



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Estatística

**APLICAÇÃO DE SÉRIES TEMPORAIS ÀS VENDAS DE
AUTOMÓVEIS EM CONCESSIONÁRIAS DO BRASIL NO
PERÍODO DE 1995 À 2021**

Manoel José do Nascimento Netto

Julho / 2021

Manoel José do Nascimento Netto

**APLICAÇÃO DE SÉRIES TEMPORAIS ÀS VENDAS DE
AUTOMÓVEIS EM CONCESSIONÁRIAS NO BRASIL NO
PERÍODO DE 1995 À 2021**

Trabalho de Conclusão de Curso retratado ao Curso de Bacharelado em Estatística na Universidade Federal da Paraíba, como requisito parcial para obtenção do Grau de Bacharelado nas Áreas de Concentração: Series Temporais e Estatística Aplicada.

Orientador: Prof^o Dr. João Agnaldo do Nascimento

**João Pessoa
Julho de 2021**

Catálogo na publicação
Seção de Catalogação e Classificação

N476a Netto, Manoel Jose do Nascimento.

Aplicação de séries temporais às vendas de automóveis em concessionárias do Brasil no período de 1995 à 2021 / Manoel Jose do Nascimento Netto. - João Pessoa, 2021. 42 f. : il.

Orientação: Joao Agnaldo Nascimento.

TCC (Graduação/Bacharelado em Estatística) - UFPB/CCEN.

1. Séries temporais. 2. ARIMA e SARIMA - Modelos de previsão. 3. Previsão de vendas - Setor automobilístico. I. Nascimento, Joao Agnaldo. II. Título.

UFPB/CCEN

CDU 519.246.8(043.2)

*Este projeto é dedicado, principalmente, à minha família
e poucos amigos, que, com certeza, me acrescentaram
muito ao longo deste curso.*

Agradecimentos

Em primeiro lugar, agradeço aos meus pais, Marta e Arimatéia, por nunca terem medido esforços para me proporcionar um ensino de qualidade durante todo o meu período escolar.

Agradeço mais uma vez aos meus pais, pela minha vida, e por me permitirem e me incentivarem a ultrapassar todos os obstáculos encontrados ao longo da realização da vida acadêmica e deste trabalho em especial.

A minha mãe em especial, que me incentivou nos momentos mais difíceis e que sempre buscou compreender a minha ausência enquanto eu me dedicava ao estudo da faculdade e à realização deste trabalho.

Aos meus irmãos, Mayara e Hugo, pelo companheirismo, pela cumplicidade e pelo apoio em todos os momentos delicados da minha vida.

Aos amigos, em especial Mateus Barreto, Fernanda Florencio, Milleny Fernandes e Clarissa Cavalcanti que sempre estiveram ao meu lado, pela amizade incondicional e pelo apoio demonstrado ao longo de todo o período de tempo em que me dediquei não somente à faculdade, mas também a este trabalho.

Agradeço a minha namorada, Sarah Lins, que nunca me negou apoio e incentivo.

A todos que participaram, direta ou indiretamente do desenvolvimento deste trabalho de pesquisa, enriquecendo o meu processo de aprendizado.

Ao meu orientador, que conduziu o trabalho com paciência e dedicação, sempre disponível a compartilhar todo o seu vasto conhecimento.

*”A paciência é a única solução para os males que não
têm solução.
(Joubert, Joseph)*

RESUMO

A previsão de vendas torna-se fundamental para obtenção de resultados positivos nos mais diversos tipos de empreendimentos. Desta forma, o presente trabalho tem por objetivo encontrar um modelo de previsão com alto poder preditivo para vendas de automóveis no Brasil utilizando dados extraídos no período de 1995 à 2021. O estudo foi realizado em três etapas: análise descritiva, modelos de previsão ARIMA e SARIMA e por fim a análise da acuidade do modelo. A base de dados consiste em 365 informações referentes as vendas realizadas pelas concessionárias de automóveis no Brasil. Através do gráfico da série foi possível observar uma certa sazonalidade e uma tendência crescente. O teste de Dikey Fuller confirmou que a série é não-estacionária. Em princípio, utilizou-se os gráficos ACF e PACF para identificação dos parâmetros do modelo ARIMA. Na próxima etapa, foram gerados dois modelos, ambos obtidos através do comando 'auto.arima'. O primeiro teve como critério de seleção, aquele que apresentasse o menor AIC, e o segundo BIC. Adicionalmente, verificamos em ambos os modelos que os resíduos padronizados são estacionários, logo a premissa de não autocorrelação dos resíduos foi respeitada. Na fase final, foi realizado um estudo comparativo das previsões utilizando os seguintes métodos: SARIMA (Modelo AIC), SARIMA (Método BIC), Suavização de Holt Winters (HW Aditivo), Suavização de Holt Winters (HW Multiplicativo), Naive, ARIMA (Drift). Porém as medidas MAE (erro médio absoluto), RMSE (desvio médio quadrático) e MAPE (erro absoluto percentual médio) revelaram que todos os modelos de predição apresentaram valores insatisfatórios, pois todos os MAPEs de previsão estavam acima de 5%, com MAEs e RMSE muito elevadas. Ainda assim, considerando as limitações, concluiu-se que o melhor modelo de previsão foi o modelo BIC, pois possui menor MAPE, menor número de parâmetros, e menores MAE e RMSE. O resultado deste estudo, sugere que outras variáveis devem ser levadas em consideração no ajuste de um modelo de previsão para vendas de carros no Brasil, afim de se obter uma maior acuracidade.

PALAVRAS CHAVE:

ARIMA. SARIMA. Previsão de vendas. Setor automobilístico, Veículos.

ABSTRACT

The sales forecast becomes fundamental to obtain positive results in the most diverse types of enterprises. Thus, the present work aims to find a forecasting model with high predictive power for car sales in Brazil, using data extracted from the period 1995 to 2021. The study was carried out in three stages: Descriptive analysis, ARIMA and SARIMA forecasting models, and finally Model accuracy analysis. The database consists of 365 pieces of information regarding sales made by car dealers in Brazil. The average sales figure was 142,982 cars per month with a standard deviation of 66,282. Through the graph of the series it was possible to observe a certain seasonality and an increasing trend. The Dikey Fuller test confirmed that the series is nonstationary. In principle, the ACF and PACF charts were used to identify the parameters of the ARIMA model. In the next step, two models were generated, both obtained using the 'auto.arima' command. The first had as selection criteria, the one with the lowest AIC, and the second BIC, generating the following models respectively: 'ARIMA(0,1,4)(1,0,0)[12] and ARIMA(0,1,2)(1,0,0)[12]. Additionally, we verified in both models that the standardized residuals are stationary, so the assumption of non-autocorrelation of the residuals was respected. In the final phase, a comparative study of the predictions was performed using the following methods: SARIMA (AIC model), SARIMA (BIC model), Holt Winters Smoothing (Additive HW) Holt Winters Smoothing (Mutiplicative HW), Naive, ARIMA (Drift). However, the measurements MAE (mean absolute error), RMSE (root mean square deviation) and MAPE (mean absolute percentage error) revealed that all prediction models showed unsatisfactory values, as all prediction MAPEs were above 5%, with very high MAEs and RMSE. Still, considering the limitations, it is concluded that the best prediction model was the BIC Model, as it has the lowest MAPE, the lowest number of parameters, and the lowest MAE and RMSE. The result of this study suggests that other variables should be taken into consideration when fitting a forecast model for car sales in Brazil, in order to obtain a higher accuracy.

KEY WORDS:

ARIMA . SARIMA. Sales forecast. Automotive sector. Vehicle.

Sumário

1	Introdução	1
2	Objetivos	3
2.1	Objetivo Geral	3
2.2	Objetivos Específicos	3
3	Referencial Teórico	4
3.1	Setor automobilístico e vendas de carros no Brasil	4
3.2	Séries Temporais	5
3.2.1	Decomposição	5
3.2.2	Tendência	6
3.2.3	Sazonalidade	6
3.2.4	Ciclos	6
3.3	Alisamento Exponencial	6
3.3.1	Alisamento Exponencial Simples	6
3.3.2	Alisamento Exponencial de Holt	7
3.3.3	Alisamento Exponencial de Holt-Winters	7
3.4	Autocovariância e Funções de Autocorrelação	8
3.4.1	Autocovariância	9
3.4.2	Funções de Autocorrelação	9
3.5	Modelos ARIMA E SARIMA	11
3.6	Previsão	12
3.7	Medição de acuracidade do modelo	13
4	Metodologia	15
4.1	Base da dados	15
4.2	Software	15
4.3	Métodos	15
4.3.1	Análise descritiva da série	15
4.3.2	Modelos de previsão ARIMA e SARIMA	15
4.3.3	Previsão	16
4.3.4	Análise da acuidade do modelo	16
5	Resultados e Discussão	17
5.1	Análises Descritiva	17
5.2	Métodos ARIMA e SARIMA	19
5.3	Previsão	24
5.4	Análise da acuidade do modelo	32

Lista de Figuras

1	Comportamento da série temporal em "autoplot"	18
2	Decomposição da série em "autoplot"	19
3	Estimativa de Autocorrelção	20
4	Estimativa de Autocorrelção Parcial	20
5	Método ARIMA usando AIC	22
6	Método ARIMA usando BIC	22
7	Ajuste do modelo AIC aos dados reais	23
8	Ajuste do modelo BIC aos dados reais	24
9	Gráfico da previsão em 3 passos através do modelo SARIMA (0,1,4)x(1,0,0) ₁₂ para a venda de carros, Brasil, Jan/1995 a Mai/2021.	26
10	Gráfico da previsão em 3 passos através do modelo SARIMA (0,1,2)x(1,0,0) ₁₂ para a venda de carros, Brasil, Jan/1995 a Mai/2021.	27
11	Gráfico da previsão em 3 passos através do modelo HW aditivo para a venda de carros, Brasil, Jan/1995 a Mai/2021.	28
12	Gráfico da previsão em 3 passos através do modelo HW multiplicativo para a venda de carros, Brasil, Jan/1995 a Mai/2021.	30
13	Gráfico da previsão em 3 passos através do modelo Naive para a venda de carros, Brasil, Jan/1995 a Mai/2021.	31
14	Gráfico da previsão em 3 passos através do modelo ARIMA (drive) para a venda de carros, Brasil, Jan/1995 a Mai/2021.	32

Lista de Tabelas

1	Modelos ARIMA(p,d,q) identificados e estimados para a série temporal	21
2	coeficientes estimados pelo modelo SARIMA (0,1,4)x(1,0,0) ₁₂ para a venda de carros, Brasil, Jan/1995 a Mai/2021.	25
3	Tabela da previsão em 3 passos através do modelo SARIMA (0,1,4)x(1,0,0) ₁₂ para a venda de carros, Brasil, Jan/1995 a Mai/2021.	25
4	Coeficientes estimados pelo modelo SARIMA (0,1,2)x(1,0,0) ₁₂ para a venda de carros, Brasil, Jan/1995 a Mai/2021.	26
5	Tabela da previsão em 3 passos através do modelo SARIMA (0,1,2)x(1,0,0) ₁₂ para a venda de carros, Brasil, Jan/1995 a Mai/2021.	27
6	Tabela da previsão em 3 passos através do modelo HW aditivo para a venda de carros, Brasil, Jan/1995 a Mai/2021.	28
7	Tabela da previsão em 3 passos através do modelo HW multiplicativo para a venda de carros, Brasil, Jan/1995 a Mai/2021.	29
8	Tabela da previsão em 3 passos através do modelo Naive para a venda de carros, Brasil, Jan/1995 a Mai/2021.	31
9	Tabela da previsão em 3 passos através do modelo ARIMA (drive) para a venda de carros, Brasil, Jan/1995 a Mai/2021.	32
10	Medidas de acurácia (MAE, RMSE e MAPE) dos modelos para a venda de carros, Brasil, Jan/1995 a Mai/2021.	33

1 Introdução

Modelos formais de previsão auxiliam grandes corporações a mesurar a quantidade de bens e serviços que serão prestados, afim de suprir as necessidades do mercado. Com recursos de previsão, pode-se realizar projeções de valores futuros baseados em valores anteriores. Utilizando as estimativas, as organizações podem estar melhor preparadas para mudanças na situação econômica ou de concorrência, analisando dados históricos da série de tempo para prever tendências de desempenho. Este método determina quais recursos necessários para produção e vendas. Nas finanças vem sendo parte importante do processo estratégico da gestão empresarial (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998; TUBINO, 2007).

Desta forma, a previsão de vendas torna-se fundamental para obtenção de resultados positivos nos mais diversos tipos de empreendimentos. Tomando como exemplo, a previsão de vendas, nota-se que, quanto maior a acurácia, ou seja, quanto maior a acertividade da previsão, menor será a falta de produtos, beneficiando assim os clientes, que não se decepcionarão com a falta de produtos; assim como para o fabricante e para seus fornecedores, que não perderão vendas por motivo de rupturas e não enfrentarão excesso de estoques por falta de demanda. O uso desta técnica estatística traz benefícios para clientes, fornecedores e fabricantes (WALTER, 2013).

Neste contexto, temos a indústria automobilística, que por sua vez é responsável, de modo indireto por 22% do PIB industrial brasileiro, tendo um papel importantíssimo na sustentação da economia. Sendo assim, é de fundamental importância estar atento em como o mercado enxerga esta esfera, e ainda quais as tendências deste setor para os anos seguintes, tendo em vista que uma crise nas vendas, pode impactar profundamente a economia Brasileira e mundial (WOLK, 2019). Além disso, segundo números do IBGE (2017), esse mercado cresceu 17,2% em 2017, ajudando o país a ter uma alta de 2,5% na produção industrial, no ano seguinte, representou 12,7%. Isso mostra o quão essenciais são, medidas que façam esse setor continuar crescendo.

Adicionalmente, prever o crescimento do número de veículos numa cidade, auxilia estrategicamente políticas públicas de expansão das vias urbanas e o tráfego viário (ECOLA, 2014). De acordo com Ipea, 2011, observa-se que o aumento das frotas de automóveis e de motocicletas se constituiu em uma tendência bastante forte no Brasil. Esse aumento decorreu em função da elevação do poder aquisitivo da população, das deficiências do transporte público e também devido ao apoio do governo federal com medidas de isenções de imposto e facilitação do crédito. Por outro lado, em 2020 devido a pandemia do Novo Corona vírus a venda de veículos no Brasil encerra com queda de 26,2% com relação a 2019 segundo os balanços divulgados pela Associação Nacional dos Fabricantes de Veículos Automotores (Anfavea), 2021. Assim, entender as circunstâncias sociais e políticas que permeiam a série histórica de vendas automotivas, são para embasar os estudos de planejamentos de vendas ou tráfego.

Portanto, análise de séries temporais é um importante instrumento no entendimento do mercado e na formulação de planos de ação e estratégias. O histórico de uma variável pode ser

utilizado na identificação de períodos de crescimento/decrescimento, sazonalidade e ainda para “prever” observações futuras. Estas sequências de observações são obtidas em intervalos regulares de tempo durante um período específico. O conjunto dessas ordenadas no tempo pode ser discreto ou contínuo. Pode-se obter uma série temporal discreta a partir de uma amostra de pontos de uma série contínua ou por meio de um parâmetro como, por exemplo, a média de períodos fixos de tempo (LATORRE CARDOSO, 2001).

Sendo assim, o presente estudo tem por objetivo utilizar métodos de predição, na série histórica: Vendas de veículos no Brasil por concessionárias de automóveis no período de 1995 à 2021, extraído do site Ipeadata. Tais dados estão agrupados no setor macroeconômico, que são dados econômicos e financeiros do Brasil em análises mensais, na mesma unidade monetária. Além disso, será utilizado os algoritmos de alisamento exponencial, nessa cenário o estudo baseia-se principalmente no de Holt-Winters, os modelos ARIMA e SARIMA e combinação de previsões. Na fase de alisamento exponencial faremos previsões de 3 passos a frente. O software estatístico utilizado é o programa R.

2 Objetivos

2.1 Objetivo Geral

Encontrar um modelo de previsão com alto poder preditivo para vendas de automóveis no Brasil utilizando dados extraídos no período de 1995 à 2021.

2.2 Objetivos Específicos

- Realizar análise descritiva da série;
- Identificar o comportamento das vendas de automóveis;
- Utilizar o método ARIMA para geração de modelos de previsão;
- Fazer previsões utilizando diferentes metodologias;
- Comparar os modelos de previsão gerados.

3 Referencial Teórico

3.1 Setor automobilístico e vendas de carros no Brasil

O mercado automobilístico é de extrema importância para a economia do Brasil. Não só por ser um setor forte e atuante, sendo responsável indiretamente por 22% do PIB industrial, mas também, porque este necessita de outras indústrias para a sua concretização (WOLK, 2019).

O surgimento da indústria de automóveis em série, por Henry Ford, no início do século XX, inaugurou uma nova era na história industrial ao introduzir a linha de montagem contínua. O mercado automotivo sempre esteve em uma posição de destaque no setor industrial de todos os países, inclusive no Brasil. Em 1950 o “Plano de Metas” do presidente Juscelino Kubitschek, incentivou a entrada de capital estrangeiro no país, principalmente para a indústria automotiva, pelo fato deste apresentar um efeito multiplicador de investimentos muito elevado. Notamos tal efeito, ao perceber que produção de um carro, estimula a criação de outras empresas, como fabricantes de pneus, amortecedores, vidros e faróis, publicidade, vendas etc (MELO, 2008).

Segundo a Federação Nacional da Distribuição de Veículos Automotores (FENABRAVE), no ano de 2020, o Brasil possui a sexta maior frota de veículos do mundo contabilizando aproximadamente 46,3 milhões automóveis em circulação que corresponde a quase 10 veículos por km^2 .

Historicamente notamos variação significativa na venda de veículos a partir do ano de 2010, onde notou-se um crescimento significativo no número de vendas de veículos, alcançando no quarto trimestre um total de 827.380 carros vendidos. Em 2012, com a economia ainda aquecida, a indústria automobilística nacional fechou com um crescimento de 4,6% sobre 2011 e o maior valor de vendas/mês encontrado de todos os anos ($n = 352.048$). Ainda neste ritmo, a economia brasileira cresceu 2,3% em 2013, acima da alta de 1% no ano anterior. Nos últimos três meses de 2013, o Produto Interno Bruto (PIB) cresceu 0,7%. Continuou crescendo o número de vendas de carros durante o ano, os meses mantiveram vendas altíssimas em torno de 256.606. A economia brasileira cresceu 0,1% em 2014. Porém os impostos já vinham há um tempo crescendo mais do que o valor adicionado, relacionado à produção (FENABRAVE, 2020).

Por outro lado, a economia já havia tido um decaimento no ano de 2014 no número de vendas de veículos. E no ano de 2015 foi ainda mais expressivo. Vendas que circulavam em torno de 256.606 passa a movimentar o mercado automobilístico nas vendas em 175.255. A economia brasileira encolheu 3,8% em 2015 na comparação com 2014. Nos anos seguintes estas vendas mantiveram-se constantes, até uma queda considerável em março de 2020, no começo da pandemia do COVID-19 no Brasil, que culminou no fechamento de concessionárias nos primeiros meses, fechando o ano com 26,6% a menos em vendas totais de automóveis e comerciais leves na comparação com 2019 (FENABRAVE, 2020).

O setor automobilístico brasileiro é dinâmico e promissor, porém incerto. Fazer projeções

em um cenário de economia volátil, geralmente, tende ao erro. Contudo os dados históricos no setor automobilístico nos fornecem muitas informações, ricas para análises, pois é possível identificar as reações do setor frente ao humor da economia, em situações diversas e projetar o que poderá ocorrer (MERLUZZI, 2021).

De acordo com MAKRIDAKIS ET AL. (1998), realizar previsões para os setores produtivos da economia é importante para auxiliar, tanto governo como as empresas que compõem esse setor, na determinação dos recursos necessários.

3.2 Séries Temporais

De acordo com DOWNING e CLARK (2006, p. 299), "As séries temporais (ou históricas) são conjuntos de medidas de uma mesma grandeza, relativas a vários períodos consecutivos". Ou seja, a série temporal consiste em uma sequência de valores de uma determinada variável observada em intervalos uniformes de tempo. A variável de controle é o tempo e as séries temporais são postas cronologicamente (SILVA; SILVA, 1999).

Segundo SOUZA, SAMOBYL e MIRANDA (2008) e HYNDAMNN e ATHANASOPOULUS (2013) as séries temporais têm três padrões elementares: tendência, sazonalidade e ciclo. A tendência se verifica quando os dados crescem ou diminuem ao longo do tempo. Por outro lado, series constantes ao longo do tempo não apresentam tendência. Além disso, a sazonalidade resulta da influência de fatores sazonais, como o semestre ou dia do ano, ou seja, o período é sempre conhecido (DOWNING; CLARK, 2006). Diferentemente da sazonalidade, as variações cíclicas acontecem em períodos não conhecidos. Segundo Silva e Silva (1999) as variações sazonais "são movimentos cíclicos que se completam em um ano" enquanto as variações cíclicas "são movimentos cíclicos que se completam em período superior a um ano".

A seguir, seguem alguns conceitos importantes retirados do livro Previsões: Métodos e aplicações de MAKRIDAKIS WHEELWRIGHT, 1982.

3.2.1 Decomposição

Modelos de Decomposição consistem em decompor o modelo que descreve o comportamento da série temporal através de suas componentes:

O valor da série temporal será igual a: padrão temporal + erro, logo o valor será:

$$Y_t = f(T_t, C_t, S_t, \varepsilon_t)$$

em que:

- Y_t : é o valor da série temporal;
- T_t : é a componente Tendência no período t;
- C_t : é a componente Ciclo no período t;
- S_t : é a componente de Sazonalidade no período t;

- ε_t : é a componente de erro ou Ruído Aleatório no período t.

3.2.2 Tendência

A tendência de longo prazo de uma série para aumentar ou diminuir (tendência para cima ou tendência para baixo).

3.2.3 Sazonalidade

Flutuação periódica na série temporal durante um determinado período. Essas flutuações formam um padrão que tende a se repetir de um período sazonal para o próximo.

3.2.4 Ciclos

Longas distâncias da tendência devido a fatores diferentes que da sazonalidade. Os ciclos, em geral, ocorrem durante um longo intervalo de tempo, e os intervalos de tempo entre picos sucessivos ou passagens de um ciclo não são necessariamente os mesmos.

3.3 Alisamento Exponencial

Métodos de previsão têm por base observações passadas que abrangem informações sobre o padrão de comportamento da série temporal. O principal ponto é diferenciar este padrão de ruídos que possam existir na série. As técnicas de suavização exponencial são as mais utilizadas frente a todas as outras técnicas de previsão de demanda devido a sua simplicidade, facilidade de ajustes e boa acurácia. Tais métodos valorizam mais os últimos valores observados na série temporal através da ponderação exponencial dos mesmos, de acordo com a proximidade ao período da previsão (PELLEGRINI, F.R.; FOLIATTO, 2001)

3.3.1 Alisamento Exponencial Simples

Este método é apropriado para séries que não retratam: tendência e sazonalidade. O nível atual da série Z_t é estimado por meio de uma média ponderada das observações anteriores, fazendo-se que os pesos decrescem exponencialmente à medida que retrogradamos no tempo. O nível da série é dado por:

$$\bar{Z} = \alpha \sum_{j=0}^{t-1} (1 - \alpha)^j Z_{t-j} + (1 - \alpha)^t Z_0, \quad t = 1, \dots, N \quad (1)$$

É possível notar ainda, que o valor de α não depende da escala em que as observações foram medidas, mas sim das propriedades da série temporal. O valor de $\alpha \in (0,1)$ deve ser definido de modo a refletir a influência das observações anteriores nas previsões. Valores pequenos ajustam

previsões que dependem de muitas observações passadas. Por outro lado, valores próximos de 1 movimentam as previsões que dependem das observações mais recentes e no caso extremo $\alpha = 1$ a previsão é a última observação. O objetivo é escolher o valor de α que minimiza a soma dos quadrados dos erros de previsão um passo à frente.

Assim, esse método consiste em um modelo de alisamento exponencial que representa uma média das previsões, corrigida pelos erros constatados em previsões passadas.

3.3.2 Alisamento Exponencial de Holt

Esta técnica é similar, em princípio, ao Simplex, o que a distingue é que ela não suaviza só o nível, ele utiliza uma nova constante de alisamento para "modelar" a tendência da série. O procedimento é análogo ao de determinação da constante de suavização de um método simples, escolhe-se o valor do vetor (α, β) que torna a soma de EQMP mínimo tal que isto ocorra (Morettin e Tolo, 1985).

Este algoritmo possibilita obter estimativas do nível e da tendência da série. A forma de recorrência é dada por:

$$\bar{Z}_t = \alpha Z_t + (1 - \alpha)(\bar{Z}_{t-1} + \hat{T}_{t-1}), \quad t = 2, \dots, N. \quad (2)$$

$$\hat{T}_t = \beta(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - \beta)\hat{T}_{t-1}, \quad t = 2, \dots, N. \quad (3)$$

Logo, o método de alisamento exponencial com deterioração na tendência que é mais uma extensão do modelo de alisamento exponencial, pois verifica o efeito das variações ou deterioração nas tendências.

3.3.3 Alisamento Exponencial de Holt-Winters

O procedimento pode ser generalizado para séries que contenham tendência e variação sazonal, α , β e γ , para cada componente da série que são em geral escolhidos no intervalo (0,1) e podem ser estimados minimizando-se a soma de quadrados dos erros de previsão. Estes parâmetros dependem das propriedades temporais do nível, tendência e sazonalidade da série e não necessariamente da escala das observações. Equivalem-se os comentários sobre estes valores refletindo a influência das observações passadas nas previsões de cada componente.

O propósito central deste algoritmo é permitir a incorporação de padrões sazonais ao algoritmo de Holt. Há dois tipos de metodologias, onde utilização depende das características da série considerada. Tais procedimentos baseiam-se em três equações que usam constantes de alisamento diferentes, cada uma correspondendo a uma das componentes do padrão da série: nível, tendência e sazonalidade.

- Método Sazonal Multiplicativo

Considere uma série sazonal com período p . A variante mais usual do método de HW considera o fator sazonal S_t como sendo multiplicativo, enquanto a tendência continua aditiva, ou seja:

$$Z_t = m_t S_t + T_t + a_t, t = 1, \dots, N. \quad (4)$$

E assim, calculamos as três equações de suavização por:

$$\hat{S}_t = \gamma \left(\frac{Z_t}{\bar{Z}_t} \right) + (1 - \gamma) \hat{S}_{t-p}, \quad 0 < D < 1. \quad (5)$$

$$\bar{Z}_t = \alpha \left(\frac{Z_t}{\hat{S}_{t-s}} \right) + (1 - \alpha) (\bar{Z}_{t-1} + \hat{T}_{t-1}), \quad 0 < \alpha < 1. \quad (6)$$

$$\hat{T}_t = \beta (\bar{Z}_t - \bar{Z}_{t-1}) + (1 - \beta) \hat{T}_{t-1}, \quad 0 < C < 1. \quad (7)$$

- **Método Sazonal Aditivo**

Neste caso, a série é composta por:

$$Z_t = m_t + T_t + S_t, \quad (8)$$

Em que, podemos calcular as equações de suavização por:

$$\hat{S}_t = \gamma (Z_t - \bar{Z}_t) + (1 - \gamma) \hat{S}_{t-s}, \quad 0 < \gamma < 1. \quad (9)$$

$$\bar{Z}_t = \alpha (Z_t - \hat{S}_{t-s}) + (1 - \alpha) (\bar{Z}_{t-1} + \hat{T}_{t-1}), \quad 0 < \alpha < 1. \quad (10)$$

$$\hat{T}_t = \beta (\bar{Z}_t - \bar{Z}_{t-1}) + (1 - \beta) \hat{T}_{t-1}, \quad 0 < \beta < 1. \quad (11)$$

3.4 Autocovariância e Funções de Autocorrelação

Um aspecto fundamental em séries temporais é o modo como observações estão relacionadas entre si no tempo. Para tanto, temos as Funções de Autocovariância e Autocorrelação. O material a seguir foi retirado do livro *Introduction to Econometrics* (MADALLA, G. S.; LAHIRI, 2009).

3.4.1 Autocovariância

Esta simboliza a covariância da variável aleatória com ela mesma defasada de k passos, onde as séries temporais são tomadas do mesmo processo estocástico. Por outra perspectiva, a covariância é definida em termos de valores defasados (lagged), denominada autocovariância, e mede o grau de variação de segunda ordem (segundo momento) entre dois elementos em dois tempos diferentes.

A covariância entre Y_t e Y_{t-1} é dada por:

$$\begin{aligned} cov[Y_t, Y_{t-1}] &= E[(Y_t - \mu)(Y_{t-1} - \mu)] \\ &= E[(\varepsilon_t + \alpha * \varepsilon_{t-1})(\varepsilon_{t-1} + \alpha * \varepsilon_{t-2})] \\ &= E[(\alpha * \varepsilon_{t-1}^2)] \\ &= \alpha * \sigma^2 \end{aligned}$$

Finalmente a covariância entre Y_t e Y_{t-2} é dada por:

$$\begin{aligned} cov[Y_t, Y_{t-2}] &= E[(Y_t - \mu)(Y_{t-2} - \mu)] = \\ &= E[(\varepsilon_t + \alpha * \varepsilon_{t-2})(\varepsilon_{t-2} + \alpha * \varepsilon_{t-3})] \\ &= 0. \end{aligned}$$

3.4.2 Funções de Autocorrelação

Autocorrelação é definida como a k - enésima ordem da autocovariância ϵ_k como:

$$\epsilon_k = cov[Y_t, Y_{t-k}] = cov[Y_t, Y_{t+k}]. \quad (12)$$

A autovariância de um processo estocástico pode ser normalizada e apresenta como função de autocorrelação, ρ_k

$$\rho_k = \frac{\gamma_k}{\sqrt{var(Z_t)}\sqrt{var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}, \quad (13)$$

Em que:

$$\text{var}(Z_t) = \text{var}(Z_{t+k}) = \gamma_0.$$

Logo, temos que a função de autocorrelação mensura o grau de correlação de uma variável, em um determinado momento, consigo mesma, em um instante de tempo posterior. Esta técnica é empregada para identificarmos um modelo adequado para a série temporal.

A função de autocorrelação, ou comumente chamada de ACF do inglês, nos auxilia na representação do desenvolvimento de Y_t ao longo do tempo. Ou seja, ela nos mostra o quão forte o valor observado hoje está correlacionado com os valores observados no passado e como choques hoje afetam valores futuros da variável estocástica. Além de nos ajudar a descrever os dados, a ACF também nos ajuda a achar raízes unitárias, escolher modelos e fazer diagnósticos sobre regressões. A ACF de um processo AR(1) é dado por:

$$\rho_k = \frac{\text{cov}[Y_t, Y_{t-k}]}{V[Y_t]} = \frac{\theta^k * \frac{\sigma_\varepsilon^2}{1-\theta^2}}{\frac{\sigma_\varepsilon^2}{1-\theta^2}}. \quad (14)$$

Para o processo MA(1) a ACF é:

$$\rho_1 = \frac{\text{cov}[Y_t, Y_{t-k}]}{V[Y_t]} = \frac{\alpha * \sigma_\varepsilon^2}{(1 + \alpha) * \sigma_\varepsilon^2} = \frac{\alpha}{1 + \alpha^2}, \quad (15)$$

com, $\rho_k = 0, k > 1$.

- Função de Autocorrelação parcial

A função de autocorrelação parcial (ou comumente PACF, do inglês) nos dá correlação entre a variável no instante t e uma de suas defasagens, retirado os efeitos das outras defasagens.

$$Y_t = \alpha + \beta_1 * Y_{t-1} + \beta_2 * Y_{t-2} + \varepsilon_t. \quad (16)$$

O parâmetro β_2 nos dá então o coeficiente de correlação entre Y_t e Y_{t-2} retirando os efeitos de Y_{t-1} . Assim, β_2 pode ser descrito como a seguir.

$$\beta_2 = r_{13,2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}}. \quad (17)$$

Assumindo estacionariedade fraca temos que:

$$r_{12} = \text{corr}(Y_t, Y_{t-1}) = \text{corr}(Y_{t-1}, Y_{t-2}) = \rho_1 r_{13} = \text{corr}(Y_t, Y_{t-2}) = \rho_2. \quad (18)$$

Logo, concluímos que

$$\beta_2 = r_{13,2} = \frac{\rho_2 - \rho_1^2}{1 - \rho^2}.$$

3.5 Modelos ARIMA E SARIMA

O modelo Autorregressivo Integrado de Media Movel – Autoregressive Integrated Moving Average (ARIMA) é um método popular entre os modelos estatísticos de análise de séries temporais (LEE; KO, 2011). Esse modelo foi apresentado na década de 1970 por Box e Jenkins e tem início nos modelos autorregressivo (AR), medias moveis (MA) e da combinação dos modelos AR e MA (ARMA), estendendo-se a modelos não estacionários (ARIMA) e sazonais (SARIMA). Cada um destes modelos podem modelar uma série individual ou combinadamente.

No modelo AR(p), o valor atual da série é expresso como um agregado linear de p valores anteriores e um ruído aleatório. Segundo Box e Jenkins (1976), Morettin e Tolo (2004), Ehlers (2009) e Lee e Ko (2011), um modelo AR(p) é dado pela equação:

$$Z_t = f_1 Z_{t-1} + f_2 Z_{t-2} + \dots + f_p Z_{t-p} + a_t. \quad (19)$$

em que p é a ordem de AR, Z_t é a série temporal observada no momento t , a_t é o ruído aleatório em uma distribuição normal com media 0 e variância igual a 1 e f_1, \dots, f_p são os parâmetros de AR(p).

Ao colocar o operador de defasagem B, que define $Z_{t-1} = BZ_t$; e, por conseguinte, $Z_{t-p} = B_p Z_t$, a equação (19) pode ser reescrita por:

$$f_p(B)Z_t = a_t, \quad (20)$$

em que $f_p(B)$ e o operador de AR(p), determinado por:

$$f_p(B) = 1 - f_1 B - f_2 B^2 - \dots - f_p B^p.$$

O modelo MA (q), que verifica a estrutura de autocorrelação dos resíduos de previsão do período atual com aqueles ocorridos em períodos anteriores e descrito pela equação (BOX; JENKINGS, 1976; MORETTIN; TOLOI, 2004; EHLERS, 2009; LEE; KO, 2011):

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}, \quad (21)$$

e pode ser reescrito pela $Z_t = \theta_q(B)a_t$ em que q e a ordem de MA(q), $\theta_1, \dots, \theta_q$ são os parâmetros do modelo MA(q), e $\theta_q(B)$ o qual é definido por $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$

Os modelos AR, MA e ARMA são usados quando a série é estacionária, em outras palavras, suas propriedades estatísticas básicas, como: media, variância e covariância, mantém-se constantes ao longo do tempo (HANKE; WICHERN; REITSCH, 2001; GUJARATI, 2006). Por outro lado, quando a série é não estacionária, por meio do processo de diferenciação de dados, ela é transformada em uma serie estacionária. Logo, além dos modelos anteriormente descritos, e utilizada a componente de integração I(d), resultando no modelo ARIMA(p,d,q) representado por:

$$f_p(B)\Delta_d Z_t = \theta_q(B)a_t. \quad (22)$$

em que $\Delta_d Z_t = (1 - B)_d Z_t$ e d e a ordem de diferenciação.

Para lidar com as series que apresentam autocorrelacao sazonal, Box e Jenkins (1976) generalizaram o modelo ARIMA e estabeleceram o modelo ARIMA sazonal multiplicativo, conhecido como SARIMA(p,d,q)(P,D,Q) e retratado por:

$$f_p(B)F_p(B_s)\Delta_d \Delta_s. \quad (23)$$

3.6 Previsão

Previsão, em estatística é o processo de estimativas em situações de incertezas. O termo previsão é semelhante, porém mais geral, e normalmente se refere a estimação de séries temporais ou dados instantâneos. Para a tomada de decisão é comum resolver problemas onde as decisões a serem tomadas serão funções de dados futuros, logo os dados serão representativos perante o que ocorre no futuro (ARMSTRONG, 1983).

Com base nos modelos de decomposição, para realizar previsões de valores futuros, basta recompô-los conforme sua forma, aditivo ou multiplicativo. Segundo Morettin e Tolo (1981), o modelo aditivo é mais adequado quando a componente de Sazonalidade não depende das outras componentes. Se a sazonalidade varia com a tendência, o modelo multiplicativo é mais adequado. Por exemplo, para construir uma previsão de vendas para o futuro idealizaremos que o padrão sazonal se manterá no período previsto. Assim, o próximo passo é repetir os índices sazonais para os meses futuros.

Espera-se, também, que a tendência se manifeste no período previsto da mesma forma que se manifestou no passado. Desta forma, o passo seguinte consiste na extrapolação da tendência futura. Podemos observar que os valores da tendência são resultado da composição de valores passados (ARMSTRONG, 1983).

3.7 Medição de acuracidade do modelo

No modelo de série temporal, precisamos verificar se o mesmo representa, ou não, adequadamente os dados. Para tanto é recomendável o uso de medidas de acuracidade, que são estatísticas que permitem avaliar o ajuste de uma previsão aos dados originais, por meio do cálculo de médias das diferenças (erros) entre os dados originais e as previsões em cada período da série temporal (LEVINE et al, 2000).

A análise residual retorna a adequação do modelo proposto as condições estabelecidas. Caso o modelo proposto represente adequadamente os dados, então, os resíduos devem possuir as características de um ruído branco. Um ruído branco é um tipo especial de série estacionária, que possui média zero, variância constante e não é serialmente correlacionado (GUJARATI; PORTER, 2011). Caso contrário os resíduos não são um ruído branco, entende-se que há informação que deve ser incorporada ao modelo proposto. Segundo Ehlers (2009), deve-se incluir no modelo informações caso haja autocorrelação nas defasagens 1 ou 2 ou, então, nos períodos sazonais. Da mesma forma, ocorre para a autocorrelação parcial. Caso a defasagem ocorra em outros pontos, não há evidências suficientes para rejeitar o modelo.

Efetuada o diagnóstico e certificada a adequação dos resíduos, realiza-se um estudo dos erros de previsão, para definição do modelo que melhor representa a série temporal. A diferença entre uma observação e seu valor ajustado é chamado de resíduo (MONTGOMERY; RUNGER, 2003).

Para análise da normalidade de resíduos pode-se utilizar, entre outros, o testes Shapiro Wilk. Já para verificação da correlação residual, pode-se realizar os testes Box-Pierce e Ljung-Box. O modelo mais adequado é aquele que possui o menor erro. As previsões não são exatas e o erro associado é maior quanto maior for o horizonte de predição (MOREIRA, 2008). De acordo com Souza, Samohyl e Miranda(2008). Espera-se que os valores variem sempre entre zero e um para indicar modelos de previsão adequados, sendo melhores as previsões que se aproximam de zero. Erro superior a um indica que o modelo proposto possui erro superior ao passeio aleatório (EHLERS, 2009).

Abaixo seguem erros são calculados usando os valores da recomposição.

Erro absoluto médio (EAM):

$$EAM = \frac{1}{n} * \sum_{t=1}^n |e_t| \quad (24)$$

Erro quadrático médio (EQM):

$$EQM = \frac{1}{n} * \sum_{t=1}^n e_t^2 \quad (25)$$

Erro percentual absoluto médio (EPAM):

$$EPAM = \frac{1}{n} * \sum_{t=1}^n \left| \left(\frac{e_t}{\hat{Y}_t} \right) * 100 \right|, \quad (26)$$

em que: $e_t = Y_t - \hat{Y}_t$

- e_t é o erro (diferença entre o valor da série, Y_t , e o valor de recomposição pelo modelo aditivo ou multiplicativo \hat{Y}_t em período genérico t). O melhor será o que apresentar os valores mais próximos de zero.

A análise do “erro” através ds medidas MAE (erro médio absoluto), RMSE (desvio médio quadrático) e MAPE (erro absoluto percentual médio) da previsão, foram a base para escolha do melhor modelo. Valores menores indicam um ajuste melhor.

4 Metodologia

4.1 Base da dados

As informações referentes a vendas de veículos pelas concessionárias de automóveis no Brasil (Janeiro do ano de 1995 a Fevereiro de 2021), foram obtidas no site do Ipeadata (Instituto de Pesquisa Econômica Aplicada.) que consiste em uma base de dados macroeconômicos, financeiros e regionais do Brasil, encontrada na página: <http://www.ipeadata.gov.br/Default.aspx>. Cada etapa do estudo está disponível em script (anexo 1), sendo totalmente reproduzível. O embasamento teórico dos métodos utilizados no processo foram discutidos no capítulo anterior “Referencial teórico”.

4.2 Software

O software estatístico utilizado foi o programa R. Este é uma versão gratuita do programa S-PLUS e é baseado na linguagem de programação S. Uma das vantagens dessa plataforma é capacidade de ceder e permitir a criação de novas funções e a possibilidade de modificações de muitas funções internas. Este programa pode ser obtido na página <http://www.r-project.org>.

4.3 Métodos

O estudo foi realizado em três etapas: Análise descritiva, Modelos de previsão ARIMA e SARIMA e por fim a Análise da acuidade do modelo.

4.3.1 Análise descritiva da série

Na primeira fase, efetuamos a análise descritiva (média, mínimo, máximo, quartis) e estudo visual do comportamento da série temporal através do gráfico da série. Ainda nesta etapa, foi realizada a decomposição da série afim de diagnosticar a presença ou não de tendência, ciclo e sazonalidade. Foi aplicado ainda o teste de DickeyFuller sob as hipóteses abaixo com o objetivo de verificar se a série é ou não estacionária.

4.3.2 Modelos de previsão ARIMA e SARIMA

Na segunda fase, o objetivo é gerar modelos de previsão utilizando o método ARIMA. Para isso, foram plotados as funções de autocorrelação e autocorrelação parcial em conjunto afim de identificar os parâmetros do modelo ARIMA. Ainda com este objetivo, aplicou-se função ndiffs com o propósito de confirmar a estacionariedade ou não da série. No caso de não estacionariedade este teste retorna ainda quantas diferenças são necessárias para tornar a série estacionária.

Na etapa seguinte, foi utilizado o comando `auto.arima` afim de se obter os modelos considerando aqueles que possuem o melhor AIC e BIC. Assim foram gerados dois SARIMAs nomeados como Modelo AIC e Modelo BIC.

Além disso, foi plotado o ajuste de cada modelo afim de avaliar visualmente suas respectivas adequabilidades a série real. De forma complementar, foram analisados os gráficos de resíduos padronizados, ACF dos resíduos e a estatística qui-quadrado de Box – Pierce (Ljung Box). Ainda neste sentido, foi aplicado ainda o teste de Box Ljung, que tem por objetivo verificar um dos pressupostos essenciais de um modelo de previsão: os resíduos devem independentes.

4.3.3 Previsão

Nesta etapa final, foi realizado um estudo comparativo das previsões utilizando os seguintes métodos:

- a) SARIMA (Modelo AIC)
- b) SARIMA (Modelo BIC)
- c) Suavização de Holt Winters (HW Aditivo)
- d) Suavização de Holt Winters (HW Multiplicativo)
- e) Naive
- f) ARIMA (Drift)

Todos os métodos foram aplicados no intuito de prever 3 passos a frente. Os gráficos com os valores da previsão foram plotados para cada modelo. Foram avaliados os MAPEs de ajuste e MAPEs de previsão, onde será considerado um modelo de previsão aqueles que possuem valores abaixo de 5%.

4.3.4 Análise da acuidade do modelo

A análise do “erro” através ds medidas MAE (erro médio absoluto), RMSE (desvio médio quadrático) e MAPE (erro absoluto percentual médio) da previsão, foram a base para escolha do melhor modelo. Valores menores indicam um ajuste melhor.

5 Resultados e Discussão

Na presente etapa apresenta-se o desenvolvimento da proposta deste trabalho, utilizando as metodologias descritas anteriormente como diretrizes.

5.1 Análises Descritiva

A base de dados consiste em 314 informações referentes as vendas realizadas pelas concessionárias de automóveis no Brasil. A série inicia no mês de janeiro do ano de 1995 e termina no mês de maio de 2021. A média no número das vendas foi de 156.231 automóveis por mês com desvio padrão de 60.201,03.

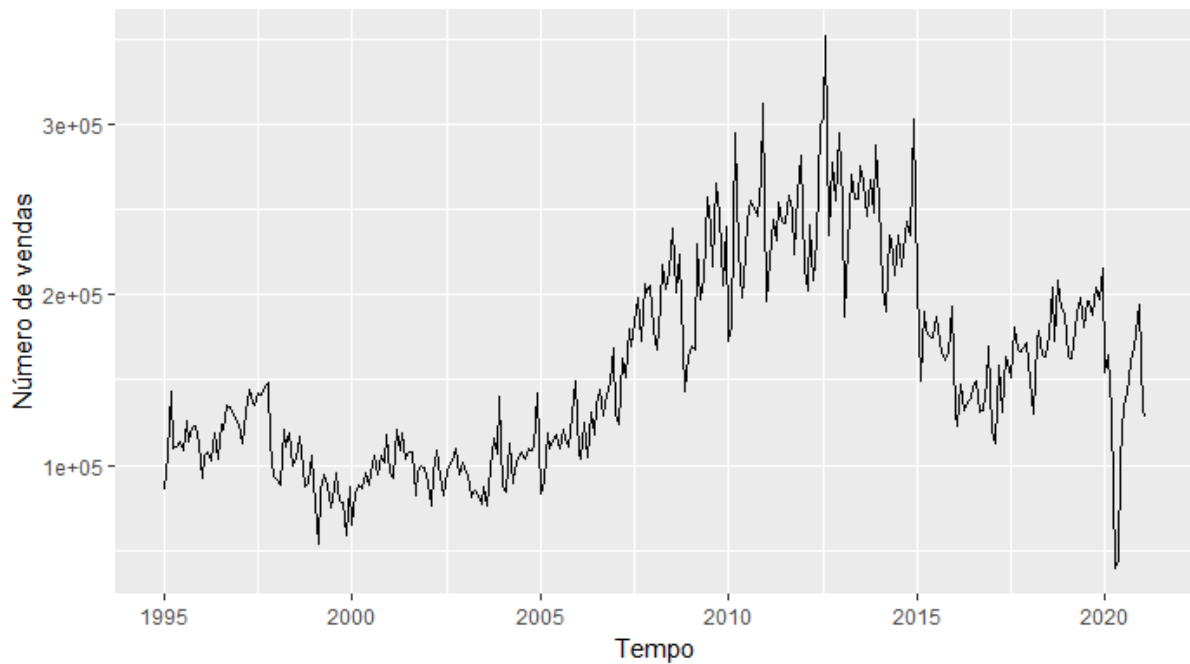
O menor valor de vendas na série foi de 39.493 no mês de abril no ano de 2020 e o maior foi de 352.048 no mês de agosto no ano de 2012. O contexto político / social pode explicar tais números. Em 2012, de acordo com IBGE (2012), o crescimento da população empregada e da massa real de salários, ao lado da expansão do crédito ao consumo, sustentaram o crescimento das vendas no comércio. Além disso, o desconto do Imposto sobre Produtos Industrializados (IPI), estimulou as vendas atingindo o recorde positivo histórico neste ano.

Por outro lado, em 2020, ocorreu a menor venda de veículos (34,493) desta sequência histórica, sendo este resultado atribuído a pandemia do Covid-19, que forçou no mês de março o início da quarentena, que consiste basicamente no isolamento social, afim de evitar a circulação do vírus. Adicionalmente, houve decretos de organizações governamentais para fechar grande parte das empresas, dessa forma parando a circulação econômica e assim, aumentando o índice de desemprego (Anfavea), 2020.

Retomando a análise, foi plotado o gráfico de série temporal, exibido na Figura 1, realizando-se assim uma análise visual dos dados, afim de identificar possíveis componentes e características específicas do comportamento das vendas ao decorrer do tempo.

Desta forma, foi possível observar variações regulares (picos e vales), que configuram uma certa sazonalidade. É possível perceber ainda, uma tendência crescente até o ano de 2015, onde apresenta um decréscimo significativo. Após este período, há um discreto crescimento, seguida da maior queda histórica da série em 2020. Mediante este contexto, a série apresenta uma leve tendência de crescimento, apesar das variações.

Figura 1: Comportamento da série temporal em "autoplot"

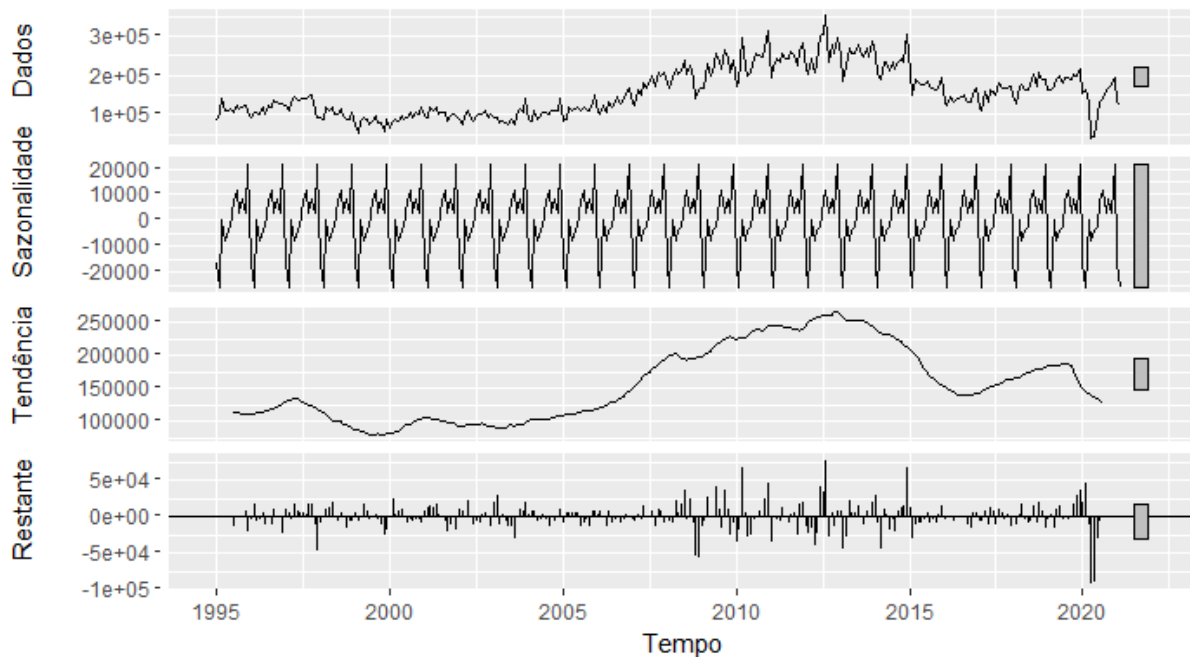


Fonte: Elaborado pelo autor (2021)

Em sequência, a série foi submetida ao processo de decomposição (Figura 2) onde podemos observar visualmente: base de dados, a tendência, a sazonalidade e aleatoriedade dos resíduos. De forma específica, podemos dizer que uma tendência existe quando há um aumento ou diminuição de longo prazo associados aos dados.

Nesta caso, há evidências que a série é levemente ascendente, corroborando com a primeira análise visual. Quanto a sazonalidade, notamos mais claramente, uma ascensão e queda nas observações que repetem-se regularmente durante o mesmo período de tempo chamado de padrão sazonal. Por outro lado, não foi possível perceber a existência de ciclos nestes.

Figura 2: Decomposição da série em "autoplot"



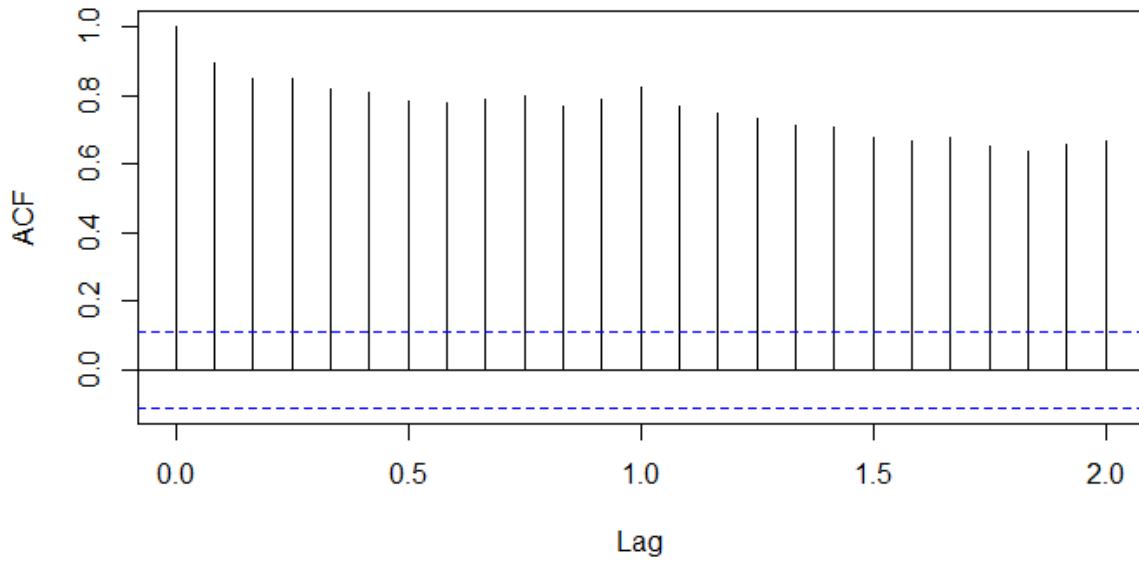
Fonte: Elaborado pelo autor (2021)

Afim de verificar a estacionariedade da série foi realizado os teste da raiz unitária de Dickey-Fuller (estatística do teste = -1,856, p -valor= 0,6371). O resultado revela que hipótese de nulidade H_0 não foi rejeitada a 5% de significância. Sendo assim, admite-se que a série é não-estacionária.

5.2 Métodos ARIMA e SARIMA

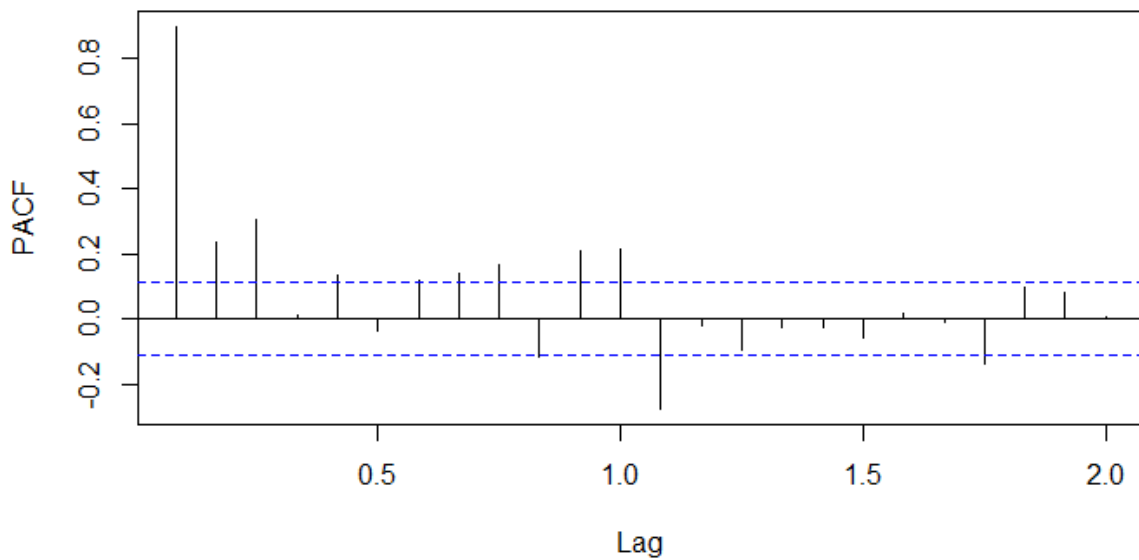
Através do método ARIMA faremos um análise de previsão para a série em estudo. Será gerado um modelo que estima melhor uma previsão para tal série. Em princípio, utilizou-se os gráficos ACF e PACF para identificação do modelo ARIMA, já que, por vezes, estes permitem determinar valores apropriados para p e q (parâmetros do modelo). Desta forma, ao analisar o gráfico da série temporal, autocorrelação total ACF (Figura 3), pode-se encontrar ondas senoides que podem indicar sazonalidade, além disso mostra que a série tem memória longa, ou seja, depende de valores antigos. A correlação parcial PACF (Figura 4) mostra que a série tem pelo menos 2 parâmetros autoregressivos.

Figura 3: Estimativa de Autocorrelção



Fonte: Elaborado pelo autor (2021)

Figura 4: Estimativa de Autocorrelção Parcial



Fonte: Elaborado pelo autor (2021)

Além disso, pela função `ndiffs` estima-se que é necessária uma diferença para tornar a série

estacionária. A hipótese de sazonalidade dos dados é reforçada pela necessidade de incluir uma diferenciação sazonal no ajuste do modelo.

Na próxima etapa, foram gerados dois modelos, ambos obtidos através da linha de comando: `auto.arima` (tabela 1).

O primeiro teve como critério de seleção, aquele que apresentasse o menor AIC, fornecendo os seguintes parâmetros $ARIMA(0,1,4)(1,0,0)[12]$. No segundo, foi considerado como critério de seleção, aquele que apresentasse o menor BIC, retornando os seguintes parâmetros $ARIMA(0,1,2)(1,0,0)[12]$. Foram gerados dois SARIMA, sem parâmetros autoregressivos ARIMA. Foi necessária apenas uma diferenciação (`sar1`) SARIMA para torná-la estacionária em ambos os modelos. E finalmente, as médias móveis ARIMA, foi o que, de fato, diferenciou um modelo do outro. Enquanto o primeiro (método AIC) possui quatro médias móveis ($ma1$, $ma2$, $ma3$, $ma4$) ARIMA, o segundo (método BIC) modelo possui apenas duas ($ma1$, $ma2$).

Tabela 1: Modelos $ARIMA(p,d,q)$ identificados e estimados para a série temporal

Modelo	AIC	BIC	MAPE	Box Ljung
$ARIMA(0,1,4)(1,0,0)[12]$	7145.86	7168.34	10.86%	p-valor = 0,995
$ARIMA(0,1,2)(1,0,0)[12]$	7148.46	7163.44	10.99%	p-valor = 0,997

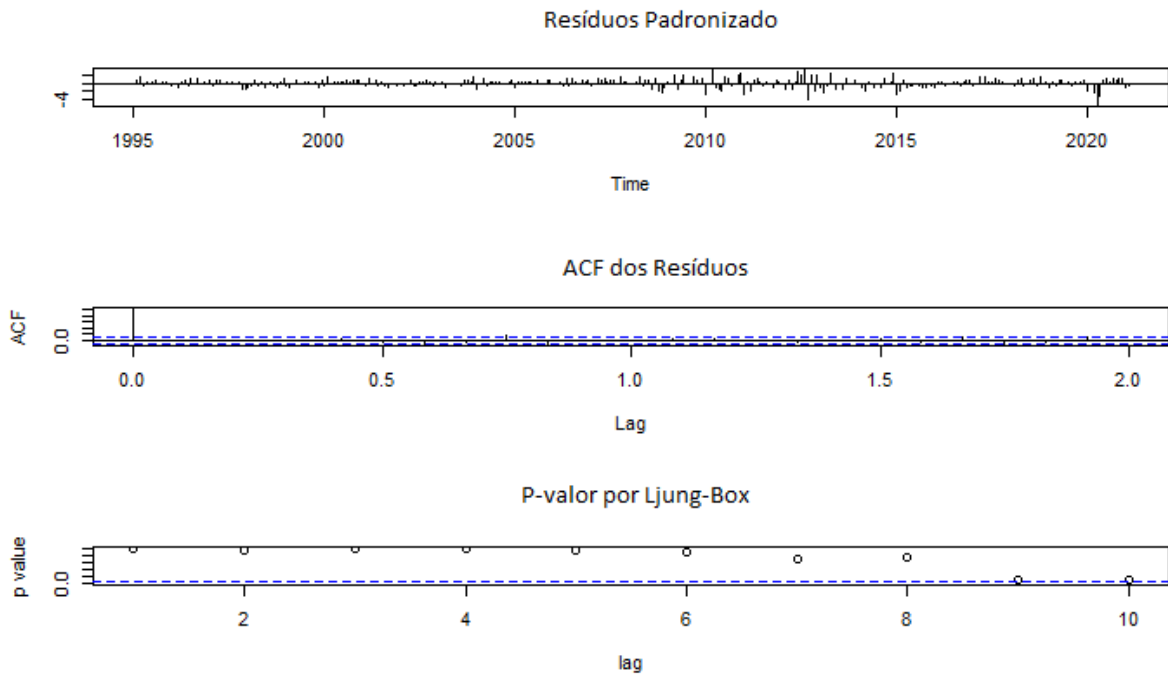
Fonte: Elaborado pelo autor (2021)

De forma complementar, listamos o erro absoluto percentual médio representado pelo MAPE (Tabela 1) afim de diagnosticar qual o melhor modelo. Este valor revela que, em média, a previsão do modelo AIC está incorreta em 10,86% e apesar da discreta melhora com relação ao modelo BIC (mape 10,99%) ainda é considerado um valor muito alto. Além disso, pelo critério da parcimônia, o acréscimo de tantos parâmetros sem melhora significativa do modelo não é indicado.

Ainda neste sentido, foi aplicado o teste de Box Ljung, que tem por objetivo verificar um dos pressupostos essenciais de um modelo de previsão: os resíduos devem independentes. Assim, o resultado dos testes para ambos os modelos (Tabela 1), com p-valor $\geq 5\%$, não rejeita a hipótese nula, logo os erros não são autocorrelacionados.

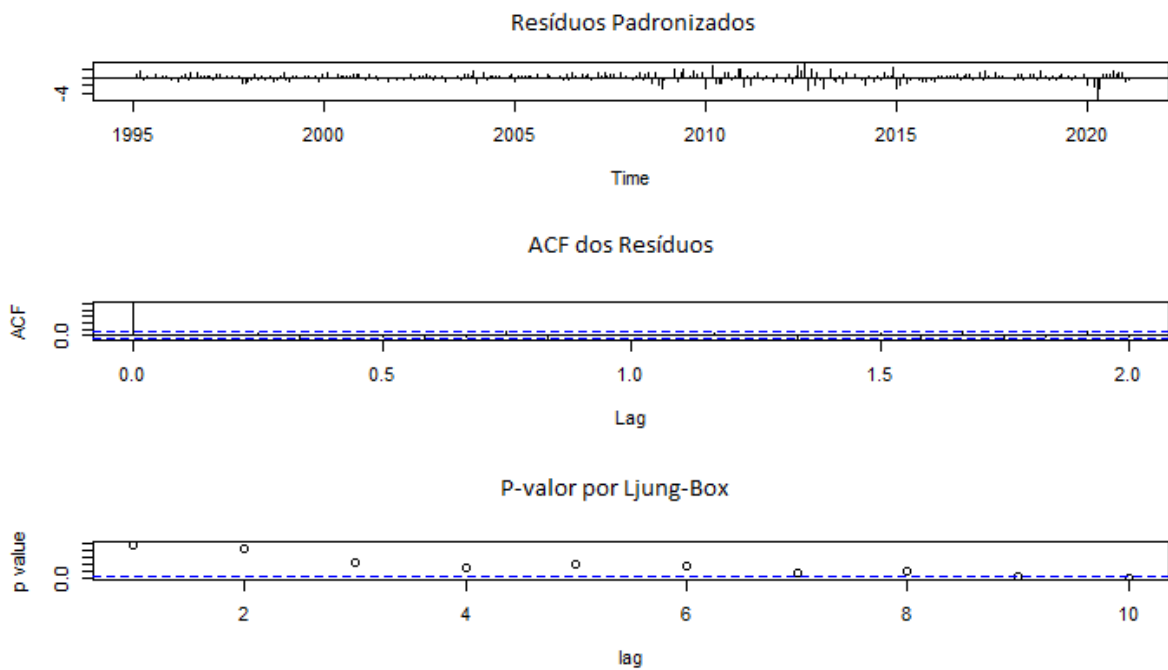
Adicionalmente, verificamos em ambos os modelos que os resíduos padronizados são estacionários pois, os erros estão normalmente distribuídos com média variando próximo a 0. Nos gráficos da autocorrelação amostral todos os lags estão dentro do intervalo de confiança de 95%, portanto os dados são não-autocorrelacionados. Pelo teste Ljung-Box para p-valor observamos que todos os pontos estão acima da linha tracejada, portanto não rejeita H_0 , ou seja, não existe dependência dos resíduos.

Figura 5: Método ARIMA usando AIC



Fonte: Elaborado pelo autor (2021)

Figura 6: Método ARIMA usando BIC

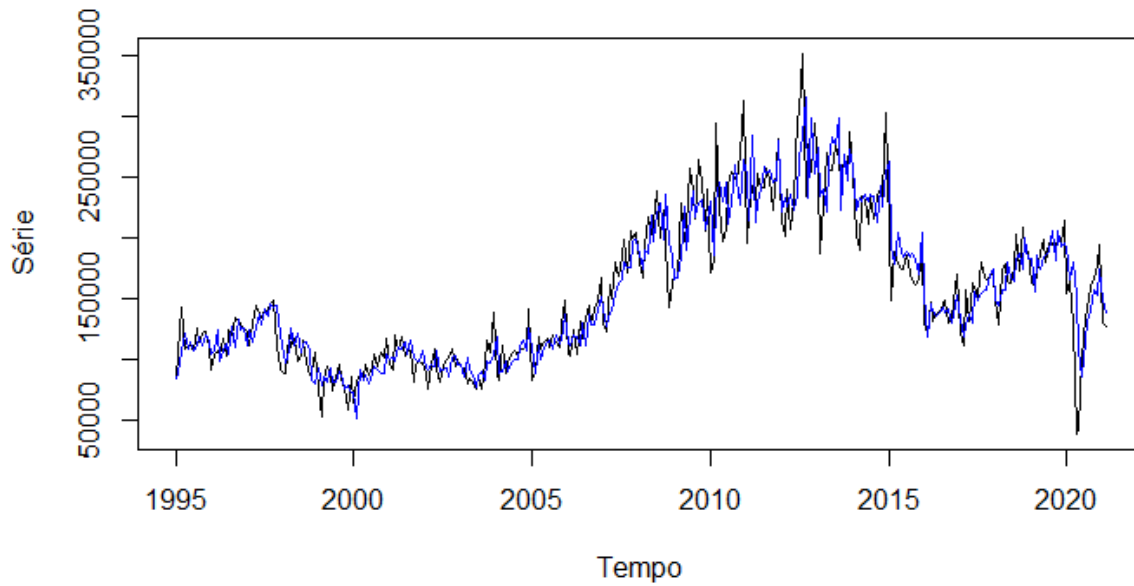


Fonte: Elaborado pelo autor (2021)

Através da Figura 5 e 6 é possível observar que as séries previstas se assemelham bastante

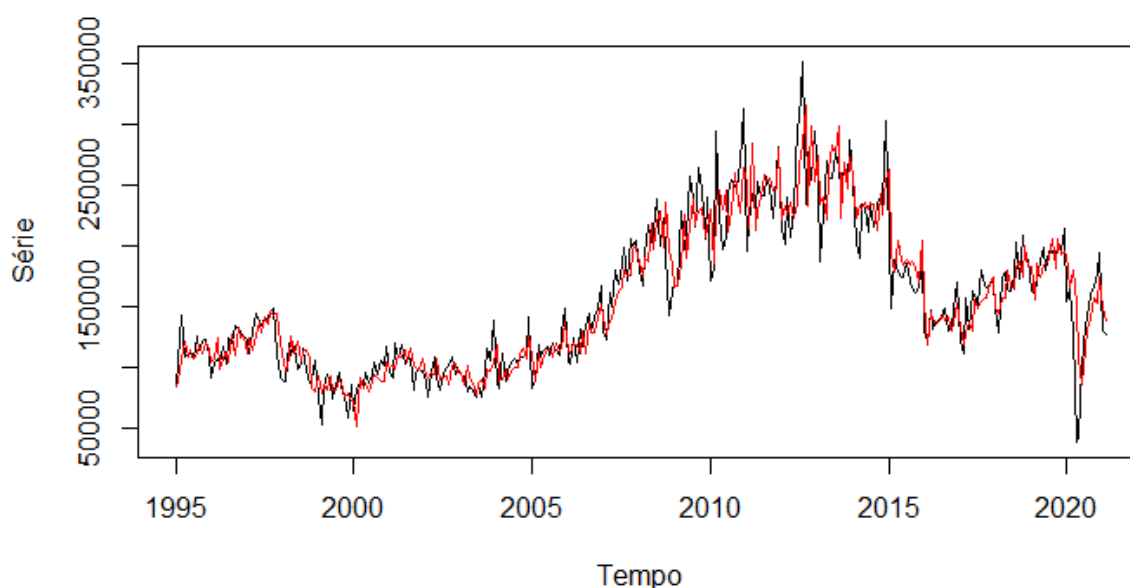
com a série original, mesmo apresentando MAPEs inadequados. Desta forma, análises visuais, podem não ser suficientes para tomada de decisão.

Figura 7: Ajuste do modelo AIC aos dados reais



Fonte: Elaborado pelo autor (2021)

Figura 8: Ajuste do modelo BIC aos dados reais



Fonte: Elaborado pelo autor (2021)

5.3 Previsão

Nesta etapa final, foi realizado um estudo comparativo das previsões utilizando os seguintes métodos:

- a) SARIMA (Método AIC)
- b) SARIMA (Método BIC)
- c) Suavização de Holt Winters (HW Aditivo)
- d) Suavização de Holt Winters HW Multiplicativo)
- e) Naive
- f) ARIMA (Drift)

Todos os métodos foram aplicados no intuito de prever 3 passos a frente.

Nesta fase, foram retirados da série os últimos três meses: dezembro de 2020, janeiro e fevereiro de 2021. Em seguida, realizou-se a previsão deste mesmo período e os resultados foram comparados aos reais.

- a) SARIMA (Modelo AIC)

O modelo SARIMA(0,1,4)(1,0,0)₁₂ foi obtido pelo software “R”, considerando como critério de seleção aquele que apresentasse melhor AIC, afim de se obter uma melhor relação entre os

seus componentes sazonais e não-sazonais auto – regressivos e médias móveis (Tabela 2) .

Tabela 2: coeficientes estimados pelo modelo SARIMA (0,1,4)x(1,0,0)₁₂ para a venda de carros, Brasil, Jan/1995 a Mai/2021.

ARIMA(0,1,4)(1,0,0) ₁₂	Coefficiente	p- valor
Média móvel 1	-0.4035	<0,01
Média móvel 2	-0.2558	<0,01
Média móvel 3	0,0986	0,166
Média móvel 4	-0,1417	<0,01
Autoregressivo sazonal 1	0.4788	<0,01

Fonte: Elaborado pelo autor (2021)

Na Tabela 3 temos os valores previstos e reais. Observamos assim, que há uma diferença significativa entre os valores. Sendo assim, foi calculado a média de erro absoluto (MAPE) afim de mensurar essa diferença em percentual. O MAPE de ajuste foi de 9,64%, e o MAPE de previsão foi de 14,44% considerados elevados, o que define inadequação do modelo como método de previsão para estes dados.

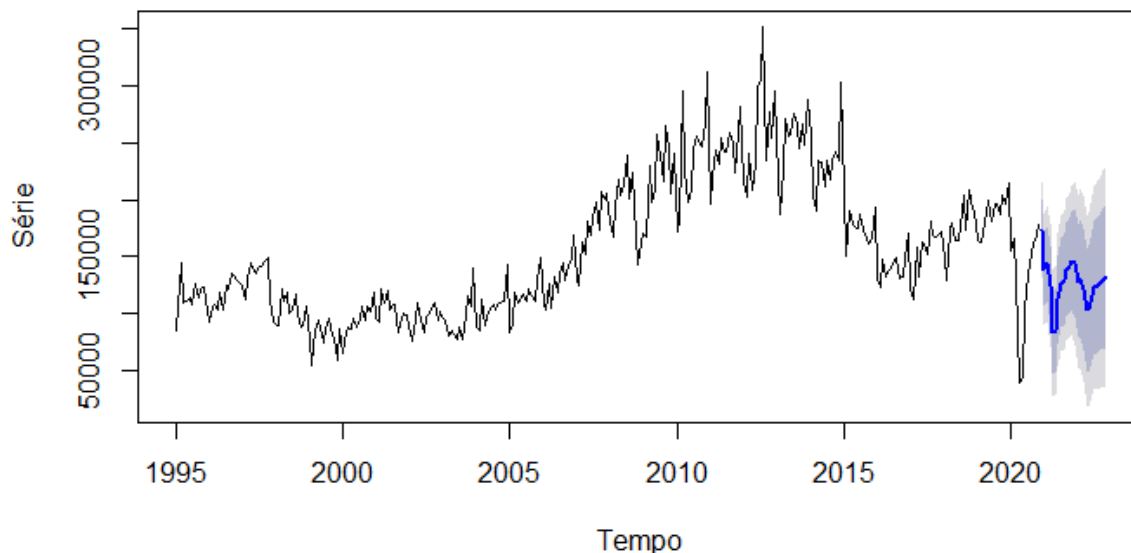
Na figura 8, temos o gráfico da série com os valores previstos representados pela linha azul, com os intervalos de confiança em cinza.

Tabela 3: Tabela da previsão em 3 passos através do modelo SARIMA (0,1,4)x(1,0,0)₁₂ para a venda de carros, Brasil, Jan/1995 a Mai/2021.

	dez/20	jan/21	fev/21
Venda Real	194668	130794	128099
Modelo AIC	170823	137940	142459

Fonte: Elaborado pelo autor (2021)

Figura 9: Gráfico da previsão em 3 passos através do modelo SARIMA (0,1,4)x(1,0,0)12 para a venda de carros, Brasil, Jan/1995 a Mai/2021.



Fonte: Elaborado pelo autor (2021)

b) SARIMA (Modelo BIC)

O modelo SARIMA(0,1,2)(1,0,0)12 foi obtido pelo software “R”, considerando como critério de seleção aquele que apresentasse melhor BIC, afim de se obter uma melhor relação entre os seus componentes sazonais e não-sazonais auto – regressivos e médias móveis (Tabela 4).

Tabela 4: Coeficientes estimados pelo modelo SARIMA (0,1,2)x(1,0,0)12 para a venda de carros, Brasil, Jan/1995 a Mai/2021.

ARIMA(0,1,2)(1,0,0)[12]	Coeficiente	p- valor
Constante	17.4792	0.9822
Média móvel 1	-0.4053	<0,01
Média móvel 2	-0.2532	<0,01
Autoregressivo sazonal 1	0.4788	<0,01

Fonte: Elaborado pelo autor (2021)

Na Tabela 5 temos os valores previstos e reais. Observamos assim, que há uma diferença significativa entre os valores. Sendo assim, foi calculado a média de erro absoluto (MAPE) afim de mensurar essa diferença em percentual. O MAPE de ajuste foi de 10,99%, e o MAPE de previsão foi de 12,71%. Este último indica que o modelo BIC é melhor como método de

previsão quando comparado com o modelo anterior. Porém as médias de erro percentual ainda são consideradas elevadas, revelando que o modelo possui baixo poder de previsão para estes dados. Além disso, notamos que os valores previstos de venda são muito próximos ao modelo AIC.

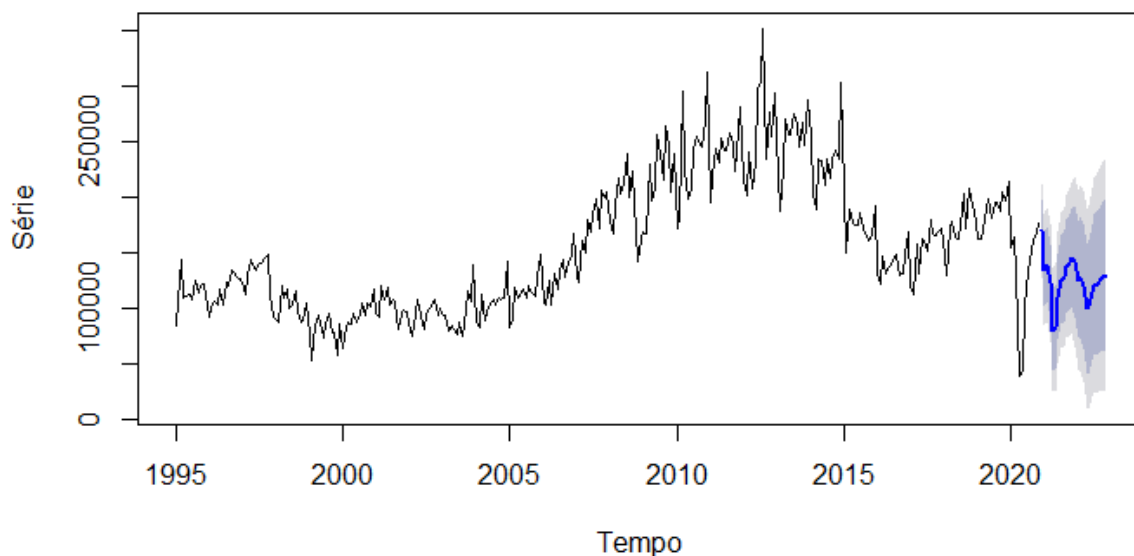
Na figura 9, temos o gráfico da série com os valores previstos representados pela linha azul, com os intervalos de confiança em cinza.

Tabela 5: Tabela da previsão em 3 passos através do modelo SARIMA (0,1,2)x(1,0,0)12 para a venda de carros, Brasil, Jan/1995 a Mai/2021.

	dez/20	jan/21	fev/21
Venda Real	194668	130794	128099
Modelo BIC	170527	135097	142459

Fonte: Elaborado pelo autor (2021)

Figura 10: Gráfico da previsão em 3 passos através do modelo SARIMA (0,1,2)x(1,0,0)12 para a venda de carros, Brasil, Jan/1995 a Mai/2021.



Fonte: Elaborado pelo autor (2021)

c) Suavização de Holt Winters (HW Aditivo)

O modelo HW Aditivo foi obtido pelo software “R”, como modelo alternativo aos modelos ARIMA tradicionais usados anteriormente. É uma extensão da suavização exponencial de Holt que captura a sazonalidade. Esse método gera valores exponencialmente suavizados para o nível

da previsão, a tendência da previsão e o ajuste sazonal da previsão. Esse é o melhor método para séries com tendência e sazonalidade que não aumentam com o tempo.

Na Tabela 6 temos os valores previstos e reais. Observamos assim, que há uma diferença significativa entre os valores. Sendo assim, foi calculado a média de erro absoluto (MAPE) afim de mensurar essa diferença em percentual. O MAPE de ajuste foi de 10,58%, e o MAPE de previsão foi de 75,40% revelendo ser o pior modelo para previsão desta série.

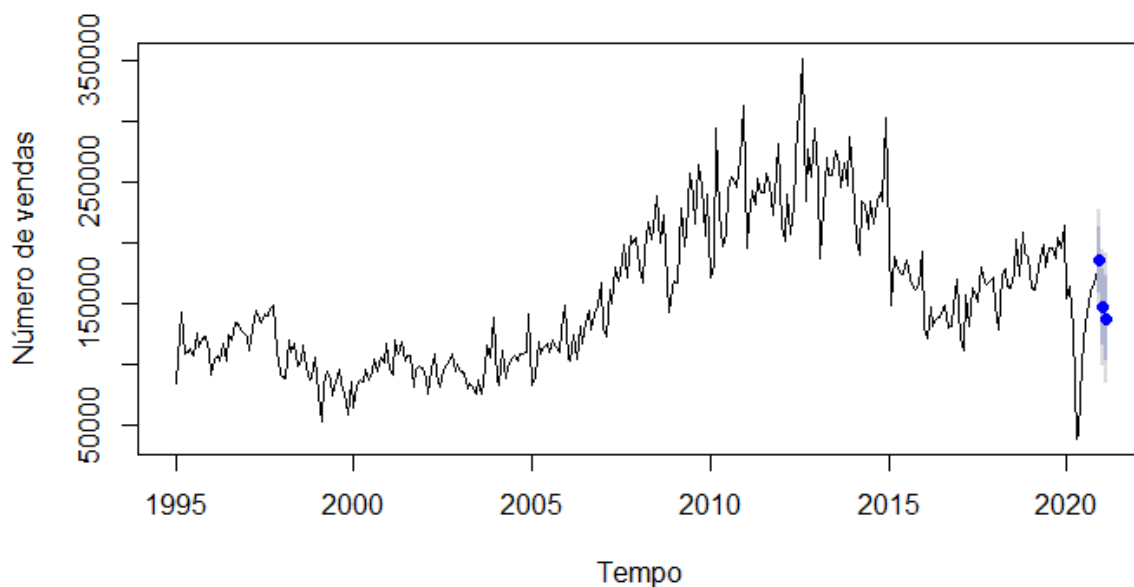
Na figura 11, temos o gráfico da série com os valores previstos representados pela linha azul, com os intervalos de confiança em cinza.

Tabela 6: Tabela da previsão em 3 passos através do modelo HW aditivo para a venda de carros, Brasil, Jan/1995 a Mai/2021.

	dez/20	jan/21	fev/21
Venda Real	194668	130794	128099
ARIMA (drive)	186181	147573	138442

Fonte: Elaborado pelo autor (2021)

Figura 11: Gráfico da previsão em 3 passos através do modelo HW aditivo para a venda de carros, Brasil, Jan/1995 a Mai/2021.



Fonte: Elaborado pelo autor (2021)

d) Suavização de Holt Winters (HW Multiplicativo)

O modelo HW (multiplicativo) foi obtido pelo software “R”, como modelo alternativo aos modelos ARIMA tradicionais usados anteriormente. É semelhante ao método aditivo de Holt-Winters. O método multiplicativo de Holt-Winters também calcula valores exponencialmente suavizados para o nível, a tendência e o ajuste sazonal da previsão. Esse método multiplicativo sazonal multiplica a previsão de tendência pela sazonalidade, produzindo a previsão multiplicativa de Holt-Winters. Esse método é melhor para dados com tendência e sazonalidade que aumentam com o tempo. Ele resulta em uma previsão com curvas, que reproduz as alterações sazonais nos dados.

Na Tabela 7 temos os valores previstos e reais. Observamos assim, que há uma diferença significativa entre os valores. Sendo assim, foi calculado a média de erro absoluto (MAPE) afim de mensurar essa diferença em percentual. O MAPE de ajuste foi de 10,60%, e o MAPE de previsão foi de 58,70% considerados elevados, o que define inadequação do modelo como método de previsão para estes dados, além disso não houve melhora significativa com relação os modelos anteriores.

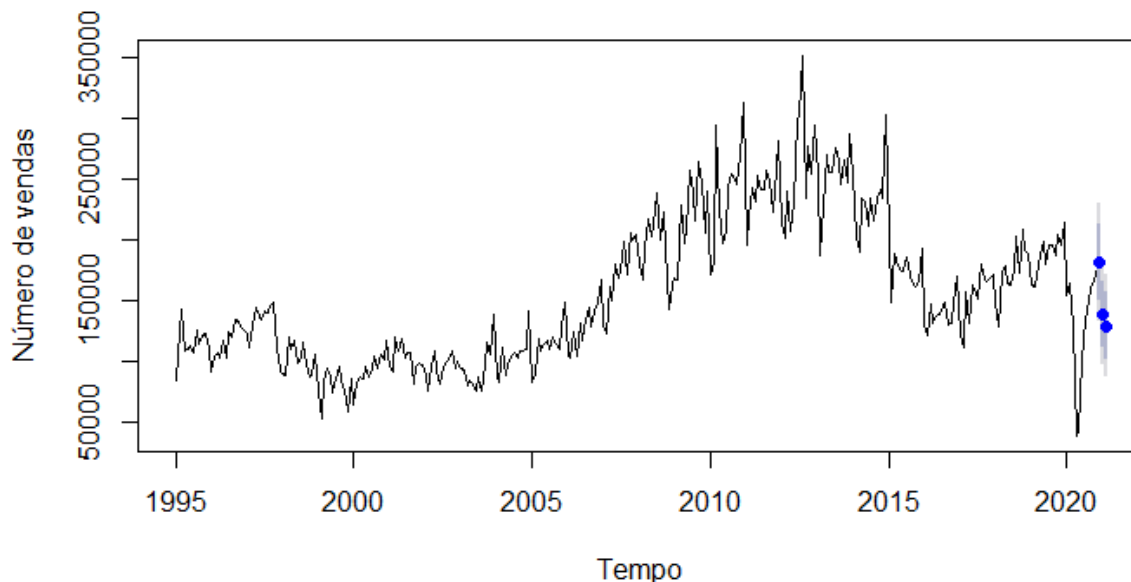
Na Figura 12, temos o gráfico da série com os valores previstos representados pela linha azul, com os intervalos de confiança em cinza.

Tabela 7: Tabela da previsão em 3 passos através do modelo HW multiplicativo para a venda de carros, Brasil, Jan/1995 a Mai/2021.

	dez/20	jan/21	fev/21
Venda Real	194668	130794	128099
ARIMA (drive)	182458	139498	129761

Fonte: Elaborado pelo autor (2021)

Figura 12: Gráfico da previsão em 3 passos através do modelo HW multiplicativo para a venda de carros, Brasil, Jan/1995 a Mai/2021.



Fonte: Elaborado pelo autor (2021)

e) NAIVE

O modelo Naive foi obtido pelo software “R”, como modelo alternativo aos modelos ARIMA tradicionais usados anteriormente. De acordo com Hanke e Witchern (2009), este é o modelo mais simples, mas, por vezes utilizado em negócios novos e sem base histórica de dados suficiente para métodos mais completos. Ele pode ser adaptado, dependendo da série temporal em questão.

Na Tabela 8 temos os valores previstos e reais. Observamos assim, que há uma diferença significativa entre os valores. Sendo assim, foi calculado a média de erro absoluto (MAPE) afim de mensurar essa diferença em percentual. O MAPE de ajuste foi de 18,92%, e o MAPE de previsão foi de 19,19% considerados elevados, o que define inadequação do modelo como método de previsão para estes dados, além disso não houve melhora significativa com relação os modelos anteriores.

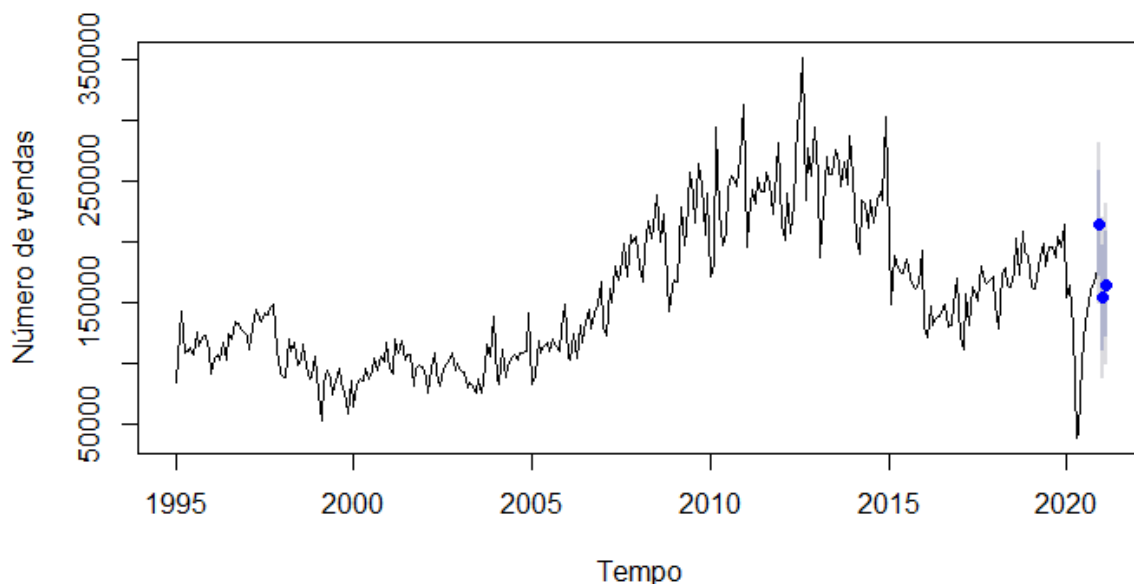
Na figura 13, temos o gráfico da série com os valores previstos representados pela linha azul, com os intervalos de confiança em cinza.

Tabela 8: Tabela da previsão em 3 passos através do modelo Naive para a venda de carros, Brasil, Jan/1995 a Mai/2021.

	dez/20	jan/21	fev/21
Venda Real	194668	130794	128099
ARIMA (drive)	215154	154556	165109

Fonte: Elaborado pelo autor (2021)

Figura 13: Gráfico da previsão em 3 passos através do modelo Naive para a venda de carros, Brasil, Jan/1995 a Mai/2021.



Fonte: Elaborado pelo autor (2021)

f) ARIMA (Drive)

O modelo Drive foi obtido pelo software “R”, como modelo alternativo aos modelos ARIMA tradicionais usados anteriormente. Como a série Y não é estacionária, o modelo mais simples possível para ela é um modelo de passeio aleatório (drift), que pode ser considerado um caso limite de um modelo AR (1) em que o coeficiente autoregressivo é igual a 1, ou seja, uma série com reversão à média infinitamente lenta. A equação de previsão para este modelo pode ser escrita como: $Y_t = Y_{t-1} + \epsilon_t$

Na Tabela 9 temos os valores previstos e reais. Observamos assim, que há uma diferença significativa entre os valores. Sendo assim, foi calculado a média de erro absoluto (MAPE) afim de mensurar essa diferença em percentual. O MAPE de ajuste foi de 13,24%, e o MAPE

de previsão foi de 19,19% considerados elevados, o que define inadequação do modelo como método de previsão para estes dados, além disso não houve melhora significativa com relação os modelos anteriores. Além disso, notamos que os valores previstos são muito próximos mês a mês, e não houve melhora significativa com relação os modelos anteriores.

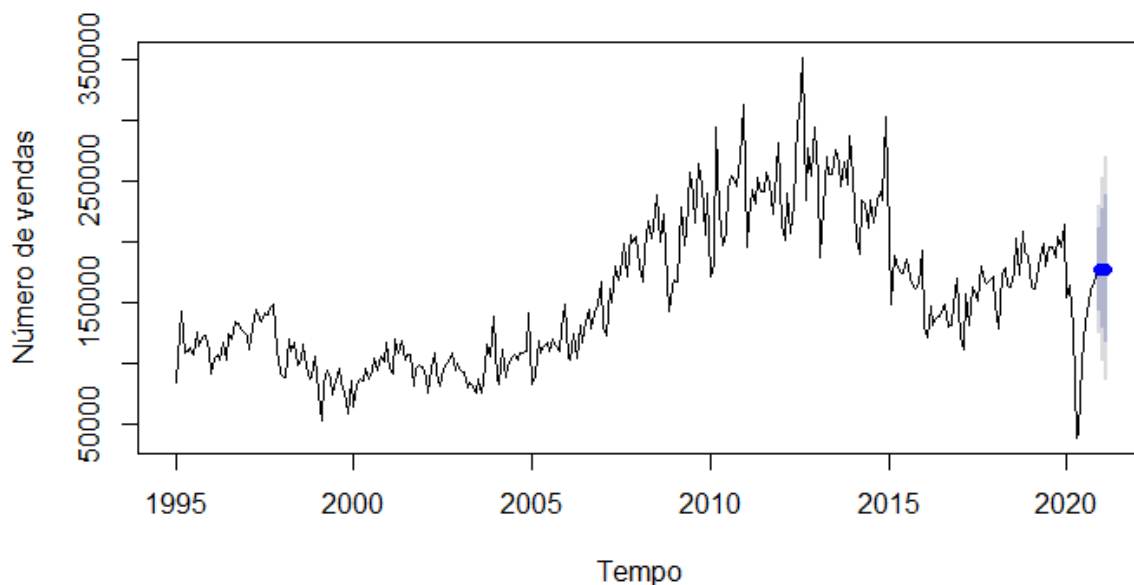
Na figura 14, temos o gráfico da série com os valores previstos representados pela linha azul, com os intervalos de confiança em cinza.

Tabela 9: Tabela da previsão em 3 passos através do modelo ARIMA (drive) para a venda de carros, Brasil, Jan/1995 a Mai/2021.

	dez/20	jan/21	fev/21
Venda Real	194668	130794	128099
ARIMA (drive)	177859	178157	178455

Fonte: Elaborado pelo autor (2021)

Figura 14: Gráfico da previsão em 3 passos através do modelo ARIMA (drive) para a venda de carros, Brasil, Jan/1995 a Mai/2021.



Fonte: Elaborado pelo autor (2021)

5.4 Análise da acuidade do modelo

A Tabela 10 apresenta as medidas MAE (erro médio absoluto), RMSE (desvio médio quadrático) e MAPE (erro absoluto percentual médio) da previsão, afim de comparar os ajustes dos diferentes modelos de séries temporais. Valores menores indicam um ajuste melhor. Porém, todos os

modelos de predição não apresentaram valores satisfatórios. Observamos que todos os MAPEs de previsão estavam acima de 5%, com MAEs e RMSE muito elevadas. Ainda assim, considerando as limitações, podemos perceber que o melhor modelo de previsão foi o Modelo BIC, pois possui menor MAPE, menor número de parâmetros, e menores MAE e RMSE.

Tabela 10: Medidas de acurácia (MAE, RMSE e MAPE) dos modelos para a venda de carros, Brasil, Jan/1995 a Mai/2021.

Modelo	MAE	RMSE	MAPE
Modelo AIC	20889,83	21055,81	14,44%
Modelo BIC	18269,00	18592,41	12,71%
HW Aditivo	11009,64	11317,34	75,40%
HW Multiplicativo	8967,15	9285,50	58,70%
Naive	27086,00	28012,19	19,19%
ARIMA (Drift)	38176,08	41075,34	28,05%

Fonte: Elaborado pelo autor (2021)

6 Conclusões

O presente estudo apresenta uma proposta de modelos para estimar a venda de veículos no Brasil para os futuros meses. Com este objetivo, foram estudados os modelos ARIMA considerando aqueles que apresentassem o melhor AIC (Modelo AIC) e melhor BIC (Modelo BIC). De um modo geral, os dois modelos estavam bem ajustados. Isso pode ser concluído através dos gráficos de probabilidade normal, gráficos de análise de resíduos, gráficos de autocorrelação e autocorrelação parcial dos resíduos.

A etapa de previsão foi realizada em três passos, utilizando o Modelo AIC e BIC. Complementarmente foram gerados outros modelos com finalidade de comparação. São eles: Suavização de Holt Winters aditivo e multiplicativo, Naive e ARIMA (Drift). Porém os indicadores de qualidade MAE, RMSE e MAPE de previsão indicaram médias de erros elevadas para todos. Neste sentido, foi possível verificar que tais modelos possuíam baixo poder de predição para esta série. Ainda assim, com todas as limitações foi possível escolher um melhor modelo geral para o estudo desta série, que foi o Modelo SARIMA BIC com média de erro absoluto percentual de 12,79%.

O resultado deste estudo, sugere que outras variáveis devem ser levadas em consideração no ajuste de um modelo de previsão para vendas de carros no Brasil. Pontos extremos, relacionados a pandemia de 2021 podem ter influenciado em tais resultados. Desta forma, recomenda-se a inclusão de outras características qualitativas em um trabalho futuro, tais como: impostos, nível de juros, custo tecnológico, volume de produção, oferta e procura, entre outras. Podendo incluir variáveis dummy, no caso de métodos clássicos, ou feitas intervenções ou descontos, quando se tratar de métodos bayesianos.

Propõe-se ainda, que pesquisas venham a ser desenvolvidas a partir deste trabalho de forma a contribuir para estudos de planejamento e controle da produção da indústria automobilística.

Referências

- [1] AKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. Forecasting: methods and applications. 3. ed. New York: John Wiley, 1998.
- [2] UBINO, D. F. Planejamento e controle da produção. São Paulo: Atlas, 2007. 190 p.
- [3] ALTER, O. M. F. C. et al. Aplicação de um modelo SARIMA na previsão de vendas de motocicletas. *Exacta*, v. 11, n. 1, 2013. Disponível em: <http://www.redalyc.org/html/810/81027458007/>; Acesso em: 14 de junho 2021.
- [4] OLK, V. Participação da indústria na economia brasileira sobe para 22%, diz CNI, 2019 Disponível em: <https://noticias.portaldaindustria.com.br/noticias/economia/participacao-da-industria-na-economia-brasileira-sobe-para-22-diz-cni/>
- [5] BGE, 2017. Síntese de indicadores sociais : uma análise das condições de vida da população brasileira : 2016. Estudos e pesquisas. Informação demográfica e socioeconômica, 146 p. ISSN 1516-3296 ; n. 36.
- [6] COLA, L.; ROHR, C.; ZMUD, J.; KUHNIMHOF, T.; PHLEPS, P. The Futures of driving in developing countries. Santa Monica, California: RAND Corporation, 2014.
- [7] PEA, 2011. Mobilidade Urbana - O automóvel ainda é prioridade. *Revista Desafios do Desenvolvimento - SBS*, 2011. Ano 8 . Edição 67.
- [8] NFAVEA, 2021. Anuário da Indústria Automobilística Brasileira 2021. Associação Nacional dos Fabricantes de Veículos Automotores. São Paulo, fevereiro de 2021.
- [9] ATORRE, M. R. D.; CARDOSO, M.R.A . Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. *Rev. bras. epidemiol.* 4 (3). Nov 2001. <https://doi.org/10.1590/S1415-790X2001000300002>.
- [10] OWNING, D CLARK, J. Estatística Aplicada.2 ed. São Paulo: Saraiva, 2006. EHLERS,R.S.Análise de Séries Temporais. 5 ed. 2009. Disponível em: . Acesso em 15 de maio 2021.
- [11] ILVA, E. MSILVA, E. M. Matemática e Estatística Aplicada. São Paulo: Atlas, 1999.
- [12] OUZA, G. P. SAMOHYL, R. W. MIRANDA, R. G. Métodos Simplificados de Previsão Empresarial. Rio de Janeiro: Ciência Moderna, 2008.
- [13] YNDMAN, R. J. ATHANASOPOLUS, G. Forecasting: principles and practice. 2013. Disponível em: . Acesso em: 14 de abril 2021.
- [14] AKRIDAKIS, Spyros; WHEELWRIGHT. Steven C.; HYNDMAN, Rob John. Forecasting: methods and applications. 3ª ed., New York: John Wiley Sons, 1998.

- [15] ORETTIN PA, TOLAI CMC. Previsão de séries temporais. 2 ed. São Paulo:Atual Editora; 1985.
- [16] ADALLA, G. S.; LAHIRI, K.Introduction to Econometrics. 4a. Edição, John Wiley Sons Ltd, UK, 2009.
- [17] EE, Cheng-Min; KO, Chia-Nan. Short-term load forecasting using lifting scheme and ARIMA models. Expert Systems with Applications, v. 38, p. 5902-5911, 2011.
- [18] OX, G. E. P.; JENKINS, G. M. Time series analysis forecasting and control. San Francisco: H. Day, 1976.
- [19] ORETTIN, P.; TOLOI, C. M. C. Análise de séries temporais. São Paulo: Edgard Blücher, 2004. 535 p.
- [20] HLERS, R. S. Análise de séries temporais. 2009. Disponível em: <https://sites.icmc.usp.br/ehlers/stemp/stemp.pdf>. Acesso em: 19 de março 2021.
- [21] RMSTRONG, J. S. Strategic planning and forecasting fundamentals. IN: ALBERT, K. The strategic management handbook. New York: MacGraw Hill, 1983.
- [22] UJARATI, D. N.; PORTER, D.C. Econometria Básica. 5 ed. Porto Alegre: AMGH, 2011.
- [23] EVINE, D. M., BERENSON, M. L. e STEPHAN, D. – Estatística: Teoria e Aplicações usando o Excel. Rio de Janeiro: LTC, 2000.
- [24] ONTGOMERY, D. C.; RUNGER, G. C. Estatística Aplicada e Probabilidade para Engenheiros. 2 ed. Rio de Janeiro: LTC, 2003.
- [25] OREIRA, D.A. Administração da Produção e Operações. 2.ed.São Paulo: Cengage, Learning, 2008.
- [26] ELO, Victor Andrade de. O automóvel, o automobilismo e a modernidade no Brasil (1891-1908). Revista Brasileira Ciência Esporte, Campinas, v. 30, n. 1, p. 187-203, set. 2008.
- [27] ENABRAVE, 2020. Federação Nacional da Distribuição de Veículos Automotores. Disponível em: <http://www.fenabrave.org.br/porta1> . Acesso em 24 de maio de 2021.
- [28] erluzzi, O. Análise do Setor Automobilístico. O que pode ocorrer nos próximos 3 anos? Indústria, tecnologia e economia, 2021. Disponível em: <https://oleodieselnaveia.com/2021/01/20/analise-do-setor-automobilistico-o-que-pode-ocorrer-nos-proximos-3-anos/>. Acesso em: 14 de Março de 2021.

Cronograma Computacional

```
rm(list=ls())

library(ggplot2)
library(ggfortify)
library(forecast)
library(urca)
library(BSDA)
library(astsa)
library(readxl)
library(fpp)
library(fma)
library(expsmooth)
library(stats)
library(tseries)
library(TTR)
#library(forecastML)
library(dplyr)
library(mafs)
library(magrittr)

vendas = read.table(file.choose(), quote="\\"", comment.char="")
#View(vendas)
attach(vendas)

dados = vendas[,2]
str(dados)

#ESTATÍSTICA DESCRITIVA

# Existe 314 observações (jan/95 - fev/21)
serie.ts <- ts(dados, start = c(1995,1), frequency=12)

# Descrição, desvio padrão e tamanho da série
```

```

summary(serie.ts)
sqrt(var(serie.ts))

length(serie.ts)

# Teste de estacionariedade
a.dif= ur.kpss(diff(serie.ts));a.dif

#A estatístico do teste foi de: 2.6084, hipótese nula é estacionariedade
adf.test(serie.ts)
#a.dif2= adf.test(diff(serie.ts));a.dif2
#Se rejeitar a hipótese nula, isso quer dizer que, a série é
#estacionária, pois  $H_0: \phi=1$ .
#Portanto atendendo a probabilidade do erro tipo I (a
probabilidade de rejeitar  $H_0$  quando
#ela é verdadeira) ao nível de significância de 5%, temos que a hipótese
#nula de não
#estacionariedade deve ser rejeitada. Dado que o p-valor, probabilidade
#de cometer o erro
#tipo I, é de 0.01. Hipótese alternativa é estacionariedade

# Gráficos: descrtivo, ACF, PACF
plot(serie.ts, xlab = "Tempo", ylab = "Número de vendas", main = "")
autoplot(serie.ts, xlab="Tempo", ylab="Número de vendas", main = "")
ggseasonplot(serie.ts, xlab="Tempo", ylab="Número de vendas", main = "")
acf(serie.ts, main="") # Ordem de um processo MA(q), todos os pontos
#estão
# fora do intervalo, portanto nada se conclui
pacf(serie.ts, xlab = "Lag", ylab = "PACF", main="") # Ordem de um
#processo AR(p), não existe um padrão e
# e um ponto de corte definido, portanto se pode se concluir nada

## Decomposição da serie temporal
serie.stl = decompose(serie.ts)
plot(serie.stl, xlab = "Tempo", ylab = "") #REVER E RVER
autoplot(serie.stl,xlab = "Tempo", ylab =

```

```
# " Restante Tendência Sazonalidade Dados", main=
#"Descomposição Aditiva da Série") # Modelo aditivo ->
#sazonal+tendencia+aleatoriedade
serie.stl$trend
```

```
#ARIMA E SARIMA
```

```
#diferença ARIMA (d)
ndiffs(serie.ts, test="adf") # 1
#diferença SARIMA (D)
nsdiffs(serie.ts, test="adf") # 0
```

```
diff(serie.ts)
```

```
#Selecionando "melhor" modelo VIA AIC
```

```
d.arima.aic <- auto.arima(serie.ts, ic="aic"); d.arima.aic
summary(d.arima.aic)
#AIC=7145.86 AICc=7146.13 BIC=7168.34
#ARIMA(0,1,4) (1,0,0) [12]
coeftest(d.arima.aic)
```

```
#Chacar qualidade do modelo
```

```
Box.test(d.arima.aic$res) # Hipótese nula de independencia
#X-squared = 2.708e-05, df = 1, p-value = 0.9958
tsdiag(d.arima.aic)
```

```
#Selecionando "melhor" modelo VIA BIC
```

```
d.arima.bic <- auto.arima(serie.ts, ic="bic"); d.arima.bic
summary(d.arima.bic)
#AIC=7148.46 AICc=7148.59 BIC=7163.44
#ARIMA(0,1,2) (1,0,0) [12]
```

```
#Checar qualidade do modelo
```

```
Box.test(d.arima.bic$res)
```

```

#X-squared = 0.0028991, df = 1, p-value = 0.9571
tsdiag(d.arima.bic)

#Ajustando os dados gerados pelo modelo aic aos reais
plot.ts(serie.ts)
lines(fitted(d.arima.aic),col=4)

#Ajustando os dados gerados pelo modelo bic aos reais
plot.ts(serie.ts)
lines(fitted(d.arima.aic),col=2)

# Ajuste do modelo selecionado e diagnostico VIA AIC E BIC
fit.aic= sarima(serie.ts, 0,1,4, 1,0,0,12); fit.aic

fit.bic= sarima(serie.ts, 0,1,2, 1,0,0,12); fit.bic

#ESTRUTUTA PARA PREVISÃO A 3 PASSOS

##### Previsão h=3 #####
serie3.ts = serie.ts[-c(312:314)]
caixa3 = serie.ts[c(312:314)]
length(serie3.ts)
length(caixa3)
serie3.ts
caixa3

serie3.ts.s = ts(serie3.ts, start=c(1995,1),frequency=12)

#PREVISÃO VIA ALGORITMO DE ALISAMENTO EXPONENCIAL

#3 PASSOS
AE.A.h3 <- hw(serie3.ts.s,seasonal="additive", h=3)
summary(AE.A.h3)

plot(AE.A.h3, main = "")
lines(serie3.ts.s, lty = 3)

```

```

AE.M.h3 <- hw(serie3.ts.s,seasonal="multiplicative", h=3)
summary(AE.M.h3)

plot(AE.M.h3, main = "")
lines(serie3.ts.s, lty = 3)

#PREVISÃO VIA SARIMA, RWF, SNAIVE

#3 PASSOS
# SARIMA (via AIC)

fit.serie3.aic = arima(serie3.ts.s, order = c(2,1,2), seasonal =
#list(order = c(0,0,2)))

pred=forecast(fit.serie3.aic)
plot(pred)
summary(fit.serie3.aic)
pred.serie3.fit.aic <- predict(fit.serie3.aic,n.ahead=3,se.fit=T)
pred.serie3.fit.aic$pred
fit.serie3.aic

plot(serie3.ts.s)
lines(pred.serie3.fit.aic$pred, lty = 3, col = "blue")

# SARIMA (via BIC)

fit.serie3.bic = arima(serie3.ts.s, order = c(0,1,2), seasonal =
#list(order = c(1,0,0)))
pred.serie3.fit.bic <- predict(fit.serie3.bic,n.ahead=3,se.fit=T)
pred.serie3.fit.bic$pred
pred=forecast(fit.serie3.bic)
plot(pred)
plot(serie3.ts.s)
lines(pred.serie3.fit.bic$pred, lty = 3, col = "blue")
summary(fit.serie3.bic)

# rwf e snaive

```

```

serie3.drift.h3 <- rwf(serie3.ts.s,h=3, drift=TRUE);serie3.drift.h3
summary(serie3.drift.h3)
serie3.snaive.h3 <- snaive(serie3.ts.s,h=3);serie3.snaive.h3
summary((serie3.snaive.h3))
plot(serie3.drift.h3)
plot(serie3.snaive.h3)

```

```
#COMPARANDO MODELOS
```

```
# Medidas para o erro
```

```

accuracy(pred.serie3.fit.aic$pred, caixa3) [1,c(3,2,5)]
accuracy(pred.serie3.fit.bic$pred, caixa3) [1,c(3,2,5)]
accuracy(AE.A.h3$mean, caixa3) [1,c(3,2,5)]
accuracy(AE.M.h3$mean, caixa3) [1,c(3,2,5)]
accuracy(serie3.snaive.h3$mean, caixa3) [1,c(3,2,5)]
accuracy(serie3.drift.h3$mean, caixa3) [1,c(3,2,5)]

```

```

conj1 <- rbind(accuracy(pred.serie3.fit.aic$pred, caixa3) [1,c(3,2,5)],
              accuracy(pred.serie3.fit.bic$pred, caixa3) [1,c(3,2,5)],
              accuracy(AE.A.h3$mean, caixa3) [1,c(3,2,5)],
              accuracy(AE.M.h3$mean, caixa3) [1,c(3,2,5)],
              accuracy(serie3.snaive.h3$mean, caixa3) [1,c(3,2,5)],
              accuracy(serie3.drift.h3$mean, caixa3) [1,c(3,2,5)])

```

```

Método <- c("Via AIC", "Via BIC", "AE Aditivo",
            #"AE Multiplicativo", "SNAIVE", "DRIFT")

```

```
results1 <- data.frame(Método,conj1); results1
```