in T$. Sem perda de generalidade, podemos supor que $\mu = 0$; caso contrário,

considere o processo $Z(t) - \mu$.

Um processo estocástico $Z = Z(t), t \in T$ pode ser classificado em "fracamente estacionário" ou estacionário de segunda ordem se e somente se para todo $t \in T$,

$$\begin{aligned} E(Z(t)) &= \mu(t) = u; \\ E(Z^2(t)) &< \infty; \\ \gamma(t_1, t_2) &= Cov(Z(t_1), Z(t_2)). \end{aligned} \tag{2.3}$$

Vale ressaltar que, para esse estudo estamos interessados somente nesta classe de processos, que serão denotados por "processos estacionários".

2.1 MODELOS DE SÉRIES TEMPORAIS

Uma série temporal é qualquer conjunto de observações geradas sequencialmente no tempo, tendo como característica a dependência entre as observações. A variável é observada em pontos temporais discretos, usualmente equidistantes, e a análise de tal comportamento temporal envolve a descrição do processo ou fenômeno que gera a sequência, sendo assim uma descrição do passado. Um procedimento lógico para realizar previsões é fazer uso desses dados históricos [11].

Considere $Z_t, t = 1, \dots, N$ uma série temporal. Podemos escrever um modelo de decomposição dessa série em Z_t resultando do total de três componentes não-observáveis,

$$Z_t = T_t + S_t + a_t, \tag{2.4}$$

sendo que T_t representa o componente de tendência e S_t o de sazonalidade, tendo a_t como um componente aleatório, de média zero e variância constante.

A questão crucial do modelo clássico é decidir como será a equação que relaciona as componentes com a variável. Existem duas opções: o modelo aditivo ou o modelo multiplicativo.

Sendo o modelo aditivo de dados no qual os efeitos dos fatores individuais são diferenciados e adicionados para modelar os dados, tem-se a expressão 2.4.

O modelo multiplicativo pressupõe que à medida que os dados aumentam o padrão sazonal também aumenta. Nesse modelo, os componentes de tendência e sazonalidade são multiplicados e depois adicionados ao componente de erro, [3]

$$Z_t = T_t * S_t * a_t. \tag{2.5}$$

Segundo Morettin (2006) [10], há vários métodos para estimar T_t , sendo que os mais frequentes consistem em:

1. Ajustar uma função do tempo, como um polinômio, uma exponencial ou outra função

suave de t ;

2. Suavizar os valores da série ao redor de um ponto, para estimar tendência naquele ponto;
3. Suavizar os valores da série através de sucessivos ajustes de retas de mínimos quadrados ponderados;

Aferindo a possibilidade de tendência através de \hat{T}_t , conseguimos escrever a série temporal ajustada, já livre de qualquer tendência que possa influenciar na previsão,

$$Y_t = Z_t - \hat{T}_t. \quad (2.6)$$

Um procedimento que é também utilizado para eliminar a tendência de uma série é aquele de tomar diferenças, sendo por exemplo, a primeira diferença dada por:

$$\Delta Z_t = Z_t - Z_{t-1}. \quad (2.7)$$

Do ponto de vista conceitual, é difícil definir o que é sazonalidade. Podemos considerar como casos de sazonalidade as situações em que é possível observar variações sazonais que ocorrem regularmente de ano para ano, de mês em mês ou resumidamente de períodos que se repetem. Como exemplos de sazonalidade temos o aumento de vendas de passagens aéreas no verão, ou o crescimento no nível de vendas no Natal.

Segundo Morettin (2006) [10], a sazonalidade estimada é subtraída no modelo da Equação 2.4, de tal forma que o ajuste sazonal possa resultar em:

1. obter estimativas \hat{S}_t de S_t ;
2. calcular $Z_t^{SA} = Z_t - \hat{S}_t$.

Se o modelo for multiplicativo, da forma 2.5 a série sazonalmente ajustada será

$$Z_t^{SA} = \frac{Z_t}{\hat{S}_t}. \quad (2.8)$$

2.2 TESTE PARA TENDÊNCIA E SAZONALIDADE

TENDÊNCIA

A componente tendência tem como definição um padrão de crescimento ou decréscimo da variável em um certo período de tempo[2].

Existem testes específicos para a identificação da tendência, como por exemplo o teste de Wald e o teste de Cox-Stuart. Uma possível alternativa é o ajuste de uma Regressão Linear Simples para a identificação visual da inclinação da reta de tendência. No entanto, a inspeção visual pode levar a resultados viesados e, sendo mais eficiente o uso de um dos testes formais disponíveis[2].

De acordo com Morettin (2006)[10], para o teste do sinal (Cox-Stuart) considerando um conjunto de observações X_1, X_2, \dots, X_n o teste baseia-se em agrupar as observações em pares $(X_1, X_{1+c}), \dots, (X_{n-c}, X_n)$, para $c = \frac{n}{2}$ se n for par e $c = \frac{n+1}{2}$ se n for ímpar, em que n é o número de observações. Para cada par (X_i, X_{i+c}) atribui-se o sinal “+” se $X_i < X_{i+c}$ e o sinal “-”, se $X_i > X_{i+c}$, eliminando-se os empates. Considerando n_t o número de pares em que $X_i \neq X_{i+c}$, a estatística do teste é dada por T_{Est} , sendo que T_{Est} é igual ao número de pares com sinal positivo. Para $T_{Est} > n_t - T_{Tab}$ rejeita-se a hipótese H_0 de não existência de tendência. O valor de T_{Tab} é encontrado numa tabela de distribuição binomial, com parâmetros $p = 0,5$ e n_t para um dado nível α de significância, se $n_t \leq 20$. Para $n_t > 20$, pode-se usar a aproximação pela distribuição normal .

Também é possível testar existência de tendência por meio do coeficiente de correlação de Spearman. Este em geral é mais poderoso que o teste de Cox-Stuart. Chamemos de R_t o posto de Z_t dentre as N observações. Se houver observações empatadas, use postos médios.

Seja,

$$T_3 = \sum_{t=1}^N [R_t - t]^2, \quad (2.9)$$

onde $t = 1, 2, \dots$, são pontos "naturais" dos instantes de tempo. Rejeitamos a hipótese de que não existe tendência se T_3 for grande ou pequeno [10].

A distribuição de T_3 é tabelada, e pode ser encontrada em (Conover, 1980). O teste pode ser feito também em função do coeficiente de correlação de Spearman, dado por [10]

$$p = 1 - \frac{6T_3}{N(N^2 - 1)}. \quad (2.10)$$

SAZONALIDADE

Uma possibilidade para verificar a existência de sazonalidade é utilizar o teste de Kruskal-Wallis. Para este, cada mês é suposto como uma amostra de uma população, isto é temos k (igual a 12, no caso de um ano completo) amostras, de tamanho n_j , ou seja as observações são

$$Y_{ij}, j = 1, \dots, k, i = 1, \dots, n_j, N = \sum_{j=1}^k n_j. \quad (2.11)$$

As observações Y_{ij} são substituídas por seus postos R_{ij} , obtidos ordenando-se todas as N observações. Seja $R_{.j}$ a soma dos postos associados à j -ésima amostra (coluna),

$$R_{.j} = \sum_{i=1}^{n_j} R_{ij}, j = 1, \dots, k. \quad (2.12)$$

A hipótese H_0 de não existência de sazonalidade é rejeitada se a estatística

$$T_1 = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_{.j}^2}{n_j} - 3(N+1) \quad (2.13)$$

for maior ou igual ao valor crítico T_{1c} , onde T_{1c} é tal que $P_H(T_1 \geq T_{1c}) = \alpha$, $\alpha =$ nível de significância do teste. Para n_j suficientemente grande, ou $k \geq 4$, sob H_0 , a distribuição de T_1 pode ser aproximada por uma variável qui-quadrado com $k - 1$ graus de liberdade.

Um ponto negativo para a aplicação de Kruskal-Wallis é que uma de suas suposições diz que as variáveis dentro de cada amostra devem ser independentes e que as amostras são independentes entre si, o que evidentemente não ocorre.

Como alternativa para estes casos o teste de Friedman para amostras relacionadas deve ser utilizado.

Neste caso, os meses são considerados "tratamentos" e os anos são considerados "blocos". A ordenação é feita por cada bloco, em vez de ordenar todas as N observações, mas mesmo aqui, os blocos são considerados independentes.

A Estatística do Teste de Friedman é

$$T_2 = \frac{12}{p_k(k+1)} \sum_{j=1}^k R_{.j}^{*2} - 3p(k+1). \quad (2.14)$$

onde $p =$ número de blocos = número de anos, $k =$ número de tratamentos = 12 e $R_{.j}^{*2}$ denota a soma dos postos da j -ésima coluna, isto é,

$$R_{.j}^* = \sum_{i=1}^p R_{ij}^*, \quad (2.15)$$

onde $R_{ij}^* =$ posto de Y_{ij} dentro do bloco i , de 1 até k . A distribuição de T_2 pode ser aproximada por uma qui-quadrado com $k - 1$ graus de liberdade.

Outra alternativa que podemos utilizar é um teste F rotineiro a uma análise de variância. O modelo subjacente é

$$Y_{i,j} = S_j + e_{ij}, i = 1, \dots, n_j, j = 1, \dots, k, \quad (2.16)$$

e supondo $e_{ij} \sim N(0, \sigma^2)$, independentes. Sob a hipótese nula $H_0 : S_1 = \dots = S_k$, a estatística

$$T_3 = \frac{N - k}{k - 1} \frac{\sum_{j=1}^k n_j (\bar{Y}_{.j} - \bar{Y})^2}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{.j})^2} \quad (2.17)$$

tem distribuição $F(k - 1, N - k)$.

Vemos, pois, que condições para se aplicar este teste incluem a validade do modelo aditivo e normalidade dos resíduos.

2.3 TIPOS DE MODELOS

Os métodos para previsão de séries temporais são classificados como métodos quantitativos, tendo como base observações passadas e o inter-relacionamento entre as observações, fornecendo previsões acuradas se o futuro tiver alguma semelhança com o passado [6].

Qualquer que seja a classificação que façamos para os modelos de séries temporais, podemos considerar vários tipos de modelos diferentes para descrever o comportamento de uma série em questão.

Os procedimentos na prática variam muito, podendo ser simples e intuitivos, com pouca análise dos dados, ou complexos e racionais, envolvendo um considerável trabalho de interpretação de séries temporais. Dentre os procedimentos estatísticos de análise de séries temporais podem ser citados: Modelos Univariados, Modelos de Função de Transferência, Modelos Multivariados.

Modelos Univariados que se baseiam em uma única série histórica. Como exemplo, a decomposição por componentes não observáveis, que foi o mais utilizado até a década de 1960; os modelos univariados de Box & Jenkins(1970) que consistem em uma classe geral de modelos lineares conhecidos como modelos ARIMA [7];

Modelos de Função de Transferência, nos quais a série de interesse é explicada não só pelo seu passado histórico, como também por outras séries temporais não correlacionadas entre si;

Modelos Multivariados, modelam simultaneamente duas ou mais séries temporais sem qualquer exigência em relação à direção da causalidade entre elas.

MODELOS AR, MA E ARIMA

Os modelos ARIMA são modelos estatísticos lineares para análise de séries temporais. A abreviação em língua inglesa refere-se a "Auto-Regressive Integrated Moving Average Model", ou seja, modelo auto-regressivo integrado de médias móveis. A notação empregada para designação do modelo é normalmente $ARIMA(p, d, q)$, em que p representa o número de parâmetros auto-regressivos, d o número de diferenciações para que a série torne-se estacionária e q o número de parâmetros de médias móveis. Casos particulares são o modelo $ARMA(p, q)$, o modelo auto-regressivo $AR(p)$ e o modelo de médias móveis $MA(q)$, todos para séries temporais estacionárias ($d = 0$) [13].

Os modelos auto-regressivo de ordem p são denotados por:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t = \phi(B) \tilde{Z}_t, \quad (2.18)$$

em que a estimativa da variável Z para um tempo t depende de uma combinação linear de p termos da série observada, incluindo o termo aleatório a_t de ruído branco. Seja Z_{t-1} equivalente ao operador de translação para o passado, denotado por B e definido por

$$BZ_t = Z_{t-1}. \quad (2.19)$$

O modelo $AR(p)$ é limitado pois assume a existência de uma relação linear entre os elementos

da sequência e baseia-se na hipótese de que a série é estacionária, isto é, a média e o desvio padrão das observações não variam com o tempo.

O modelo de médias móveis $MA(q)$ assume que a série modelada é gerada através de uma combinação linear de q sinais de ruídos $a(t - i)$, aleatórios e independentes entre si,

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (2.20)$$

e sendo $\tilde{Z}_t = Z_t - \mu$, teremos

$$\tilde{Z}_t = Z_t - \mu = (1 - \theta_1 B - \dots - \theta_q B^q) a_t = \theta(B) a_t, \quad (2.21)$$

onde

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (2.22)$$

é o operador de médias móveis de ordem q .

A combinação dos modelos $AR(p)$ e $MA(q)$ dá então origem ao modelo $ARMA(p, q)$, conforme equação a seguir:

$$\phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (2.23)$$

No caso da série ser não estacionária, a realização de um diferença na série é necessária para torná-la estacionária conduzindo ao modelo $ARIMA(p, d, q)$.

Várias séries econômicas e financeiras, por exemplo, são não-estacionárias, mas quando diferenciadas tornam-se estacionárias. Por exemplo, Z_t é não-estacionária, mas W_t é estacionária, sendo:

$$W_t = Z_t - Z_{t-1} = (1 - B)Z_t = \Delta Z_t. \quad (2.24)$$

Se $W_t = \Delta^d Z_t$ for estacionária, podemos representar W_t por um modelo $ARMA(p, q)$, ou seja,

$$\phi(B)W_t = \theta(B)a_t. \quad (2.25)$$

Se W_t for uma diferença de Z_t , então Z_t é uma integral de W_t , daí dizermos que Z_t segue um modelo auto-regressivo, integrado, de médias móveis, ou modelo $ARIMA$,

$$\phi(B)\Delta^d Z_t = \theta(B)a_t, \quad (2.26)$$

de ordem (p, d, q) e escrevemos $ARIMA(p, d, q)$, se p e q são as ordens de $\phi(B)$ e $\theta(B)$, respectivamente.

2.4 IDENTIFICAÇÃO DO MODELO (ARIMA)

Com o objetivo de determinar os valores de p , d e q do modelo $ARIMA(p, d, q)$, e estimativas para serem utilizadas na fase de estimações, os procedimentos de identificação consistem de duas

partes [10].

Inicialmente diferencia-se a série temporal original o número de vezes necessárias para torná-la estacionária, de modo a possibilitar a análise do processo com o modelo $ARMA(p, q)$; o número de diferenciações d é aquele necessário para que a função de autocorrelação amostral (ACF) da série transformada decresça rapidamente para zero.

Identificar o processo $ARMA(p, q)$ resultante, através da análise das autocorrelações e autocorrelações parciais estimadas.

Um ponto importante na identificação do modelo ARIMA é manter ele o mais parcimonioso possível, ou seja, com o melhor ajuste, mas com menor número de parâmetros possível. Assim

1. um processo $AR(p)$ tem fac que decai de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão;
2. um processo $MA(q)$ tem fac finita, no sentido que ela apresenta um corte após o "lag" q ;
3. um processo $ARMA(p, q)$ tem fac infinita em extensão, a qual decai de acordo com exponenciais e/ou senóides amortecidas após o "lag" $q - p$.

2.5 ANÁLISE DE INTERVENÇÃO

Segundo Morettin (2006) [10], por uma intervenção entendemos a ocorrência de algum tipo de evento em dado instante de tempo T , conhecido a priori. Tal ocorrência pode manifestar-se por um intervalo de tempo subsequente e que afeta temporariamente, ou permanentemente, a série em estudo. A análise de intervenção tem por objetivo avaliar o impacto de tal evento no comportamento da série.

Usualmente, as séries indicadoras de intervenções podem ser representadas por dois tipos de variáveis binárias:

Função degrau ("step function"),

$$\begin{aligned} X_{j,t} = S_t^{(T)} &= 0, t < T, \\ X_{j,t} = S_t^{(T)} &= 1, t \geq T; \end{aligned} \tag{2.27}$$

função impulso

$$\begin{aligned} X_{j,t} = I_t^{(T)} &= 0, t \neq T, \\ X_{j,t} = I_t^{(T)} &= 1, t = T. \end{aligned} \tag{2.28}$$

No caso da função 2.27, o efeito da intervenção é permanente após o instante T ao passo que, para a função 2.28, o efeito é temporário.

Uma classe geral de modelos, que leva em conta a ocorrência de múltiplas intervenções, é dada por

$$Z_t = \sum_{j=1}^k v_j(B)X_{j,t} + N_t \quad (2.29)$$

em que

- $X_{j,t}$, $j = 1, 2, \dots, k$, são variáveis de intervenção do tipo 2.27 ou 2.28;
- $v_j(B)$, $j = 1, \dots, k$ são funções racionais da forma $\frac{w_j(B)B^{b_j}}{\delta_j(B)}$, onde

$$w_j(B) = w_{j,0}B - w_{j,1}B - \dots - w_{j,s}B^s$$

e

$$\delta(B) = 1 - \delta_{j,1}B - \dots - \delta_{j,r}B^r$$

são polinômios em B , b_j é a defasagem no tempo para o início do efeito da j -ésima intervenção.

- N_t é a série temporal livre do efeito das intervenções e é denominada série residual.

O modelo da série residual é um modelo SARIMA representado de forma geral por

$$\varphi(B)N_t = \theta(B)a_t. \quad (2.30)$$

Em geral, o efeito de uma intervenção é mudar o nível da série ou, então, a inclinação. Entretanto, há três fontes de ‘ruídos’ que podem obscurecer o efeito da intervenção:

- tendência;
- sazonalidade;
- erro aleatório.

O fato de existir tendência numa série pode levar a falsas conclusões. De fato, se esta existir e uma intervenção ocorrer no instante T , o fato do nível pós-intervenção ser maior do que nível pré-intervenção pode ser devido simplesmente à tendência. A utilidade dos modelos SARIMA é que os três elementos citados anteriormente são levados em conta quando a componente residual N_t de (2.29) é modelada [10].

2.6 CRITÉRIOS DE SELEÇÃO DE MODELOS

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

explicar o fenômeno sob estudo. Akaike (1974) utilizou a Informação de Kullback-Leibler para testar se um dado modelo é adequado. Porém seu uso é limitado, pois depende da distribuição g (modelo verdadeiro), que é desconhecida [10].

Se uma boa estimativa para a log verossimilhança esperada puder ser obtida através dos dados observados, esta estimativa poderá ser utilizada como um critério para comparar modelos. Assim um modo de comparar n modelos, $g_1(x|\theta_1), g_2(x|\theta_2), \dots, g_n(x|\theta_n)$, é simplesmente comparar as magnitudes da função suporte maximizada $L(\hat{\theta}_i)$. Mas tal método não fornece uma verdadeira comparação, haja vista que, em não conhecendo o verdadeiro modelo $g(x)$, primeiramente o método da máxima verossimilhança estima os parâmetros θ_i de cada modelo $g_i(x)$, $i = 1, 2, \dots, n$, e posteriormente são utilizados os mesmos dados para estimar $E_G[\log f(x|\hat{\theta})]$, isto introduz um viés em $L(\hat{\theta}_i)$, sendo que, a magnitude deste viés varia de acordo com a dimensão do vetor de parâmetros [10].

Deste modo, os critérios de informação são construídos para avaliar e corrigir o viés ($b(G)$) da função suporte.

CRITÉRIO DE INFORMAÇÃO DE AKAIKE

O Critério de Informação de Akaike (AIC) é definido como

$$AIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{2(p+q)}{N} \quad (2.31)$$

em que $\hat{\sigma}_{p,q}^2$ é uma estimativa da variância residual, p é o número de parâmetros do modelo autoregressivo, q é o número de parâmetros do modelo de médias móveis e N é o número de observações da série. O critério de informação de Akaike (AIC) pode assumir qualquer valor, inclusive valor negativo, já que ele depende da forma da função de verossimilhança. Para se comparar o AIC, é necessário que os modelos ajustados sejam feitos no mesmo período amostral[4]. O modelo que apresentar o menor valor de AIC é considerado o melhor, de acordo com esse critério [10].

CRITÉRIO DE INFORMAÇÃO BAYESIANO

O Critério de Informação Bayesiano (BIC), proposto por Schwarz (1978) é dado por:

$$BIC = \ln \hat{\sigma}_{p,q}^2 + \frac{(p+q)\ln(N)}{N}, \quad (2.32)$$

em que $\hat{\sigma}_{p,q}^2$ é uma estimativa da variância residual, p é o número de parâmetros do modelo autoregressivo, q é o número de parâmetros do modelo de médias móveis e N é o número de observações da série [14]. O melhor modelo é aquele que apresentar menor valor de BIC [10].

ERRO PERCENTUAL MÉDIO ABSOLUTO

O Erro Percentual Médio absoluto de Previsão, representado pela sigla MAPE (da expressão em inglês *mean absolute percentual error*) é definido por:

$$MAPE_{t,s} = \frac{\sum_{i=1}^s (y_{t-s+i} - \hat{y}_{t-s+i})/y_{t-s+i}}{s} \quad (2.33)$$

em que s é o horizonte de previsão, y_{t-s+i} é o valor real das s ultimas observações, \hat{y}_{t-s+i} é o valor previsto em um modelo feito a partir dos dados sem as s últimas observações, e n é igual ao número de observações. Assim compara-se os s valores retirados com os valores previstos a partir dos dados sem as s últimas observações.

O MAPE é uma medida de erro comumente usada para aferir a qualidade do ajuste de um modelo, sendo considerado o melhor modelo, para fins de previsão, aquele que apresentar menor valor desta medida [10].

2.7 TESTES DE ADEQUAÇÃO DO MODELO

Após estimar o modelo temos que verificar se ele representa, ou não, adequadamente, os dados. Para tal objetivo, vários testes estão disponíveis na literatura, alguns deles citados a seguir.

TESTE DE AUTOCORRELAÇÃO RESIDUAL

Estimados ϕ e θ , as quantidades

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)W_t, \quad (2.34)$$

são chamadas resíduos estimados ou simplesmente resíduos. Se o modelo for adequado, os \hat{a}_t deverão estar próximos dos a_t e, portanto, deverão ser aproximadamente não-correlacionados. Se indicarmos por \hat{r}_k as autocorrelações dos resíduos \hat{a}_t , então deveríamos ter $\hat{r}_k \simeq 0$. Em geral, deveríamos ter, aproximadamente,

$$\hat{r}_k \sim N(0, 1/n), \quad (2.35)$$

sempre sob a suposição que o modelo ajustado é apropriado. As autocorrelações \hat{r}_k são calculadas por [10][11]

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^n \hat{a}_t^2}. \quad (2.36)$$

A comparação de \hat{r}_k com os limites $\pm 2/\sqrt{n}$ fornece uma indicação geral de possível quebra de comportamento de ruído branco em a_t , com a condição de que seja lembrado que, para pequenos valores de k , estes limites subestimarão a significância de qualquer discrepância.

TESTE DE BOX-PIERCE

Box e Pierce (1970) sugeriram um teste para as autocorrelações dos resíduos estimados, que, apesar de não detectar quebras específicas no comportamento de ruído branco, pode indicar se esses valores são muito altos [7].

Se o modelo for apropriado, a estatística

$$Q(K) = n(n+2) \sum_{j=1}^k \frac{\hat{r}_j^2}{(n-j)}, \quad (2.37)$$

terá aproximadamente uma distribuição qui-quadrado com $K - p - q$ graus de liberdade. A distribuição assintótica é obtida sob a hipótese que $K = K(n) \rightarrow \infty$, quando $n \rightarrow \infty$.

A hipótese de ruído branco para os resíduos é rejeitada para valores grandes $Q(K)$.

TESTE DE LJUNG-BOX

O teste de Ljung-Box (1978) é uma ferramenta diagnóstica usada para testar a falha do ajuste de um modelo de série temporal.

O teste é aplicado aos resíduos de uma série temporal após o ajustamento de um modelo $ARMA(p, q)$ aos dados. O teste examina as autocorrelações dos resíduos. Se as autocorrelações forem muito pequenas, concluímos que o modelo não exibe falha significativa de ajuste.

2.8 MODELO SAZONAL

Muitas séries temporais apresentam padrões repetitivos que aparecem regularmente a cada intervalo de tempo, como é o caso das indústrias, em que vendas e produção seguem uma sazonalidade forte em determinados períodos do ano. Para lidar com as séries que apresentam autocorrelação sazonal, (Box e Jenkins, 1976) generalizaram o modelo ARIMA e definiram o modelo ARIMA sazonal multiplicativo, conhecido como SARIMA(p, d, q)(P, D, Q) [7].

Nesse caso, considerando sazonalidade mensal de ordem 12, a modelagem segue a equação a seguir.

$$\phi(B)\Phi(B^{12})\Delta^d\Delta_{12}^D Z_t = \theta(B)\Theta(B^{12})a_t. \quad (2.38)$$

em que:

$\Theta(B^{12}) = 1 - \Theta_1(B^{12}) - \Theta_2(B^{12 \cdot 2}) - \dots - \Theta_Q(B^{12 \cdot Q})$ o operador de médias móveis sazonal de ordem Q , definido pela equação 2.22;

$\Phi(B^{12}) = 1 - \Phi_1(B^{12}) - \Phi_2(B^{12 \cdot 2}) - \dots - \Phi_P(B^{12 \cdot P})$ é o operador autorregressivo sazonal de ordem P ;

$\Delta^d = (1 - B)^d$ o operador diferença no qual d indica o número de diferenças de ordem 1 necessárias para retirar a tendência da série;

$\Delta_{12}^D = (1 - B^{12})^D$ o operador diferença sazonal no qual D indica o número de diferenças de ordem 12 necessárias;

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

$\phi(B) = 1 - \phi_1(B^1) - \phi_2(B^2) - \dots - \phi_p(B^p)$ o polinômio autoregressivo de ordem p ;

Z_t a variável resposta do modelo;

a_t representa o erro aleatório e não observável do modelo ou ruído branco.

O modelo com esta estrutura é denominado $SARIMA(p, d, q)(P, D, Q)$. Ressalta-se que o procedimento de obtenção deste modelo segue os mesmos passos empregados para achar o modelo ARIMA não sazonal. Isto quer dizer que, no SARIMA, faz-se também a observância do comportamento da FAC e da FACp, entretanto, olha-se para as defasagens sazonais.

Quando Z_t exhibe um comportamento sazonal determinístico com período 12, um modelo que pode ser útil é

$$Z_t = \mu_t + N_t, \tag{2.39}$$

onde μ_t é uma função determinística periódica, satisfazendo $\mu_t - \mu_{t-12} = 0$, ou

$$(1 - B^{12})\mu_t = 0 \tag{2.40}$$

e N_t é um processo estacionário que pode ser modelado por um $ARMA(p, q)$.

Dessa maneira, N_t satisfaz à equação

$$\phi(B)N_t = \theta(B)a_t. \tag{2.41}$$

Assim, para um modelo sazonal determinístico, aplicando a diferença sazonal $(1 - B^{12})$ obtemos

$$(1 - B^{12})Z_t = (1 - B^{12})\mu_t + (1 - B^{12})N_t, \tag{2.42}$$

onde temos

$$(1 - B^{12})Z_t = (1 - B^{12})N_t. \tag{2.43}$$

Substituindo, obtemos

$$\phi(B)(1 - B^{12})Z_t = \theta(B)(1 - B^{12})a_t, \tag{2.44}$$

ou

$$\phi(B)W_t = \theta(B)(1 - B^{12})a_t \tag{2.45}$$

onde $W_t = (1 - B^{12})Z_t$.

A inclusão dos parâmetros sazonais, de ordem (P, D, Q) se faz necessária mesmo após eliminar a componente sazonal determinística, ainda que reste autocorrelação significativa em:

1. "lags" de baixa ordem, demonstrando que os resíduos ainda permanecem correlacionados.
2. "lags" sazonais, isto é, múltiplos de período s . Visto isto devemos cogitar a possibilidade de uma sazonalidade estocástica já comentada anteriormente, podendo assim ajustar um modelo sazonal (SARIMA).

3. METODOLOGIA

A base de dados utilizada neste trabalho se refere ao faturamento mensal de uma empresa atacadista de Uberlândia, que ao longo da sua trajetória e com a diversificação e amplitude de seu atendimento em todas as regiões do país, assumiu o papel de integradora da cadeia de consumo, gerando um aumento constante do faturamento da empresa. Os registros das receitas diárias foram agrupadas em registros mensais e a série foi constituída de 168 valores para os totais mensais. Com o objetivo de não divulgar os dados originais, a série foi multiplicada por uma constante. Considerou-se o período de Janeiro de 2005 a Dezembro de 2017 para ajustar os modelos e as previsões foram feitas para o ano de 2018.

Para identificar o comportamento da série foi construído o gráfico para o total mensal, para inspeção visual.

Como a suposição mais frequente que se faz a respeito de uma série temporal é a estacionariedade, deve-se verificar a presença dos componentes tendência e sazonalidade na série, os quais devem ser observadas e retirados no intuito de manter a série estacionária. O teste de Cox Stuart (teste do sinal) visa verificar a presença ou ausência da componente de tendência, o qual foi utilizado no *software R* com o pacote *randtests* e o comando *cox.stuart.test()* [1].

A verificação da existência de sazonalidade foi feita inicialmente por meio da análise gráfica da função de autocorrelação. A função autocorrelação, que é a correlação da série observada no tempo com a mesma série defasada em relação ao tempo inicial, apresenta os picos de sazonalidade com o respectivo intervalo de confiança e os períodos da sazonalidade da série.

Para confirmar estatisticamente a existência da componente de sazonalidade, foi utilizado o teste de Fisher, realizado no *software R* com o pacote *GeneCycle* e o comando *fisher.g.test()*.

Dada uma série temporal Z_t , o objetivo é construir um modelo estatístico que capture toda a informação estatística sistemática contida no conjunto de dados. Em outras palavras, considerando os dados em análise, o objetivo é identificar padrões não aleatórios na série temporal da variável faturamento, supondo que, a observação deste comportamento passado possibilitará obter previsões sobre o futuro. Espera-se que as previsões obtidas possam orientar a tomada de decisões, produzindo resultados que facilitem a análise da série em conjunto com a área financeira desta empresa, de forma que se possa ter a comparação, a nível analítico, através de uma projeção futura.

A metodologia utilizada neste estudo é a identificação e estimação de um modelo conforme definido por Box e Jenkins mais especificamente, um modelo sazonal representado por $SARIMA(p, d, q)(P, D, Q)$. Sendo possível que vários modelos sejam ajustados para o mesmo

conjunto de dados, utilizou-se de técnicas de comparação de modelos para escolher o melhor modelo possível, podendo ser considerados os critérios de Informação de Akaike AIC e do erro percentual médio absoluto de previsão $MAPE$ ou a raiz quadrada deste índice [7].

O cálculo da função autocorrelação dos resíduos foi utilizado para verificar se as autocorrelações estão dentro do intervalo de confiança, indicando não significância de correlação dos resíduos. Os testes utilizados para confirmar estatisticamente a hipótese de que os resíduos sejam ruído branco são a estatística de Ljung-Box e o teste de Box-Pierce.

Também foi utilizado para diagnóstico dos modelos os testes de Kolmogorov Smirnov com a função *lillie.test()* para suposição de normalidade dos resíduos, sendo essa uma das condições exigidas para a realização de inferências válidas.

Foram obtidas previsões mensais com os modelos sazonais ajustados, gerando os gráficos da série completa e das previsões para o ano de 2018.

Todas as análises foram executadas utilizando o *software R*.

4. RESULTADOS

O gráfico da série histórica do Faturamento mensal de uma empresa de atacado é apresentado na Figura 4.1. É possível observar pela análise gráfica que os dados aparentemente possuem sazonalidade evidenciada. Também pode-se afirmar que a série é não estacionária, pois apresenta uma tendência nítida, característica comum em séries econômicas. Além disso, é possível verificar uma queda brusca no nível da série, aproximadamente em 2009 e em 2010.

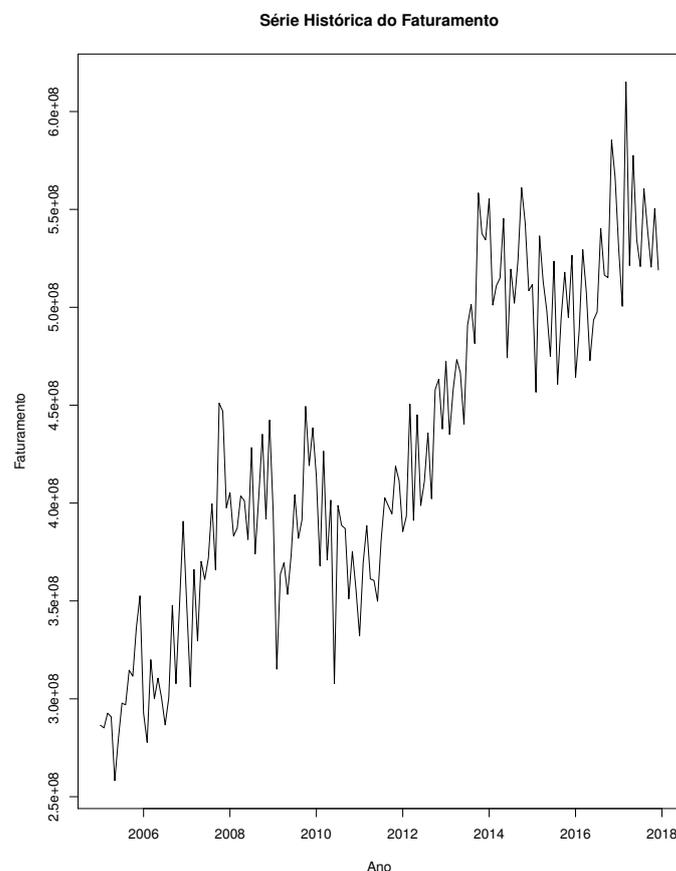


Figura 4.1: Faturamento Mensal dos anos de 2005 à 2018.

Na Figura 4.2 tem-se o gráfico da função de autocorrelação (FAC) da série, que pode ser útil para identificar séries temporais não estacionárias. Para uma série temporal estacionária, o gráfico da FAC mostrará um decaimento para zero relativamente rápido. Percebe-se pelo gráfico que este não é o caso da série de faturamento mensal considerada.

Na Tabela 4.1 tem-se o resultado do teste de Cox-Stuart, o qual foi realizado para testar a hipótese nula de que a série histórica possui média constante, ou seja, não possui tendência. De

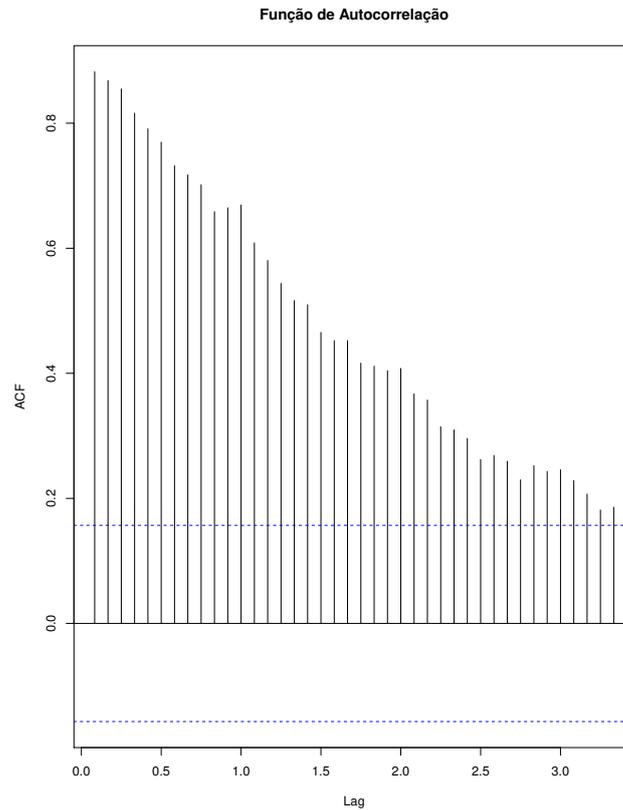


Figura 4.2: FAC da série original de faturamento.

Tabela 4.1: Teste de Cox-Stuart para tendência

Série	Estatística	P-valor
Faturamento Mensal do Atacado	78	2,2e-16

acordo com o apresentado nesta tabela, a hipótese nula foi rejeitada ao nível de significância de 5%. Sendo assim, é necessário realizar uma diferença na série para eliminar a tendência identificada.

O gráfico da série com a primeira diferença pode ser visualizado na Figura 4.3, sendo possível notar uma média constante com a variabilidade baixa. Dessa forma, aparentemente não há necessidade de realizar uma segunda diferença na série. Realizando-se o teste de Cox-Stuart na série com uma diferença (Tabela 4.2) verificou-se que foi possível retirar a tendência e torna-la uma série estacionária, fato que pode ser confirmado observando o decaimento imediato dos lags para zero Figura (4.4).

Tabela 4.2: Teste de Cox-Stuart para tendência na Série com Diferença

Série	Estatística	P-valor
Faturamento Mensal do Atacado	41	0,6488

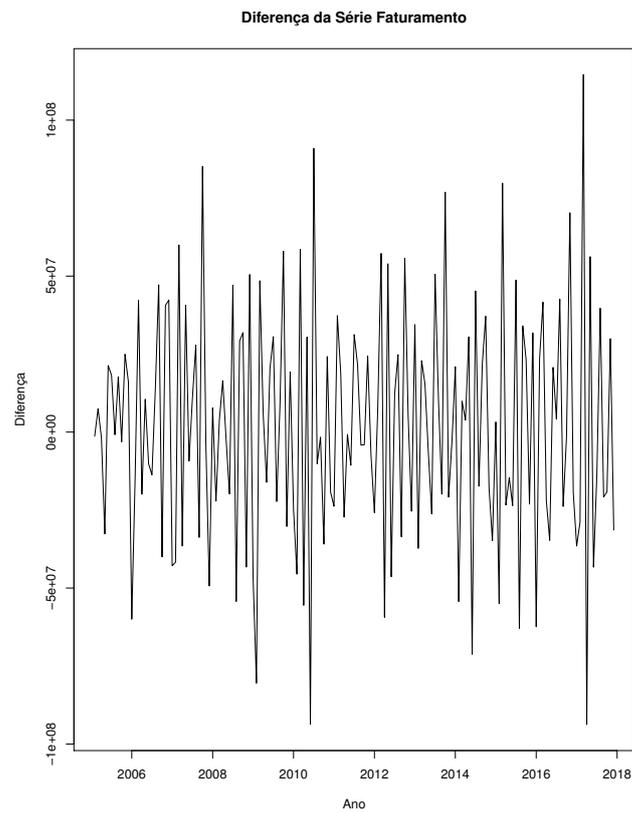


Figura 4.3: Gráfico da série original com uma diferença.

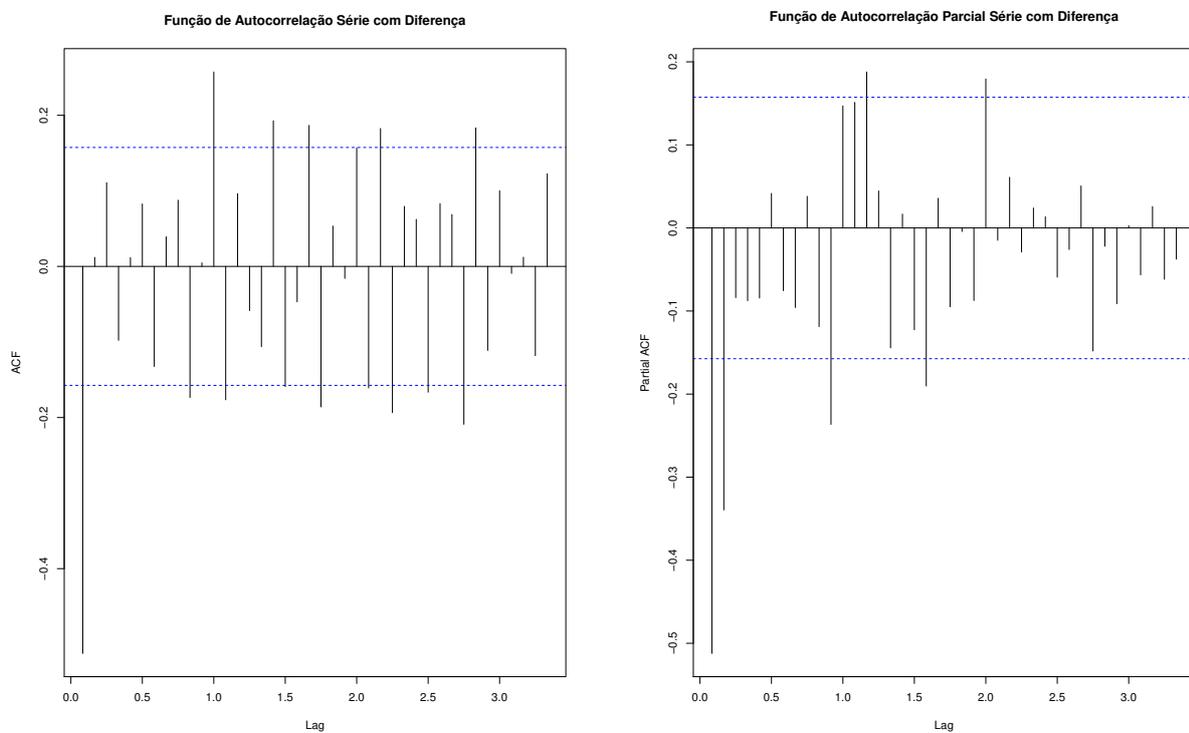


Figura 4.4: FAC e FACP da série com uma diferença.

Para a série do Faturamento Mensal com uma diferença, a FAC e a função de autocorre-

lação parcial (FACP) estão apresentadas na Figura 4.4. De acordo com a FAC e a FACP da série original e da série com uma diferença, observou-se uma sugestão de modelo para a série, tendo em vista que a FAC pode sugerir a ordem do componente de Médias Móveis no modelo. Considerando a FACP nesta Figura, e os "lags" 1 e 2 significativos, pode-se ter um indicativo da ordem do componente auto-regressivo do modelo. Dessa forma, um possível modelo para a série teria o componente auto-regressivo de ordem dois e o componente de médias móveis de ordem 1.

Com base no conceito abordado de uma série temporal sazonal, podemos ver evidências na Figura 4.1, de que a série possui comportamentos aproximadamente periódicos, ou seja os fenômenos se repetem em períodos equivalentes de tempo.

A sazonalidade determinística pressupõe um padrão sazonal regular e estável no tempo, o que permite prever o comportamento sazonal a partir dos dados. Para verificar a existência de sazonalidade na série, utilizou-se o teste formal de Fisher utilizando o comando (*fisher.g.test()*) conforme Tabela 4.3.

Tabela 4.3: Teste de Fisher para Sazonalidade

Teste Sazonalidade	P-valor
Fisher	3,027676e-20

De acordo com o apresentado na Tabela 4.3, a hipótese nula de não sazonalidade foi rejeitada ao nível de significância de 5%. Confirmando-se a existência de sazonalidade na série, deve-se então resultar na adição de um componente sazonal ao modelo.

Considerando que a série é resultante da combinação de fatores como tendência e sazonalidade, podendo resultar em um modelo aditivo ou multiplicativo, deve-se verificar qual destes tipos de modelos pode ser usado para representar a série em estudo. Uma estratégia comum para verificar essa hipótese é dividir a série em grupos de observações consecutivas e calcular para cada grupo a média e a amplitude ou o desvio padrão. Examinando o gráfico da média pelo desvio padrão ou amplitude, é possível verificar se a variabilidade depende da média, sendo necessário, em caso afirmativo, utilizar uma transformação para estabilizar a variância. A independência entre esses fatores pode também indicar o efeito aditivo da sazonalidade.

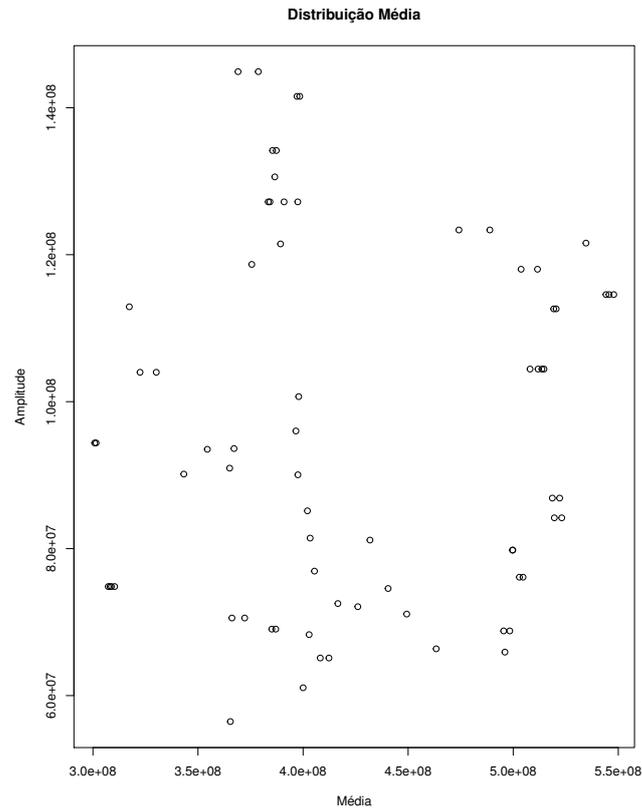


Figura 4.5: Média pela Amplitude de grupos de observações da série.

Observando a Figura 4.5, que representa a média pela amplitude de um grupo de 2 observações da série original sem diferenças, nota-se que o modelo aditivo deve ser adotado pois os componentes da série atuam isoladamente, ou seja de modo absoluto e independente entre si, sendo que a escolha adequada do modelo correto é fundamental na sensibilidade das variações sazonais. Além disso, verifica-se que não é necessário o uso de transformações, já que a amplitude não depende da média.

Após identificados os componentes da série, iniciou-se a etapa de estimação do modelo, observando a necessidade de realizar uma diferença de ordem 1 na série para retirar a tendência, e uma diferença de ordem 12 para retirar a sazonalidade determinística identificada pelo teste de Fisher. As ordens dos modelos sugeridas pela FAC e FACP também foram consideradas.

Inicialmente, ajustou-se à série um modelo autorregressivo sazonal, já que o decaimento exponencial da FAC assim sugere. A representação de tal modelo é um SARIMA(2,1,0)(0,1,0). Os parâmetros do modelo foram todos significativos de acordo com o teste t a um nível de significância de 0,05 (Tabela 4.4).

De acordo com os resíduos do modelo, apresentados na Figura 4.6, tem-se indicativo de que os resíduos possuem média zero e homocedasticidade na variação ao longo do tempo. O histograma apresentado na Figura 4.7 indica que os resíduos seguem uma distribuição normal. Observando-se a FAC dos resíduos, (Figura 4.8), é possível verificar visualmente que os resíduos são ruído branco, considerando que apenas um "lag" da FAC foi significativo. As verificações gráficas foram confirmadas por meio dos testes apresentados na Tabela 4.5, em que o teste de Kolmogorov-Smirnov permitiu aceitar a hipótese de que os resíduos apresentam distribuição

normal. Os testes de Box Pierce e Ljung-Box apresentaram pvalor maior que o nível de significância de 0,05, permitindo assim aceitar a hipótese de que os resíduos são ruído branco. Sendo assim, pode-se concluir que o ajuste deste modelo aos dados é adequado.

Tabela 4.4: Estimativas dos parâmetros do Modelo $(2,1,0)(0,1,0)$ e do teste t.

Modelo	Parâmetro	Estimativa	Estatística de Teste	P-valor
$(2,1,0)(0,1,0)$	ϕ_1	-0,7895	-10,213	$1,392692e - 18$
	ϕ_2	-0,3781	-4,891	$5,442346e - 06$

Tabela 4.5: Testes de Diagnóstico Modelo $(2,1,0)(0,1,0)$

Modelo	Teste	Estatística de Teste	P-valor
$(2,1,0)(0,1,0)$	Kolmogorov	0,060353	0,1782
	Box - Pierce	0,09796	0,7543
	Ljung - Box	0,099856	0,752

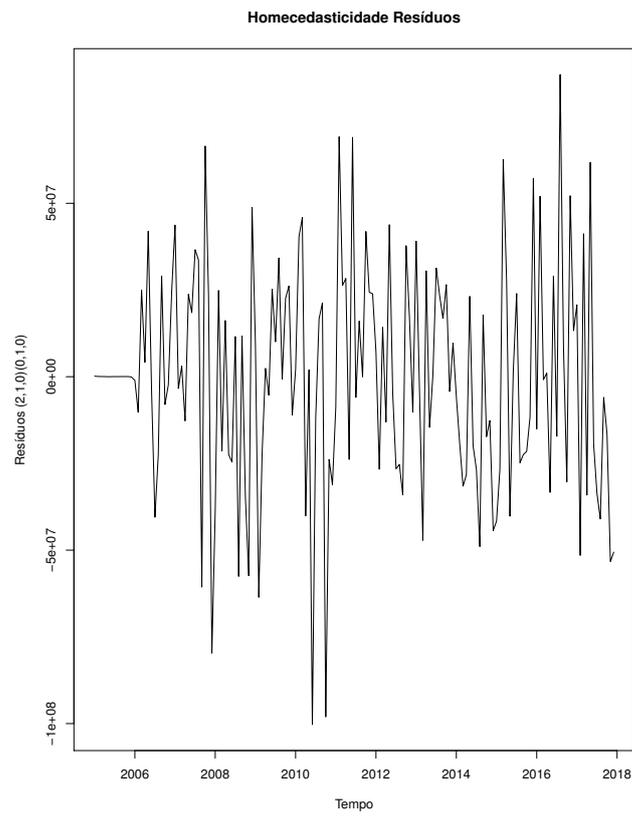


Figura 4.6: Gráfico de linhas dos Resíduos Modelo $(2,1,0)(0,1,0)$

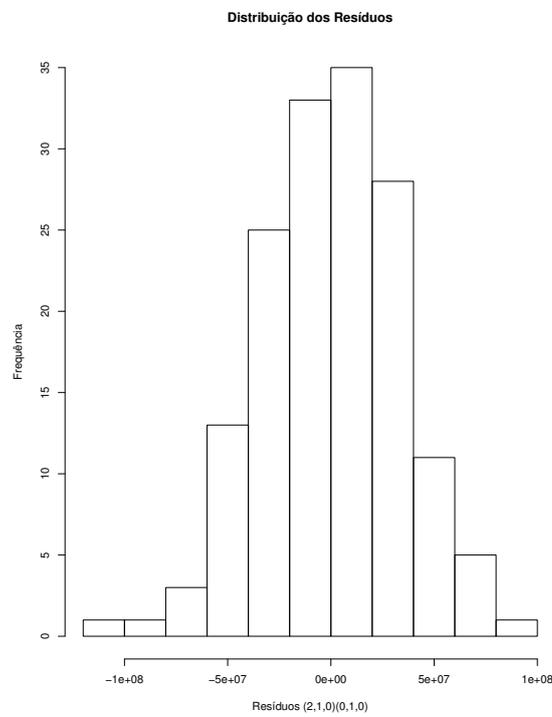


Figura 4.7: Histograma dos Resíduos do Modelo $(2,1,0)(0,1,0)$.

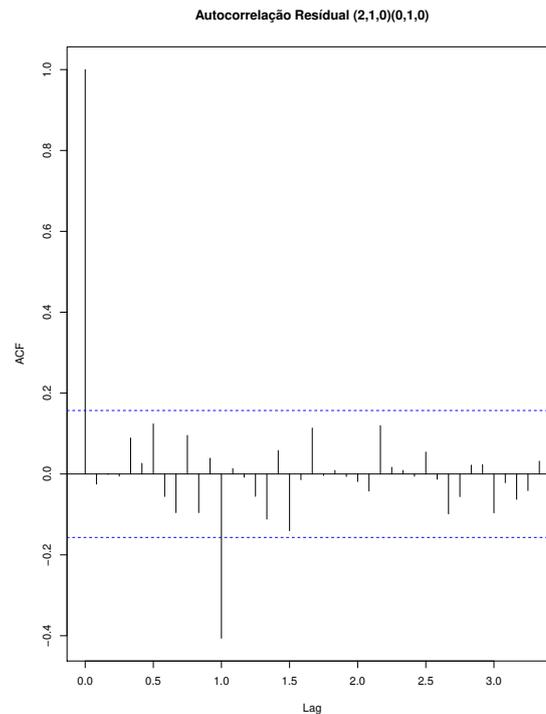


Figura 4.8: Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(0,1,0)$.

Considerando que na representação gráfica da série, (Figura 4.1), verificou-se uma queda brusca no faturamento em dois momentos (observação 50 e observação 70), retornando à sua tendência de crescimento logo em seguida, decidiu-se investigar a possibilidade de haver um efeito de intervenção significativo na série. As observações 50 e 70 referem-se respectivamente ao mês de fevereiro de 2009 e ao mês de outubro de 2010. Uma possível intervenção significativa nestes dois momentos pode ser justificada pela crise mundial na economia, ocorrida a partir de 2008, fazendo com que várias séries econômicas analisadas nesse período apresentassem esse comportamento. Sendo assim, para cada modelo significativo encontrado, ajustou-se também uma versão do modelo com efeito de intervenção em cada uma destas observações.

O segundo modelo ajustado foi o modelo de representação SARIMA(2,1,0)(0,1,0) com intervenção na casa 50. Os parâmetros do modelo também foram todos significativos de acordo com o teste t a um nível de significância de 0,05 (Tabela 4.6).

A representação gráfica dos resíduos desse modelo nas Figuras 4.9 e 4.10 apresentam indícios de que estes seguem distribuição normal com média zero e possuem homocedasticidade na variação ao longo do tempo. A FAC dos resíduos apresentou apenas um "lag" significativo, conforme Figura 4.11, sugerindo assim que os resíduos são ruído branco. Na Tabela 4.7 tem-se o teste de Kolmogorov-Smirnov que permite aceitar a hipótese nula de normalidade dos resíduos, e nesta mesma Tabela, os testes de Box Pierce e Lyjung-Box permitiram aceitar a hipótese de que os resíduos são ruído branco. Considerando estes resultados, o ajuste deste modelo aos dados pode ser considerado adequado.

Tabela 4.6: Estimativas dos parâmetros do Modelo $(2,1,0)(0,1,0)$ c/ interv.50 e do teste t.

Modelo	Parâmetro	Estimativa	Estatística de Teste	P-valor
$(2,1,0)(0,1,0)$ Interv. 50	ϕ_1	-0,8352	-10,58689	1,384985e-19
	ϕ_2	-0,3876	-4,885304	1,384985e-19
	Ω_1	-595000	-3,251366	0,002316672
	Δ_1	0,71576	5,592312	2,349747e-07

Tabela 4.7: Testes de diagnóstico modelo $(2,1,0)(0,1,0)$ c/ interv.50

Modelo	Teste	Estatística de Teste	P-valor
$(2,1,0)(0,1,0)$ Interv. 50	Kolmogorov	0,052908	0,4222
	Box - Pierce	0,12343	0,7253
	Ljung - Box	0,12604	0,7226

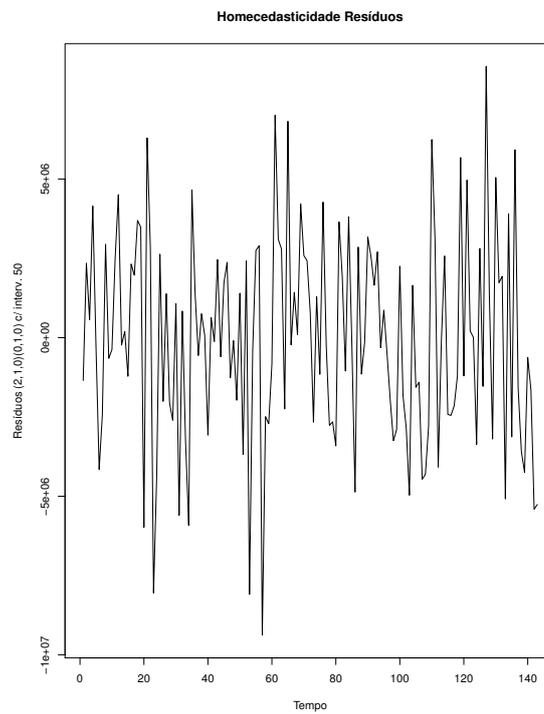


Figura 4.9: Gráfico de linhas dos Resíduos Modelo $(2,1,0)(0,1,0)$ c/ interv. 50.

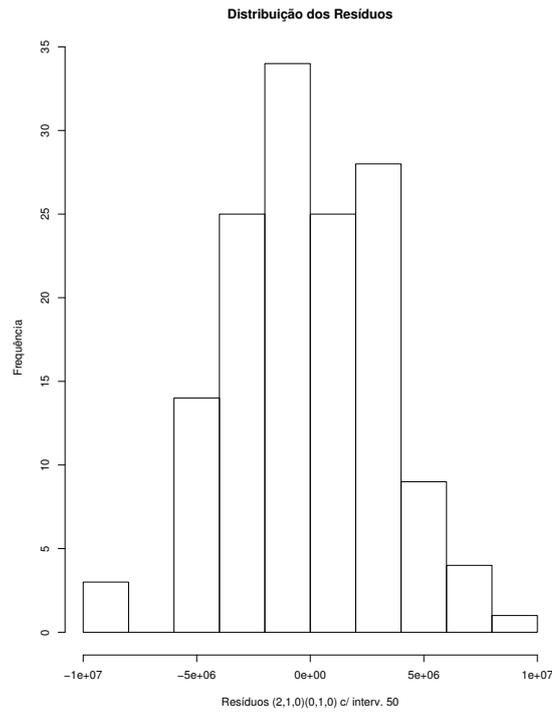


Figura 4.10: Histograma dos Resíduos do Modelo $(2,1,0)(0,1,0)$ c/ interv. 50.

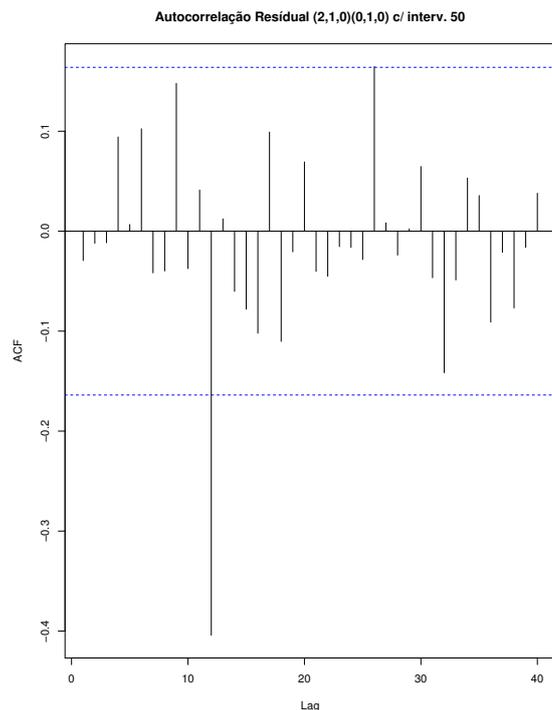


Figura 4.11: Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(0,1,0)$ c/ interv. 50.

O terceiro modelo ajustado foi o modelo de representação SARIMA(2,1,0)(0,1,0) com intervenção na casa 70. Os parâmetros do modelo também foram todos significativos de acordo com o teste t a um nível de significância de 0,05 (Tabela 4.8).

De acordo com o gráfico de linhas e o histograma apresentados nas Figuras 4.12 e 4.13 verificam-se indicativos de que os resíduos deste modelo seguem distribuição normal com mé-

dia zero e possuem homocedasticidade na variação ao longo do tempo. A FAC dos resíduos, apresentada na Figura 4.14, sugere que os resíduos do modelo são ruído branco, sendo esta hipótese também aceita pelos testes de Box Pierce e Lyjung-Box, conforme p-valor apresentado na Tabela 4.9. Nesta mesma Tabela, o teste de Kolmogorov Smirnov também apresentou pvalor maior que 0,05, permitindo assim aceitar a hipótese de normalidade dos resíduos. O modelo ajustado pode portanto ser considerado adequado para representar a série em estudo.

Tabela 4.8: Estimativas dos parâmetros do modelo (2,1,0)(0,1,0) c/ interv.70 e do teste t.

Modelo	Parâmetro	Estimativa	Estatística de Teste	P-valor
(2,1,0)(0,1,0) Interv. 70	ϕ_1	-0,8313	-10,73892	5,390298e-20
	ϕ_2	-0,4259	-5,48275	3,914114e-07
	Ω_1	-6090000	-3,530435	0,0009543181
	Δ_1	0,87838	7,352306	2,865002e-11

Tabela 4.9: Testes de diagnóstico modelo (2,1,0)(0,1,0) c/ interv.70.

Modelo	Teste	Estatística de Teste	P-valor
(2,1,0)(0,1,0) Interv. 70	Kolmogorov	0,052463	0,4359
	Box - Pierce	0,14012	0,7082
	Ljung - Box	0,14308	0,7052

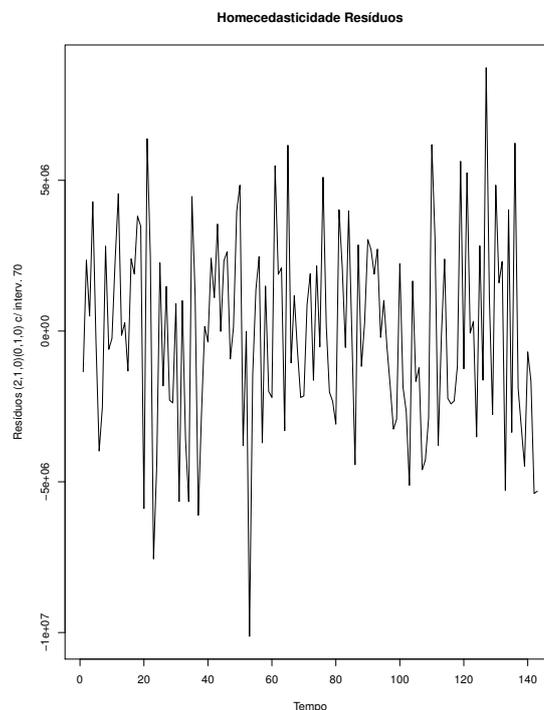


Figura 4.12: Gráfico de linhas dos Resíduos Modelo (2,1,0)(0,1,0) c/ interv. 70.

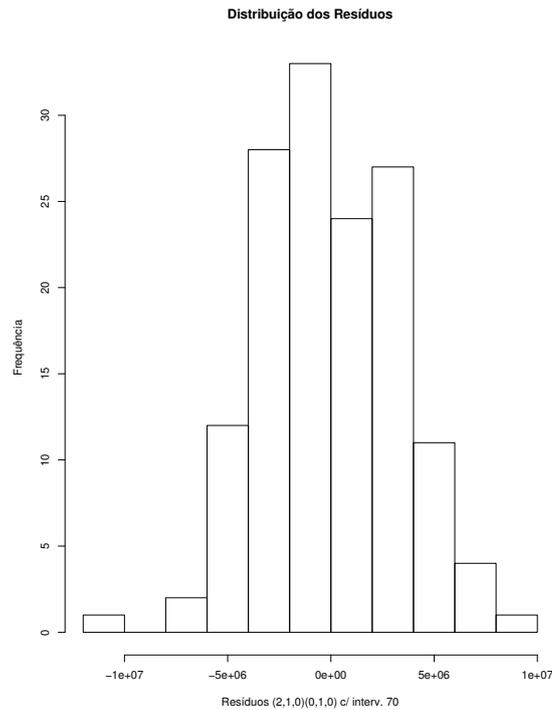


Figura 4.13: Histograma dos Resíduos do Modelo $(2,1,0)(0,1,0)$ c/ interv. 70.

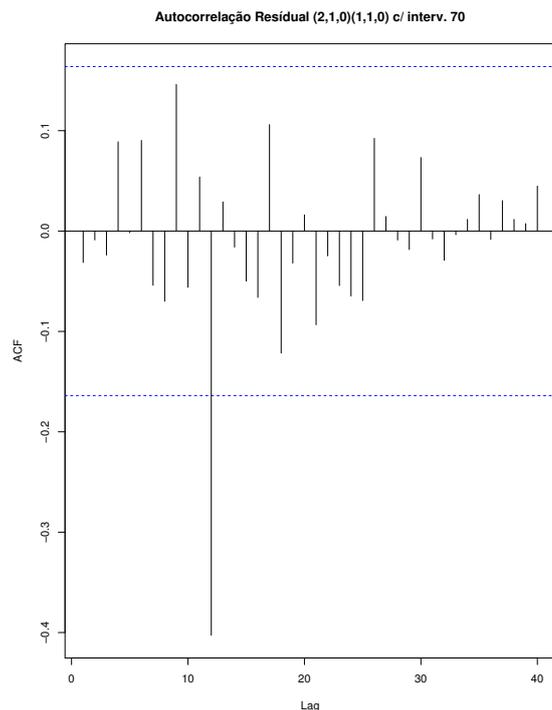


Figura 4.14: Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(1,1,0)$ c/ interv. 70.

O quarto modelo ajustado foi o modelo de representação SARIMA(2,1,0)(1,1,0), incluindo no modelo anterior um componente sazonal autorregressivo. Esse acréscimo foi sugerido por verificar na FAC indícios da presença de sazonalidade estocástica, após realizar a diferença de 12 na série. Os parâmetros do modelo também foram todos significativos de acordo com o teste t a um nível de significância de 0,05 (Tabela 4.10).

Observando-se a FAC dos resíduos do modelo (Figura 4.17) verifica-se que os resíduos são ruído branco, pois não se tem muitas autocorrelações fora do intervalo de confiança.

As Figuras 4.15 e 4.16 indicam que o resíduos seguem um distribuição normal, com média zero.

Os testes de Kolmogorov-Smirnov, Box Pierce e Lyjung-Box, apresentados na Tabela 4.11, permitem aceitar as hipóteses de normalidade dos resíduos e de ruído branco, sendo portanto possível concluir que o modelo apresentou um bom ajuste aos dados.

Tabela 4.10: Estimativas dos parâmetros do modelo (2,1,0)(1,1,0) e do teste t.

Modelo	Parâmetro	Estimativa	Estatística de Teste	P-valor
(2,1,0)(1,1,0)	ϕ_1	-0,7820	-10,496	$2,422609e - 19$
	ϕ_2	-0,4501	-5,922	$4,85389e - 08$
	Φ_1	-0,4606	-5,882	$5,89060e - 08$

Tabela 4.11: Testes de diagnóstico modelo (2,1,0)(1,1,0)

Modelo	Teste	Estatística de Teste	P-valor
(2,1,0)(1,1,0)	Kolmogorov	0,06091	0,1685
	Box - Pierce	0,004241	0,9481
	Ljung - Box	0,0043231	0,9476

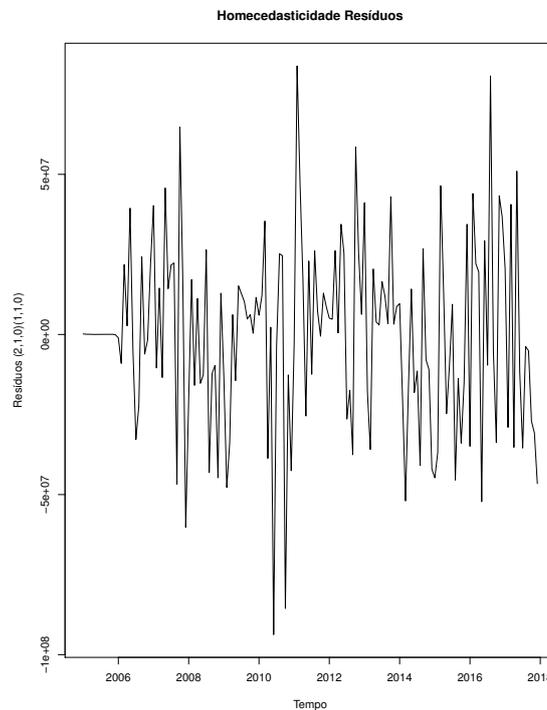


Figura 4.15: Gráfico de linhas dos Resíduos do Modelo (2,1,0)(1,1,0).

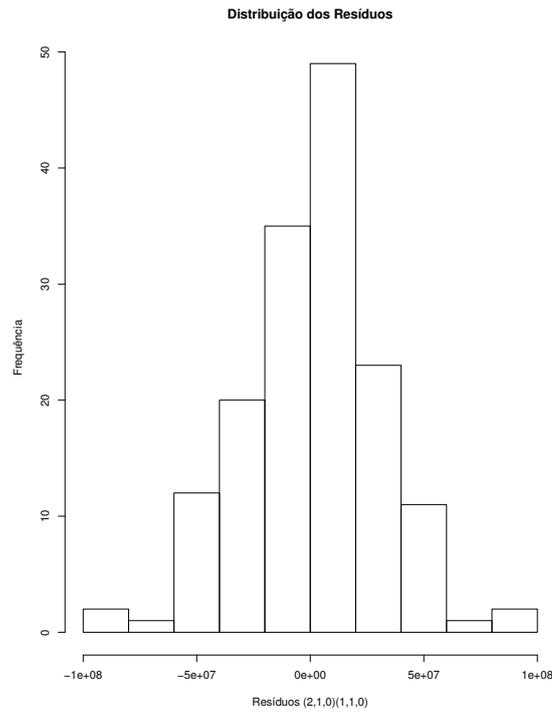


Figura 4.16: Histograma dos Resíduos do Modelo $(2,1,0)(1,1,0)$.

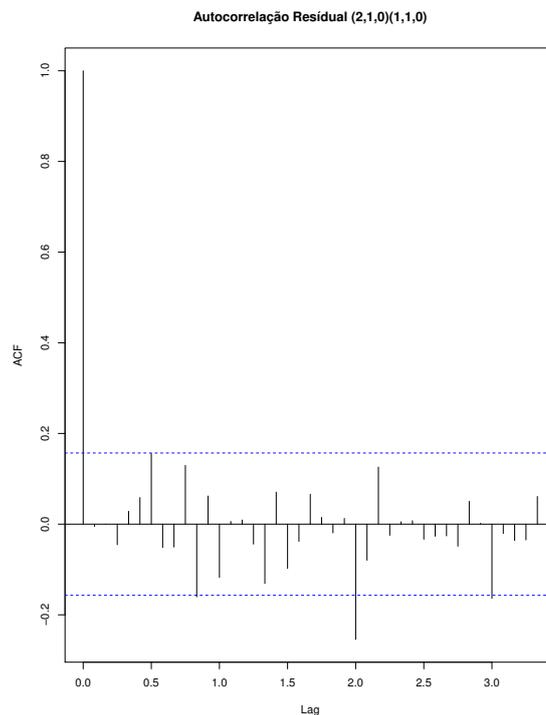


Figura 4.17: Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(1,1,0)$.

O quinto modelo ajustado foi o modelo de representação SARIMA(2,1,0)(1,1,0) com intervenção na casa 50. Os parâmetros do modelo também foram todos significativos de acordo com o teste t a um nível de significância de 0,05 (Tabela 4.12).

A representação gráfica dos resíduos desse modelo nas Figuras 4.18 e 4.19 apresentam indícios de que estes seguem distribuição normal com média zero e possuem homocedasticidade na

variação ao longo do tempo. A FAC dos resíduos apresentou poucos "lags" significativos, conforme Figura 4.20, sugerindo assim que os resíduos são ruído branco. Na Tabela 4.13 tem-se o teste de Kolmogorov-Smirnov que permite aceitar a hipótese nula de normalidade dos resíduos, e nesta mesma Tabela, os testes de Box Pierce e Lyjung-Box permitiram aceitar a hipótese de que os resíduos são ruído branco. Considerando estes resultados, o ajuste deste modelo aos dados pode ser considerado adequado.

Tabela 4.12: Estimativas dos parâmetros do modelo (2,1,0)(1,1,0) c/ interv.50 e do teste t.

Modelo	Parâmetro	Estimativa	Estatística de Teste	P-valor
(2,1,0)(1,1,0) Interv. 50	ϕ_1	-0,8209	-10,77014	4,439668e-20
	ϕ_2	-0,4618	-5,930397	4,667837e-08
	Φ_1	-0,4871	-5,752244	1,102346e-07
	Ω_1	-6140000	-3,238397	0,002410437
	Δ_1	0,73284	5,050238	2,740053e-06

Tabela 4.13: Testes de diagnóstico modelo (2,1,0)(1,1,0) c/ interv.50

Modelo	Teste	Estatística de Teste	P-valor
(2,1,0)(1,1,0) Interv. 50	Kolmogorov	0,042139	0,7728
	Box - Pierce	0,04166	0,8383
	Ljung - Box	0,04254	0,8366

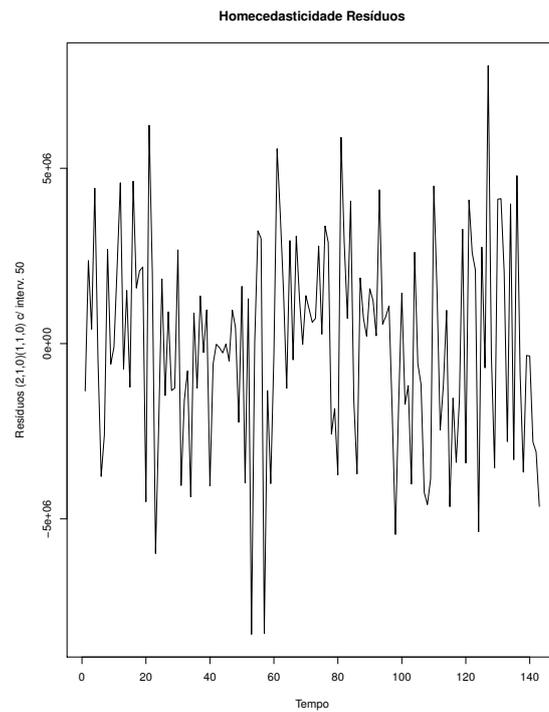


Figura 4.18: Gráfico de linhas dos Resíduos do Modelo $(2,1,0)(1,1,0)$ c/ interv. 50.

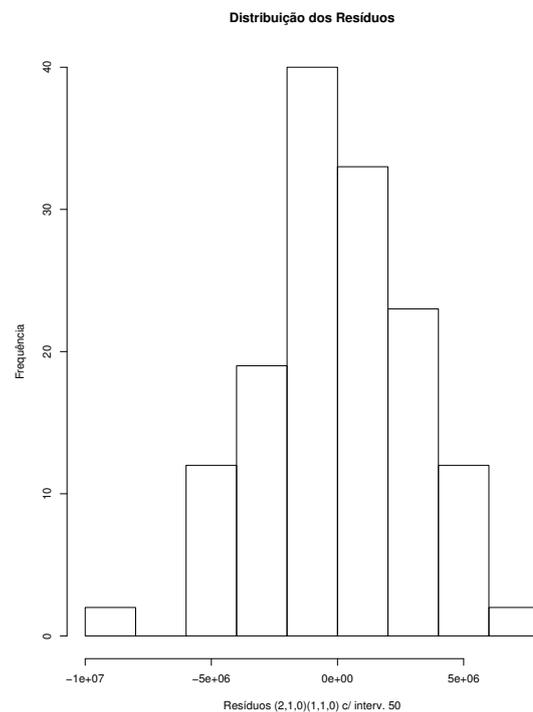


Figura 4.19: Histograma dos Resíduos do Modelo $(2,1,0)(1,1,0)$ c/ interv. 50.

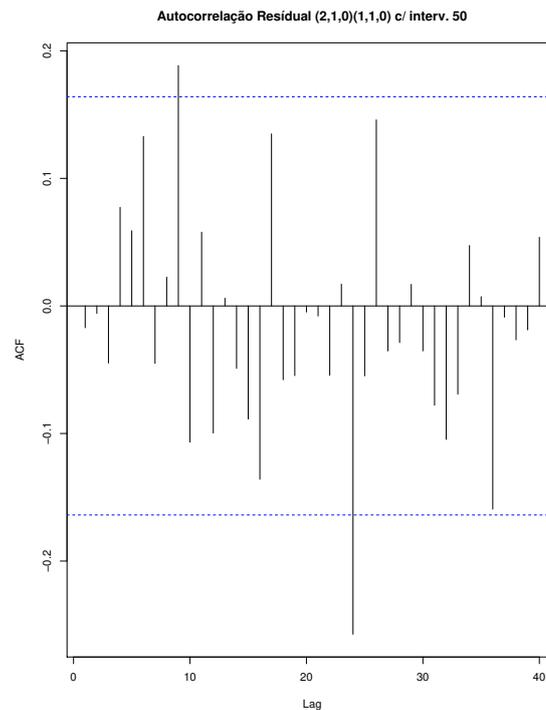


Figura 4.20: Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(1,1,0)$ c/ interv. 50.

O sexto modelo ajustado foi o modelo de representação SARIMA(2,1,0)(1,1,0) com intervenção na casa 70. Os parâmetros do modelo também foram todos significativos de acordo com o teste t a um nível de significância de 0,05 (Tabela 4.14).

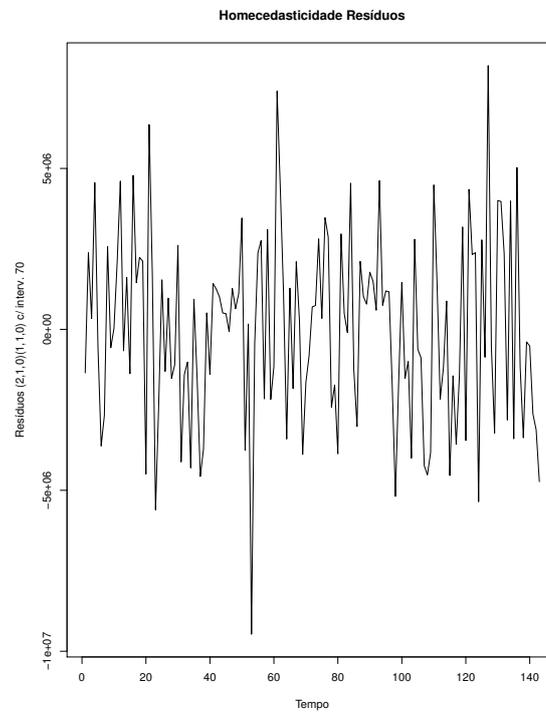
De acordo com o gráfico de linhas e o histograma apresentados nas Figuras 4.21 e 4.22 verificam-se indicativos de que os resíduos como os anteriores seguem distribuição normal com média zero e possuem homocedasticidade na variação ao longo do tempo. A FAC dos resíduos (Figura 4.23) sugere que os resíduos do modelo são ruído branco, conforme p-valor apresentado na Tabela 4.15. Assim o modelo ajustado é considerado adequado para representar a série em estudo.

Tabela 4.14: Estimativas dos parâmetros do modelo $(2,1,0)(1,1,0)$ c/ interv.70 e do teste t.

Modelo	Parâmetro	Estimativa	Estatística de Teste	P-valor
$(2,1,0)(1,1,0)$ Interv. 70	ϕ_1	-0,8103	-10,82131	3,229512e-20
	ϕ_2	-0,4928	-6,483357	2,927096e-09
	Φ_1	-0,4907	-5,834703	7,421412e-08
	Ω_1	-6270000	-3.530405	0,0009544104
	Δ_1	0,92606	10,3865	4,78825e-19

Tabela 4.15: Testes diagnóstico modelo $(2,1,0)(1,1,0)$ c/ interv.70

Modelo	Teste	Estatística de Teste	P-valor
$(2,1,0)(1,1,0)$ Interv. 70	Kolmogorov	0,045554	0,6628
	Box - Pierce	0,0025555	0,9597
	Ljung - Box	0,0026095	0,9593

Figura 4.21: Gráfico de linhas dos Resíduos do Modelo $(2,1,0)(1,1,0)$ c/ interv. 70.

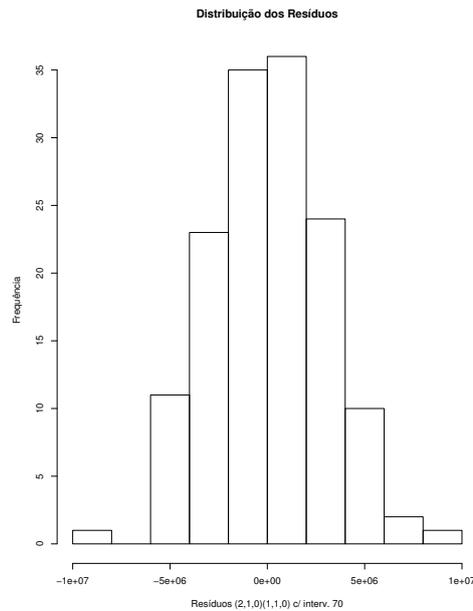


Figura 4.22: Histograma dos Resíduos do Modelo $(2,1,0)(1,1,0)$ c/ interv. 70.

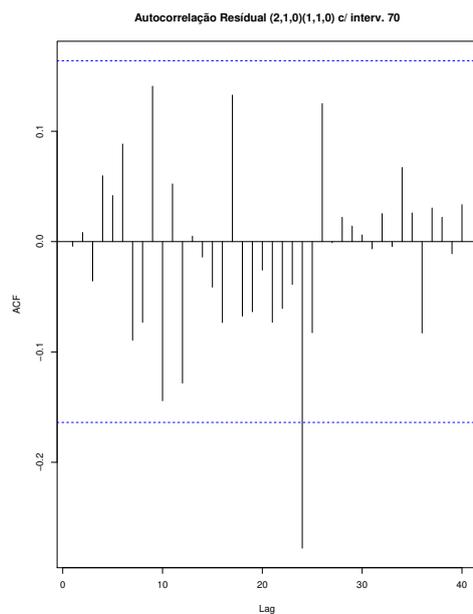


Figura 4.23: Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(1,1,0)$ c/ interv. 70.

Um sétimo modelo incluindo um componente de médias móveis e médias móveis sazonal, $SARIMA(1,1,1)(0,0,2)$, também foi ajustado aos dados sendo que todos os parâmetros foram significativos de acordo com a Tabela 4.16. Conforme modelos anteriores para este caso os resíduos indicam média zero e homocedasticidade na variação ao longo do tempo, conforme Figura 4.24 e Figura 4.25.

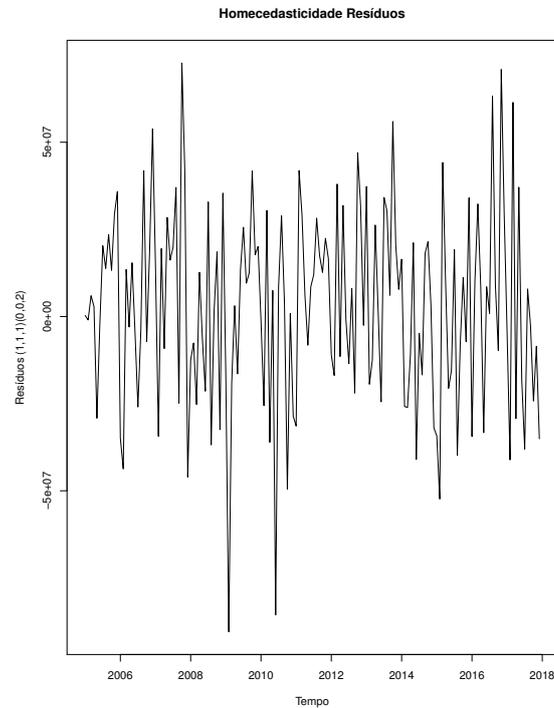
Observando-se a FAC dos resíduos (Figura 4.26) é possível verificar visualmente que os resíduos são ruído branco, onde a hipótese de ruído branco pode ser confirmada pela Tabela 4.17. Assim é possível dizer que o ajuste deste modelo aos dados é adequado.

Tabela 4.16: Estimativas dos parâmetros do modelo $(1,1,1)(0,0,2)$ e do teste t.

Modelo	Parâmetro	Estimativa	Estatística de Teste	P-valor
$(1,1,1)(0,0,2)$	ϕ_1	-0,2317	-2,1102	0,0437221
	θ_1	-0,5236	-5,7601	$1,061347e - 07$
	Θ_1	0,3203	3,9300	0,00024078
	Θ_2	0,2270	2,5250	0,01716248

Tabela 4.17: Testes de diagnóstico modelo $(1,1,1)(0,0,2)$

Modelo	Teste	Estatística de Teste	P-valor
$(1,1,1)(0,0,2)$	Kolmogorov	0,049316	0,4665
	Box - Pierce	0,14447	0,7039
	Ljung - Box	0,14726	0,7012

Figura 4.24: Gráfico de linhas dos Resíduos Modelo $(1,1,1)(0,0,2)$

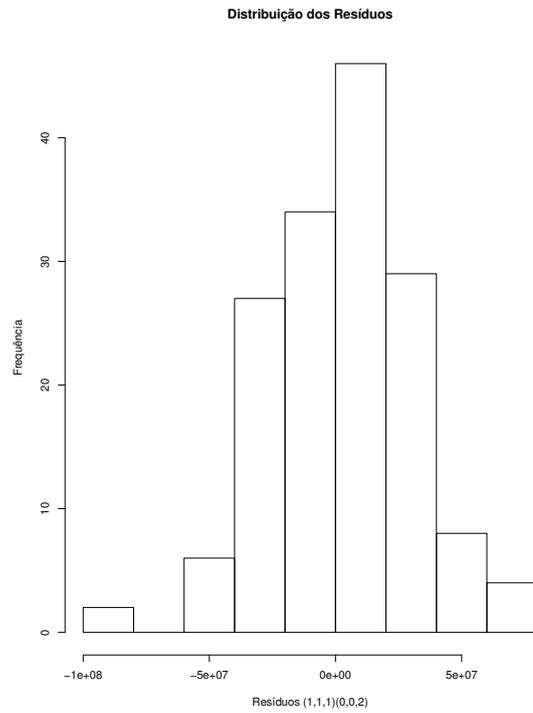


Figura 4.25: Histograma dos Resíduos do Modelo $(1,1,1)(0,0,2)$.

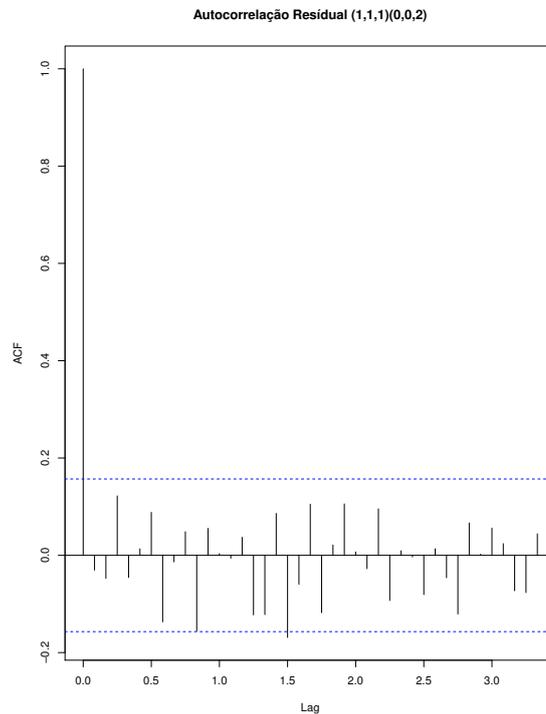


Figura 4.26: Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(0,1,0)$.

Um oitavo modelo ajustado é o SARIMA(1,1,1)(0,0,2) incluindo o efeito de intervenção na casa 50. Os parâmetros do modelo foram todos significativos de acordo com o teste t a um nível de significância de 0,05 (Tabela 4.18)

A representação gráfica dos resíduos desse modelo nas Figura 4.27 e 4.28 apresentam indícios

de que estes seguem distribuição normal com média zero e possuem homocedasticidade na variação ao longo do tempo.

Observando-se a FAC dos resíduos (Figura 4.29) não apresentou nenhum "lag" significativo sugerindo assim que os resíduos são ruído branco. Na Tabela 4.19 tem-se o teste de Kolmogorov-Smirnov que permite aceitar a hipótese nula de normalidade dos resíduos, e nesta mesma Tabela, os testes de Box Pierce e Ljung-Box permitiram aceitar a hipótese de que os resíduos são ruído branco. Com estes resultados, este modelo é adequado para o ajuste dos dados.

Tabela 4.18: Estimativas dos parâmetros do modelo $(1,1,1)(0,0,2)$ c/ interv. 50 e do teste t.

Modelo	Parâmetro	Estimativa	Estatística de Teste	P-valor
$(1,1,1)(0,0,2)$ Interv. 50	ϕ_1	-0,2665	-2,492052	0,01859234
	ϕ_2	0,53447	6,162458	1,48695e-08
	θ_1	-0,3267	-3,821947	0,0003537427
	θ_2	-0,2260	-2,470215	0,01959365
	Ω_1	-7400000	-3,392939	0,001488959
	Δ_1	0,69786	4,765827	9,251621e-06

Tabela 4.19: Testes diagnóstico modelo $(1,1,1)(0,0,2)$ c/ interv. 50.

Modelo	Teste	Estatística de Teste	P-valor
$(1,1,1)(0,0,2)$ Interv. 50	Kolmogorov	0,045184	0,6137
	Box - Pierce	0,13103	0,7174
	Ljung - Box	0,13358	0,7147

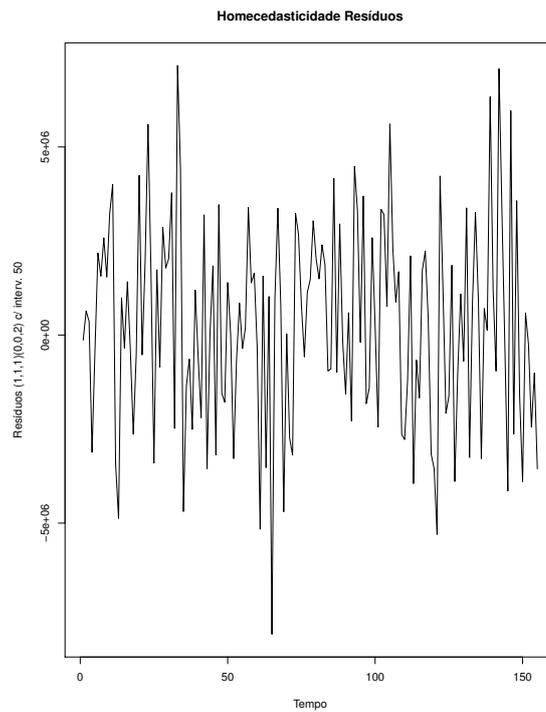


Figura 4.27: Gráfico de linhas dos Resíduos Modelo $(1,1,1)(0,0,2)$ c/ interv. 50

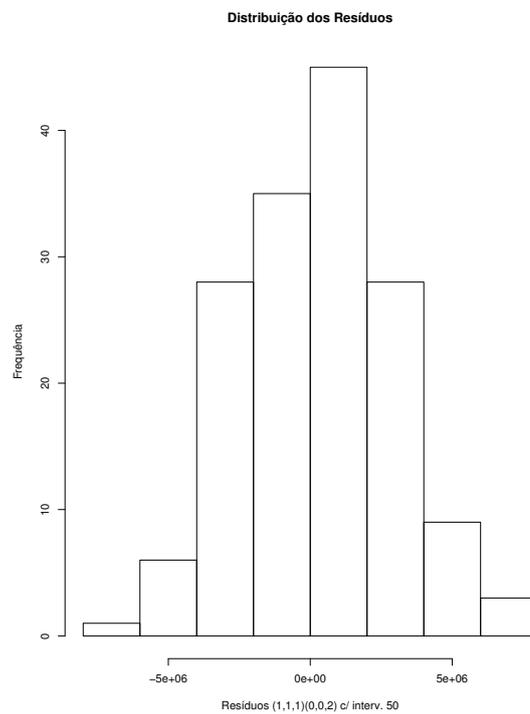


Figura 4.28: Histograma dos Resíduos do Modelo $(1,1,1)(0,0,2)$ c/ interv. 50.

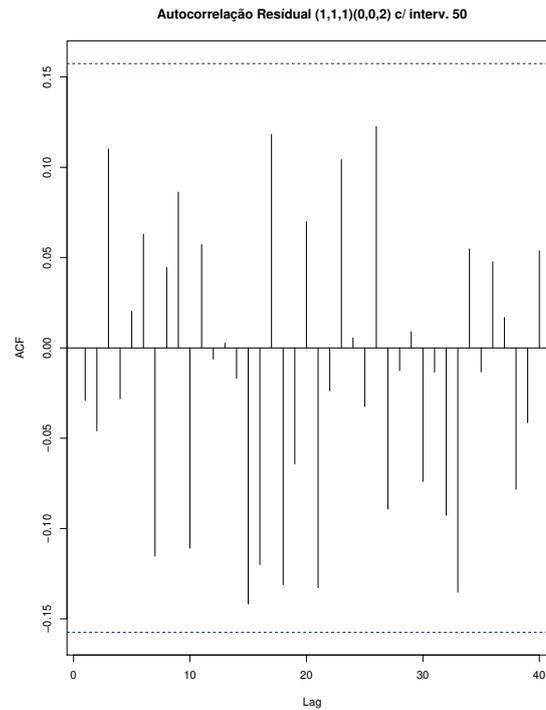


Figura 4.29: Gráfico da Autocorrelação Residual do Modelo $(2,1,0)(0,1,0)$ c/ interv. 50.

Um nono modelo ajustado foi o modelo de representação SARIMA(1,1,1)(0,0,2) com intervenção na casa 70. Os parâmetros do modelo também foram todos significativos de acordo com o teste t a um nível de significância de 0,05 (Tabela 4.20).

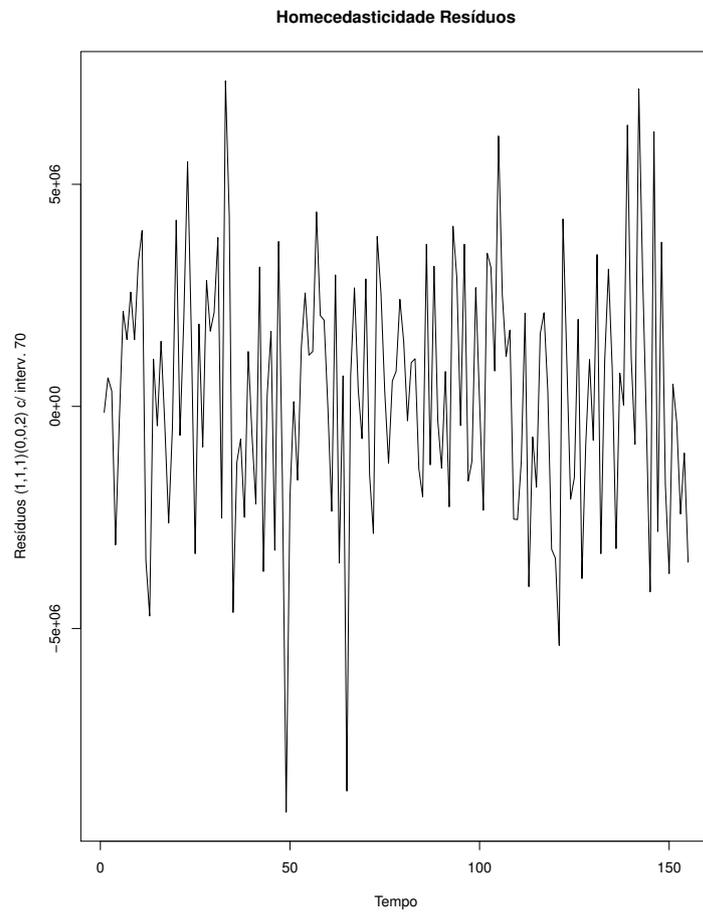
De acordo com o gráfico de linhas e o histograma apresentados nas Figuras 4.30 e 4.31 verificam-se indicativos de que os resíduos deste modelo seguem distribuição normal com média zero e possuem homocedasticidade na variação ao longo do tempo. A FAC dos resíduos, apresentada na Figura 4.32, sugere que os resíduos do modelo são ruído branco, sendo esta hipótese também aceita pelos testes de Box Pierce e Lyjung-Box, conforme p-valor apresentado na Tabela 4.21. Nesta mesma Tabela, o teste de Kolmogorov Smirnov também apresentou pvalor maior que 0,05, permitindo assim aceitar a hipótese de normalidade dos resíduos. O modelo ajustado pode portanto ser considerado adequado para representar a série em estudo.

Tabela 4.20: Estimativas dos parâmetros do modelo $(1,1,1)(0,0,2)$ c/ interv. 70 e do teste t.

Modelo	Parâmetro	Estimativa	Estatística de Teste	P-valor
$(1,1,1)(0,0,2)$ Interv. 70	ϕ_1	-0,2311	-2,056233	0,04881012
	ϕ_2	0,54577	-5,930349	4,668928e-08
	θ_1	-0,3361	-3,923652	0,0002463951
	θ_2	-0,2061	-2,242655	0,03299266
	Ω_1	-4260000	-2,092338	0,04535809
	Δ_1	0,93357	9,317066	3,291471e-16

Tabela 4.21: Testes de diagnóstico modelo $(1,1,1)(0,0,2)$ c/ interv. 70

Modelo	Teste	Estatística de Teste	P-valor
$(1,1,1)(0,0,2)$ Interv. 70	Kolmogorov	0,036552	0,8806
	Box - Pierce	0,10765	0,7428
	Ljung - Box	0,10974	0,7404

Figura 4.30: Gráfico de linhas dos Resíduos Modelo $(1,1,1)(0,0,2)$ c/ interv. 70

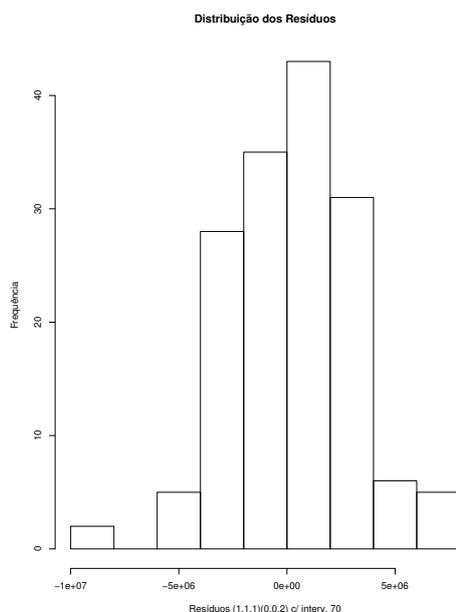


Figura 4.31: Histograma dos Resíduos do Modelo $(1,1,1)(0,0,2)$ c/ interv. 70.

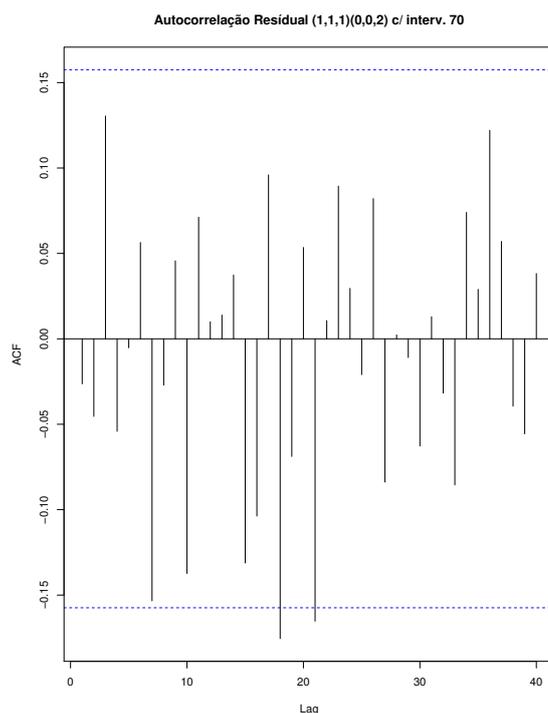


Figura 4.32: Gráfico da Autocorrelação Residual do Modelo $(1,1,1)(0,0,2)$ c/ interv. 70.

Decidiu-se analisar o erro percentual absoluto médio de previsão (MAPE), para verificar quais dos modelos apresenta maior capacidade de previsão, considerando que este é um dos principais objetivos ao se ajustar um modelo de séries temporais. Para finalidade de previsão, os dados de faturamento da empresa, referentes ao ano de 2018, foram reservados para comparação com os valores preditos pelos modelos. Na Tabela 4.22 verifica-se o valor de MAPE para cada um dos modelos.

Os modelos com intervenção apresentaram um ajuste melhor, considerando que seus valores de MAPE foram menores que as versões dos mesmos modelos sem intervenção. Sendo assim, os modelos com intervenção apresentaram as melhores previsões, portanto deve-se utilizar como modelo mais adequado aos dados, aquele modelo com intervenção que apresentou o menor erro percentual absoluto médio. De acordo com esse critério, entre esses 9 modelos, o mais adequado é o modelo 8, SARIMA(1,1,0)(0,0,2) com intervenção na casa 50, por apresentar menor valor de MAPE, sendo que o modelo SARIMA(1,1,0)(0,0,2) também apresentou o menor valor de AIC.

Tabela 4.22: Critério de Seleção do Modelo

Modelo	MAPE
(2,1,0)(0,1,0)	0,0850
(2,1,0)(0,1,0) c/ interv. 50	0,0850
(2,1,0)(0,1,0) c/ interv. 70	0,0838
(2,1,0)(1,1,0)	0,0487
(2,1,0)(1,1,0) c/ interv. 50	0,0478
(2,1,0)(1,1,0) c/ interv. 70	0,0476
(1,1,1)(0,0,2)	0,0435
(1,1,1)(0,0,2) c/ interv. 50	0,0429
(1,1,1)(0,0,2) c/ interv. 70	0,0442

Na Figura 4.33 e Tabela 4.23 tem-se o valor real mensal de faturamento do ano de 2018, comparado com as previsões das três variações do modelo SARIMA(1, 1, 1)(0, 0, 2) que foram significativos e selecionados por serem os mais adequados para previsão, já que apresentaram o menor valor de MAPE.

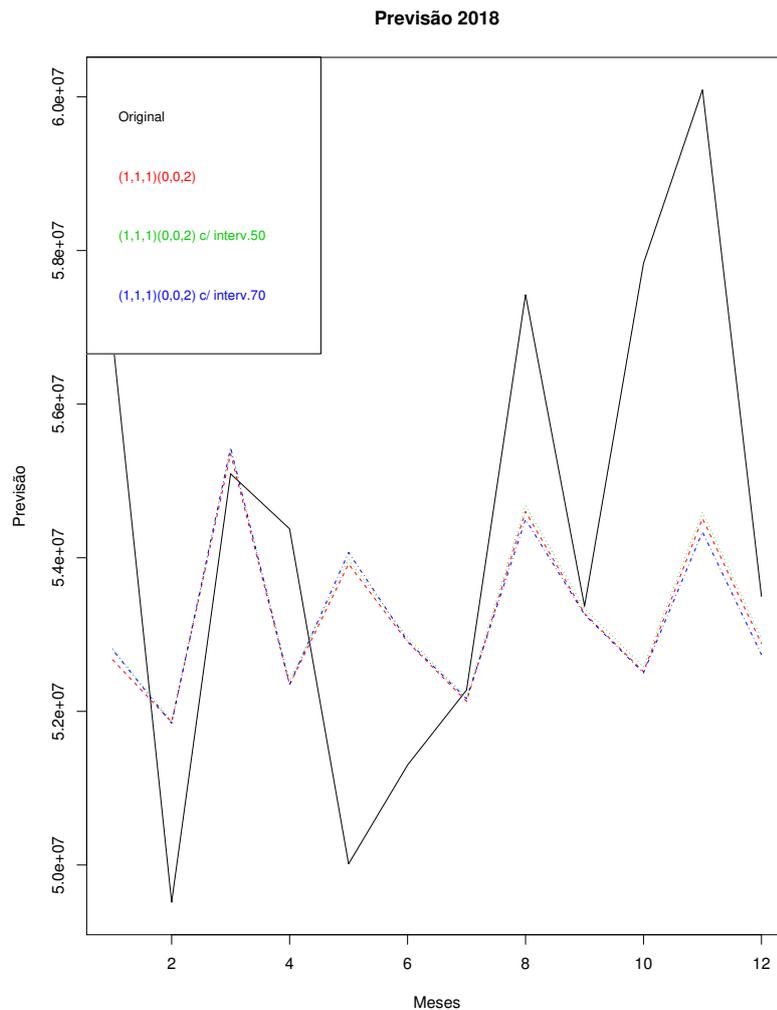


Figura 4.33: Comparação dos três modelos significativos com a série original.

É possível notar que o modelo não acompanhou significativamente a queda no faturamento mensal em Maio de 2018, refletido pelo início da greve dos caminhoneiros em 21 de Maio. A greve resultou no bloqueio das estradas impossibilitando a entrega dos produtos vendidos pela empresa, impactando diretamente no faturamento até o início da “Black Friday”, época em que faturamento voltou a crescer à níveis normais devido a grande procura natural nesta época do ano, por conta de promoções e descontos. No período destinado a “Black Friday” o volume de vendas aumenta consideravelmente, porém esse aumento não foi previsto pelo modelo, já que os dados históricos antes de 2012 não tinham esse comportamento.

Tabela 4.23: Faturamento Observado e Previstos.

Mês	Original	(1,1,1)(0,0,2)	^{0.80} (1,1,1)(0,0,2) c/ interv. 50	(1,1,1)(0,0,2) c/ interv.70
Jan/18	56844622	52671544	52817451	52802921
Fev/18	49514856	51861938	51874656	51845475
Mar/18	55093023	55351527	55429473	55434148
Abr/18	54376641	52348186	52398920	52350475
Mai/18	50014893	53918607	53976289	54069207
Jun/18	51301773	52903730	52962563	52917907
Jul/18	52277156	52131677	52185463	52164216
Ago/18	57424968	54602182	54680183	54498437
Set/18	53369569	53274311	53338543	53261217
Out/18	57832845	52517722	52578905	52500775
Nov/18	60091388	54511865	54593106	54336426
Dez/18	53495882	52882646	52951200	52739323

Assim, de acordo com os resultados, decide-se por escolher o modelo SARIMA(1,1,1)(0,0,2) com intervenção na casa 50 para representar a série em questão. A expressão do modelo ajustado é dada pela seguinte equação.

$$Z_t = \frac{-7400000}{1 - 0,69786B} X_{1,t} + \frac{(1 - 0,53447B)(1 + 0,3267B^{12} + 0,2260B^{2*12})a_t}{(1 + 0,2665B)(1 - B)}, \quad (4.1)$$

em que a variável dummy é:

$$X_{1,t} = \begin{cases} 0, & t \neq 50 \\ 1, & t = 50 \end{cases}$$

Na Tabela 4.24 tem-se a previsão do modelo com seus limites de confiança.

Mês	Previsão	Realizado	Lim. Inf. 95%	Lim. Sup. 95%
Jan/18	52817451	56844622	48177874	57457029
Fev/18	51874656	49514856	47144056	56605256
Mar/18	55429473	55093023	50326431	60532514
Abr/18	52398920	54376641	47035791	57762048
Mai/18	53976289	50014893	48344040	59608538
Jun/18	52962563	51301773	47078924	58846201
Jul/18	52185463	52277156	46059356	58311571
Ago/18	54680183	57424968	48321202	61039164
Set/18	53338543	53369569	46754829	59922258
Out/18	52578905	57832845	45777903	59379907
Nov/18	54593106	60091388	47581540	61604672
Dez/18	52951200	53495882	45735214	60167186

Após as análises informais e os testes de hipóteses pode-se afirmar que os erros do modelo escolhido para representar a série, tem média zero com variação constante ao longo do tempo

e não são autocorrelacionados, ou seja, é ruído branco, concluindo então que o modelo foi bem ajustado.

5. CONCLUSÕES

Neste trabalho, cumpriu-se a proposta de estudar o comportamento e as características da série de faturamento mensal de uma empresa atacadista de uma grande cidade, utilizando a metodologia de séries temporais, proposta por Box e Jenkins, para modelar e fazer previsão para o faturamento da mesma empresa para o ano de 2018.

O faturamento dessa empresa apresentou sazonalidade e tendência crescente no período analisado, entre 2005 e 2017, com uma queda brusca no nível da série nos anos de 2009 e 2011, voltando entretanto, ao nível original e à tendência de crescimento normais da série.

Na etapa de estimação, foi possível selecionar adequadamente três modelos SARIMA(p,d,q)(P,D,Q) que apresentaram bom ajuste aos dados, e para cada modelo, ajustou-se duas novas versões incluindo efeito de intervenção para as observações fevereiro de 2009 e outubro de 2011, sendo todos os parâmetros significativos e o ajuste também conduziu ao ruído branco.

O modelo selecionado como o mais adequado para previsões, de acordo com o critério do erro percentual médio absoluto, modelo SARIMA (1,1,1)(0,0,2) com intervenção na casa 50, ou seja com intervenção em Fevereiro de 2009. A significância desse modelo com intervenção sugere que os efeitos da crise mundial ocorrida a partir de 2008 possivelmente refletiu nos resultados da empresa considerada, como é comumente observado em séries econômicas.

A previsão ainda permanece como uma das ferramentas mais poderosas e importantes no auxílio do planejamento de qualquer empresa, independente da sua localização e tamanho. Com essa previsão a empresa é capaz de construir uma reserva financeira e planejar o fluxo de caixa para imprevistos internos e externos ao negócio.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] *RStudio: Integrated Development for R*. RStudio Inc., 2015. <http://www.rstudio.com/>, acessado em 10/08/2018.
- [2] *ABG Consultoria.*, 2019. <https://www.abgconsultoria.com.br/blog/caracteristicas-das-series-temporais/>, acessado em 11/02/2019.
- [3] *Suporte ao Minitab*, 2019. <https://support.minitab.com/ptbr/minitab/17/topic-library/modeling-statistics/time-series/time-series-models/what-aremapemad-and-msd/>, acessado em 11/02/2019.
- [4] Akaike, H.: *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 19(06):716–723, 1974.
- [5] Antunes, J.L.F. e Cardoso, M.R.A.: *Uso da análise de séries temporais em estudos epidemiológicos*. Monografia, 2015.
- [6] Becker, M.H. e Bisognin, C.: *Modelos para previsão em séries temporais: uma aplicação para a taxa de desemprego na Região Metropolitana de Porto Alegre*. Monografia, 2010.
- [7] Box, G.E.P., Jenkins, G.M. e Reinsel, G.C.: *Time Series Analysis: Forecasting and Control*, vol. 1. Wiley, 5ª ed., 2015.
- [8] Jacobs, W., Zanini, R.R. e Costa, M.: *Estudo Comparativo de Séries Temporais para Previsão de Vendas de um Produto*. 2014.
- [9] Martin, A.C., Henning, E., Walter, O.M.F.C. e Konrath, A.C.: *Análise de séries temporais para previsão da evolução do número de automóveis no Município de Joinville*. Espacios, 37(06):37, 2016.
- [10] Morettin, P.A. e Toloi, C.M.C.: *Análise de Séries Temporais*, vol. 1. ABE - Projeto Fisher, 2ª ed., 2006.
- [11] Nogueira, F.: *Modelagem e Simulação - Modelos de Previsão*, 2009. <https://www.ufjf.br/epd042/files/2009/02/previsao1.pdf>, acessado em 12/02/2019.
- [12] Pinho, F.C., Castro Agra, M. de e Salles, A.A. de: *Previsão de Vendas nos Mercados de Vestuário e Alimentos*. Monografia, 2015.

-
- [13] Rezende, L.F.: *Modelagem via séries temporais da produção de milho no Brasil*. Monografia, 2016.
- [14] Schwarz, G.: *Estimating the Dimension of a Model*. Institute of Mathematical Statistics, 6(02):461–464, 1978.
- [15] Spanhol, C.P., Benites, A.T. e Neto, L.F.F.: *Modelo de Previsões de Holt - Winters Aplicado ao Índice de Faturamento Real do Comércio Varejista de Alimentos da RMSP*. pp. 93–107, 2004.
- [16] Wheelwright, S.C. e Makridakis, S.: *Forecasting Methods for Management*, vol. 1. Wiley, 5ª ed., 1989.
- [17] Zhang, G.P.: *Time series forecasting using a hybrid ARIMA and neural network model*. Em *Neurocomputing*, pp. 159–175. Elsevier Science B.V, 2003.