

ANÁLISE MULTIVARIADA APLICADA AO PREÇO DO ETANOL HIDRATADO PRATICADO NO BRASIL

MULTIVARIATE ANALYSIS APPLIED TO THE PRICE OF HYDROUS ETHANOL PRACTICED IN BRAZIL

Aline Ferrarini Carassai

Engenheira de Produção (UFPR)
Universidade Federal do Paraná (UFPR)
Curitiba-PR, Brasil
alinecarassai@gmail.com

Marcos Augusto Mendes Marques

Doutor em Métodos Numéricos em Engenharia (UFPR)
Universidade Federal do Paraná (UFPR)
Curitiba-PR, Brasil
marquesammarcos@gmail.com

Wiliam de Assis Silva

Mestre em Engenharia de Produção (UFPR)
Universidade Federal do Paraná (UFPR)
Curitiba-PR, Brasil
wiliamdeassis@gmail.com

Mariana Kleina

Doutora em Métodos Numéricos em Engenharia (UFPR)
Universidade Federal do Paraná (UFPR)
Curitiba-PR, Brasil
marianakleina11@gmail.com

Daniel Ayub

Mestre em Engenharia de Produção (UFPR)
Universidade Federal do Paraná (UFPR)
Curitiba-PR, Brasil
ayub.daniel@gmail.com

* **Recebido em: 10/08/2020**

* **Aceito em: 26/02/2021**

RESUMO

Desde a década de 70, quando a produção de biocombustíveis começou a ser incentivada no Brasil, a demanda por etanol vem crescendo como uma alternativa mais acessível ao consumidor. Este consumo progressivo é responsável por guiar diversas políticas ambientais e econômicas adotadas no país e, conseqüentemente, o preço deste produto. Assim, o objetivo deste artigo foi estimar um modelo de regressão linear múltipla para previsão do preço médio do etanol hidratado vendido nos postos de combustíveis brasileiros. Foram utilizados dados mensais de preços informados pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) do período de janeiro de 2014 a janeiro de 2018, aos quais foi aplicado o método de *Stepwise* para estimação do modelo estatístico de regressão linear

múltipla. Desta forma, foram selecionadas quatro variáveis preditoras para o modelo de previsão do preço médio do etanol hidratado: volume de produção de petróleo, preço médio de revenda da gasolina, câmbio USD/BRL e preço médio do açúcar no mercado nacional. Com a estimação do modelo de regressão após a retirada de *outliers*, obteve-se um modelo com R^2 ajustado de 98,67% e critérios dos testes estatísticos de significância atendidos para um nível de significância de 5%.

Palavras-chave: Etanol; Análise Multivariada; Regressão Linear Múltipla.

ABSTRACT

Since the 70's, when the production of biofuels started being encouraged in Brazil, the demand for ethanol has grown as a more accessible alternative for the consumers. This growing consumption has been responsible for determining many of the adopted environmental and economic policies in the country, as well as the product's price. Thus, the article's aim was to estimate a model of multiple linear regression to forecast the average hydrous ethanol price, sold in Brazilian gas stations. Monthly data provided by the National Agency of Petrol, Natural Gas and Biofuels (ANP) from January 2014 to January 2018 was used, to which was applied the Stepwise method to estimate the multiple linear regression statistical model. By doing so, four predictor variables were selected to the forecasting model of average hydrous ethanol price: the amount of produced petrol, the average price of gas, the USD/BRL rate and the average price of sugar. Hence, with the estimation of the linear regression model after the withdrawal of outliers, a model with an adjusted R^2 of 98,67% was obtained. All the statistical criteria were met for a significance level of 5%.

Keywords: Ethanol; Multivariate Analysis; Multiple Linear Regression.

1. INTRODUÇÃO

De acordo com o Ministério da Agricultura, Pecuária e Abastecimento do Brasil (2018), a participação do agronegócio no Produto Interno Bruto (PIB) brasileiro chega a 24%, envolvendo as atividades de produção agropecuária, transformação e distribuição. Somente no ano de 2017, o crescimento da agropecuária foi de 14,5%, impulsionado por produtos como a soja, milho, algodão, arroz e cana-de-açúcar. O etanol, no entanto, produzido a partir da transformação da cana, retratou queda de 2,73% em sua produção no ano de 2017 (CEPEA, 2017). Tais dados demonstram não só a volatilidade dos preços e demanda do etanol, mas também sua representatividade na consolidação do Produto Interno Bruto (PIB) do Brasil, já que o país é o segundo maior produtor deste combustível no mundo (ANP, 2018).

O consumo de etanol vem apresentando maior força no mercado de combustíveis desde a introdução de veículos *flex fuel* (veículos que têm a flexibilidade de usar tanto a gasolina quanto o etanol para o funcionamento dos motores) no Brasil, concedendo a este produto importância econômica no país e posição de destaque mundial em sustentabilidade. Observou-se em 2017 que mais de 70% dos veículos leves licenciados vendidos no país possuíam a tecnologia *flex fuel*. Neste tipo de veículo, o etanol hidratado (combustível) concorre diretamente com a gasolina e a competitividade de preços entre os dois combustíveis afeta pontualmente o poder de compra do consumidor (COSTA *et al.*, 2017).

Além disso, estima-se que a demanda por etanol combustível continue a evoluir nos próximos anos, primeiramente pela oferta de gasolina, cada vez mais cara em função do volume importado e falta de investimentos na capacidade de refino, e também pelos crescentes incentivos governamentais para a utilização de biocombustíveis. Por fim, cada vez mais países procuram pela opção de veículos *flex* ou híbridos buscando economia nos gastos com combustível e alternativas ambientalmente mais sustentáveis de transporte (COPERSUCAR, 2018).

Assim, avaliar o mercado de etanol hidratado e realizar estudos estatísticos a respeito de sua precificação são atitudes com significativa importância para ajudar a projetar a necessidade de importação de combustíveis, conduzir a política de produção de empresas do setor automotivo e até mesmo estimar o poder de compra do consumidor e sua demanda por diversos serviços. Por estes motivos, o objeto de estudo escolhido para análise no presente artigo foi o etanol hidratado, vendido diretamente nos postos de combustível.

Observando a literatura, verificou-se que foram encontrados diversos estudos que objetivaram prever os preços de *commodities* com relevante importância para a economia brasileira. No entanto, a maior parte destes estudos abordou modelos univariados para a previsão, em que o valor futuro de uma variável é previsto apenas relacionando-o com valores passados da mesma. Silva (2019), por exemplo, utiliza o modelo univariado ARIMA para previsão do preço do etanol hidratado no estado de São Paulo, obtendo uma equação cuja variável independente é o preço já verificado do etanol hidratado em períodos anteriores. Numa tentativa de envolver mais variáveis e tornar os modelos preditivos mais completos pode-se aplicar a análise multivariada que por sua vez permite a percepção da influência de um número amplo de variáveis externas sobre o objeto de estudo. Dessa forma, organizações podem embasar suas decisões e compreender oscilações de mercado a partir da observação destas variáveis.

Com este intuito, o presente estudo objetiva oferecer a instituições interessadas ou dependentes do mercado de etanol hidratado uma solução que possibilite uma previsão confiável do preço deste produto para facilitar decisões como volume a ser produzido, comprado ou vendido, bem como trazer ao conhecimento destas instituições outros produtos e variáveis que influenciem nestas decisões. Além disso, a disseminação deste conhecimento entre os consumidores finais do etanol hidratado poderia evitar a prática de preços abusivos por parte dos postos de combustíveis.

Desta maneira, o propósito desta pesquisa é determinar um modelo preditivo de regressão linear múltipla para a previsão de preços de revenda do etanol hidratado no Brasil, baseado em séries históricas de preços praticados (variável resposta) em conjunto com a análise da relação de variáveis explicativas previamente selecionadas.

Como parâmetro inicial foi utilizada a série de preços médios de revenda de etanol hidratado disponibilizada pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP, 2018) do período de janeiro de 2014 a janeiro de 2018.

Já as variáveis escolhidas como parâmetro para estimação do modelo de regressão foram determinadas com base nas análises conjunturais e setoriais da Companhia Nacional de Abastecimento (CONAB, 2017), Fundação Getúlio Vargas (FGV, 2017) e ANP (2018).

2. HISTÓRICO DO SETOR SUCROALCOOLEIRO E ANÁLISE MULTIVARIADA

Esta seção tem por objetivo avaliar a cronologia dos eventos significativos relacionados ao setor sucroalcooleiro, desde o surgimento do programa Proálcool, na década de 70, até as projeções futuras de consumo desta *commodity*. Além disso, nesta seção serão exploradas definições fundamentais acerca do estudo da Análise Multivariada, que neste contexto atuará como uma ferramenta de apoio à decisão sobre investimentos e custos relacionados ao etanol praticado no mercado brasileiro.

2.1. SETOR SUCROALCOOLEIRO

O acelerado aumento do consumo de derivados de petróleo para atender a demanda nacional por combustíveis levou a uma grande dependência do petróleo importado, este chegando a cerca de 80% de todo o petróleo consumido no país (ALMEIDA, LONGHI e SANTOS, 2017). Por este motivo, na década de 70, o governo brasileiro deu início à

estimulação do uso de biocombustíveis por meio do programa Proálcool, que regulamentou a adição de etanol anidro na gasolina (KOHLHEPP, 2010).

Na segunda fase do Proálcool, no final da década de 70, desencadeou-se também o incentivo à comercialização de etanol hidratado, com maior quantidade de água em sua composição, utilizado em veículos com motores preparados que começaram a ser vendidos no país nesta época. Os veículos movidos a etanol chegaram a representar 90% das vendas em 1983 (TONIN e TONIN, 2014).

Conforme FGV (2017) o país passou por uma séria crise de abastecimento de etanol no final da década de 80 em função da queda do preço internacional do petróleo, aumento da produção nacional e valorização do açúcar no mercado internacional, que acabou consumindo a maior parte da cana produzida no Brasil. No entanto, após esta fase, os incentivos à produção e consumo de etanol no país foram aumentando por parte do governo (FGV, 2017).

Em 2003 surge no país a tecnologia *flex fuel* (carros bicombustíveis ou híbridos), que representou 71% da frota de veículos em 2015 e fez com que o consumo de etanol por veículos automotores aumentasse, em substituição à gasolina (COSTA *et al.*, 2017).

Em 2016 o governo lança o programa Renovabio, procurando estimular a produção de biocombustíveis no país e visando a ampliar a utilização de combustíveis renováveis na matriz energética brasileira, determinando ações como: certificações para os biocombustíveis, adições de biocombustíveis aos combustíveis fósseis e incentivos fiscais e creditícios (FARINA e RODRIGUES, 2017).

No ano de 2018, visando a atingir os objetivos do Acordo de Paris, o Brasil comprometeu-se a reduzir suas emissões de gases de efeito estufa em 43% até 2030 por meio do emprego de energias renováveis e ampliação do consumo de biocombustíveis (FGV, 2017).

Com relação a perspectivas futuras, conforme Schutz, Massuquetti e Alves (2013) estima-se que, em 2030, o etanol chegue a representar 22% da demanda total de combustíveis líquidos no Brasil. Tais iniciativas voltadas para a promoção de biocombustíveis devem-se principalmente à posição de destaque ocupada pelo Brasil em termos de condições climáticas e área agrícola disponível para este tipo de produção. Além disso, o fomento do plantio com fins energéticos visa a impulsionar a geração de empregos e a inovação tecnológica no setor, bem como reduzir impactos ambientais e à saúde humana.

Em face desta análise cronológica observa-se que o etanol brasileiro tem significativo potencial de crescimento produtivo considerando sua característica renovável, dos incentivos governamentais e das necessidades relacionadas à demanda por energia para os próximos anos. No entanto, é preciso observar os aspectos de volatilidade quanto ao preço de venda deste produto, oscilações de mercado e de seus impactos ao planejamento de investimentos e tomada de decisões econômicas. Assim, de modo a propor uma ferramenta de apoio à tomada de decisão quanto aos preços praticados deste combustível, a seguir será proposto um modelo matemático de regressão linear múltipla que terá por finalidade prever os preços médios praticados para o etanol em postos de combustíveis em função de variáveis macroeconômicas relacionadas ao setor.

2.2. ANÁLISE MULTIVARIADA

O fundamento para embasar a tomada de decisão de empresas atualmente é, principalmente, a coleta de informações e gerenciamento das mesmas. A Análise Multivariada refere-se a um conjunto de técnicas que tem como objetivo analisar parâmetros relacionados a um determinado objeto de estudo, facilitando o acesso a informações sobre o mesmo (HAIR *et al.*, 2010).

De acordo com Marques (2005), a análise multivariada de dados é aplicada quando uma ou mais variáveis de interesse são descritas por diversas variáveis preditoras

simultaneamente. Assim, esta análise procura mensurar e explicar a relação entre diversas variáveis que influenciam, de alguma forma, o objeto estudado.

De acordo com Hair *et al.* (2010), entre as técnicas para estabelecer relação entre as variáveis estudadas estão a regressão linear múltipla e a correlação linear de Pearson, as quais serão utilizadas para estabelecer o modelo de previsão para este objeto de estudo.

2.2.1. Regressão linear múltipla

A Regressão Linear Múltipla é uma técnica estatística usada para prever o valor de uma variável dependente (variável resposta) a partir diversas variáveis independentes (variáveis preditoras). Este modelo determina que a variável Y depende de diversas variáveis X_i mais um erro aleatório (ϵ), determinado a partir de erros de medição e influência de outras variáveis não discriminadas no estudo (JOHNSON e WICHERN, 2007).

Desta forma, a equação de um modelo de regressão linear múltipla é dada pela equação 1:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_r X_r + \epsilon \quad (1)$$

Este modelo corresponde a um conjunto de r variáveis preditoras X que, combinadas, melhor modelam o comportamento da variável resposta Y. A equação modelada é estimada pelo método dos mínimos quadrados (MMQ), que determina os valores dos r parâmetros β (SASSI *et al.*, 2012).

2.2.2. Stepwise

No processo de seleção do conjunto de variáveis preditoras X para a estimação do modelo de regressão, podem ser utilizadas variáveis com correlações não significativas e, portanto, não adequadas para o modelo de previsão. Assim, se faz necessário utilizar-se de técnicas para se minimizar esse efeito negativo sobre o modelo, e uma dessas técnicas é denominada de *Stepwise*. A técnica *Stepwise* é um método de busca sequencial que visa a selecionar variáveis preditoras com a maior relação possível com a variável resposta, reduzindo assim, a quantidade de variáveis preditoras no modelo de regressão estimado, ou seja, acontece uma redução da dimensão do modelo estimado. Este método adiciona ou remove variáveis preditoras a partir do resultado do coeficiente de determinação R^2 , calculado para cada variável ainda existente no modelo (NCSS, 2018).

Primeiramente, o modelo escolhe como variável preditora inicial para a equação de regressão aquela com maior correlação com a variável resposta Y. Após a escolha da primeira variável, o modelo busca uma segunda variável que explique mais significativamente o erro remanescente na primeira equação de regressão encontrada. A equação de regressão é recalculada usando-se as duas variáveis, e o valor da estatística F (Teste de significância da existência da regressão) é examinado para verificar se a variável resposta definida para a primeira equação de regressão ainda é significativa, dada a adição da segunda variável (HAIR *et al.*, 2010).

Desta forma, a técnica *Stepwise* adiciona as variáveis preditoras em ordem decrescente de correlação com a variável resposta uma a uma, testando sua significância a cada nova inclusão e mantendo-a ou não na equação de regressão de acordo com o resultado encontrado. O processo é continuado até que não reste variável preditora a ser incluída na equação que resulte em melhoria da previsão.

2.2.3. Significância do modelo de regressão linear múltipla

Para verificar se o modelo de regressão estimado é estatisticamente significativo, deve-se realizar testes de significância como a ANOVA e o Teste t:

- **ANOVA (Teste de significância para existência da regressão):** A Análise de Variância (ANOVA), ou teste de significância para existência da regressão, é um teste para confirmar a existência de relação linear entre as variáveis independentes e a variável dependente do modelo, mostrando que o coeficiente β_r da variável X_r é diferente de zero (MARQUES e MARQUES, 2009). Primeiramente, formula-se a hipótese H_0 , em que $\beta_r = 0$. A partir do quadro montado pelo desenvolvimento da ANOVA é possível rejeitar ou aceitar a hipótese H_0 por meio da estatística F.

- **Testes de Significância dos Coeficientes da Equação de Regressão (Teste t):** O teste t representa a significância da correlação de uma variável independente com a variável dependente. Este valor indica se o coeficiente desta variável é ou não nulo, e pode ser utilizado no processo de *Stepwise* para determinar se esta variável será descartada após a adição de outra variável independente no modelo (HAIR *et al.*, 2010).

2.2.4. Análise dos resíduos

Além dos testes da ANOVA e do teste t, o modelo de regressão linear múltipla deve ser avaliado quanto a algumas suposições, em relação ao resíduo, para verificação da validade da equação. O resíduo é a principal medida de avaliação da variável estatística em relação à qualidade do ajuste, e é analisado quanto à sua homocedasticidade, independência e normalidade conforme HAIR *et al.* (2010).

- **Homocedasticidade:** A homocedasticidade garante que a variância do erro seja constante para todas as variáveis. Para o teste de homocedasticidade, é aplicado o teste de Breusch-Pagan e deseja-se não rejeitar a hipótese H_0 , ou seja, o resultado para o p-valor do teste deve ser maior que o nível de significância considerado.

- **Independência:** O teste de independência, busca assumir que os resíduos do modelo de regressão linear múltipla estimado, não pode ser modelado por nenhuma outra técnica estatística preditiva. Graficamente, a aleatoriedade dos resíduos, é demonstrada por uma distribuição aleatória no gráfico de resíduos ou por meio do teste de hipótese de Durbin-Watson. Deseja-se que o teste gere um resultado para o p-valor maior que o nível de significância estabelecido, ou seja, não se pode rejeitar H_0 para se garantir a independência dos resíduos.

- **Normalidade:** A normalidade dos resíduos, assim como os outros testes de significância para os resíduos, deve ser testada para garantir a confiabilidade do ajuste do modelo de regressão. Um dos testes mais utilizados para verificação da normalidade de um conjunto de dados é o teste de Shapiro-Wilk, que deve resultar em um p-valor maior que o nível de significância prefixado, ou seja, deseja-se não rejeitar a hipótese H_0 para se verificar a normalidade.

2.2.5. Multicolinearidade

A avaliação da multicolinearidade visa a garantir que as variáveis independentes incluídas no modelo não possuam alta correlação entre si, gerando redundância no modelo e assim tornando as inferências baseadas no modelo de regressão estimado não confiáveis. A multicolinearidade também pode ser avaliada a partir do valor do fator de inflação de

variância (VIF), que deve assumir valores abaixo de 10 para garantir baixa correlação entre as variáveis preditoras (HAIR *et al.*, 2010).

3. METODOLOGIA

Para o início do estudo, foram utilizados dados históricos do preço do etanol fornecidos pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP, 2018). Esta série histórica corresponde a um levantamento da média de preços de revenda de etanol praticados no país entre janeiro de 2014 e janeiro de 2018, ponderados pela média de vendas em cada estado.

Como fonte de dados econômicos foram utilizados o site do Instituto Brasileiro de Geografia e Estatística (IBGE, 2018a; IBGE, 2018b). Já para os dados referentes a *commodities*, foram utilizadas principalmente informações do Ministério da Agricultura, Pecuária e Abastecimento (2018), da Companhia Nacional de Abastecimento (CONAB, 2018), do Centro de Estudos Avançados em Economia Aplicada (CEPEA, 2018) e da União da Indústria de Cana-de-açúcar (ÚNICA, 2018).

Para o estudo foram selecionadas 26 variáveis preditoras iniciais escolhidas a partir da análise de relatórios setoriais elaborados pela Fundação Getúlio Vargas (FGV, 2017) e de boletins da ANP (2018), principalmente. Estes documentos possibilitaram a apuração de indicadores que costumam exercer influência sobre a composição do preço do etanol hidratado. Para cada variável foi utilizado um grupo de 49 observações mensais, compreendendo o período de janeiro de 2014 a janeiro de 2018. As variáveis selecionadas inicialmente podem ser observadas na tabela 1.

Tabela 1 - Variáveis preditoras e correlações

Variáveis	Correlação
X1 Preço do petróleo Brent (BRL)	-0,575486
X2 Preço do etanol EUA (USD)	-0,47113
X3 Preço do açúcar EUA (USD)	0,274864
X4 Preço do milho EUA (USD)	-0,418051
X5 Volume produção petróleo (barris)	0,518447
X6 Importação de etanol hidratado (m ³)	-0,438535
X7 Exportação de etanol hidratado (m ³)	-0,076615
X8 Volume de importação de gasolina (barris)	0,386529
X9 Preço médio de revenda da gasolina	0,922095
X10 Estoque público de milho (ton)	-0,832155
X11 Temperaturas médias (SP)	0,206395
X12 Preço médio do milho nacional (ton)	0,304955
X13 IPCA (y/y %)	-0,222593
X14 Câmbio USD/BRL	0,577884
X15 Selic anual (%)	0,113127
X16 Licenciamento total de automóveis e comerciais leves flex (Un.)	-0,73606
X17 Licenciamento total de automóveis e comerciais elétricos (Un.)	0,477226
X18 Preço médio do açúcar no mercado nacional (BRL)	0,702577
X19 Área de cana colhida (hec)	-0,308131
X20 IPP - Fabricação de derivados de petróleo e biocombustíveis (% mês anterior)	0,13078
X21 Vendas gasolina (barris)	0,117011
X22 Vendas de etanol hidratado (barris)	-0,12588

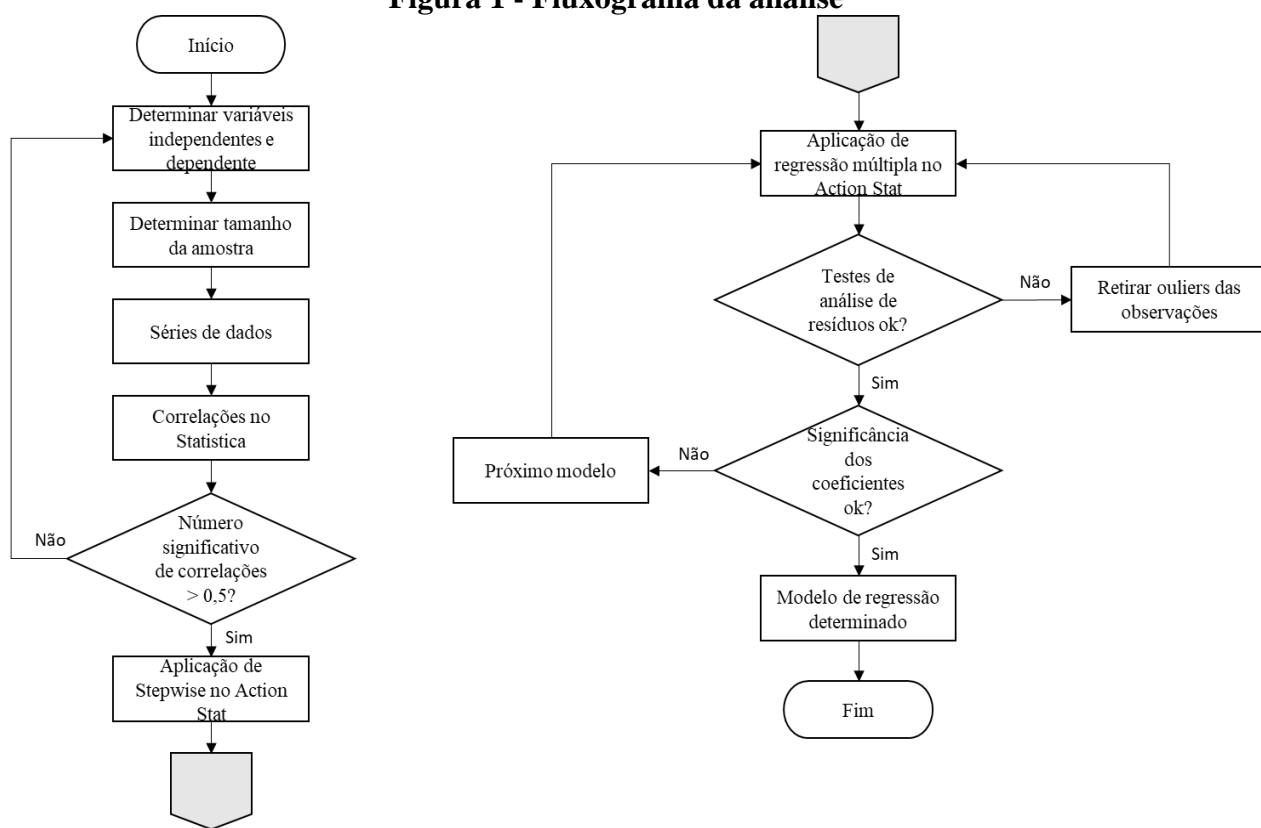
Tabela 1 (Continuação) - Variáveis preditoras e correlações

Variáveis	Correlação
X23 Volume de exmportação de açúcar (kg)	0,171319
X24 PIB EUA (y/y %)	0,035034
X25 Preço do petróleo WTI (USD)	-0,557393
X26 Volume de produção de cana (ton)	0,065794

Fonte: Os Autores (2018)

Acerca do tratamento dos dados coletados, para melhor visualização da metodologia utilizada até a obtenção do modelo final, elaborou-se um fluxograma que sumariza as informações sobre o método de trabalho aplicado (FIGURA 1).

Figura 1 - Fluxograma da análise



Fonte: Os Autores (2018)

A análise da tabela 1 e figura 1 mostra que a correlação entre as variáveis preditoras iniciais e o preço do etanol foi analisada na matriz de correlações estimada no *software* Statistica. A partir da matriz, foi possível observar oito variáveis com correlação significativa (acima de 0,5, em módulo) com o preço do etanol, mostrando indícios que estas deveriam ser as possíveis variáveis utilizadas para se prosseguir com o modelo de regressão.

O modelo de regressão foi estimado no *software* Action Stat pelo método de *Stepwise*, o qual realiza todas as combinações possíveis entre as variáveis preditoras para escolha do modelo com melhores resultados para os testes estatísticos considerados.

Em seguida, o modelo gerado foi analisado quanto à significância dos coeficientes da equação, ao R^2 ajustado, à multicolinearidade e quanto aos resultados da análise de resíduos para normalidade, homocedasticidade e independência. Todos os testes foram realizados no *software* Action Stat.

A partir dos resultados da análise dos coeficientes e de resíduos, alguns *outliers* das séries utilizadas foram retirados com o objetivo de descartar observações discrepantes influenciadas por fatores não mensuráveis e, assim, melhorar os resultados dos testes estatísticos realizados. As observações que ultrapassassem os limites de controle do gráfico de dispersão dos resíduos padronizados versus valores padronizados eram retiradas e consideradas *outliers*. O *software* Action Stat foi utilizado para uma nova análise de coeficientes e de resíduos a cada *outlier* retirado.

Durante a análise, uma das variáveis do modelo inicial gerado pelo Action Stat não apresentou significância estatística para o nível previamente fixado de 5% e, portanto, foi excluída do modelo original.

4. RESULTADOS E DISCUSSÕES

A partir da utilização da técnica *Stepwise* no *software* Action Stat, obteve-se um modelo inicial com cinco variáveis preditoras, sendo elas: X_5 , X_6 , X_9 , X_{14} e X_{18} . Entretanto, durante a análise de resíduos e dos coeficientes, a variável X_6 apresentou resultados estatísticos não significativos e foi excluída do modelo.

Assim, após a retirada da variável X_6 e de *outliers* para o melhor ajuste do modelo, o modelo de regressão linear múltipla estimado é dado pela equação 2:

$$Y_j = -0,2456 - 4,56 \times 10^{-9} \times X_{5j} + 0,8210 \times X_{9j} - 0,1908 \times X_{14j} + 0,7320 \times X_{18j} + \varepsilon_i \quad (2)$$

Sendo:

Y_j = Preço médio previsto do etanol hidratado para o mês j ;

X_{5j} = Volume de produção de petróleo (barris) para o mês j ;

X_{9j} = Preço médio de revenda da gasolina (BRL) para o mês j ;

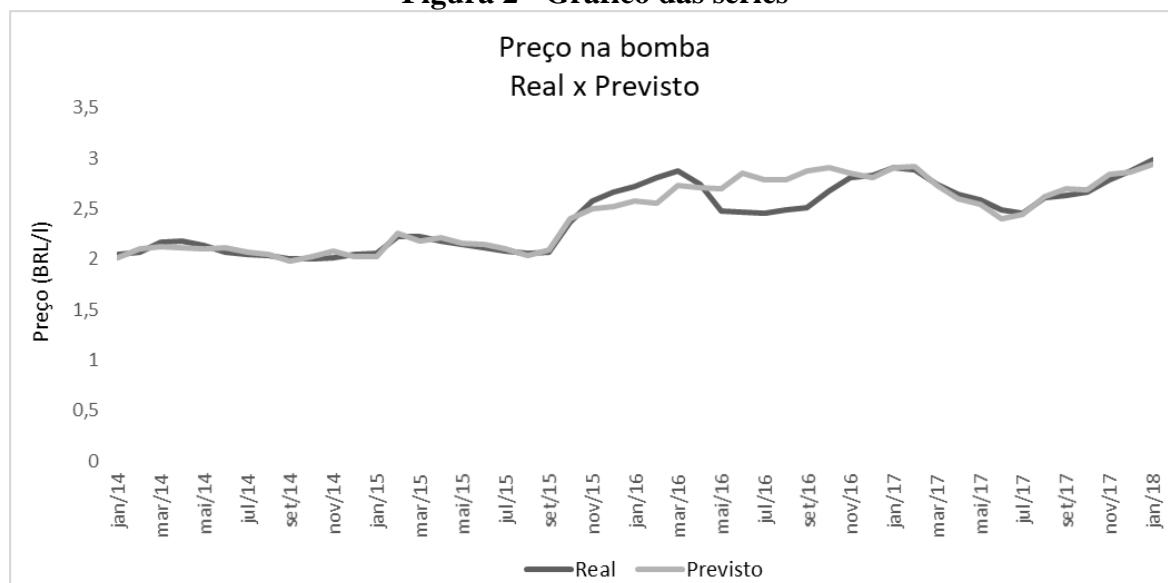
X_{14j} = Câmbio USD/BRL para o mês j ;

X_{18j} = Preço do açúcar no mercado nacional (BRL) para o mês j .

Dessa forma, substituindo os valores reais das variáveis preditoras para o mês j , é possível estimar o preço do etanol para este mês, com uma determinada margem de erro. Para a série histórica de 49 meses considerada para o modelo, o erro médio percentual absoluto foi de 3,12%. O R^2 ajustado do modelo, que representa a proporção da variabilidade da variável resposta que é explicada pelas variáveis preditoras, foi de 98,67%. Vale ressaltar que tal valor foi alcançado após a retirada de *outliers* do conjunto de observações utilizado.

Para uma melhor observação da adequação do modelo de regressão atingido, plotou-se um gráfico da série de preços reais do etanol contrastando com outro da série de preços previstos utilizando a equação obtida (FIGURA 2).

Figura 2 - Gráfico das séries



Fonte: Os Autores (2018)

Os testes estatísticos realizados visando a confirmação da significância dos coeficientes estão ilustrados nas tabelas 2 e 3, considerando um nível de significância de 5%:

Tabela 2 – ANOVA

Fatores	p-valor	Critério
X ₅	1,1697E-26	Atendido
X ₉	1,90008E-28	Atendido
X ₁₄	1,3147E-09	Atendido
X ₁₈	3,14085E-17	Atendido

Fonte: Os Autores (2018)

Tabela 3 - Significância dos coeficientes

Preditor	p-valor	Critério
Intercepto	0,003628939	Atendido
X ₅	0,003509475	Atendido
X ₉	1,4449E-24	Atendido
X ₁₄	2,68092E-12	Atendido
X ₁₈	3,14085E-17	Atendido

Fonte: Os Autores (2018)

Para se confirmar multicolinearidade aceitável entre as variáveis preditoras, as mesmas devem assumir valores para VIF menores que 10, o que se comprova para o modelo obtido, conforme a tabela 4.

Tabela 4 – Multicolinearidade

Preditor	VIF	Critério
X ₅	2,068232399	Atendido
X ₉	3,051903105	Atendido
X ₁₄	1,945983803	Atendido
X ₁₈	1,748571236	Atendido

Fonte: Os Autores (2018)

Os resultados da análise de resíduos quanto à normalidade, homocedasticidade e independência dos resíduos foram os seguintes:

- Normalidade: Utilizando-se do teste de Shapiro-Wilk, obteve-se um p-valor de 0,115479335, não podendo se rejeitar a hipótese H_0 , ou seja, o critério de normalidade foi atendido.
- Homocedasticidade: Utilizando-se do teste de Breusch-Pagan, obteve-se um p-valor de 0,772227407, não podendo se rejeitar a hipótese H_0 , ou seja, o critério de homocedasticidade foi atendido.
- Independência: Utilizando-se do teste de Durbin-Watson, obteve-se um p-valor de 0,106921317, não podendo se rejeitar a hipótese H_0 , ou seja, o critério de independência foi atendido.

Desta forma, a partir dos resultados dos testes de significância realizados, confirmou-se a significância estatística dos coeficientes estimados para modelo de regressão linear múltipla estimado por meio do *software* Action Stat e, assim, atestar a validade do modelo e a relação das variáveis independentes com o preço do etanol (variável resposta).

Analisando a matriz de correlação das variáveis, observa-se que todas as variáveis independentes descritas pelo modelo possuem relação direta com a variável dependente (preço do etanol). Ou seja, uma alteração de valor das variáveis preditoras corresponde a uma subsequente alteração do preço do etanol.

A variável X_{5j} , que corresponde ao volume de petróleo produzido no mês j , é a variável do modelo que apresenta menor relação com o preço do etanol hidratado. É possível constatar esta relação a partir do valor assumido por seu coeficiente na equação de regressão.

De acordo com a FGV (2017), a queda do preço do petróleo, estimulada pelo aumento de sua produção, é responsável por desestimular investimentos no setor sucroenergético, já que é acompanhada por uma maior demanda pela gasolina, derivada deste produto. Apesar disso, a influência do preço do petróleo sobre o preço da gasolina ainda é mínima, visto que este último é mais impactado por tributos federais e estaduais. Assim, é possível confirmar a correlação entre a produção do petróleo e o preço do etanol, atestada pela variável X_{5j} .

Observando o coeficiente da variável X_{9j} (preço de revenda da gasolina) estimado no modelo de regressão, é possível constatar uma relação maior com o preço do etanol. Um aumento no preço da gasolina nos postos de combustíveis acarreta em uma demanda maior por seu substituto, o etanol, resultando em uma elevação dos preços do mesmo, conforme atestado por Bacchi (2005) em seu estudo utilizando a metodologia VEC. A autora, em sua análise, constata que o preço da gasolina exerce grande influência sobre o preço do etanol hidratado.

Já a variável X_{14j} , que corresponde ao câmbio BRL/USD para o mês j , apresenta grande influência sobre o preço do etanol devido a dois fatores: primeiramente, grande parte da gasolina consumida no Brasil é importada e, quando a moeda brasileira apresenta desvalorização frente ao dólar, este aumento de preços em função do câmbio é repassado ao consumidor. O preço do etanol, como constatado anteriormente, acompanha a elevação de preços da gasolina. Além disso, o Brasil exporta grandes quantidades de açúcar principalmente para Bangladesh, Emirados Árabes e Índia (ÚNICA, 2018). Com a valorização do dólar, a exportação deste produto se torna ainda mais atrativa para as usinas de cana-de-açúcar que, como consequência, produzem um volume menor de etanol a partir da cana. A relação entre o mercado externo de açúcar e o preço doméstico de etanol hidratado foi estudada por Campos (2010). A autora, em sua pesquisa, atesta que o preço internacional de açúcar, influenciado pelo valor do câmbio, tem relação direta com o preço deste produto no mercado doméstico, o que causa efeito sobre o preço do etanol hidratado no Brasil.

Da mesma forma que a observada na variável X_{14j} , o aumento do preço do açúcar no mercado nacional, representado pela variável X_{18j} , é seguido por um volume de produção

menor do etanol, reduzindo sua oferta e acarretando na elevação de seu preço. Esta relação também é observada por Bacchi (2005) quando a autora corrobora a influência do preço de outros produtos do setor sucroalcooleiro sobre o etanol.

5. CONSIDERAÇÕES FINAIS

Com o intuito de gerenciar políticas ambientais e econômicas e identificar a viabilidade da produção do etanol hidratado em detrimento ao açúcar no Brasil, faz-se necessário um estudo rigoroso das variações do preço e da demanda deste produto no país. Este objetivo pode ser atingido utilizando métodos de análise multivariada, considerando séries históricas do preço do etanol e os diversos fatores que possuem relação com o mesmo.

Utilizando a técnica *Stepwise*, foi obtido um modelo de regressão linear múltipla, tendo como variável reposita o preço do etanol e, como variáveis preditoras, as variáveis que apresentaram maior correlação com este. Este modelo foi testado quanto à significância de seus coeficientes e seus resíduos, assim atendendo todos os requisitos para atestar sua validade. Foram obtidas como variáveis preditoras principais o volume de produção de petróleo, o preço médio de revenda da gasolina, o câmbio USD/BRL e o preço do açúcar no mercado nacional.

Avaliando a qualidade da série de preços ajustada, foi obtido um erro médio percentual absoluto de 3,12% quando comparada com a série real. O R^2 ajustado resultou em 98,67%, após a retirada dos *outliers*, demonstrando que o modelo representa de maneira satisfatória a variabilidade dos preços do etanol. Por fim, após análise do setor sucroalcooleiro, foi possível concluir que as variáveis preditoras do modelo descrevem com boa acuracidade a variável prevista.

As variáveis preditoras escolhidas conforme o modelo apresentado são: Volume de produção de petróleo em barris mensais (X_5), preço médio de revenda da gasolina em reais (X_9), câmbio mensal USD/BRL (X_{14}) e preço do açúcar no mercado nacional em reais (X_{18}).

Observando as variáveis obtidas pelo modelo e comparando-as com estudos de caso de variações de preços do setor sucroalcooleiro, é possível atestar a acuracidade dos resultados. Silva (2018) realizou um estudo para explorar as possíveis causas de um aumento expressivo no preço do etanol hidratado revendido na cidade de Goiânia (GO) em 2017. Entre as principais variáveis explicativas encontradas pelo autor estão o preço médio do barril do petróleo, o preço médio do açúcar e o preço médio de revenda da gasolina comum. O autor também identificou como uma das causas principais desta variação a elevação da margem de lucro de alguns postos de combustíveis do município.

Desta forma, é possível confirmar que as variáveis independentes do modelo obtido pelo presente estudo explicam, de fato, as variações no preço do etanol hidratado. Além disso, pode-se afirmar que o conhecimento destas variáveis explicativas permite ao consumidor final uma avaliação mais profunda dos preços do etanol nos postos de combustíveis, desencorajando a prática de preços abusivos.

Como complemento à análise multivariada apresentada nesta pesquisa e às conclusões obtidas, poderá ser realizado um estudo de viabilidade da previsão do preço do etanol hidratado, bem como a verificação da acuracidade da equação descrita a partir de observações futuras das variáveis preditoras escolhidas.

REFERÊNCIAS

ALMEIDA, V. P.; LONGHI, G. M.; SANTOS, L. R. Etanol: 40 anos de evolução do mercado de combustíveis e automóveis no Brasil. **Revista Teoria e Evidência Econômica**, v. 23. n. 49, 2017.

ANP - Agência Nacional Do Petróleo, Gás Natural E Biocombustíveis. Superintendência de Defesa da Concorrência, Estudos e Regulação Econômica. **Sistema De Levantamento De**

- Preços.** Relatório de Defesa da Concorrência, 2018. Disponível em: <http://www.anp.gov.br/images/Preços/Mensal2013/MENSAL_BRASIL-DESDE_Jan2013.xlsx>. Acesso em: 13 abril 2018.
- BACCHI, M. R. P. Formação de preços no setor sucroalcooleiro da região centro-sul do Brasil: relação com o mercado de combustível fóssil. **XXXIII Encontro Nacional de Economia**. Associação Nacional dos Centros de Pós-Graduação em Economia, Natal – RN, 2005.
- CAMPOS, S. K. **Fundamentos econômicos da formação do preço internacional de açúcar e dos preços domésticos de açúcar e etanol**. 2010. Tese (Doutorado em Ciências) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2010.
- CEPEA – Centro de Estudos Avançados em Economia Aplicada. **Boletim Cepea do Agronegócio Brasileiro**, referência 11/2017. Piracicaba, 2017. Disponível em: <https://www.cepea.esalq.usp.br/upload/kceditor/files/Relatorio%20PIBAGRO%20Brasil_NO_VEMBRO_CEPEA.pdf>. Acesso em: 21 abril 2018.
- CEPEA – Centro de Estudos Avançados em Economia Aplicada. **Agromensal Cepea/Esalq**: informações de mercado, referência 02/2018. São Paulo, 2018. Disponível em: <<https://www.cepea.esalq.usp.br/upload/revista/pdf/0469352001520532213.pdf>>. Acesso em: 21 abril 2018.
- CONAB – Companhia Nacional de Abastecimento. **Perfil do setor do açúcar e do etanol no Brasil – Edição para a safra 2014/2015**. Brasília, 2017. Disponível em: <https://www.conab.gov.br/info-agro/safras/cana/perfil-do-setor-sucroalcooleiro/item/download/1345_e86575e80b25295c05a1c14127bcc3a2>. Acesso em: 21 abril 2018.
- CONAB – Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de cana-de-açúcar**. Brasília, 2018. Disponível em: <https://www.conab.gov.br/component/k2/item/download/16105_76d98d8ca570d1c62e187ee00aafb9a2>. Acesso em: 21 abril 2018.
- COPERSUCAR. **Motores flex: 15 anos de benefícios para o consumidor**, 2018. Disponível em: <<https://www.copersucar.com.br/noticias/motores-flex-15-anos-de-beneficios-para-o-consumidor/>>. Acesso em: 21 abril 2018.
- COSTA, C. C.; BURNQUIST, H. L.; SOUZA, M. J. P.; VALDES C. Modelagem da demanda de combustível no Brasil: a importância dos veículos *flex*. **Revista Econômica do Nordeste**, Fortaleza, v. 48, n. 3, p. 113-127, jul./set. 2017.
- FARINA, E.; RODRIGUES, L. Desenvolvimento Sustentável. *Agroanalysis*, 2017. Disponível em: <<http://bibliotecadigital.fgv.br/ojs/index.php/agroanalysis/article/download/73724/70808>>. Acesso em: 22 set. 2018.
- FGV – Fundação Getúlio Vargas. **Caderno de Biocombustíveis**, referência 08/2017. Rio de Janeiro, 2017. Disponível em: <https://fgvenergia.fgv.br/sites/fgvenergia.fgv.br/files/caderno_biocombustivel_-_baixa.pdf>. Acesso em: 21 abril 2018.
- HAIR J, BLACK W, BABIN B, ANDERSON R, TATHAM R. **Análise Multivariada de Dados**. 6 ed. Porto Alegre: Bookman, 2010.
- IBGE – Instituto Brasileiro de Geografia e Estatística. Diretoria de Pesquisas, Coordenação de Índices de Preços. **Sistema Nacional de Índices de Preços ao Consumidor**, 2018a. Disponível em: <ftp://ftp.ibge.gov.br/Preços_Indices_de_Precos_ao_Consumidor/IPCA/Serie_Historica/ipca_SerieHist.zip>. Acesso em: 13 abril 2018.

-
- IBGE- Instituto Brasileiro de Geografia e Estatística. Sistema IBGE de Recuperação Automática (SIDRA). **Levantamento Sistemático da Produção Agrícola**, 2018b. Disponível em: < <https://sidra.ibge.gov.br/tabela/6588>>. Acesso em: 24 abril 2018.
- JOHNSON R. A.; WICHERN D. W. **Applied Multivariate Statistical Analysis**. 6 ed. Upper Saddle River: Pearson Prentice Hall, 2007.
- KOHLHEPP, G. **Análise da situação da produção de etanol e biodiesel no Brasil**. Estudos avançados, v. 24. n. 68. p. 223-253, 2010.
- MARQUES, J.; MARQUES M.A.M. **Estatística básica para os cursos de engenharia**. 2 ed. Curitiba: Domínio do Saber, 2009.
- MARQUES, M. A. M. **Aplicação da Análise multivariada no estudo da infra-estrutura dos serviços de saúde dos municípios paranaenses**. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas da Universidade Federal do Paraná, Curitiba, 2005.
- MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. **Agropecuária puxa o PIB de 2017**. 2018. Disponível em: <<http://www.agricultura.gov.br/noticias/agropecuaria-puxa-o-pib-de-2017>>. Acesso em: 22 abril 2018.
- NCSS – Statistical Software. **Stepwise Regression**, 2018. Disponível em: <https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Stepwise_Regression.pdf>. Acesso em: 22 jun. 2018.
- SASSI, C. P; PEREZ, F. G; MIYAZATO, L. et al. Modelos de regressão linear múltipla utilizando os softwares R e statistica: uma aplicação a dados de conservação de frutas. **ICMC – USP**. São Carlos. 2012. Disponível em: <http://repositorio.usp.br/directbitstream/1710e431-9f8a-471e-9b1c-4eba205beafd/Relat%C3%B3rios%20T%C3%A9cnicos_377...> . Acesso em: 15 out. 2018.
- SILVA, A. B. **Análise Descritiva das Possíveis Causas da Elevação de Preço do Etanol Hidratado no Município de Goiânia, Goiás, Brasil**. 2018. Monografia (Especialização em Gestão de Recursos Hídricos, Ambientais e Energéticos) - Universidade da Integração Internacional da Lusofonia Afro-Brasileira, São Francisco do Conde, 2018.
- SILVA, C. A. G.. Análise da previsão do preço do etanol hidratado no estado de São Paulo : uma aplicação do modelo arima. **Brazilian Journal of Development**. Curitiba, v. 5. n. 10. p. 17763-17778, 4 out. 2019.
- SCHUTZ, F.; MASSUQUETTI, A.; ALVES, T. W. Demanda e oferta energética: uma perspectiva mundial e nacional para o etanol. **Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental**. v. 16. n. 16. p. 3167-3186, 2013.
- TONIN, J. R.; TONIN, J. M. Do Proálcool ao “Próetanol”: novos desafios na produção do etanol brasileiro. **Informe Gepec**, v. 18. n. 1. p. 61-76, 2014.
- ÚNICA – União da Indústria de Cana-de-açúcar. **Histórico de exportação mensal de açúcar pelo Brasil, por região**. São Paulo, 2018. Disponível em: <<http://www.unicadata.com.br/listagem.php?idMn=65>>. Acesso em: 7 maio 2018.