

ANÁLISE EXPLORATÓRIA DOS PREÇOS DOS COMBUSTÍVEIS NO BRASIL

Dennis de Sousa Farias¹

Jacqueline Akina Nakagawa¹

Maurício Duarte²

João Ricardo Favani²

RESUMO

Este estudo analisa a evolução dos preços dos combustíveis no Brasil — gasolina comum, etanol hidratado, óleo diesel e gás de cozinha (GLP) — entre 2001 e 2023, com base nos dados abertos da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). Foram aplicadas técnicas de estatística descritiva, correlação de Spearman e regressão linear simples, aliadas a recursos de visualização de dados com a linguagem R. O objetivo é identificar variações regionais e temporais, bem como relações entre os diferentes combustíveis. Os resultados revelam disparidades relevantes entre as regiões do país, com a região Norte apresentando os maiores preços medianos, atribuídos a fatores logísticos. A análise temporal mostra tendência de alta nos preços, especialmente após 2020, refletindo eventos macroeconômicos como a pandemia e a valorização do petróleo no mercado internacional. Observou-se uma correlação positiva e forte entre gasolina e etanol ($r = 0,97$), e entre gasolina e GLP, com a regressão linear indicando que 93% da variação no preço do GLP pode ser explicada pelo da gasolina ($R^2 = 0,93$). A pesquisa destaca a interdependência entre os combustíveis e a importância de ferramentas estatísticas para interpretar o comportamento do setor energético. Os achados oferecem subsídios para políticas públicas mais eficazes, apoio a decisões estratégicas no setor e ampliam a compreensão acadêmica sobre a dinâmica dos preços de combustíveis no Brasil.

Palavras-chave: Preços dos combustíveis. Estatística descritiva. Visualização de dados.

1 INTRODUÇÃO

O mercado de combustíveis no Brasil apresenta significativas variações de preços influenciadas por múltiplos fatores, incluindo políticas governamentais, flutuações no mercado internacional de petróleo, custos logísticos e condições

¹ Discentes em Big Data no Agronegócio na FATEC Pompeia, Pompeia-SP,

² Docentes do curso Big Data no Agronegócio, FATEC Pompeia, Pompeia-SP.

agrícolas (especialmente para o etanol). Tais oscilações têm impacto direto na economia nacional e no cotidiano da população, destacando a importância de análises robustas que identifiquem padrões e tendências neste setor estratégico.

A investigação das dinâmicas de preços nos combustíveis brasileiros demanda uma abordagem multidimensional capaz de integrar padrões temporais, disparidades geográficas e interações setoriais. Como demonstra a ANP (2023) em suas séries históricas, "a volatilidade dos preços apresenta comportamentos distintos entre as regiões do país, com variações que superam 15% em períodos de crise internacional, enquanto a sazonalidade do etanol segue ciclos agrícolas bem definidos" (ANP, 2023, p. 27). Essa complexidade sistêmica justifica a aplicação conjunta de técnicas estatísticas descritivas e inferenciais propostas neste estudo, que permitirão não apenas mapear tendências históricas (2001-2023), mas também testar hipóteses sobre a interdependência entre os mercados de gasolina, etanol, diesel e GLP - particularmente relevante num cenário de transição energética.

Dessa forma, este estudo tem como objetivo a investigação da evolução dos preços dos principais combustíveis no Brasil — gasolina comum, etanol hidratado, óleo diesel e gás liquefeito de petróleo (GLP), também conhecido como gás de cozinha — entre 2001 e 2023, com foco em três dimensões: variações temporais, diferenças regionais e correlações entre os tipos de combustíveis. Para isso, emprega técnicas de estatística descritiva e inferencial, aliadas a recursos de visualização de dados, a fim de oferecer uma análise abrangente da dinâmica de preços no setor energético.

A estrutura do trabalho segue uma progressão lógica: após esta introdução contextualizadora, o capítulo de Material e Métodos descreve detalhadamente as técnicas estatísticas e as ferramentas utilizadas, bem como a base de dados, fornecendo o embasamento necessário para a compreensão das análises realizadas. A seção de Resultados e Discussão expõe os principais achados, incluindo as análises regionais, temporais e correlacionais, além da interpretação crítica dos resultados, contextualizando-os com estudos anteriores e destacando suas implicações práticas para o setor. Após isso, as considerações finais, apresentando também ideias para aplicação futura e continuidade da pesquisa.

Esta organização facilita a compreensão tanto dos aspectos metodológicos quanto das conclusões da pesquisa.

Os resultados desta pesquisa podem contribuir de forma relevante para diferentes públicos: oferecendo subsídios para a formulação de políticas públicas, apoiando empresas do setor energético em suas decisões estratégicas e ampliando a base de conhecimento disponível para estudos acadêmicos. A utilização de técnicas de Big Data e análise exploratória de dados evidencia o potencial dessas abordagens para gerar insights sobre a dinâmica dos preços de combustíveis no Brasil, promovendo uma compreensão mais clara e fundamentada desse mercado essencial.

2 MATERIAL E MÉTODOS

Para assegurar a consistência e a confiabilidade dos resultados obtidos, é fundamental que a pesquisa esteja embasada em uma sólida fundamentação teórica e em procedimentos metodológicos bem definidos. A teoria oferece suporte conceitual às escolhas realizadas, enquanto a metodologia descreve de forma clara o caminho seguido na análise dos dados. A seguir, apresentam-se os principais fundamentos e métodos que nortearam a condução deste estudo.

2.1 Fundamentação Teórica

Com o crescente volume de informações e a consolidação dos dados como um dos ativos mais valiosos da atualidade, torna-se imprescindível a realização de estudos robustos, que contemplem dados atualizados, visualizações eficazes e análises integradas. A capacidade de correlacionar essas informações é essencial para garantir maior precisão, eficácia e profundidade analítica, possibilitando previsões e interpretações acerca de uma ampla gama de temas.

Nesse cenário, destaca-se o conceito de Big Data, frequentemente representado pelos chamados “Vs”: volume, variedade, velocidade, veracidade e valor. Conforme cita Loh (2019), o volume refere-se à quantidade massiva de dados gerados continuamente; a variedade diz respeito aos diversos formatos como textos, números, imagens e vídeos; a velocidade trata da rapidez com que os dados são produzidos e processados; a veracidade está relacionada à qualidade e confiabilidade das informações; e o valor diz respeito à utilidade extraída desses dados para tomada de decisão. Esses atributos impõem novos desafios e exigem abordagens analíticas sofisticadas, como as utilizadas neste estudo, para que os dados deixem de ser apenas acumulados e passem a gerar conhecimento relevante.

Nesse contexto, foi utilizado o conjunto de dados abertos da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), intitulado "Combustíveis Brasil", um dataset que abrange informações coletadas entre os anos de 2001 e 2023. As variáveis contemplam os preços médios da gasolina comum, etanol hidratado, óleo diesel e GLP. Para fins desta pesquisa, foi considerada a variável "região", presente na base de dados, com o objetivo de possibilitar análises regionais.

Dessa forma, para que seja apresentada uma análise mais assertiva, é necessário o entendimento de alguns conceitos que são imprescindíveis na estatística descritiva e, de acordo com Sassi (2020), “podemos afirmar que existe duas maneiras para chegarmos a conclusões: usando inferência dedutiva e usando inferência indutiva” (Sassi, 2020, p. 3). A inferência dedutiva é aquela baseada a partir de premissas. A inferência indutiva é aquela que generaliza dados específicos para toda uma população, que é a utilizada neste trabalho.

Segundo Silvestre (2007), a estatística consiste em um conjunto de métodos apropriados para coletar, explorar, descrever e interpretar dados numéricos. Portanto, trata-se de uma ferramenta essencial para a condução de análises coerentes e alinhadas ao propósito do estudo.

Complementando a pesquisa, de acordo com Oliveira e Oliveira (2011), as medidas de dispersão ajudam a entender a variação dos dados em uma distribuição. Elas complementam as medidas de tendência central, fornecendo uma

visão mais completa sobre como os valores se espalham, ou seja, apesar da moda, média, e mediana serem úteis, estas medidas podem ser insuficientes para representar de forma fidedigna todos os valores de uma variável quantitativa.

A correlação é uma técnica estatística que avalia a relação entre duas variáveis, permitindo identificar se essa associação é linear (positiva ou negativa, quando as variáveis variam na mesma direção ou em sentidos opostos) ou não linear (seguindo padrões mais complexos), além de medir a intensidade dessa relação. Segundo Fröhlich et al. (2016), essa análise fornece informações cruciais sobre o comportamento conjunto das variáveis, possibilitando inferências sobre a direção (crescente ou decrescente) e o tipo de dependência entre elas, o que é fundamental para interpretar adequadamente a natureza dessa interdependência.

A regressão linear simples, assim como a correlação, baseia-se na análise de conjuntos de dados amostrais. De acordo com Queiroz (2020), esse modelo fornece uma equação que descreve o comportamento de uma variável em função da outra, sendo representado graficamente por uma reta de regressão que melhor se ajusta aos pares de dados observados.

Portanto, a análise de correlação e regressão linear simples são ferramentas estatísticas complementares que permitem explorar a relação entre variáveis de maneira robusta. Enquanto a correlação identifica a direção e a força da associação entre duas variáveis, a regressão linear vai além, modelando matematicamente essa relação por meio de uma equação que descreve como uma variável influencia a outra. Essa abordagem não só quantifica a interdependência, como também possibilita previsões e inferências sobre o comportamento dos dados, fundamentais para embasar decisões em pesquisas empíricas. Dessa forma, a integração dessas técnicas amplia a compreensão da dinâmica entre as variáveis, garantindo uma análise mais precisa e interpretável, conforme destacado por Fröhlich et al. (2016) e Queiroz (2020).

Dando prosseguimento à investigação, após a apresentação dos conceitos estatísticos fundamentais para a compreensão dos dados, é necessário a explicação de algumas técnicas utilizadas para a representação visual das informações extraídas do conjunto de dados. Para tanto, foram empregadas

técnicas consagradas de visualização estatística por meio da linguagem de programação R, incluindo recursos como boxplots, gráficos de linhas e diagramas de dispersão. Tais representações, que serão detalhadas ao longo deste capítulo, são essenciais para evidenciar padrões, variações e comportamentos dos dados, além de facilitar a interpretação analítica e a comunicação dos resultados obtidos.

É necessário salientar que R “é uma linguagem de programação interpretada que foi desenvolvida para análises estatísticas e seu interpretador é disponível gratuitamente” (Fröhlich et al., 2016, p. 112). Além de ser uma linguagem em alto nível, também facilita o uso, pela semelhança com a linguagem natural. O R foi a tecnologia utilizada como interpretador dos dados para visualizar os gráficos.

De acordo com Bohnert et al. (2023), para entender os dados de forma clara, é essencial apresentá-los visualmente, e isso é o que chamamos de representação de dados. Essa representação envolve a seleção de formatos visuais — como linhas, barras, gráficos circulares ou outros elementos gráficos — que transformam os dados brutos em uma imagem capaz de destacar suas características, aproveitando a capacidade do cérebro de interpretar informações visuais. Além disso, a apresentação dos dados abrange todas as demais escolhas que definem a estrutura da visualização, como a inclusão de interatividade, anotações, paleta de cores e organização geral do layout. No fim, o propósito da visualização de dados é simplificar o entendimento, permitindo que o usuário absorva as informações de maneira mais eficiente e precisa.

Além do gráfico de linhas, que é um dos mais utilizados para visualização de dados, é importante abordar o método de visualização de dados como o Box Plot. Segundo Capela e Capela (2011):

O Box-plot, também conhecido como gráfico de caixa, é um gráfico estatístico que possibilita representar a distribuição de um conjunto de dados com base em alguns parâmetros descritivos. Existem algumas variações quanto à quantidade de estatísticas representadas nesse tipo de gráfico, mas de uma forma geral todos incluem a mediana, o 1º e o 3º quartil, os valores mínimos e máximos e eventuais outliers e extremos (CAPELA; CAPEILA, 2011, p. 361).

Também foi utilizada, como técnica de visualização de dados, a construção do diagrama de dispersão. Conforme expõe Couto (2018), esse tipo de gráfico, também denominado scatterplot, permite representar visualmente um conjunto de dados por meio de coordenadas cartesianas nos eixos X e Y, possibilitando a observação da disposição dos valores ao longo dos eixos horizontal e vertical.

A escolha adequada das técnicas de visualização — como box-plots, gráficos de linhas e diagramas de dispersão — é fundamental para transformar dados brutos em informações inteligíveis, conforme demonstrado pelas referências teóricas. Essas ferramentas não apenas facilitam a identificação de padrões e outliers, mas também potencializam a comunicação dos resultados, alinhando-se ao propósito da estatística descritiva de tornar complexidade acessível (BOHNERT et al., 2023; CAPELA; CAPELA, 2011). Assim, a combinação de métodos quantitativos e representações visuais, como as implementadas em R, garante uma análise multidimensional capaz de abordar as nuances do dataset "Combustíveis Brasil".

Com base nos fundamentos teóricos apresentados — desde medidas de tendência central e dispersão até técnicas de correlação e visualização —, o próximo capítulo dedicar-se-á à exposição e discussão dos resultados obtidos. Por meio de gráficos, serão analisadas as variações temporais e regionais dos preços dos combustíveis, as correlações entre os diferentes tipos (como gasolina e etanol) e os padrões identificados via regressão linear. Essa etapa visa não apenas validar as hipóteses levantadas, mas também contextualizar os achados à luz de estudos anteriores, destacando contribuições práticas para o setor energético.

2.2 Procedimentos Metodológicos

A presente pesquisa caracteriza-se como um estudo quantitativo, de natureza exploratória e descritiva, com ênfase na análise de dados secundários. O conjunto de dados utilizado foi extraído da plataforma Kaggle, dataset denominado "Combustíveis Brasil", onde a fonte principal é o repositório de dados abertos da ANP (Agência Nacional do Petróleo, Gás Natural e Biocombustível do Brasil) o qual contempla informações sobre os preços médios de comercialização de gasolina

comum, etanol hidratado, óleo diesel e gás liquefeito de petróleo (GLP). Os dados são disponibilizados em formato .csv e organizados por período, localidade, tipo de combustível, entre outras variáveis.

Todas as análises foram realizadas utilizando o software RStudio. A escolha do R justifica-se por ser uma linguagem gratuita, amplamente utilizada em análises estatísticas, com vasta documentação e bibliotecas voltadas à manipulação de dados e construção de gráficos, conforme destacado por Fröhlich et al. (2016).

Após o download dos dados, foram realizados os seguintes procedimentos de tratamento e organização:

- Importação e limpeza dos dados: utilizando a linguagem de programação R (R Core Team, 2024), foram carregados os arquivos .csv para o ambiente de análise. As etapas de limpeza incluíram: remoção de colunas com valores ausentes e eliminação de possíveis inconsistências nas entradas categóricas;
- Filtragem das variáveis: a análise foi restrita às seguintes variáveis: *referência, ano, mês, região, produto (combustível) e preço médio de revenda (R\$/litro ou R\$/kg)*.

Para a análise estatística descritiva, foram aplicadas medidas de tendência central (média, mediana e moda) e de dispersão, conforme proposto por Oliveira e Oliveira (2011). Essas medidas possibilitaram a compreensão inicial da distribuição dos preços ao longo do período estudado.

Para investigar a associação estatística entre variáveis, foram utilizados dois métodos principais:

- Como os dados não são normalmente distribuídos, foi utilizada a Correlação de Spearman, com o objetivo de identificar e quantificar a intensidade e a direção da relação linear entre os preços de diferentes combustíveis. O coeficiente foi interpretado com base nas diretrizes de Fröhlich et al. (2016);
- Regressão linear simples, aplicada para modelar a dependência de uma variável em relação à outra (Queiroz, 2020).

Além disso, para representar graficamente os dados e facilitar a interpretação, foram utilizadas técnicas de visualização estatística, como:

- Boxplots para comparação regional (Figura 1);
- Gráficos de linhas para análise da evolução dos preços (Figura 2);
- Diagramas de dispersão para relações entre combustíveis (Figura 3);
- Gráficos de regressão linear (Figura 4).

3 RESULTADOS E DISCUSSÃO

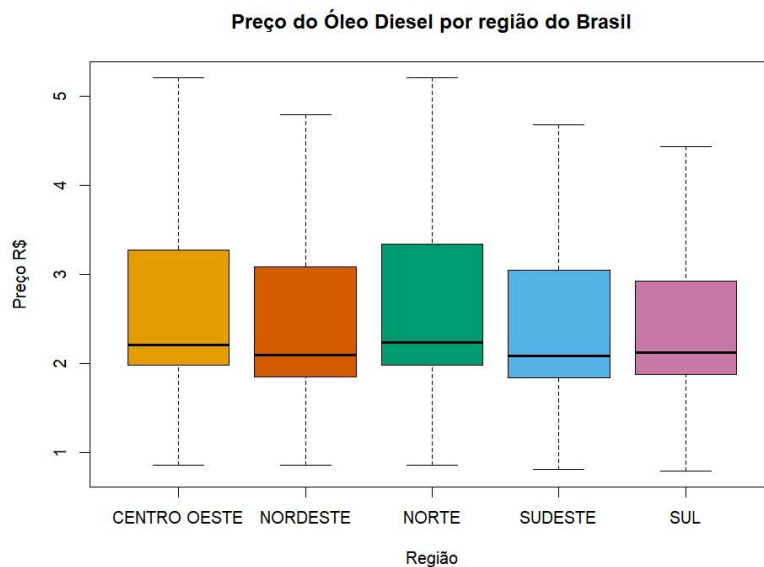
Esta seção apresenta os principais resultados obtidos por meio da análise exploratória dos preços dos combustíveis no Brasil, com base no conjunto de dados "Combustíveis Brasil", que compreende o período de 2001 a 2023. Foram utilizadas representações gráficas de caráter descritivo, como boxplots e diagramas de dispersão, com o objetivo de identificar padrões de variação regional, comportamentos temporais e relações entre os diferentes combustíveis analisados.

3.1 Análise Regional dos Preços do Óleo Diesel

A análise inicial concentrou-se nas variações regionais dos preços. Para isso, foi elaborado um boxplot referente ao óleo diesel, permitindo observar a distribuição dos valores nas cinco grandes regiões do país.

A Figura 1 apresenta o boxplot com a distribuição dos preços do óleo diesel nas cinco regiões do Brasil. Nota-se que a região Norte apresenta a mediana mais elevada, com maior dispersão de valores. A região Sudeste apresenta os menores preços medianos, com menor variação entre os quartis. As regiões Centro-Oeste, Sul e Nordeste encontram-se em faixas intermediárias.

Figura 1 – Distribuição dos preços do óleo diesel por região do Brasil.



Fonte: Elaborado pelos autores (2025).

A predominância de preços mais altos na região Norte pode ser explicada pela complexidade logística e pelas longas distâncias até os centros de distribuição, o que encarece significativamente o transporte. Segundo Gonçalves et al. (2023), os custos logísticos relacionados ao transporte e à infraestrutura impactam diretamente o valor final dos combustíveis, sobretudo em regiões com baixa densidade de rodovias pavimentadas. Em contrapartida, a região Sudeste, com maior proximidade das refinarias e infraestrutura consolidada, apresenta os menores valores médios.

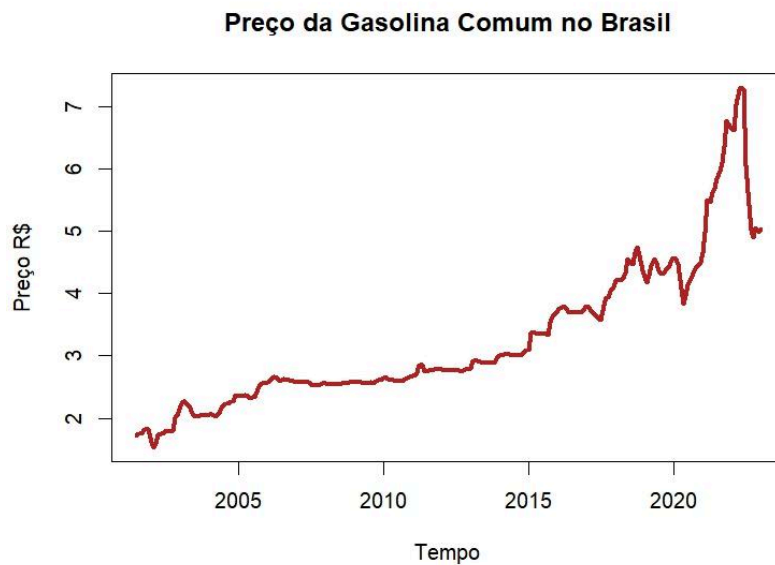
Esses dados reforçam a influência de fatores estruturais e logísticos na formação dos preços do óleo diesel, refletindo desigualdades regionais na distribuição de combustíveis no Brasil.

3.2 Evolução Temporal dos Preços

A Figura 2 apresenta a evolução dos preços da gasolina comum no Brasil ao longo do período de 2001 a 2023. Nota-se uma tendência geral de crescimento, com períodos de estabilidade seguidos por aumentos acentuados, especialmente a partir de 2020. Os anos de 2021 e 2022 registram os maiores picos de preço,

reflexo direto da pandemia de COVID-19, da desvalorização cambial e da alta no mercado internacional do petróleo.

Figura 2 – Evolução dos preços da gasolina comum no Brasil (2001–2023).



Fonte: Elaborado pelos autores (2025).

Esse comportamento é corroborado por Andrade et al. (2022), que destacam que as oscilações no preço da gasolina no Brasil são fortemente influenciadas por variáveis macroeconômicas como o câmbio e o preço do barril de petróleo Brent, além das políticas de preços da Petrobras. Além disso, alterações na carga tributária estadual (ICMS) e a volatilidade do mercado global contribuíram para os aumentos registrados no período pós-pandêmico.

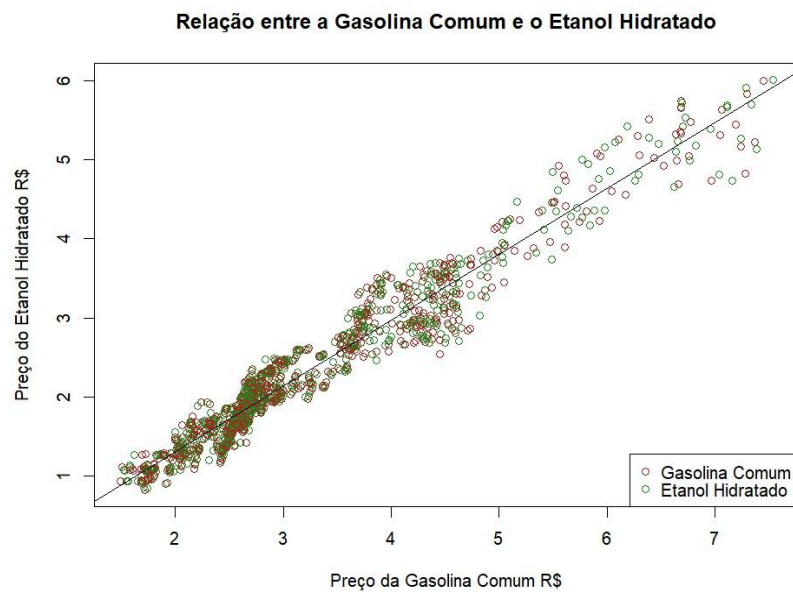
A análise evidencia que o preço da gasolina no Brasil é sensível a fatores externos e macroeconômicos, o que reforça a importância de estratégias de mitigação de volatilidade para garantir maior previsibilidade ao consumidor final.

3.3 Correlação entre Gasolina e Etanol

A Figura 3 apresenta o gráfico de dispersão entre os preços da gasolina comum e do etanol hidratado no Brasil. É possível observar uma correlação positiva

entre os dois combustíveis: à medida que o preço da gasolina aumenta, o do etanol também tende a subir. Essa relação é visualmente evidenciada pela concentração dos pontos ao longo de uma linha ascendente. O coeficiente de correlação de Spearman entre as duas variáveis foi de 0,97, indicando uma correlação forte e positiva.

Figura 3 – Correlação entre os preços da gasolina comum e do etanol hidratado.



Fonte: Elaborado pelos autores (2025).

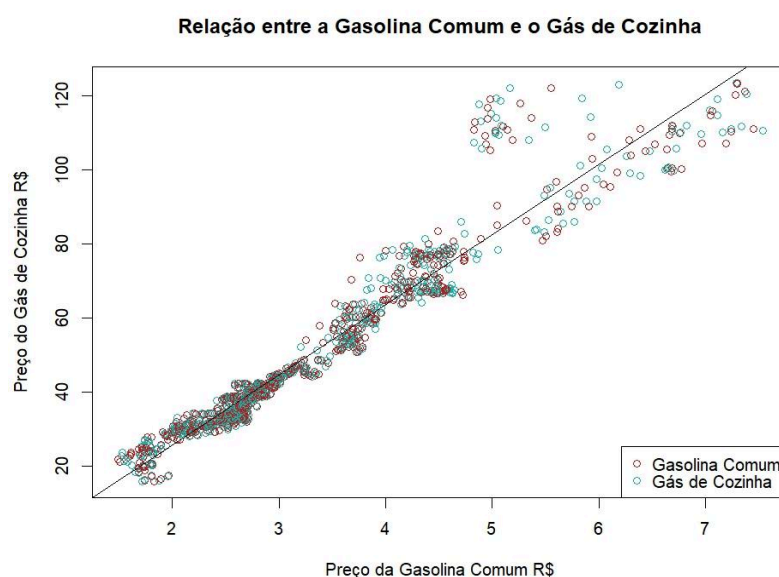
Esse comportamento reflete a substitutibilidade parcial entre os dois combustíveis, especialmente no contexto brasileiro, onde a escolha do consumidor entre gasolina e etanol depende da relação de preços e do rendimento energético de cada um. Segundo Losekann e Castro (2012), os preços tendem a se alinhar devido à concorrência direta no mercado e à reação das distribuidoras aos ajustes de mercado.

A análise reforça a interdependência entre os preços da gasolina e do etanol no Brasil, evidenciando que oscilações em um dos produtos tendem a influenciar diretamente o outro.

3.4 Regressão Linear entre Gasolina e Gás de Cozinha

A Figura 4 apresenta a regressão linear entre os preços da gasolina comum e do gás liquefeito de petróleo (GLP) no Brasil. A linha de tendência traçada sobre os pontos da amostra demonstra uma relação positiva entre as duas variáveis: verifica-se que aumentos no preço da gasolina comumente acompanham elevações nos preços do gás de cozinha.

Figura 4 – Regressão linear entre os preços da gasolina comum e do GLP.



Fonte: Elaborado pelos autores (2025).

O coeficiente de determinação (R^2) obtido foi de 0,93, o que indica que aproximadamente 93% da variação no preço do gás de cozinha pode ser explicada pelas variações no preço da gasolina. Essa associação estatisticamente significativa sugere que o comportamento do GLP pode, em parte, ser explicado pelas variações no preço da gasolina. De acordo com a Agência Brasil (2021), o aumento do preço do gás de cozinha acompanha a valorização do petróleo no mercado internacional e a variação cambial, reforçando a interdependência entre os derivados do petróleo.

Ainda que o GLP e a gasolina sejam utilizados para finalidades diferentes, a correlação evidenciada na regressão linear demonstra que a cadeia de formação de preços sofre influência de fatores macroeconômicos semelhantes, como o valor do barril de petróleo Brent, a cotação do dólar e os custos de refino e transporte. Esses

resultados reforçam a necessidade de políticas públicas que considerem a interdependência dos combustíveis na formulação de estratégias de estabilização de preços.

4 CONSIDERAÇÕES FINAIS

O presente trabalho realizou uma análise exploratória dos preços dos combustíveis no Brasil entre os anos de 2001 e 2023, com base em dados da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). Por meio de técnicas estatísticas e representações gráficas, foi possível compreender padrões de comportamento dos principais combustíveis — gasolina comum, etanol hidratado, óleo diesel e gás de cozinha (GLP) — em diferentes regiões do país e ao longo do tempo.

Além da contribuição acadêmica, os achados do estudo podem ser úteis para aplicações práticas em setores como o transporte rodoviário. Por exemplo, empresas que atuam nacionalmente podem usar os padrões regionais identificados para reavaliar pontos de abastecimento, priorizando regiões com preços historicamente mais baixos de diesel. Isso permite otimizar rotas e reduzir custos logísticos, especialmente em períodos de alta volatilidade. O conhecimento sobre a interdependência entre os combustíveis também pode apoiar decisões estratégicas de compra e armazenamento em grandes frotas.

Como próximos passos, o estudo poderá ser aprofundado com o uso de métodos estatísticos adicionais, como a Análise de Variância (ANOVA) para avaliar diferenças estatísticas significativas entre as regiões, bem como o teste de comparações múltiplas de Tukey, a fim de identificar pares de regiões com médias significativamente distintas. Essas ferramentas podem contribuir para uma análise ainda mais robusta das variáveis envolvidas.

Os resultados obtidos reforçam a importância de considerar fatores estruturais, logísticos e macroeconômicos na formulação de políticas públicas que visem à estabilização dos preços dos combustíveis no Brasil. Além disso,

destaca-se a relevância do uso de análises estatísticas e exploratórias para orientar decisões estratégicas no setor energético.

REFERÊNCIAS

AGÊNCIA BRASIL. **Gás de cozinha sobe mais que o dobro da inflação em 2020.**

Agência Brasil, 12 jan. 2021. Disponível em:

<https://agenciabrasil.ebc.com.br/economia/noticia/2021-01/gas-de-cozinha-sobe-mais-que-o-dobro-da-inflacao-em-2020>. Acesso em: 12 mai. 2025.

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS (ANP).

Preços de combustíveis nas refinarias. Brasília, DF: ANP, 2023. Disponível em:

<https://www.gov.br/anp/pt-br/centrais-de-conteudo/dados-abertos/serie-historica-de-precos-de-combustiveis>. Acesso em: 12 mai. 2025.

ANDRADE, V. N.; MUNTASER, J. G. S.; PRADO, T. A. R. **Influência de variáveis macroeconômicas no preço das ações do setor financeiro da B3.** Revista de

Economia Mackenzie, v. 19, n. 1, p. 170–190, 2022. Disponível em:

<https://editorarevistas.mackenzie.br/index.php/rem/article/view/14944>. Acesso em: 12 mai. 2025.

CAPELA, M. V.; CAPELA, J. M. V. Elaboração de gráficos box-plot em planilhas de cálculo. In: **Congresso de Matemática Aplicada e Computacional da Região Sudeste – CNMAC Sudeste**, 2011. Disponível em:

<https://cursos.unipampa.edu.br/cursos/engenhariadeproducao/files/2016/08/sim-artigo-01-elaboracao-de-graficos-box-plot-em-planilhas-de-calculo.pdf>. Acesso em: 20 abr. 2025.

COUTO, J. M. C. **Técnicas de visualização de dados em gerenciamento de projetos de desenvolvimento de software:** proposta de extensão do PMBOK. 2018. 120 f.

Dissertação (Mestrado em Ciência da Computação) – Pontifícia Universidade

Católica do Rio Grande do Sul, Porto Alegre, 2018. Disponível em:
<https://repositorio.pucrs.br/dspace/handle/10923/12462>. Acesso em: 20 abr. 2025.

FRÖHLICH, A.; TURCATO, A. B.; TIEZZI, D. G. Metodologia científica para estatística descritiva utilizando a linguagem R. **ULAKES Journal of Medicine**, [S. l.], v. 3, n. 2, 2023. Disponível em:
<https://revistas.unilago.edu.br/index.php/ulakes/article/view/865>. Acesso em: 20 abr. 2025.

GONÇALVES, L. L. R.; DUARTE, S. L.; BORGES JUNIOR, D. R.; DUARTE, D. L. **O impacto do preço do diesel no transporte externo e no custo total do café**. *Congresso de Iniciação Científica da UFU, Uberlândia, 2023*. Disponível em:
https://eventos.ufu.br/sites/eventos.ufu.br/files/documentos/contufu2023_completo_0052.pdf. Acesso em: 12 mai. 2025.

LOH, S. **Volume, Velocidade, Variedade, Veracidade e Valor: Como os 5 Vs do Big Data estão impactando as Organizações e a Sociedade**. Porto Alegre, 2019, 166p. Disponível em: <https://www.intext.com.br/5vs-big-data.pdf>. Acesso em: 30 jun. 2025.

LOSEKANN, L.; CASTRO, N. J. de. **Impactos dos preços da gasolina e do etanol sobre a demanda de combustíveis no Brasil**. *Revista de Economia Contemporânea*, v. 16, n. 3, p. 451–478, 2012. Disponível em:
<https://www.scielo.br/j/rec/a/GpRWHVbwJYjnjchVSdygWfb/>. Acesso em: 12 mai. 2025.

OLIVEIRA, A.; OLIVEIRA, T. A. **Elementos de estatística descritiva**. Lisboa: Universidade Aberta, 2011. Disponível em: <http://hdl.handle.net/10400.2/1986>. Acesso em: 4 abr. 2025.

QUEIROZ, M. V. F.. **Correlação linear e regressão linear simples no conteúdo de matemática do ensino médio**. 2020. 95 f. Dissertação (Mestrado em Matemática) – Universidade Federal de Campina Grande, Campina Grande, 2020. Disponível em:

https://mat.ufcg.edu.br/profmat/wp-content/uploads/sites/5/2020/06/TCC-PROFMAT_MATHEUS_QUEIROZ.pdf. Acesso em: 05 mai. 2025.

SASSI, G. P. Introdução à Estatística Descritiva para Pesquisas em Informática na Educação. In: JAQUES, P. et al. (Org.). **Metodologia de Pesquisa Científica em Informática na Educação**. Porto Alegre: SBC, 2020. v. 2, p. 00-00. Disponível em: https://ceie.sbc.org.br/metodologia/wp-content/uploads/2019/05/livro2_cap9.pdf. Acesso em: 27 mar. 2025.

SILVESTRE, António Luís. **Análise de Dados e Estatística Descritiva**. Lisboa: Escolar Editora, 2007. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=mzu4j2SUKzMC&oi=fnd&pg=PA1&dq=SILVESTRE,+Ant%C3%B3nio+Lu%C3%ADs.+An%C3%A1lise+de+Dados+e+Estat%C3%ADstica+Descritiva.+Lisboa:+Escolar+Editora,+2007.&ots=AM9NrFHVbd&sig=0U_qWITJBi4zS5Gw2HDx3Lc2slk&redir_esc=y#v=onepage&q=SILVESTRE%2C%20Ant%C3%B3nio%20Lu%C3%ADs.%20An%C3%A1lise%20de%20Dados%20e%20Estat%C3%ADstica%20Descritiva.%20Lisboa%3A%20Escolar%20Editora%2C%202007.&f=false. Acesso em: 4 abr. 2025