



**Universidade Federal de Uberlândia
Instituto de Matemática e Estatística**

Bacharelado em Estatística

**ANÁLISE DOS PREÇOS MÉDIOS DE
COMBUSTÍVEIS EM UBERLÂNDIA-MG
E PREVISÃO DA GASOLINA VIA
MODELOS DE SÉRIES TEMPORAIS**

Thainá Faustina Silva

Uberlândia-MG

2025

Thainá Faustina Silva

**ANÁLISE DOS PREÇOS MÉDIOS DE
COMBUSTÍVEIS EM UBERLÂNDIA-MG
E PREVISÃO DA GASOLINA VIA
MODELOS DE SÉRIES TEMPORAIS**

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.

Orientador: Prof^ª. Dra. Nádia Giaretta Biase

**Uberlândia-MG
2025**



**Universidade Federal de Uberlândia
Instituto de Matemática e Estatística**

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, 19 de Setembro de 2025

BANCA EXAMINADORA

Prof^a. Dra. Nádia Giaretta Biase

Prof. Dr. Ednaldo Carvalho Guimarães

Prof. Dr. Lúcio Borges de Araújo

**Uberlândia-MG
2025**

RESUMO

Este trabalho teve como objetivo analisar a evolução dos preços médios semanais dos combustíveis (Diesel S10, Etanol e Gasolina) na cidade de Uberlândia-MG, entre julho de 2021 e dezembro de 2024, utilizando modelos de séries temporais para prever valores futuros do combustível Gasolina da bandeira Ipiranga. Inicialmente, foram tratadas lacunas nos dados com técnicas de imputação sazonais, seguidas de análises descritivas e comparativas entre diferentes bandeiras de distribuição. A série com variabilidade intermediária foi selecionada para modelagem, sendo aplicados testes estatísticos para verificar tendência e sazonalidade. O modelo SARIMA foi ajustado com base nos menores valores de AIC, BIC, RMSE e MAPE. Os resultados indicaram estabilidade e menor volatilidade nos preços em 2024. O estudo contribui para a compreensão do comportamento do mercado de combustíveis e oferece suporte à tomada de decisões por consumidores, empresas e formuladores de políticas públicas.

Palavras-chave: Análise preditiva, AIC, BIC, SARIMA, Sazonalidade.

ABSTRACT

This study aimed to analyze the weekly average fuel prices (Diesel S10, Ethanol, and Gasoline) in Uberlândia, MG, from July 2021 to December 2024, using time series models to forecast future prices of Ipiranga-branded Gasoline. Data gaps were initially addressed through seasonal imputation techniques, followed by descriptive and comparative analyses across different fuel brands. The series exhibiting intermediate variability was selected for modeling, and statistical tests were conducted to assess trends and seasonality. The SARIMA model was fitted based on the lowest AIC, BIC, RMSE, and MAPE values. The results revealed price stability and reduced volatility in 2024. This study contributes to a better understanding of fuel market dynamics and offers valuable insights to support decision-making by consumers, businesses, and policymakers.

Keywords: Predictive analysis, AIC, BIC, SARIMA, Seasonality.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	II
Lista de Abreviações e Símbolos	III
1 Introdução	1
2 Fundamentação Teórica	3
2.1 Séries Temporais	3
2.1.1 Estacionariedade	3
2.1.2 Componentes de uma Série Temporal	4
2.1.3 Modelos ARIMA e SARIMA	7
2.1.4 Critérios de Seleção	8
2.1.5 Métricas de Comparação	9
2.2 Previsão de dados ausentes	10
3 Metodologia	11
4 Resultados	14
5 Conclusões	24
Referências Bibliográficas	25

LISTA DE FIGURAS

4.1	Evolução dos Preços Médios dos Combustíveis (2021-2024)	15
4.2	Distribuição dos Preços Médios dos Combustíveis por Bandeira	16
4.3	Preços médios do combustível gasolina da bandeira Ipiranga de julho de 2021 a dezembro de 2024 em Uberlândia.	17
4.4	Função de Autocorrelação (ACF)	17
4.5	Função de Autocorrelação Parcial (PACF)	18
4.6	Média em função do desvio padrão	18
4.7	Funções ACF e PACF dos resíduos	20
4.8	Periodograma acumulado	20
4.9	Previsão dos seis primeiros meses de 2025 para o modelo 4	21

LISTA DE TABELAS

3.1	Tabela de preços médios por semana com valores ausentes	11
4.1	Tabela de preços médios por semana com valores preenchidos	14
4.2	Estatísticas descritivas dos preços médios dos combustíveis (2021–2024)	15
4.3	Tabela de variabilidade com médias, desvios padrão (Sd) e coeficiente de variação (CV)	16
4.4	Modelos propostos e resultados	19
4.5	Métricas RMSE e MAPE para os modelos selecionados	19
4.6	Previsões pontuais e intervalos de 95% comparados com os valores reais e suas diferenças	22

LISTA DE ABREVIACÕES E SÍMBOLOS

LISTA DE ABREVIACÕES

ACF AutoCorrelation Function

AIC Akaike Information Criterion

ARFIMA AutoRegressive Fractionally Integrated Moving Average

ARIMA AutoRegressive Integrated Moving Average

BIC Bayesian Information Criterion

MAPE Mean Absolute Percentage Error

PACF Partial AutoCorrelation Function

RMSE Root Mean Square Error

SARIMA Seasonal AutoRegressive Integrated Moving Average

TDEM Time-Dependence Effect Method

1. INTRODUÇÃO

Os combustíveis derivados de fontes não renováveis são combustíveis formados por processos naturais ao longo do tempo, contendo altas quantidades de carbono que favorecem a combustão, convertendo energia química em energia térmica. Essas fontes de energia podem ser encontradas a partir do petróleo, do gás natural e do carvão mineral, conhecidos como combustíveis fósseis [11]. Segundo Ribeiro (2014) [23], o carvão mineral teve grande presença, no século XVIII, na revolução industrial originada na Inglaterra. Em seguida, a exploração do petróleo foi impulsionada no século XIX sendo utilizada por indústrias de diferentes portes. Já no século XX, o gás natural passou a ser explorado.

Visto que combustíveis derivados de fontes não renováveis não se recompõem a curto prazo e são altamente poluentes à natureza, começou a ser utilizada novas tecnologias com os combustíveis derivados de fontes renováveis que são formados por recursos naturais, como a cana-de-açúcar proveniente da agricultura, que se renovam em um curto período e são mais sustentáveis. Comparado com os combustíveis fósseis que demoram muito para se formar e não tem a capacidade de renovação rápida.

Segundo Silva e Freitas (2008) [25], os combustíveis de fontes renováveis como o álcool teve presença significativa na década de 1970, visto que, na tentativa de diminuir o consumo de gasolina, o governo brasileiro criou o Programa Nacional do Álcool (PróÁlcool). Assim, tendo um desempenho onde 90% dos carros produzidos no Brasil eram movidos a álcool entre 1983 e 1988 [1].

Com a popularização dos automóveis, tornou se mais acessível a aquisição deles por parte da população e esse aumento resultou em grande demanda por combustíveis, pois são fontes de energia utilizadas no dia a dia, como no transporte, na indústria e na produção de aquecimento. Dado o aumento da frota de veículos, houve também um maior consumo desse produto elevando assim seu preço.

O preço dos combustíveis é, sem dúvida, um dos aspectos mais determinantes na economia de um país. No Brasil foi possível observar expressivas oscilação nos preços da gasolina, etanol e diesel, o que gerou muita preocupação aos cidadãos, empresários e ao governo. Essas oscilações foram determinadas por vários fatores, como a pandemia da Covid-19, a cotação do dólar, a oferta e demanda de petróleo, as políticas de preços das distribuidoras e a carga tributária.

Em Uberlândia, uma das principais cidades do interior de Minas Gerais, sendo um grande polo logístico, da mesma forma que no Brasil, sentiu esse impacto econômico com maior velocidade do que nas outras regiões.

Nesse sentido, avaliar o comportamento dos preços desses produtos ao longo dos anos pode ser de grande valia para compreender a estrutura dos dados, verificando a existência de sazonalidade e/ou tendência, bem como identificar e estimar modelos de previsão aos quais os dados se ajustem e fazer previsões dos preços para períodos subsequentes, a partir desses modelos. Dentre os métodos estatísticos com esta finalidade, a análise de séries temporais permite analisar e modelar observações geradas sequencialmente no tempo, igualmente espaçadas e que apresentam dependência serial.

Diversos estudos já utilizaram modelos de séries temporais para analisar e prever o comportamento dos preços de produtos em diferentes contextos. Brito (2019) [8] realizou um estudo aplicando modelos ARIMA, ARFIMA e de Suavização Exponencial para prever o preço do etanol no Brasil, identificando que o modelo ARIMA (1,1,1) apresentou melhor acurácia preditiva para previsões diárias, considerando o período de 2010 a 2014. Da mesma forma, Araujo et al. (2012) [5] aplicaram modelos ARIMA e SARIMA para prever os preços do boi gordo, constatando que o modelo SARIMA apresentou um destaque devido à forte presença de sazonalidade, sendo eficaz e podendo auxiliar na tomada de decisões dos pecuaristas.

Em função disso, este trabalho teve como objetivo, analisar a evolução da média semanal dos preços dos combustíveis em Uberlândia, no período de julho de 2021 a dezembro de 2024. Para isso, foram utilizados modelos de séries temporais, a fim de prever valores futuros da série com variabilidade intermediária. Além disso, buscou-se compreender como essas oscilações afetam a sociedade local e de que forma modelos preditivos podem auxiliar na antecipação do preço, permitindo que consumidores, empresas e formuladores de políticas públicas possam tomar melhores decisões.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 SÉRIES TEMPORIAS

As séries temporais são fundamentais na análise de dados que variam ao longo do tempo, pois permitem identificar padrões não aleatórios e tendências capazes de auxiliar na previsão de eventos futuros. Essas séries consistem em conjuntos de observações registradas em intervalos regulares, como, por exemplo, o preço médio do combustível automotivo analisado neste estudo.

A análise de séries temporais parte da suposição de que há um sistema causal relativamente constante ao longo do tempo, o qual influencia os dados e gera padrões que podem ser identificados e modelados [22].

Ao compreender esses padrões, torna-se possível realizar projeções que apoiam a tomada de decisões, possibilitando avaliar os impactos das variações no preço médio dos combustíveis sobre a sociedade local e contribuindo para estratégias mais eficazes de planejamento econômico e mobilidade urbana.

Segundo Morettin e Toloi [18], uma série temporal pode ser definida como uma sequência de variáveis aleatórias indexadas no tempo, denotadas por $\{Z_t\}_t \in T$, onde t representa o tempo, e Z_t é o valor observado no instante t . A análise estatística de séries temporais busca modelar a estrutura de dependência entre as observações, com o objetivo de compreender o comportamento dinâmico do fenômeno estudado e realizar previsões futuras.

2.1.1 ESTACIONARIEDADE

A estacionariedade é um requisito fundamental para a aplicação da maioria dos modelos estatísticos em séries temporais. Uma série estacionária evolui de forma aleatória ao longo do tempo, mantendo uma média constante, variância estável e uma estrutura de dependência que não varia com o tempo. No entanto, muitas séries observadas na prática apresentam características de não-estacionariedade, como tendência, sazonalidade ou mudanças de nível.

Segundo Morettin e Toloi [18], uma série é considerada estacionária em sentido amplo quando satisfaz três condições: a média $E(Z_t)$ é constante ao longo do tempo; a variância $Var(Z_t)$ é finita e constante; e a função de autocovariância $Cov(Z_t, Z_{t+k})$ depende apenas da defasagem k , e não do tempo t .

Em alguns casos, a série pode ser estacionária apenas em determinados intervalos, apresentando variações estruturais em outros — fenômeno conhecido como estacionariedade local. Quando a série não é estacionária, técnicas de transformação, como a diferenciação (diferença entre valores consecutivos), são empregadas para estabilizar a média e a variância. Geralmente, uma ou duas aplicações dessa técnica são suficientes para tornar a série adequada à modelagem estatística.

A primeira diferença de $Z(t)$, conforme definida por Morettin e Toloi [18], é dada pela seguinte equação:

$$\Delta Z(t) = Z(t) - Z(t-1), \quad (2.1)$$

a segunda diferença é

$$\Delta^2 Z(t) = \Delta[\Delta Z(t)] = \Delta[Z(t) - Z(t-1)], \quad (2.2)$$

ou seja,

$$\Delta^2 Z(t) = Z(t) - 2Z(t-1) + Z(t-2), \quad (2.3)$$

De modo geral, a n -ésima diferença de $Z(t)$ é

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)]. \quad (2.4)$$

2.1.2 COMPONENTES DE UMA SÉRIE TEMPORAL

Segundo Box et al. (2015) [7], uma série temporal pode ser decomposta em componentes como tendência, sazonalidade, ciclos e ruído aleatório. A identificação e separação desses elementos é essencial para a construção de modelos preditivos eficazes, pois cada componente representa uma característica distinta do comportamento da série ao longo do tempo.

Nessa mesma linha, Morettin e Toloi [18] detalham esses componentes e destacam a importância de considerá-los individualmente na análise estatística de séries temporais:

- **Tendência (T):** Componente de longo prazo que pode indicar uma variação persistente no tempo, como um aumento ou declínio progressivo nos preços.
- **Sazonalidade (S):** Flutuações regulares e recorrentes, geralmente associadas a padrões específicos de tempo, como variações semanais, mensais ou anuais.
- **Ciclo (C):** Diferencia-se da sazonalidade por não ter periodicidade fixa. Representa movimentos econômicos ou estruturais de maior duração.
- **Erro ou Ruído (E):** Componente residual não explicado pelos demais, frequentemente tratado como ruído branco.

No contexto dos preços de combustíveis, a sazonalidade semanal assume relevância, uma vez que padrões de consumo e abastecimento podem variar ao longo do ano, influenciados por fatores logísticos, feriados, variações climáticas e decisões de natureza política.

A decomposição de uma série pode ser representada por meio de dois modelos distintos, conforme a relação entre os componentes:

$$X_t = T_t + S_t + C_t + E_t \quad (\text{Modelo Aditivo}) \quad (2.5)$$

$$X_t = T_t \times S_t \times C_t \times E_t \quad (\text{Modelo Multiplicativo}) \quad (2.6)$$

A escolha entre os modelos aditivo e multiplicativo geralmente é orientada pela inspeção visual da série, especialmente pela análise da relação entre a média e o desvio padrão ao longo do tempo.

TENDÊNCIA

Segundo Morettin e Toloi [18] existe métodos para estimar a tendência. Os mais utilizados consistem em:

- ajustar uma função do tempo, como um polinômio, uma exponencial ou outra função suave de t ;
- suavizar (ou filtrar) os valores da série ao redor de um ponto, para estimar a tendência naquele ponto;
- suavizar os valores da série através de sucessivos ajustes de retas de mínimos quadrados ponderados ("lowess").

Para verificar a presença de tendência em uma série temporal, podem ser aplicados testes não paramétricos, que são particularmente úteis quando não se deseja assumir hipóteses fortes sobre a distribuição dos dados. Neste estudo, será utilizado o teste do sinal, também conhecido como teste de Cox-Stuart [18].

Esse teste consiste em agrupar as observações da série em pares da seguinte forma:

$$(Z_1, Z_{1+c}), (Z_2, Z_{2+c}), \dots, (Z_{N-c}, Z_N)$$

onde $c = \frac{N}{2}$, se N for par e $c = \frac{N+1}{2}$, se N for ímpar.

A cada par (Z_i, Z_{i+c}) associamos o sinal $+$ se $Z_i < Z_{i+c}$ e o sinal $-$ se $Z_i > Z_{i+c}$, eliminando os empates. Seja n o número de pares onde $Z_i \neq Z_{i+c}$.

$$\begin{cases} H_0 : P(Z_i < Z_{i+c}) = P(Z_i > Z_{i+c}) \\ H_1 : P(Z_i < Z_{i+c}) \neq P(Z_i > Z_{i+c}) \end{cases} \quad (2.7)$$

A hipótese nula (H_0) pressupõe que não há tendência na série, ou seja, que a série é simétrica em torno de sua mediana. Nessa situação, espera-se que a quantidade de sinais positivos e negativos seja aproximadamente equilibrada. A hipótese alternativa (H_1), por sua vez, sugere a existência de uma tendência significativa, caracterizada por excesso de sinais positivos (indicando crescimento) ou negativos (indicando queda).

O critério de decisão varia conforme o número de pares comparados (m). Para $m \leq 20$, a estatística de teste segue a distribuição binomial $B(m; p = 0,05)$ ao nível de significância α , enquanto para $m > 20$ adota-se a aproximação pela distribuição normal $N(np; npq)$. A rejeição da hipótese nula ocorre quando $T_c \geq m - t$, em que T_c representa o número de pares com sinal positivo, N corresponde ao número total de observações e t é o valor tabelado de referência. Dessa forma, quando a proporção de sinais positivos e negativos se afasta significativamente do equilíbrio esperado, conclui-se pela presença de uma tendência na série temporal.

SAZONALIDADE

No caso da sazonalidade, Morettin e Toloi [18] destacam que existem diversos procedimentos para sua estimativa. Os mais comuns são:

- **Método de regressão:** consiste em ajustar um modelo com variáveis indicadoras (dummies) que representam os diferentes períodos do ciclo sazonal (por exemplo, meses ou semanas). Esse método permite quantificar a contribuição de cada período para o comportamento da série.
- **Método de médias móveis:** baseia-se na suavização da série temporal por meio de médias centradas. A componente sazonal pode ser estimada como a diferença entre a série original e a série suavizada, uma vez que esta última elimina, em grande parte, os efeitos sazonais.

Para testar a presença de periodicidade na série, Morettin e Toloi citam o procedimento proposto por **Fisher**, que testa a seguinte hipótese nula:

$$H_0 : R_i \approx 0, \text{ todo } i$$

em que R_i representa a autocorrelação no defasagem (lag) i . A hipótese alternativa considera a existência de pelo menos uma defasagem i com autocorrelação significativa, ou seja:

$$H_1 : \exists i \text{ tal que } R_i \neq 0$$

Em termos espectrais, a rejeição da hipótese nula indica a presença de uma frequência dominante, caracterizada por picos significativos no periodograma da série, o que evidencia componentes cíclicas ou sazonais relevantes. Assim, enquanto H_0 aponta para ausência de periodicidade, a aceitação de H_1 sugere comportamento periódico na série.

A estatística de teste de Fisher é calculada a partir da decomposição espectral da série, utilizando o periodograma como estimador da densidade espectral, verificando a existência de frequências dominantes que indiquem ciclos ou padrões sazonais significativos.

A estatística de Fisher é dada por:

$$g = \frac{\max I_j}{\sum_{j=1}^{n/2} I_j}, \quad (2.8)$$

onde I_j apresenta o valor do periodograma na frequência j e n o número total de observações da série. Valores elevados de g indicam que uma única frequência domina a variabilidade da série, o que leva à rejeição da hipótese nula e à conclusão de que existe periodicidade significativa.

Para comparar o valor observado de g , utiliza-se o valor crítico Z_α , definido por:

$$Z_\alpha = 1 - \left(\frac{\alpha}{2}\right)^{\frac{1}{(n/2)-1}}, \quad (2.9)$$

em que α representa o nível de significância do teste. Se $g > Z_\alpha$, rejeita-se a hipótese nula, concluindo-se que a série apresenta periodicidade significativa na frequência correspondente ao maior pico do periodograma.

2.1.3 MODELOS ARIMA E SARIMA

O modelo ARIMA (AutoRegressive Integrated Moving Average) é um dos principais instrumentos da análise de séries temporais, tendo sido consolidado a partir dos trabalhos de Box e Jenkins [6]. Ele tem origem nos modelos autorregressivo (AR), de médias móveis (MA) e na combinação de ambos (ARMA). Quando a série apresenta não estacionariedade, aplica-se o processo de diferenciação para transformá-la em estacionária. Dessa forma, além das componentes AR e MA, utiliza-se a componente de integração $I(d)$, resultando no modelo $ARIMA(p, d, q)$. Conforme Morettin e Toloi [18] explicam esse modelo é adequado para séries temporais não estacionárias e é caracterizado pelos parâmetros:

- p : ordem da parte autorregressiva (AR);
- d : ordem de diferenciação (responsável por tornar a série estacionária);
- q : ordem da média móvel (MA).

Sua formulação geral é dada por:

$$\phi_p(B)\Delta^d Z_t = \theta_q(B)a_t \quad (2.10)$$

em que $\Delta^d Z_t = (1 - B)^d Z_t$ e d é a ordem de diferenciação.

Quando a série apresenta sazonalidade, utiliza-se o modelo SARIMA, que estende o ARIMA com componentes sazonais (P, D, Q) associados a um período s , conforme a notação:

$$SARIMA(p, d, q)(P, D, Q)_s \quad (2.11)$$

Esse modelo pode ser generalizado pela inclusão de operadores sazonais, sendo representado por:

$$\phi(B)\Phi(B^s)\Delta^d\Delta_S^D Z_t = \theta(B)\Theta(B^s)a_t \quad (2.12)$$

em que:

- $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ é o operador auto-regressivo de ordem p .
- $\Phi(B^s) = (1 - \Phi_1 B^s - \dots - \Phi_P B^{sP})$ é o operador auto-regressivo sazonal de ordem P .
- $\Theta(B^s) = (1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ})$ é o operador de médias móveis de ordem Q .
- $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ é o operador de médias móveis sazonal de ordem q .
- $\Delta^d = (1 - B)^d$ é o operador diferença simples, d indicando o número de diferenças.
- $\Delta_S^D = (1 - B^s)^D$ é o operador diferença sazonal, D indicando o número de diferenças sazonais.

Esse modelo é particularmente adequado para séries com sazonalidade semanal, como no caso dos preços de combustíveis analisados neste estudo, considerando $s = 52$. Os autores, Morettin e Toloi [18], destacam a importância da análise da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP) para a identificação dos parâmetros mais adequados do modelo.

Além disso, Morettin e Toloi [18] ressaltam que, após o ajuste, é fundamental analisar os resíduos do modelo. Estes devem se comportar como ruído branco, o que pode ser avaliado por meio de testes estatísticos, como o Box-Pierce, bem como pela inspeção visual dos resíduos padronizados.

2.1.4 CRITÉRIOS DE SELEÇÃO

Os critérios de seleção AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion) foram propostos por Akaike em 1973 e 1978, respectivamente. Ambos são amplamente utilizados na escolha de modelos estatísticos, sendo que o modelo ideal é aquele que apresenta o menor valor para o critério de informação considerado [10].

CRITÉRIO DE INFORMAÇÃO DE AKAIKE (AIC)

O Critério de Informação de Akaike [3] é dado por:

$$AIC = -2\log L(q) + 2n, \quad (2.13)$$

onde $L(q)$ é a máxima verossimilhança e n é o número de parâmetros ajustados independentemente dentro do modelo.

CRITÉRIO DE INFORMAÇÃO BAYESIANO (BIC)

O Critério de Informação Bayesiano [18] é dado por:

$$BIC = -2\log L(q) + n\log(N), \quad (2.14)$$

onde $L(q)$ é a máxima verossimilhança, n é o número de parâmetros ajustados no modelo e N é o número de observações utilizadas na estimação do modelo.

2.1.5 METRÍCAS DE COMPARAÇÃO

O MAPE (Erro Absoluto Percentual Médio) e o RMSE (Raiz do Erro Quadrático Médio) são métricas complementares utilizadas para avaliar a qualidade das previsões em séries temporais. O MAPE expressa o erro médio em termos percentuais, permitindo interpretar o quão distante, proporcionalmente, as previsões estão dos valores reais. Já o RMSE mede o erro em unidades reais da variável analisada e pode ser interpretado como o desvio médio das previsões em relação aos valores observados.

ERRO ABSOLUTO PERCENTUAL MÉDIO (MAPE)

O Erro Absoluto Percentual Médio (MAPE) [14] é dado por:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - p_i}{y_i} \right|, \quad (2.15)$$

onde n é o número total de amostras, y_i é o valor real para a amostra i e o p_i é o valor previsto para a amostra i .

Um valor baixo de MAPE indica que as previsões estão próximas dos valores observados, ou seja, o modelo apresenta bom desempenho. Por outro lado, um valor elevado de MAPE revela que as previsões se distanciam significativamente dos dados reais, sinalizando a necessidade de aprimoramento do modelo [14].

RAIZ DO ERRO QUADRÁTICO MÉDIO (RMSE)

A Raiz do Erro Quadrático Médio [15] é dada por:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2}, \quad (2.16)$$

onde n é o número total de amostras, y_i é o valor real para a amostra i e o p_i é o valor previsto para a amostra i .

2.2 PREVISÃO DE DADOS AUSENTES

O impacto dos dados ausentes na análise estatística depende tanto da causa da ausência dos dados quanto da forma como o analista de dados tratara eles. Dessa maneira, o objetivo é definir critérios adequados para substituir essas lacunas por valores apropriados [20].

O método TDEM (Time-Dependence Effect Method), proposto por Nogarotto et al. [19], é utilizado quando há apenas uma série de dados disponível ao longo de vários anos. No entanto, como no estudo em questão há nove séries disponíveis, os cálculos foram realizados separadamente para cada uma delas, com o objetivo de estimar os valores ausentes por meio da equação 2.17.

$$\hat{x}_{masd} = \bar{x}_{.asd} + \frac{1}{2} \left(\bar{x}_{ma..} - \sum_{m=1}^M \frac{\bar{x}_{ma..}}{M} \right) + \frac{1}{2} \left(\bar{x}_{m.s.} - \sum_{m=1}^M \frac{\bar{x}_{m.s.}}{M} \right) + \frac{1}{2} \left(\bar{x}_{m..d} - \sum_{m=1}^M \frac{\bar{x}_{m..d}}{M} \right) \quad (2.17)$$

onde, $m = \text{mês}(1, 2, \dots, 12)$; $a = \text{ano}(2010, 2011, \dots, 2014)$; $s = \text{semana do mês}(1, 2, 3, 4)$ e $d = \text{dia da semana}(1, 2, \dots, 7)$. Os valores $\bar{x}_{.asd}$, $\bar{x}_{ma..}$, $\bar{x}_{ma..}$ e $\bar{x}_{m..d}$ são valores médios considerando os efeitos do tempo.

3. METODOLOGIA

Ao realizar esse estudo, foi utilizado o conjunto de dados da Série Histórica de Preços de Combustíveis disponível no portal de Dados Abertos do Governo Federal [12], os dados foram fornecidos pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP).

A ANP acompanhou os preços semanais praticados por revendedores de combustíveis automotivos e a partir desses dados foram extraídas as informações referentes aos combustíveis automotivos: Diesel S10, Etanol e Gasolina das bandeiras: Branca, Ipiranga e Raizen no período de julho de 2021 a dezembro de 2024.

Para o desenvolvimento desse trabalho foi analisado o preço médio semanal dos combustíveis de cada bandeira ao longo do período mencionado. No entanto, ao realizar a apuração dos dados foi constatado que no ano de 2022 não houve a coleta dos preços entre os dias 19/09/2022 e 09/10/2022 de nenhuma das bandeiras, não sendo possível calcular as médias das semanas 38, 39 e 40, o que pode ser visualizado na Tabela 3.1.

Tabela 3.1: Tabela de preços médios por semana com valores ausentes

Semana	Diesel S10			Etanol			Gasolina		
	Branca	Ipiranga	Raizen	Branca	Ipiranga	Raizen	Branca	Ipiranga	Raizen
36	6,81	6,95	6,96	3,18	3,21	3,22	4,77	4,80	4,79
37	6,74	7,00	6,96	3,00	3,04	3,05	4,63	4,70	4,69
38	-	-	-	-	-	-	-	-	-
39	-	-	-	-	-	-	-	-	-
40	-	-	-	-	-	-	-	-	-
41	6,51	6,59	6,49	3,33	3,39	3,35	4,73	4,77	4,79
42	6,56	6,61	6,49	3,36	3,43	3,35	4,72	4,77	4,79

Os dados foram importados para o software R Studio com o objetivo de preencher valores ausentes em colunas específicas, utilizando a abordagem sazonal (TDEM) [19], conforme apresentado na equação 2.17. Para isso, foram criadas colunas auxiliares e utilizada a função *function()* [13] para calcular as médias sazonais, aplicando uma fórmula de imputação nos dados ausentes, considerando o mês, a semana do mês e o dia da semana. Dessa forma, cada série passou a conter 182 observações, totalizando um conjunto com 9 séries no total.

Após o preenchimento dos valores ausentes, realizou-se uma análise descritiva das séries por meio do comando *summarise()* [29] do pacote *dplyr* [27] no R. Esse comando é utilizado para

criar resumos e calcular estatísticas sobre os dados, sendo combinado com a função *group_by()* [28] para realizar cálculos agregados por grupo. Além disso, foram elaborados gráficos com o pacote *ggplot2* [26], com o intuito de analisar a variação da média anual dos preços dos combustíveis ao longo do período. Também foram gerados gráficos de boxplots, permitindo a comparação visual da distribuição dos preços entre as diferentes bandeiras.

Antes de iniciar o ajuste do modelo estatístico, foi realizada uma análise comparativa entre as séries disponíveis, com o objetivo de avaliar o grau de variabilidade de cada uma. Optou-se por utilizar a série que apresentou variabilidade intermediária, buscando um equilíbrio entre estabilidade e sensibilidade às flutuações, de modo a garantir maior robustez e precisão nas estimativas do modelo. Para isso, foi desenvolvida uma *function()* [13] personalizada que calcula a média, o desvio padrão e o coeficiente de variação (CV), sendo o CV utilizado como medida de variabilidade ao longo dos anos.

Com a série definida, foram feitas verificações de algumas suposições, de modo a otimizar o ajuste do modelo no software e possibilitar a previsão dos dados para o ano seguinte.

Para verificar a presença de tendência na série temporal, aplicou-se o teste de Cox-Stuart, conforme descrito por Morettin e Toloi (2006) [18], utilizando o comando *cox.stuart.test()* do pacote *randtests* [9]. Esse teste parte da hipótese nula de ausência de tendência nos dados. Caso a hipótese nula seja rejeitada, considera-se que há tendência presente na série. Nessas situações, utiliza-se a diferenciação da série como método para remover a tendência.

Em seguida, para verificar a presença de sazonalidade, foi aplicado o teste de Fisher G, por meio do comando *fisher.g.test*, disponível no pacote *GeneCycle* [2]. Esse teste avalia a hipótese nula de que não há componente sazonal significativa na série. A rejeição da hipótese nula indica a presença de sazonalidade. Quando a sazonalidade é identificada, ajusta-se um modelo do tipo autorregressivo Integrado de Média Móvel Sazonal (SARIMA) [6], caso contrário, opta-se pelo ajuste de um modelo Autorregressivo Integrado de Média Móvel (ARIMA) [6].

Para realizar o ajuste, foi utilizado o comando *auto.arima()* [24], disponível no pacote *forecast* [16], uma das principais ferramentas do R para análise de séries temporais. Foram testados diferentes modelos com diferentes combinações de parâmetros, com o objetivo de identificar aquele que melhor se adequasse aos dados.

A escolha do modelo mais apropriado baseou-se no Critério de Informação de Akaike (AIC), no Critério de Informação Bayesiano (BIC) e nas métricas RMSE (Root Mean Square Error), utilizada para comparar diferentes modelos aplicados à mesma amostra, e MAPE (Mean Absolute Percentage Error), que avalia o desempenho preditivo do modelo, obtidas por meio da função *accuracy()* [17] do pacote *forecast*. Foi selecionado, portanto, o modelo que apresentou os menores valores nesses critérios e métricas.

Também foi realizada a análise dos resíduos após o ajuste do modelo. Espera-se, nesse contexto, que os resíduos se comportem como ruído branco, ou seja, que não apresentem tendência, sazonalidade, autocorrelação ou variação não constante. Esse comportamento indica que o modelo conseguiu capturar adequadamente os padrões presentes na série temporal. Assim, a análise dos resíduos constitui uma etapa essencial de validação, assegurando a qualidade

das inferências e previsões realizadas a partir do modelo.

Em seguida, a função *forecast()* foi aplicada ao modelo selecionado para gerar previsões dos primeiros seis meses de 2025 com intervalos de confiança que serão posteriormente comparados aos valores reais. Essa etapa reforça a utilidade do pacote *forecast* como uma ferramenta prática e eficiente para análises preditivas em séries temporais.

Para investigar a possibilidade de melhorar a acurácia das previsões, foram ajustados modelos com intervenções, considerando eventos externos ou mudanças estruturais que possam ter impactado os preços, como alterações na política de preços, tributos ou oscilações abruptas no mercado de combustíveis. O objetivo dessa abordagem foi avaliar se a inclusão de intervenções, que modelam efeitos pontuais e extraordinários, proporciona melhores resultados preditivos em comparação aos modelos padrão.

4. RESULTADOS

Considerando que, inicialmente, as séries apresentavam dados ausentes, conforme mostrado na Tabela 3.1, a aplicação da equação 2.17 permitiu estimar as médias de preço para as semanas 38, 39 e 40, cujos resultados estão apresentados na Tabela 4.1.

Tabela 4.1: Tabela de preços médios por semana com valores preenchidos

Semana	Diesel S10			Etanol			Gasolina		
	Branca	Ipiranga	Raízen	Branca	Ipiranga	Raízen	Branca	Ipiranga	Raízen
36	6,81	6,95	6,96	3,18	3,21	3,22	4,77	4,80	4,79
37	6,74	7,00	6,96	3,00	3,04	3,05	4,63	4,70	4,69
38	6,35	6,47	6,43	3,39	3,45	3,38	5,23	5,27	5,24
39	5,98	6,09	6,05	3,67	3,74	3,66	5,43	5,51	5,47
40	6,12	6,23	6,11	3,67	3,84	3,68	5,33	5,45	5,40
41	6,51	6,59	6,49	3,33	3,39	3,35	4,73	4,77	4,79
42	6,56	6,61	6,49	3,36	3,43	3,35	4,72	4,77	4,79

Em seguida, foi realizada uma análise descritiva dos preços médios dos combustíveis. Vale destacar que a média de 2021 contempla apenas os últimos seis meses do ano, refletindo um período mais curto, o que pode impactar a comparação com os anos subsequentes. Dessa forma, essa média pode não ser totalmente representativa de uma tendência anual, considerando que os preços normalmente apresentam variações ao longo do ano.

As estatísticas descritivas, apresentadas na Tabela 4.2, revelaram uma leve recuperação dos preços em 2023 e 2024, após a queda observada no ano de 2022. O preço médio do Diesel S10 subiu de R\$ 5,72 em 2023 para R\$ 5,85, em 2024, indicando um pequeno aumento, acompanhado de uma expressiva redução na volatilidade, já que o desvio-padrão caiu de 0,53 para 0,06, sugerindo uma estabilização dos preços. O Etanol também apresentou alta, passando de R\$ 3,50 para R\$ 3,72, com a mediana seguindo comportamento semelhante, o que indica uma distribuição relativamente equilibrada ao longo do período. A Gasolina, por sua vez, evoluiu de R\$ 5,17 para R\$ 5,72, mantendo um desvio-padrão estável de 0,21, o que demonstra que, apesar da elevação nos preços, não houve oscilações significativas.

Além disso, os valores mínimo e máximo também demonstram menor variação em 2024, sugerindo um mercado mais previsível. Em 2023, o Diesel S10 variou entre R\$ 4,78 e R\$ 6,37, enquanto em 2024 os valores ficaram entre R\$ 5,72 e R\$ 6,00. O Etanol apresentou comportamento semelhante, com um intervalo de R\$ 3,11 a R\$ 3,96 em 2023 e de R\$ 3,04 a R\$ 4,02 em 2024. Já a Gasolina mostrou uma leve redução na dispersão dos preços, variando

de R\$ 4,69 a R\$ 5,57 em 2023 e de R\$ 5,17 a R\$ 5,99 em 2024.

Tabela 4.2: Estatísticas descritivas dos preços médios dos combustíveis (2021–2024)

Combustível	Ano	Média (R\$)	Mediana (R\$)	Desvio-padrão	Mínimo	Máximo
Diesel S10	2021	5,05	4,96	0,31	4,67	5,46
Diesel S10	2022	6,61	6,68	0,62	5,37	7,66
Diesel S10	2023	5,72	5,87	0,53	4,78	6,37
Diesel S10	2024	5,85	5,84	0,06	5,72	6,00
Etanol	2021	4,72	4,72	0,39	4,10	5,34
Etanol	2022	4,22	4,25	0,68	3,03	5,49
Etanol	2023	3,50	3,44	0,22	3,11	3,96
Etanol	2024	3,72	3,82	0,25	3,04	4,02
Gasolina	2021	6,49	6,47	0,35	5,98	6,98
Gasolina	2022	6,10	6,49	1,10	4,67	7,57
Gasolina	2023	5,17	5,22	0,21	4,69	5,57
Gasolina	2024	5,72	5,80	0,21	5,17	5,99

Essas variações podem ser visualizadas na Figura 4.1, que mostra a evolução dos preços médios dos combustíveis entre 2021 e 2024, evidenciando a tendência de recuperação e maior estabilidade ao final do período analisado.

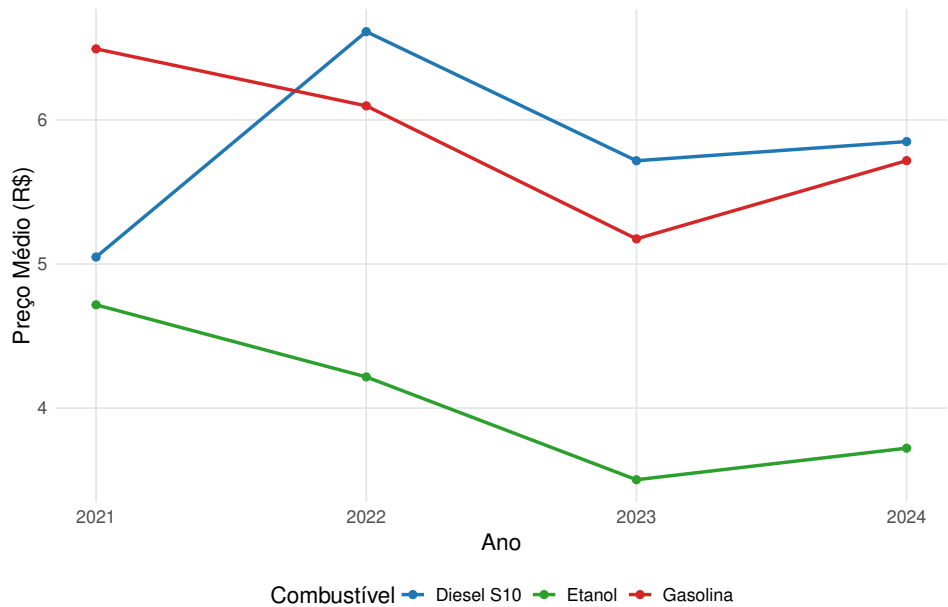


Figura 4.1: Evolução dos Preços Médios dos Combustíveis (2021-2024)

Fatores como a volatilidade do petróleo, ajustes na política de preços da Petrobras e a reoneração parcial de tributos influenciaram esse processo de recuperação. Apesar dos aumentos nos preços, a redução da volatilidade e a menor dispersão dos dados indicam um mercado mais estável, reflexo da recuperação econômica e de uma inflação mais controlada.

Além da análise temporal, foi realizada uma comparação visual, por meio de boxplots, entre os preços praticados pelas diferentes bandeiras, com o objetivo de compreender a influência da marca na formação dos preços, conforme pode ser observado na Figura 4.2.

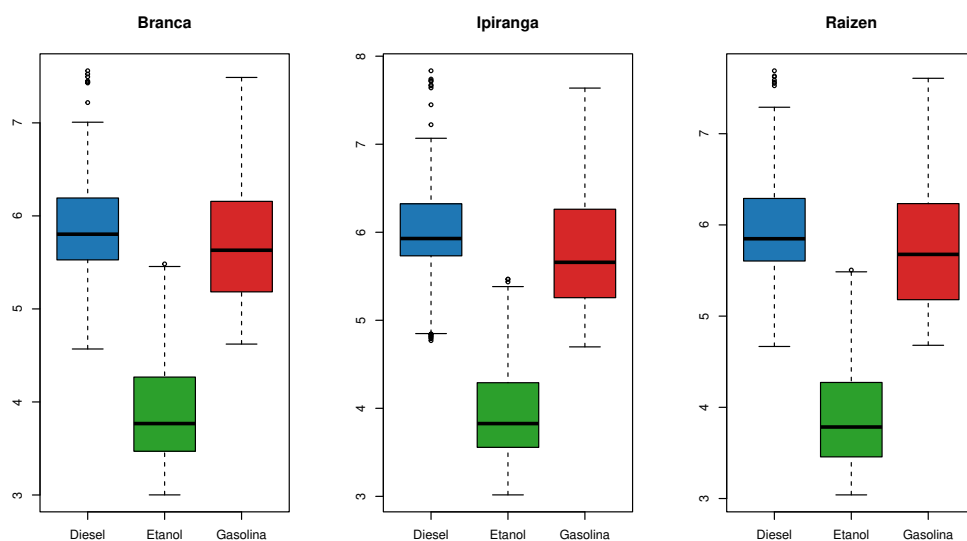


Figura 4.2: Distribuição dos Preços Médios dos Combustíveis por Bandeira

Os resultados apontam que a bandeira Branca apresenta, de modo geral, preços medianos mais baixos para Diesel, Etanol e Gasolina, embora com maior variabilidade nos valores. Por outro lado, as bandeiras Ipiranga e Raízen praticam preços ligeiramente mais elevados, mas com distribuições mais estáveis, o que sugere maior previsibilidade. Assim, a escolha da bandeira pode influenciar tanto o custo quanto a estabilidade dos preços enfrentados pelo consumidor.

Na Tabela 4.3, são apresentados os cálculos de média, desvio padrão e coeficiente de variação (CV) para os preços médios de cada combustível. Os resultados indicam que os preços do Diesel S10 da bandeira Ipiranga apresentaram a menor variação relativa com CV de 11,30%, seguida pela Branca com 11,72% e Raízen com 11,68%, reforçando a ideia de menor variabilidade para esta categoria. O Etanol indicou a maior variabilidade com relação aos demais combustíveis, a Ipiranga apresentou a menor variação dessa categoria com CV de 14,87%, seguida pela Raízen com 15,41% e Branca com 15,51%. Enquanto a Gasolina apresentou a variabilidade intermediária com o menor CV de 13,18% na bandeira Branca, seguida pela Ipiranga com 13,27% e Raízen 13,46%. No geral, as médias e os desvios padrão foram próximos entre as bandeiras, sugerindo uma certa uniformidade nos preços praticados.

Tabela 4.3: Tabela de variabilidade com médias, desvios padrão (Sd) e coeficiente de variação (CV)

	Diesel S10			Etanol			Gasolina		
	Branca	Ipiranga	Raízen	Branca	Ipiranga	Raízen	Branca	Ipiranga	Raízen
Media	5,84	6,00	5,91	3,91	3,98	3,93	5,74	5,82	5,79
Sd	0,68	0,68	0,69	0,61	0,59	0,61	0,76	0,77	0,78
CV (%)	11,72	11,30	11,68	15,51	14,87	15,41	13,18	13,27	13,46

Considerando que a Gasolina da bandeira Ipiranga apresentou variabilidade intermediária, optou-se por analisar a série a fim de realizar os ajustes necessários e gerar as previsões. A Figura 4.3 apresenta o comportamento da média dos preços da Gasolina da bandeira Ipiranga, no município de Uberlândia, de julho de 2021 a dezembro de 2024. Observa-se uma leve tendência de crescimento até meados de 2022, seguida por uma queda acentuada no segundo semestre, possivelmente decorrente da redução do ICMS sobre os combustíveis. Apesar dessas variações, não se identifica um padrão sazonal claramente definido ao longo do período, indicando que variações pontuais podem ter sido influenciadas por políticas tributárias e fatores específicos do mercado.

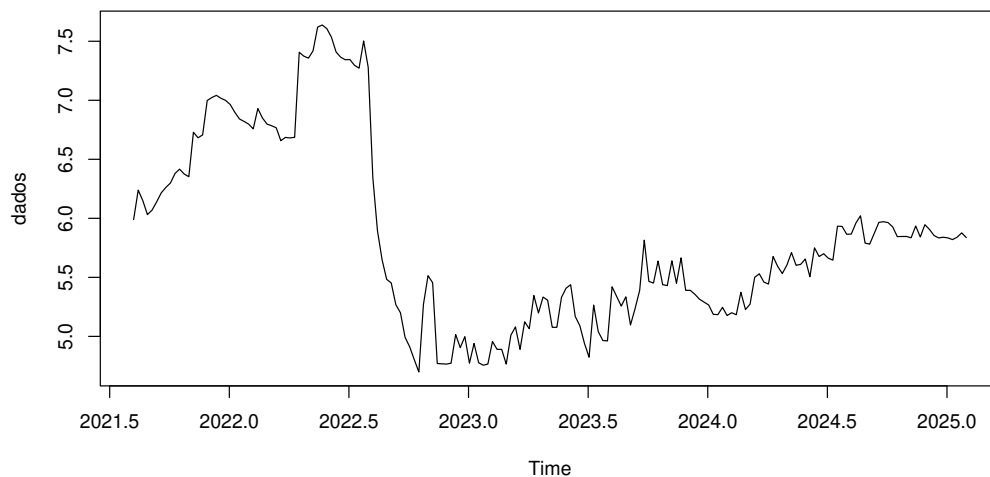


Figura 4.3: Preços médios do combustível gasolina da bandeira Ipiranga de julho de 2021 a dezembro de 2024 em Uberlândia.

Por meio da função de autocorrelação (ACF), é possível analisar esse padrão. A Figura 4.4 mostra um decaimento lento ao longo dos lags, com valores iniciais significativamente positivos e convergência para zero. Esse comportamento sugere que a série possivelmente não é estacionária, indicando a necessidade de diferenciação antes da modelagem, caso o teste de Cox-Stuart aponte para rejeição da hipótese nula.

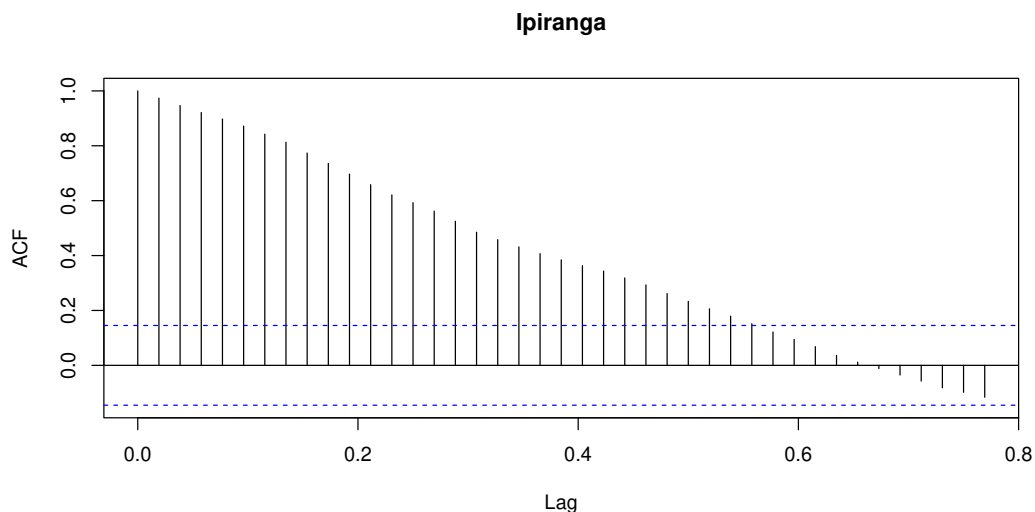


Figura 4.4: Função de Autocorrelação (ACF)

Na Figura 4.5, que apresenta a função de autocorrelação parcial (PACF), não se observa evidência clara de sazonalidade, uma vez que não há um padrão repetitivo de picos significativos em defasagens múltiplas (como 12, 24, etc., no caso de dados mensais). A ausência desses picos sugere que, se houver alguma sazonalidade, ela não é predominante na estrutura da série.

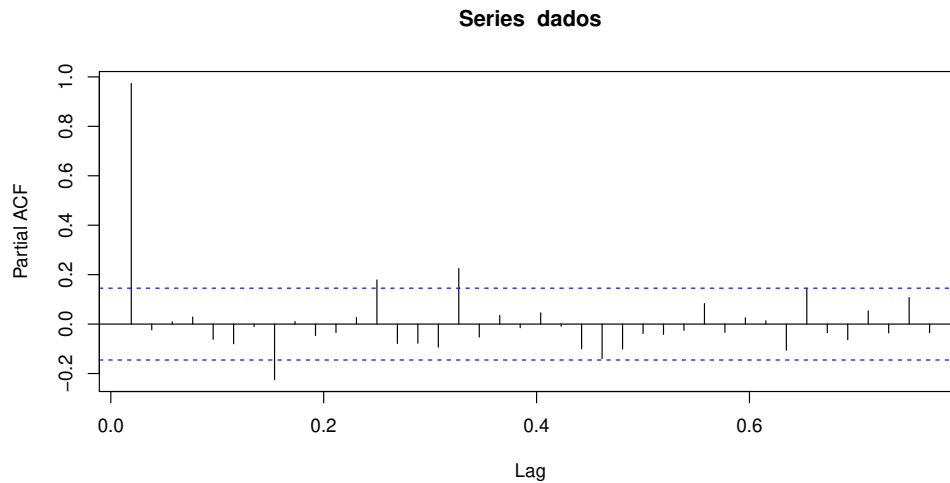


Figura 4.5: Função de Autocorrelação Parcial (PACF)

Outra análise complementar consiste em observar a dispersão da média em função do desvio padrão. Para isso, a série original foi dividida em pares de observações e, para cada grupo, foram calculados a média e o desvio padrão. Os resultados estão representados na Figura 4.6, onde se observa que o desvio padrão não varia em função da média. Isso indica um modelo aditivo, no qual não há necessidade de aplicar transformações nos dados para estabilizar a variância.

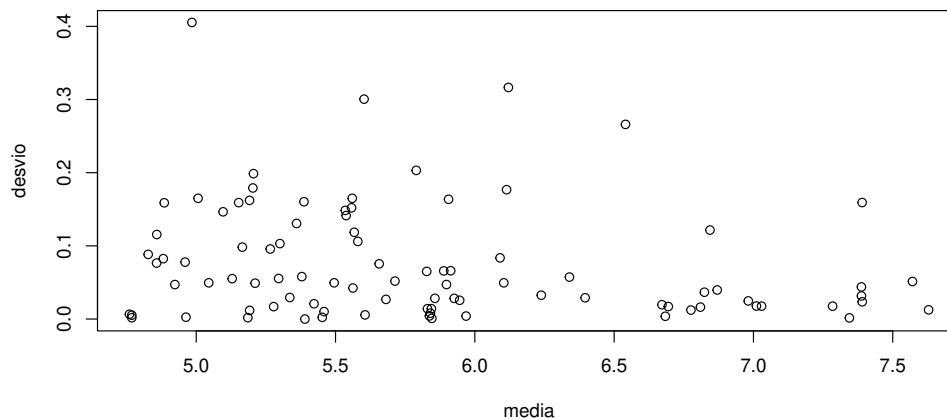


Figura 4.6: Média em função do desvio padrão

Como não foi identificado a necessidade de transformar os dados, prosseguiu-se com o teste de Cox-Stuart, cujo resultado apresentou um p-valor de 0.05857, ligeiramente superior ao nível de significância de 5%, levando à não rejeição da hipótese nula. Dessa forma, não há evidência estatisticamente significativa de tendência na série. Contudo, é importante considerar que a ausência de tendência aparente pode ser influenciada pelo curto período analisado e pela queda acentuada observada no segundo semestre de 2022, que possivelmente reduziu a significância do teste, refletindo-se no valor-p próximo ao limiar crítico.

Após a verificação da ausência da componente tendência, aplicou-se o teste de Fisher para investigar a presença de sazonalidade na série. O teste resultou em um valor de $p < 0.01$, inferior a 5%, o que levou a rejeição da hipótese nula. Dessa forma, conclui-se que há presença de sazonalidade, permitindo prosseguir com o ajuste do modelo SARIMA, incorporando componentes sazonais.

Diante disso, ao utilizar o comando *auto.arima()*, foi inicialmente sugerido o modelo $\text{ARIMA}(1,1,1)(1,0,0)_{[52]}$ como o mais adequado para o ajuste. No entanto, considerando que os testes indicaram ausência de tendência e presença de sazonalidade, esse modelo não se mostrou apropriado, pois incorpora componentes que não se justificam estatisticamente. Assim, com o objetivo de identificar o modelo mais compatível com a estrutura da série, foram ajustados outros cinco modelos, cujos parâmetros e critérios de avaliação estão apresentados na Tabela 4.4.

Tabela 4.4: Modelos propostos e resultados

Modelo	Parâmetros	AIC	BIC
Modelo 1	$(1,0,0)(0,1,0)_{[52]}$	48,61	54,3405
Modelo 2	$(1,0,0)(1,1,0)_{[52]}$	14,51	23,1111
Modelo 3	$(1,0,0)(0,1,1)_{[52]}$	6,8	15,4046
Modelo 4	$(1,0,0)(2,1,0)_{[52]}$	-12,65	-1,1816
Modelo 5	$(0,0,6)(1,1,0)_{[52]}$	72,52	95,4607

Com a identificação dos modelos com parâmetros estatisticamente significativos, foram selecionados os três que apresentaram os menores valores nos critérios AIC e BIC — especificamente, os modelos 2, 3 e 4. Para definir, entre eles, aquele com melhor desempenho preditivo, foram utilizadas as métricas RMSE (Root Mean Square Error) e MAPE (Mean Absolute Percentage Error), cujos resultados estão apresentados na Tabela 4.5. Essas medidas permitiram avaliar a acurácia das previsões de forma mais robusta.

Tabela 4.5: Métricas RMSE e MAPE para os modelos selecionados

Modelo	Parâmetros	RMSE	MAPE
Modelo 2	$(1,0,0)(1,1,0)_{[52]}$	0,1910	2,2102
Modelo 3	$(1,0,0)(0,1,1)_{[52]}$	0,1585	1,8277
Modelo 4	$(1,0,0)(2,1,0)_{[52]}$	0,0959	1,1241

Com isso, o Modelo 4 apresentou os menores valores tanto nos critérios de informação (AIC e BIC) quanto nas métricas de avaliação preditiva (RMSE e MAPE), sendo assim, considerado o mais adequado para representar a série.

Em seguida, realizou-se a análise dos resíduos do modelo ajustado. Foram examinadas graficamente as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) dos resíduos, conforme apresentado na Figura 4.7. Observa-se que, em ambas, no máximo dois lags apresentam significância, o que indica ausência de correlação relevante entre os resíduos.

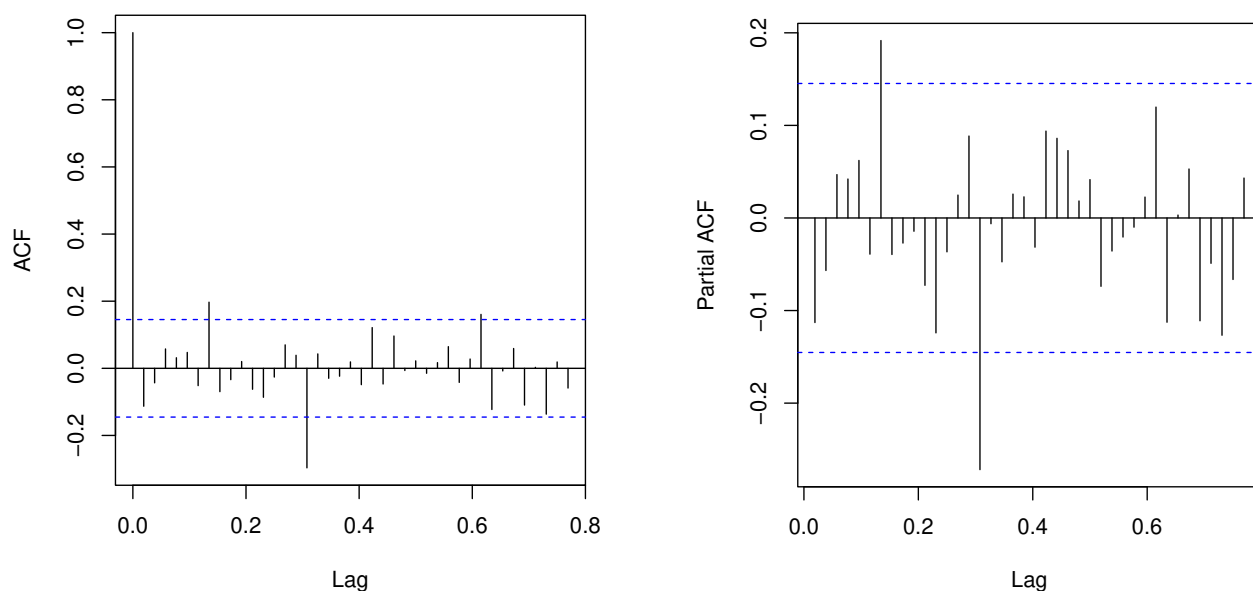


Figura 4.7: Funções ACF e PACF dos resíduos

Para confirmar esse diagnóstico, aplicou-se o teste de Box-Pierce, cuja hipótese nula considera os resíduos como ruído branco. O teste resultou em um p-valor de 0,1593, valor superior a 5%, o que levou à não rejeição da hipótese nula. Dessa forma, conclui-se que os resíduos apresentam comportamento compatível com ruído branco. De maneira complementar, foi plotado o periodograma acumulado dos resíduos exibido na Figura 4.8, reforçando visualmente esse resultado.

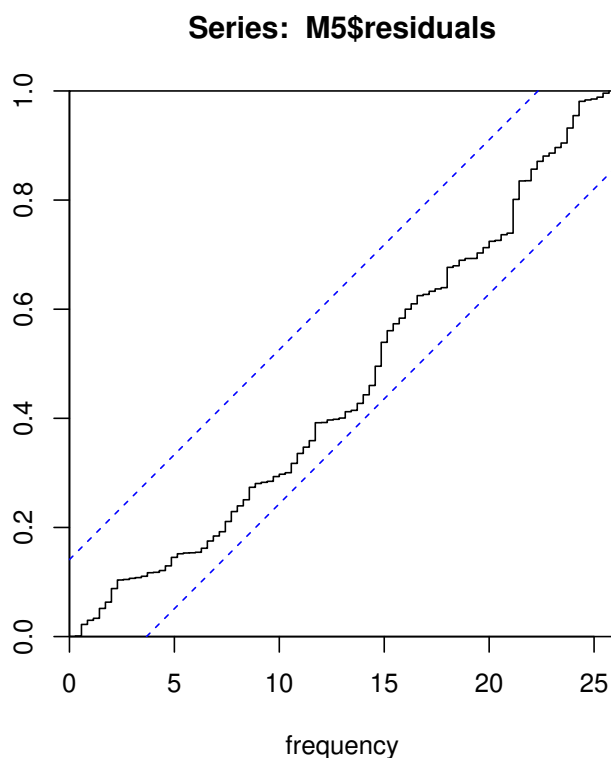


Figura 4.8: Periodograma acumulado

Por fim, aplicou-se a função *forecast()*, que gerou a previsão para as semanas dos primeiros seis meses de 2025, conforme apresentado na Figura 4.9.

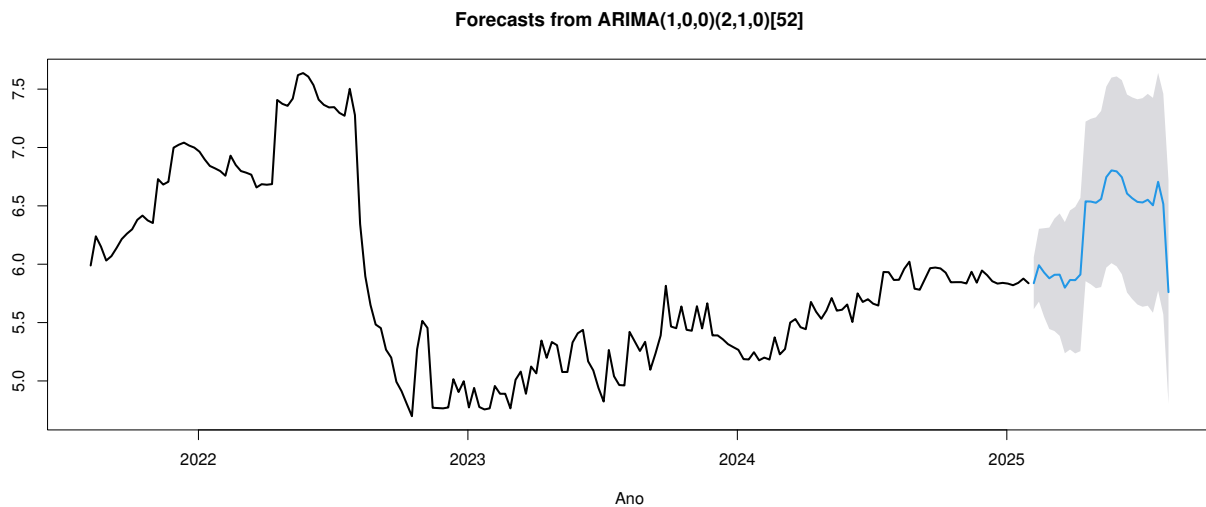


Figura 4.9: Previsão dos seis primeiros meses de 2025 para o modelo 4

Os valores previstos, acompanhados de seus respectivos intervalos de confiança, foram comparados aos valores reais, permitindo avaliar a acurácia do modelo ajustado a partir das diferenças observadas.

A Tabela 4.6 apresenta os valores previstos pelo modelo para as primeiras 27 semanas de 2025, acompanhados de seus respectivos intervalos de confiança de 95%, em comparação com os valores reais observados, bem como as diferenças entre previsões e observações. Essa análise permite avaliar a acurácia do modelo ajustado e a confiabilidade de suas previsões.

Observa-se que, nas dez primeiras semanas, as diferenças entre os valores previstos e os valores reais foram relativamente pequenas, variando entre -0,14 e 0,41, indicando boa capacidade preditiva do modelo nesse período. Além disso, os preços reais apresentaram média de 6,03, desvio-padrão de 0,15 e amplitude de 0,47, indicando maior nível de preços e dispersão em comparação com o restante do período.

A partir da 11^a semana, em meados de março de 2025, nota-se um aumento sistemático das diferenças, com predominância de valores negativos, atingindo -0,87 na 25^a semana. Nesse intervalo, a média dos preços reais caiu para 5,98, com menor variabilidade (desvio-padrão de 0,11 e amplitude de 0,40). Essa estabilidade, em patamar ligeiramente inferior, fez com que o modelo superestimasse os valores, possivelmente devido à queda brusca observada em 2022 e ao curto período de calibração, que influenciaram sua sensibilidade.

Já na 27^a semana, ao final de junho, observa-se uma recuperação na precisão do modelo, com diferença de apenas 0,14 em relação ao valor real. O preço efetivo de 5,90 superou levemente o previsto de 5,76, confirmando a tendência de estabilização em torno de um novo patamar e indicando retomada da capacidade preditiva. Dessa forma, as discrepâncias observadas refletem tanto ajustes do mercado após choques anteriores quanto as limitações do curto período de análise, sem comprometer a confiabilidade geral do modelo.

Tabela 4.6: Previsões pontuais e intervalos de 95% comparados com os valores reais e suas diferenças

Semana	Valores Previstos	Lo 95	Hi 95	Valores Reais	Diferenças
1	5,83	5,61	6,05	5,89	0,06
2	5,99	5,67	6,30	5,85	-0,14
3	5,92	5,54	6,30	6,22	0,30
4	5,87	5,44	6,31	5,95	0,08
5	5,90	5,42	6,39	5,98	0,08
6	5,91	5,38	6,43	6,32	0,41
7	5,79	5,23	6,36	6,13	0,34
8	5,86	5,26	6,46	6,01	0,15
9	5,86	5,23	6,49	5,96	0,09
10	5,91	5,25	6,56	5,98	0,07
11	6,53	5,85	7,22	5,95	-0,58
12	6,53	5,82	7,24	5,92	-0,61
13	6,52	5,79	7,25	5,98	-0,54
14	6,55	5,80	7,31	6,00	-0,56
15	6,74	5,97	7,52	5,93	-0,81
16	6,80	6,00	7,59	5,94	-0,86
17	6,79	5,98	7,60	6,00	-0,79
18	6,74	5,91	7,57	5,99	-0,75
19	6,60	5,75	7,45	5,99	-0,61
20	6,56	5,70	7,42	6,22	-0,34
21	6,53	5,65	7,41	6,15	-0,38
22	6,52	5,63	7,42	6,15	-0,37
23	6,55	5,64	7,45	5,97	-0,59
24	6,50	5,58	7,42	5,88	-0,62
25	6,70	5,77	7,63	5,83	-0,87
26	6,51	5,57	7,46	5,82	-0,69
27	5,76	4,80	6,71	5,90	0,14

Além disso, ao ajustar modelos com intervenção, os resultados obtidos apresentaram desempenho muito semelhante ao do Modelo 4 ajustado sem intervenção, tanto em termos de ajuste quanto de capacidade preditiva. Diante dessa equivalência, optou-se por apresentar e detalhar os resultados dos modelos sem intervenção, uma vez que fornecem explicações claras, simplicidade interpretativa e mantém a robustez das previsões, sem comprometer a fidelidade à série histórica observada.

De forma comparativa, Almeida (2010) [4] verificou que modelos SARIMA ajustaram-se de maneira satisfatória às séries de vendas de gasolina e diesel no Rio Grande do Sul, evidenciando a aplicabilidade desta abordagem em combustíveis regionais. Por outro lado, Reis (2024) [21] observou que, embora o modelo SARIMA tenha se ajustado bem às séries de diesel (com biodiesel), apresentou desempenho inferior para o etanol hidratado, reforçando a ideia de que

cada série temporal deve ser analisada individualmente para a escolha do modelo mais adequado.

Dessa forma, os resultados deste estudo indicam que o ajuste foi consistente e robusto para a série da Gasolina da bandeira Ipiranga, mas também destacam que a adequação de modelos de previsão em séries de combustíveis pode variar conforme as características de cada produto e contexto local.

5. CONCLUSÕES

Este trabalho analisou a série temporal dos preços médios dos combustíveis Diesel S10, Etanol e Gasolina em Uberlândia, no período de julho de 2021 a dezembro de 2024, com previsão para a Gasolina em 2025. Após o tratamento de dados ausentes, observou-se uma recuperação dos preços em 2023 e 2024, acompanhada de maior estabilidade, especialmente para o Diesel S10.

A comparação entre bandeiras revelou que a Branca apresenta preços médios mais baixos, porém com maior variabilidade, enquanto Ipiranga e Raízen exibem preços ligeiramente superiores, porém mais estáveis. A análise da série temporal da Gasolina da bandeira Ipiranga revelou ausência de tendência, mas presença de sazonalidade, justificando a adoção do modelo SARIMA para capturar adequadamente os padrões periódicos observados.

O modelo SARIMA $(1,0,0)(2,1,0)_{[52]}$ apresentou o melhor ajuste e capacidade preditiva, com resíduos compatíveis com ruído branco. As previsões para os primeiros meses de 2025 mostraram boa aderência aos valores reais, evidenciando a eficácia do modelo em antecipar flutuações da série e permitindo a avaliação da variabilidade semanal dos preços.

Dessa forma, os resultados do estudo atendem aos objetivos propostos, fornecendo uma análise detalhada da evolução dos preços médios semanais dos combustíveis, possibilitando a previsão do comportamento da Gasolina e oferecendo subsídios para que consumidores, empresas e formuladores de políticas públicas possam tomar decisões mais informadas e estratégicas frente às oscilações do mercado. Além disso, o trabalho reforça a importância do uso de modelos SARIMA em séries temporais com padrões sazonais, contribuindo para o avanço da análise preditiva de preços de combustíveis em contextos regionais.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] A. Brasil, M. da A. P. e: *Plano Nacional de Agroenergia 2006-2011*. Embrapa Informação Tecnológica, 2ª ed., 2006.
- [2] Ahdesmaki, M., Fokianos, K. e Strimmer, K.: *GeneCycle: Identification of Periodically Expressed Genes*, 2025. <https://CRAN.R-project.org/package=GeneCycle>, acessado em 13/04/2025.
- [3] Akaike, H.: *A New Look at the Statistical Model Identification*. Em *IEEE Transactions On Automatic Control*, vol. AC-19, pp. 716–723, December 1974. <https://pdfs.semanticscholar.org/ad58/58a2bc9ac4f7602a586c274482ef137751e5.pdf>.
- [4] Almeida, S.G. de: *Análise do volume de vendas de óleo diesel e gasolina no RS, 2000 a 2010*. Monografia de Especialização, Centro de Ciências Naturais e Exatas, 2010.
- [5] Araujo, S.J. de, Aredes, A.F. de e Santos, V.F. dos: *Previsão De Preços Do Boi Gordo Com Modelos Arima E Sarima*. Revista De Economia Da UEG, 8(2):27–44, 2012. <https://www.revista.ueg.br/index.php/economia/article/view/440>.
- [6] Box, G. E. P. e Jenkins, G.: *Time Series Analysis: Forecasting and Control*. Holden-Day, 2ª ed., 1970, ISBN 978-08-162-1094-7.
- [7] Box, G. E. P., Jenkins, G. e Ljung, M. G.: *Time Series Analysis: Forecasting and Control*. Holden-Day, 5ª ed., 2015, ISBN 978-11-186-7502-1.
- [8] Brito, F.B.S. de: *Etanol no Brasil: uma abordagem preditiva de preços*. Dissertação de Mestrado, Centro Universitário SENAI CIMATEC, 2019.
- [9] Caeiro, F. e Mateus, A.: *Testing Randomness in R*, 2024. <https://CRAN.R-project.org/package=randtests>, acessado em 11/04/2025.
- [10] Chechi, L. e Bayer, F. M.: *Modelos univariados de séries temporais para previsão das temperaturas médias mensais de Erechim, RS*. Revista Brasileira de Engenharia Agrícola e Ambiental, 16(12):1321–1329, 2012. <https://doi.org/10.1590/S1415-43662012001200009>.
- [11] Educação, M.: *Combustíveis*, s.a. <https://mundoeducacao.uol.com.br/quimica/combustiveis.htm>, acessado em 08/02/2025.

- [12] Federal, G.: *Série Histórica de Preços de Combustíveis e de GLP*, 2018. <https://dados.gov.br/dados/conjuntos-dados/serie-historica-de-precos-de-combustiveis-e-de-glp>, acessado em 16/12/2024.
- [13] Filgueira, L.: *Criando funções no R*, 2016. <http://www.estadisticacomr.uff.br/?p=224>, acessado em 17/03/2025.
- [14] Filho, M.: *MAPE (Erro Absoluto Percentual Médio) em Machine Learning*, 2023. <https://mariofilho.com/mape-erro-absoluto-percentual-medio-em-machine-learning/>, acessado em 18/07/2025.
- [15] Filho, M.: *RMSE (Raiz Do Erro Quadrático Médio) Em Machine Learning*, 2023. <https://mariofilho.com/rmse-raiz-do-erro-quadratico-medio-em-machine-learning/>, acessado em 18/07/2025.
- [16] Hyndman, R., Athanasopoulos, G. et al.: *forecast: Forecasting functions for time series and linear models*, 2025. <https://pkg.robjhyndman.com/forecast/>, R package version 8.24.0.9000.
- [17] Hyndman, R. J. et al.: *Package ‘forecast’*, 2025. <https://cran.r-project.org/web/packages/forecast/forecast.pdf>, R package version 8.24.0.
- [18] Morettin, P. A. e Toloi, C. M. C.: *Análise de séries temporais*. Egard Blucher, 2ª ed., 2006, ISBN 978-85-212-0389-6.
- [19] Nogarotto, D. C., Rissi, N. M. e Pozza, S. A.: *Um novo método de preenchimento de dados faltantes aplicado a séries temporais de concentração de MP10*. Revista Tecnologia e Sociedade, 15(37):275–296, 2019. <https://periodicos.utfpr.edu.br/rts/article/view/8594>.
- [20] Plaia, A. e Bondi, A. L.: *Single imputation method of missing values in environmental pollution data sets*. Em *Atmospheric Environment*, vol. 40, pp. 7316–7330. ScienceDirect, 2006. <https://doi.org/10.1016/j.atmosenv.2006.06.040>.
- [21] Reis, L. L. C.: *Mercado de biodiesel e etanol no Brasil : uma abordagem de ciência de dados*. Monografia de Graduação, Faculdade de Ciências Econômicas, 2024.
- [22] Reis, M. M.: *INE 7001 Análise de Séries Temporais*. <https://www.inf.ufsc.br/~marcelo.menezes.reis/INE7001.html>, notas de aula utilizados na disciplina INE 7001 Estatística Para Administradores I, 2015.
- [23] Ribeiro, D.: *Combustíveis fósseis*. Revista de Ciência Elementar, 2(2):172, 2014. <http://doi.org/10.24927/rce2014.172>.
- [24] Rosse, V. P.: *Ajuste do ARIMA para análise e previsões de séries temporais*, 2020. https://rpubs.com/Possato/ARIMA_ajuste_tutorial, acessado em 13/04/2025.

-
- [25] Silva, P.R.F. da e Freitas, T.F.S. de: *Biodiesel: o ônus e o bônus de produzir combustível*. Ciência Rural, 38(3):843–851, 2008. <https://doi.org/10.1590/S0103-84782008000300044>.
- [26] Wickham, H.: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016, ISBN 978-3-319-24277-4. <https://ggplot2.tidyverse.org>.
- [27] Wickham, H., François, R. et al.: *dplyr: A Grammar of Data Manipulation*, 2023. <https://dplyr.tidyverse.org>, R package version 1.1.4, <https://github.com/tidyverse/dplyr>.
- [28] Wickham, H., François, R. et al.: *Group by one or more variables*, 2025. https://dplyr.tidyverse.org/reference/group_by.html, acessado em 11/04/2025.
- [29] Wickham, H., François, R. et al.: *Summarise each group down to one row*, 2025. <https://dplyr.tidyverse.org/reference/summarise.html>, acessado em 17/03/2025.